Site-selective characterization of the dynamics of the nucleic acid chaperone properties of HIV-1 nucleocapsid protein using innovative fluorescent probes

Marianna Sholokh

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

UMR CNRS 7312 Laboratoire de Biophotonique et Pharmacologie

THÈSE présentée par :

Marianna SHOLOKH

soutenue le : **12 juillet 2016**

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

Discipline/ Spécialité : Sciences du vivant/Biophysique

Caractérisation site-sélective de la dynamique des propriétés chaperonnes de la protéine de la nucléocapside de VIH-1 vis-à-vis de ses cibles nucléiques, à l'aide de sondes fluorescentes innovantes

THÈSE dirigée par :

M. MELY Yves	Professeur, université de Strasbourg
Mme ZAPOROZHETS Olga	Professeur, université nationale Taras Chevtchenko de Kyiv

RAPPORTEURS :

Mme TISNÉ Carine	Docteur, université Paris Descartes
M. DEPREZ Eric	Docteur, école normale supérieure Cachan

AUTRES MEMBRES DU JURY :

Mme TEULADE-FICHOU Marie-Paule	Docteur, institut Curie
M. RUFF Marc	Docteur, université de Strasbourg

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

UMR CNRS 7312 Laboratory of Biophotonics and Pharmacology

THESIS presented by :

Marianna SHOLOKH

on : **12 July 2016**

to obtain the degree of : **Doctorate of University of Strasbourg**

Discipline/ Specialization : Life sciences/Biophysics

Site-selective characterization of the dynamics of the nucleic acid chaperone properties of HIV-1 nucleocapsid protein using innovative fluorescent probes

THESIS directed by :

M. MELY Yves	Professeur, université de Strasbourg
Mme ZAPOROZHETS Olga	Professeur, université nationale Taras Chevtchenko de Kyiv

RAPPORTEURS :

Mme TISNÉ Carine	Docteur, université Paris Descartes
M. DEPREZ Eric	Docteur, école normale supérieure Cachan

AUTRES MEMBRES DU JURY :

Mme TEULADE-FICHOU Marie-Paule	Docteur, institut Curie
M. RUFF Marc	Docteur, université de Strasbourg

Table of contents

Acknowledgements 4
List of abbreviations 6
Chapter 1. Bibliographical review 7
Part 1. Overview on the human immunodeficiency virus type 1 (HIV-1) 8
1.1.1. Discovery of HIV-1 8
1.1.2. Structure of the HIV-1 viral particle 9
1.1.3. Genetic organization of the HIV-1 genome 10
1.1.4. Virus replication cycle 18
1.1.5. Antiretroviral therapies 22
Part 2. Nucleocapsid protein as a potential target for anti-HIV therapy 24
1.2.1. Structural characteristics of the nucleocapsid protein (NC) 24
1.2.2. Binding of NC to nucleic acids 27
1.2.3. Chaperone properties of NC 30
1.2.4. Role of NC in HIV-1 reverse transcription 32
1.2.5. Role of NC in HIV-1 assembly 39
Part 3. Fluorescent amino acid analogs as markers for peptide-nucleic acid interactions 44
1.3.1. Introduction to fluorescence 44
1.3.2. Fluorescent probes 46
1.3.3. Fluorescence approaches to study protein – nucleic acid interactions 50
1.3.4. Natural fluorescent amino acids 53
1.3.5. Environment-sensitive probes 54
1.3.6. Ratiometric environment-sensitive probes based on 3-hydroxychromones 59
1.3.7. Applications of 3HC probes and 3HC-based fluorescent amino acid analogs for monitoring biomolecular interactions 65
Part 4. Fluorescent nucleic acid analogs: diversity, properties and applications 70
1.4.1. Introduction to nucleic acids 70
1.4.2. Emissive nucleoside analogs and their classification 73
1.4.3. Applications of fluorescent nucleoside analogs 83
Research objectives 86
Chapter 2. Materials and Methods 89
Part 1. Materials 90
2.1.1. Synthesis of fluorescent amino acid and nucleoside analogs 90
2.1.2. Peptide synthesis 91
2.1.3. Preparation of Zn-bound NC(11-55) peptides 94
2.1.4. Peptide activity tests 94
2.1.5. Oligonucleotide sequences used in this work 95
2.1.6. Synthesis of the labeled 3HCnt-(−)PBS DNA and d^{th}G7(−)PBS DNA 96

2.1.7. Preparation of the double-stranded ODNs ... 97
2.1.8. Lipids ... 97
2.1.9. Experiments with free nucleosides ... 99
2.1.10. Preparation of M3HFaa in a polystyrene film ... 99
Part 2. Physical measurements and procedures ... 99
2.2.1. UV/visible absorption and steady-state fluorescence measurements ... 99
2.2.2. Thermal denaturation experiments ... 101
2.2.3. Competition experiments ... 102
2.2.4. Quenching experiments ... 102
2.2.5. Time-resolved fluorescence spectroscopic measurements ... 103
2.2.6. Time-resolved fluorescence anisotropy measurements ... 105
2.2.7. Fluorescence lifetime imaging microscopy (FLIM) ... 106
2.2.8. Kinetic measurements ... 108
2.2.9. Cell cultures ... 112
2.2.10. Microinjection ... 113
2.2.11. Deconvolution of the spectra ... 113
Chapter 3. Results and Discussion ... 115
Part 1. Interaction of NC with its molecular partners using fluorescent amino acid analogs based on 3-hydroxychromones ... 116
Publication 1. Fluorescent amino acid undergoing excited state intramolecular proton transfer for site-specific probing and imaging of peptide interactions. ... 121
Part 2. Dynamics of (–)/(+)PBS annealing reaction and its promotion by the HIV-1 nucleocapsid protein using fluorescent isomorphic and environment-sensitive nucleoside analogs ... 123
3.2.1. Full mechanism of (–)/(+)PBS annealing revealed by site-selective labeling of (–)PBS by thienyl-3-hydroxychromone nucleoside analog ... 123
Publication 2. Full mechanism of (–)/(+)PBS annealing and its promotion by the HIV-1 nucleocapsid protein revealed by site-selective fluorescence labeling ... 131
3.2.2. Photophysical characterization of the isomorphic guanosine surrogate thienodeoxyguanosine as a potential G7 mimic in (–)PBS ... 133
Publication 3. Tautomers of a fluorescent G surrogate and their distinct photophysics provide additional information channels ... 137
3.2.3. Monitoring the dynamics and conformations of the G residue in (–)PBS stem-loop and (–)/(+)PBS duplex substituted by thienodeoxyguanosine and their side-by-side comparison with 2-aminopurine ... 139
Publication 4. Conquering 2-aminopurine's deficiencies: highly emissive isomorphic guanosine surrogate faithfully monitors guanosine conformation and dynamics in DNA ... 145
Chapter 4. Conclusions and Perspectives ... 147
Résumé ... 151
References ... 167

Acknowledgements

This thesis work is an important part of my life, a real adventure and challenge that will leave many-many bright memories. It would never be possible without the answer of Professor Yves Mély one day to my naive email sent from Ukraine. I am sincerely grateful to you for supervising my work, your continuous guidance and encouragement given to me along the whole period in the Lab, and for the motivation boosted hundred times after each our meeting. Thank you for teaching me, dedicating your precious time and your understanding that finally conducted us to reaching the objectives of the work. I'm grateful to my co-supervisor Professor Olga Zaporozhets for your wisdom and warm support all over the PhD period.

I'd like to thank the jury members Dr. Carine Tisné, Dr. Marie-Paule Teulade-Fichou, Dr. Marc Ruff, and Dr. Eric Deprez for accepting my request to evaluate this work.

I'm deeply thankful to Dr. Guy Duportail for your care and advices, to Dr. Andrey Klymchenko for your help and support, to Dr. Hugues de Rocquigny, Dr. Manuel Boutant and Dr. Eléonore Real for consulting me in complex biological issues, to Dr. Christian Boudier, Dr. Nicolas Humbert, Dr. Ludovic Richert, Dr. Julien Godet, Dr. Frederic Przybilla, Dr. Pascal Didier, Dr. Youri Arntz, Dr. Halina Anton, Dr. Andreas Reisch and Dr. Mayeul Collot for scientific discussions and help. Moreover, I'd like to thank Dr. Jean-Luc Darlix for careful reading of my manuscript introduction. I have my thanks to Marlyse Wernert, Ingrid Barthel, Ludovic Four and Michel Morciano for your availability and help.

I thank all my colleagues from the Laboratory of Biophotonics and Pharmacology for a friendly atmosphere in the Lab, useful discussions and happy moments. I highly appreciate the time spent with Lesia, Iryna, Katya, Liliana, Redouane, Rajhans, Sasha, Bogdan, Manu, Kamal, Yosuke, Krishna, Vika, Alex, Taras, Evgen, Natalia, Avisek, Sarwat, Nedal, Hassan, Wassem and others. Thank you guys for creating pleasant atmosphere and good memories of the PhD life!

Many thanks to our numerous collaborators, namely Professor Vasyl Pivovarenko for your "parental" attitude to me and our fruitful work together, to Dr. Dmytro Dziuba, Dr. Nicolas Barthes, Dr. Benoit Michel and Dr. Alain Burger from Nice, to Dr. Dongwon Shin and Professor Yitzhak Tor from the US, to Dr. Mattia Mori and Dr. Roberto Improta from

Italy, to Dr. Ranjan Das from India, to Dr. Karine Voltz and Dr. Roland Stote from Strasbourg. It was a pleasure to work together on different projects and to learn a lot from it.

Thank you Iuliia for being my neighbor during the most part of the PhD period, for your encouragement, scientific discussions and active leisure time together. I highly appreciate the support and care of my friends here and in Ukraine. Thank you for being next to me in more or less good moments.

Finally, I'm grateful to my husband Sergii and my beloved family in Kyiv for your patience, enormous support, and your love that I'm feeling all the time.

Last but not the least, I want to acknowledge Prof. Serge Potier, Mélanie Muser and Ecole Doctorale des Sciences de la Vie et de la Santé for providing me the financial support; the Dean of the Faculty of Pharmacy Professor Jean-Pierre Gies, Professor Eric Marchioni, the Faculty of Pharmacy and the University of Strasbourg for giving me an opportunity to gain the teaching experience and finalize my thesis manuscript in parallel.

Thank you very much!

Marianna Sholokh

List of abbreviations

AIDS – acquired immunodeficiency syndrome
ALADAN/DANA – 6-(2-dimethylaminonaphthoyl) alanine
ANS – 1,8-anilinonaphtalene sulfonate
2Ap – 2-aminopurine
AZT – azidothymine
BDF – base discriminating fluorescent probe
BMFC – 6-bromomethyl-2-(2-furanyl)-3-hydroxychromone
CA – capsid protein
CTD – C-terminal domain
dG – deoxyguanosine
4-DAPA – 4-(N,N-dimethylamino)-phtalimide propionic acid
DIS – dimerization initiation sequence
DLS – dimer linkage structure
4-DMAP – 4-*N,N*-dimethylaminophtalimide
4-DMN – 4-*N,N*-dimethylamino-1,8-naphtalimide
6-DMN – 6-*N,N*-dimethylamino-2,3-naphtalimide
DNA – deoxyribonucleic acid
dNTP – deoxynucleotide triphospate
ESCRT – endosomal sorting complex required for transport
ESIPT – excited-stated intramolecular proton transfer
FCS – fluorescence correlation spectroscopy
FITC – fluorescein isothiocyanate
FLIM – fluorescence lifetime imaging microscopy
FRET – Förster resonance energy transfer
G – guanine
GFP – green fluorescent protein
gRNA – genomic ribonucleic acid
HAART – highly active antiretroviral therapy
3HC – 3-hydroxychromone
3HCaa – 2-furyl-3-hydroxychromone amino acid analog
3HCnt – 2-thienyl-3-hydroxychromone nucleoside analog
3HF – 3-hydroxyflavone
HIV-1 – human immunodeficiency virus type 1
HTLV – human T-cell lymphotropic virus
IC – internal conversion
ICS – intersystem crossing
IN – integrase
IRES – internal ribosome entry segment
MA – matrix protein
3MI – 3-methyl-isoxanthopterin
6MI – 6-methyl-isoxanthopterin
mRNA – messenger RNA
Myr – myristic acid moiety
NA – nucleic acid
NC – nucleocapsid protein
NMR – nuclear magnetic resonance
NPC – nuclear pore complex
NTD – N-terminal domain
ODN – oligonucleotide
PAH – polycyclic aromatic hydrocarbons
PAS – primer activation signal
PBS – primer binding site
Phe – phenylalanin
PIC – pre-integration complex
PIP_2 – phosphatidylinositol-(4,5)-biphosphate phospholipids
PPT – polypurine tract
PRODAN – 6-propionyl-2-(dimethylamino)naphthalene
PR – protease
2PyG – 8-(2-pyridyl)-2′-deoxyguanosine
QD – quantum dot
QY – fluorescence quantum yield
RNA – ribonucleic acid
RRE – Rev response element
RT – reverse transcriptase
RTC – reverse transcription complex
SD – splice donor site
SL – stem-loop
SP – spacer peptide
(–)/(+)ssDNA – minus/plus strong-stop DNA
SU – surface protein
TAMRA – carboxytetramethylrhodamine
TAR – transactivation response element
Tat – transcription trans-activator
^{th}G – thienoguanosine
^{tz}G – isothiazologuanosine
TM – transmembrane protein
TMR – tetramethylrhodamine
Trp – tryptophan
tRNA – transfer RNA
UTR – untranslated region
8VdG – 8-vinyl-2′-deoxyguanosine
Vif – viral infectivity factor
Vpr – virus protein R
ZF – zinc finger

Chapter 1. Bibliographical review

Part 1. Overview on the human immunodeficiency virus type 1 (HIV-1)

1.1.1. Discovery of HIV-1

The story of HIV began more than 35 years ago, when the first clinical observations of a new alarming epidemic were made in the United States in 1981. The only known human retrovirus at that time was the Human T-cell lymphotropic virus (HTLV) that causes transformation of the T cells. However, the HTLV was not assumed as an etiological agent causing AIDS (Acquired Immunodeficiency Syndrome), and in 1983 the group of Luc Montagnier, Françoise Barré-Sinoussi and Jean-Claude Chermann from the Institut Pasteur in Paris first reported the isolation of a new human retrovirus (at that time known as LAV, lymphoadenopathy virus or HTLV-III, human T-cell lymphotropic virus type III) (Barré-Sinoussi, Chermann et al. 1983; Popovic, Sarin et al. 1983) possibly responsible for the AIDS disease. In 1984 Gallo et al (Gallo, Salahuddin et al. 1984) and Levy et al (Levy, Hoffman et al. 1984) published reports that confirmed the identification of the causative agent of AIDS. Françoise Barré-Sinoussi and Luc Montagnier in 2008 were awarded the Nobel Prize in Physiology and Medicine for the discovery of the HIV.

HIV is a lentivirus, a member of the large family of retroviruses. $CD4^+$ T lymphocytes and cells of the monocyte/macrophage lineage are the predominant targets of the virus (Dalgleish, Beverley et al. 1984; Klatzmann, Champagne et al. 1984; Silvin and Manel 2015). Less abundant cell populations such as dendritic cells can also be infected by HIV-1 (Altfeld, Fadda et al. 2011). Macrophages and TCD4+ cells constitute the major reservoirs for HIV-1 virus and are involved in various aspects of the AIDS disease (Venzke and Keppler 2006; Pan, Baldauf et al. 2013). When the TCD4+ cells number declines below a critical level, the cell-mediated immunity is lost, and infected persons become progressively susceptible to opportunistic infections. HIV-1 is transmitted by sexual contacts, drug addiction, through blood and from infected mother to child.

HIV-1 is characterized by extensive and dynamic genetic diversity, generating several molecular subtypes as well as recombinant forms. HIV-1 phylogenetic classifications are currently based either on nucleotide sequences derived from multiple subgenomic regions (*gag*, *pol* or *env*) of the same isolates or on the full length genome sequence analysis. Three major phylogenetic groups of the HIV-1 can be distinguished: group M (main), group O (outlier), and group N (non-M/non-O). The majority of infections worldwide belongs to the

group M and can be further subdivided into 10 subtypes (from A to K). Among them, the subtypes A (12% of infections), B (10% of infections), and C (50% of infections) are the most prevalent HIV-1 genetic forms worldwide. There is a specific geographic distribution for the subtypes, except sub-Saharan Africa, where almost all subtypes and recombinant forms have been detected. For instance, viruses of the subtype A are spread in central and eastern Africa as well as in eastern Europe; subtype B has been found in central and western Europe, North and South America, Australia, Southeast Asia, northern Africa, and Middle East. The subtype C is responsible for the HIV-1 infections mainly in southern Africa and India (Buonaguro, Tornesello et al. 2007).

The origin of HIV-1 was reported to be the cross-species transmission of the simian immunodeficiency virus SIVcpz from the chimpanzee subspecies *Pan troglodytes troglodytes* to humans likely due to direct exposure to animal blood upon hunting, butchering or consumption of raw meat (Gao, Bailes et al. 1999; Keele, Van Heuverswyn et al. 2006; Buonaguro, Tornesello et al. 2007).

In 1986 the second causative agent of AIDS, called HIV-2, was isolated by Luc Montagnier's group (Clavel, Guetard et al. 1986; Guyader, Emerman et al. 1987). HIV-1 and HIV-2 have similar genomic organization, replication cycle, modes of transmission and both results in AIDS. However, HIV-2 is characterized by a much lower transmissibility and reduced progression of AIDS. Geographically, HIV-2 is predominantly present in western Africa and more recently in southern India (Nyamweya, Hegedus et al. 2013).

According to the recent data of UNAIDS, in 2014 36.9 million people have been living with HIV-1, about 2 million people have been infected and further 1.2 million people died of AIDS. These statistics demonstrate that HIV/AIDS still remain a crucial global health issue. However, the new HIV infections has fallen by 35% since 2000, and AIDS-related deaths decreased by 42% since the peak in 2004 mainly due to the development of efficient therapies, and early diagnostics. These achievements further encourage scientists and medicine workers all over the world to consolidate their research and finally eradicate the HIV.

1.1.2. Structure of the HIV-1 viral particle

The mature HIV-1 viral particle has a spherical shape of 110 – 130 nm in diameter (Figure 1.1.1). The internal part of the viral particle is composed of two copies of homologous

single-stranded genomic RNA (gRNA) extensively coated by the nucleocapsid protein (NC). About 1500-2000 NC molecules are present in the viral core and interact with the gRNA via electrostatic, H-bonding and π-stacking interactions (Darlix, Lapadat-Tapolsky et al. 1995; Muriaux and Darlix 2010). The viral enzymes, namely reverse transcriptase (RT), protease (PR) and integrase (IN) are also enrolled in the viral core called capsid. Additionally, three accessory proteins Vpr (100-200 molecules), Vif and Nef are found in the viral particle (Darlix, Lapadat-Tapolsky et al. 1995; Darlix, Garrido et al. 2007; Muriaux and Darlix 2010). The capsid is composed of the capsid protein (CAp24) surrounded by a viral shell of the matrix protein (MAp17). The latter is bound to the outer phospholipid envelope originating from the host cell membrane. In the envelope a small number of viral glycoproteins in trimeric form are anchored, namely the surface SUgp120 and the transmembrane TMgp41 proteins (Ganser-Pornillos, Yeager et al. 2008; Muriaux and Darlix 2010; Sundquist and Krausslich 2012; Campbell and Hope 2015).

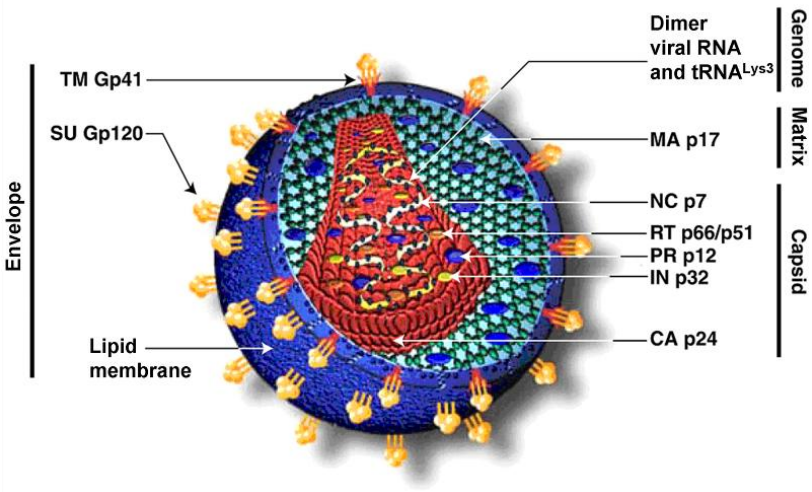


Figure 1.1.1. Schematic illustration of the mature HIV-1 virion (Adapted from Godet 2010). The viral capsid formed of the capsid protein CAp24 (red disks) enrolls two copies of viral RNA coated by the nucleocapsid protein NC and condensed with the viral enzymes. The conical capsid is surrounded by the matrix protein MAp17 (light blue spheres), bound to the envelope anchored with viral glycoproteins SUgp120 and TMgp41 (in orange).

1.1.3. Genetic organization of the HIV-1 genome

Viral RNA

The HIV-1 genome is packaged in the viral particle in the form of a 60S complex containing two tightly associated 9200 nucleotide-long molecules of the full length single-

stranded viral RNA with a 5' cap and 3' polyadenylation signal (Yedavalli and Jeang 2010). The dimeric RNA results from non-covalent interactions involving the 5′ dimer linkage structure (DLS), the dimerization initiation sequence (DIS) as well as many other contacts. During the reverse transcription step, the gRNA serves as a template for reverse transcriptase to generate a linear viral DNA duplex flanked with terminal redundant long terminal repeats (LTR).

The HIV-1 genome contains two types of genetic elements: coding and non-coding. The coding elements are three genes, namely *gag*, *pol* and *env* responsible for structural, enzymatic and envelope proteins, and the genes responsible for the regulatory (Tat, Rev) and accessory proteins (Nef, Vif, Vpr and Vpu) (Figure 1.1.2).

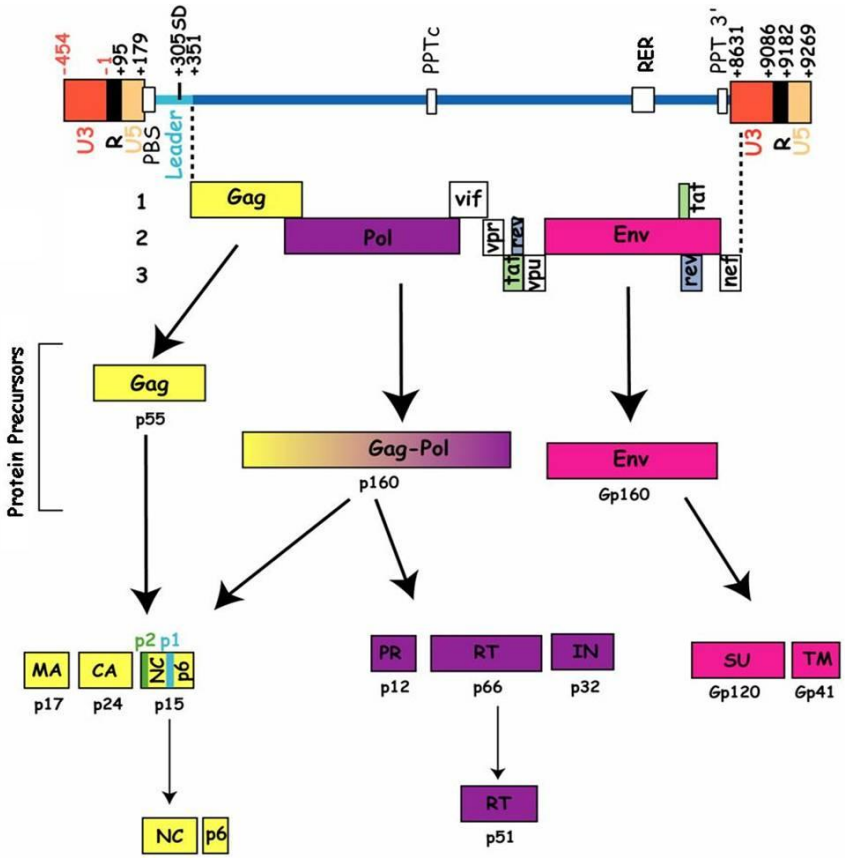


Figure 1.1.2. HIV-1 genetic organization. Three major genes: *gag*, *pol* and *env*, and genes encoding regulatory and accessory proteins are present in the viral RNA. Adapted from Goldschmidt 2004.

The 5′-untranslated region (UTR) is highly conserved in the HIV-1 RNA genome. It adopts complex structures involved in the key steps of the viral replication process. It is composed of the R, U5, PBS, and the leader regions. Recent NMR studies show that the 5′-UTR exists in equilibrium between two structures: one that promotes translation of the *gag*

gene by the internal ribosome entry segment (IRES) mechanism (Brasey, Lopez-Lastra et al. 2003), and the second that promotes dimerization of the gRNA (Berkhout and Van Wamel 2000; Berkhout, Ooms et al. 2002; Abbink and Berkhout 2003; Lu, Heng et al. 2011; Deforges, Chamond et al. 2012; Heng, Kharytonchyk et al. 2012) (Figure 1.1.3). The conformational change allows switching between these two functions during virus assembly.

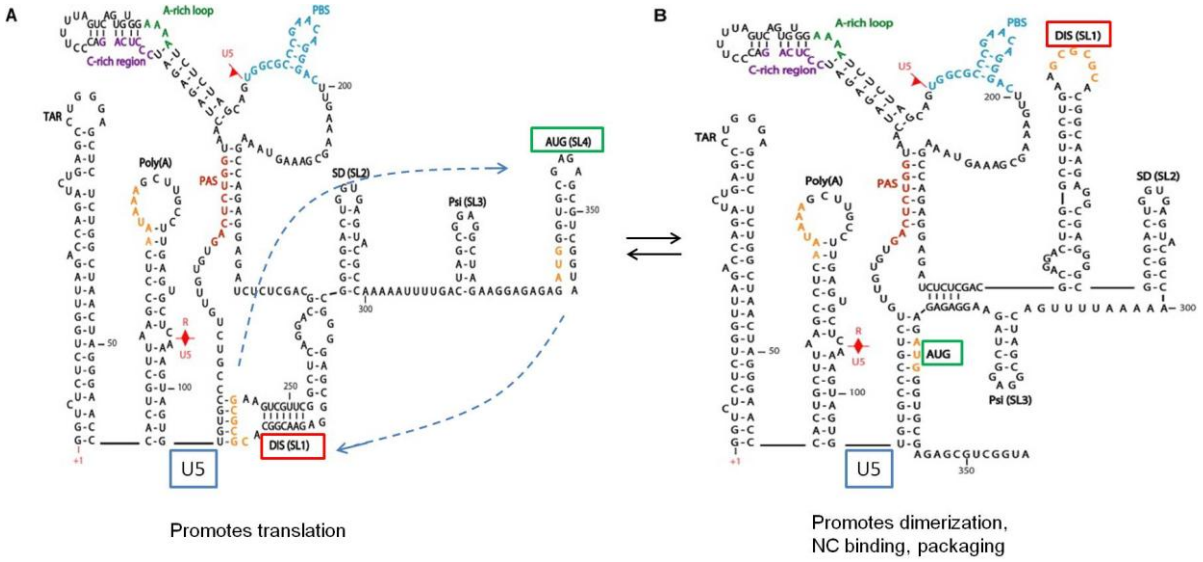


Figure 1.1.3. Secondary structures of the 5′-UTR of the HIV-1 gRNA (NL4.3 strain). (A) Secondary structure that promotes translation, where the unique 5′-end (U5) region base pairs with the dimerization initiation site (DIS) sequence; (B) Secondary structure associated with U5:AUG base pairing that exposes the DIS and high-affinity NC binding sites in order to promote the gRNA dimerization and packaging. The primer binding site (PBS) sequence is represented in blue, the primer activation signal (PAS) in red, the A-rich region in green, and the C-rich region in purple. The polyadenylation signal, the nucleotides involved in the dimerization signal (SL1), and the starting codon for the translation of *gag* are in orange. Adapted from Sleiman, Goldschmidt et al. 2012.

The **region R** (redundant/repeat) counts over a hundred nucleotides and is present as two identical copies at the 5′ and 3′ ends of the viral RNA. It contains two stem-loops: the transactivation response element (TAR) involved in regulation of the synthesis of the proviral DNA via the Tat protein (Peterlin, Luciw et al. 1986), and a polyA hairpin containing the AAUAAA polyadenylation signal (Feng and Holland 1988; Berkhout, Klaver et al. 1995). This motif is functional only at the 3′-end of the gRNA where it allows cleaving the viral RNA and addition of the poly-A tail (Klasens, Thiesen et al. 1999).

The **unique U5** region is located at the 5′-end of the gRNA and is the first part of the genome to be reverse-transcribed. U5 as well as the **U3** unique region contain sequences involved in integration of the proviral DNA into the host cell DNA.

The **primer binding site** (**PBS**) region, immediately following the U5 region, is an 18 nucleotide-long sequence strictly complementary to the 18 3′-terminal nucleotides of $tRNA^{Lys3}$ (G59-A76). PBS acts as a primer for the HIV-1 reverse transcription process.

The **leader region** is located between PBS and the starting codon for *gag* translation and is composed of four stem-loops (SL):

SL1 contains the primary dimerization initiation site (DIS) that promotes gRNA dimerization (Paillart, Marquet et al. 1994; Skripkin, Paillart et al. 1994; Gotte, Li et al. 1999) and may be linked to the gRNA packaging (Russell, Liang et al. 2004). It interacts with the U5 region of the gRNA and promotes translation of the gRNA (Figure 1.1.3A);

SL2 contains the splice donor site (SD) critical for splicing of the viral RNA for the synthesis of the viral accessory proteins (Purcell and Martin 1993; Oreilly, Mcnally et al. 1995);

SL3 is one of the main determinants of viral RNA encapsidation (ψ encapsidation signal), a critical step of HIV-1 assembly (Paillart, Shehu-Xhilaga et al. 2004; Lu, Heng et al. 2011);

SL4 contains an AUG initiation codon of translation. The AUG base pairs with the residues of the U5 region that promotes packaging of gRNA (Figure 1.1.3B).

The non-coding **polypurine tracts 3′ PPT and cPPT** are purine-rich sequences resistant to RNase H activity of RT that serve as primers for synthesis of the plus-strand DNA [(+)ssDNA] towards the 3′ end of gRNA.

The **Rev response element (RRE)** is a ~ 350 nucleotide-long highly structured element embedded in the *env* gene in unspliced and singly spliced viral RNA transcripts. RRE is the binding site of the HIV-1 Rev protein, a ~ 13 kDa accessory protein expressed during the early stage of the virus replication. After translation, Rev enters the nucleus, binds the RRE, and recruits cellular proteins to form larger complexes that are exported from the nucleus. Once in the cytoplasm, the complexes dissociate giving the unspliced and singly-spliced viral

RNAs that are packaged into the new virions or translated into the viral structural proteins and enzymes, respectively (Rausch and Le Grice 2015).

Proviral DNA

The HIV-1 proviral DNA encodes for three polyprotein precursors: Gag ($Pr55^{gag}$), GagPol and Env, whose cleavage during the maturation step results in mature infectious virions. Gag and GagPol are cleaved by viral PR after virion assembly, while Env is cleaved by the cellular protease during its trafficking from endoplasmic reticulum to plasma membrane (Adamson and Freed 2007). The viral DNA contains also two LTRs, the non-coding regions involved in the integration of viral DNA and transcription of viral RNA.

Gag gene encodes the 55 kDa polyprotein precursor $Pr55^{Gag}$ that forms the viral particle. It contains the matrix protein p17 (MA), capsid protein p24 (CA), nucleocapsid protein p7 (NC), p6 domain, and two spacer peptides, SP1 and SP2 (Freed 2015) (Figure 1.1.4).

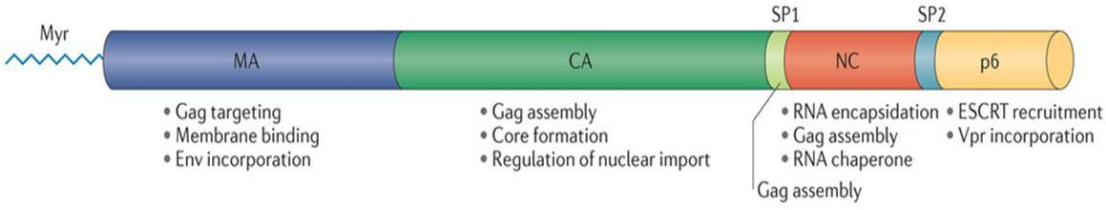


Figure 1.1.4. Structure and major functions of $Pr55^{Gag}$ polyprotein precursor: N-terminal myristic acid moiety (Myr) is covalently attached to the matrix domain (MA), capsid domain (CA), spacer protein 1 (SP1), nucleocapsid (NC), spacer protein 2 (SP2) and p6 domain. Adapted from Freed 2015.

The **MA** is folded into a highly globular structure composed of five α-helices and a three-stranded mixed β-sheet (Hill, Worthylake et al. 1996; Massiah, Starich et al. 1994). A number of basic residues at the top of MA allow the protein to interact electrostatically with the negatively charged phospholipids of the inner leaflet of the plasma membrane. Additionally, the myristic acid moiety (Myr) is inserted into the lipid bilayer in order to stabilize the Gag-membrane attachment during the assembly step. Moreover, MA assists the incorporation of Env glycoproteins into the forming virions (Alfadhli and Barklis 2014; Freed 2015).

The **CA** consists of two domains: a 150 amino acid-long N-terminal domain (CA_{NTD}) and an 80 amino acid-long C-terminal domain (CA_{CTD}) connected by a flexible linker (Campbell and Hope 2015). When assembled, CA_{NTD} is located on the outer surface of the capsid core and CA_{CTD} is oriented towards the interior of the capsid. A proline-rich loop of the CA_{NTD} is involved in binding of the host protein cyclophilin A (CYPA) that may help to protect viral capsid against the antiviral innate immune defenses during virus entry (Gamble, Vajdos et al. 1996; Freed 2015). The CA_{CTD} promotes Gag-Gag multimerization as a result of CA-CA interactions during viral assembly (Briggs and Krausslich 2011), and contains a highly conserved major homology region (MHR) important for stabilization of Gag oligomers associated to plasma membrane (Tanaka, Robinson et al. 2015).

The **NC** is a 55 amino acid protein with two conserved zinc fingers. NC serves as a nucleic acid (NA) chaperone, notably during the gRNA reverse transcription. As a part of $Pr55^{Gag}$, NC is critical for recognition and selection of the viral RNA during viral assembly and participates in RNA encapsidation. The structure and functions of NC will be discussed in details below.

The spacer proteins **SP** are located upstream of NC domain in Gag (SP1) and between NC and p6 (SP2) (Figure 1.1.4). The SP1 is involved in Gag assembly (Datta, Temeselew et al. 2011), while the importance of **SP2** is less defined. It is believed to be dispensable for virus maturation and viral infectivity (de Marco, Heuser et al. 2012).

The **p6** domain is largely unstructured (Fossen, Wray et al. 2005) and was shown to recruit cellular ESCRT (endosomal sorting complex required for transport) machinery to facilitate viral budding (Votteler and Sundquist 2013).

Pol gene encodes a 160 kDa polyprotein $Pr160^{GagPol}$, synthesized as a result of a programmed frameshifting event during the translation of Gag-encoding viral mRNA that brings *pol* sequence in the same reading frame as *gag* (Jacks, Power et al. 1988). It is an interesting fact that during HIV-1 replication, a large number of Gag molecules must be generated to serve as precursors to the structural proteins of the virions, while the enzymes encoded by *pol* gene are needed in smaller amount. The ribosomal frameshifting event occurs at a frequency of 5-10% during translation of Gag or Gag-Pol mRNA resulting in a Gag/Gag-Pol polyprotein ratio of around 20:1 in cell. Therefore, to express Gag protein at high levels in comparison with the *pol*-encoded proteins, the virus uses the same initiation codon in the same mRNA to express *gag* and *pol* genes. Translation of these RNAs leads to the synthesis

of a fusion protein that is usually called Gag-Pol precursor (Swanstrom and Wills 1997). Thus, $Pr160^{GagPol}$ polyprotein contains MA, CA, NC domains from Gag, and three viral enzymes, namely RT, PR and IN.

RT is a heterodimer composed of the two polypeptide chains p66 (66 kDa, 560 amino acids) and p51 (51 kDa, 440 amino acids). The p66 contains catalytically active N-terminal DNA polymerase domain and C-terminal RNase H domain. When two p66 chains dimerize, viral protease cleaves the RNase H moiety of one chain to produce a stable p66/p51 heterodimeric RT (Das and Arnold 2013). RT uses viral gRNA as a template and a host-cell tRNA as a primer to synthesize the minus-strand DNA, producing an RNA-DNA hybrid. This complex becomes a substrate of RNase H domain of RT, which cleaves the RNA strand at numerous points producing short RNA segments hybridized to the nascent DNA. Among these RNAs, two specific purine-rich sequences PPTs play roles of primers that initiate synthesis of the plus-strand DNA, ultimately synthesizing double-stranded DNA viral genome. Then RNase H removes the tRNA and PPT primers from minus- and plus-strand DNAs, respectively.

PR is a 99 amino acid aspartyl protease that functions as a dimer with an active site located at the dimer interface (Sundquist and Krausslich 2012). PR triggers virus maturation by cleaving Gag and GagPol polyproteins at specific sites to produce functional proteins. Protease-mediated Gag and GagPol cleavage is activated during virus assembly that is accompanied by significant changes in the virion morphology (Figure 1.1.5). In the immature virion, Gag molecules are assembled in a radial way. Following protease cleavage, the newly processed proteins, namely MA, CA, SP1, NC, SP2 and p6 reassemble to form the distinct layers of the mature virion. MA remains bound to the inner viral membrane, NC coats viral RNA, and CA assembles into a conical capsid to envelope the ribonucleoprotein complex.

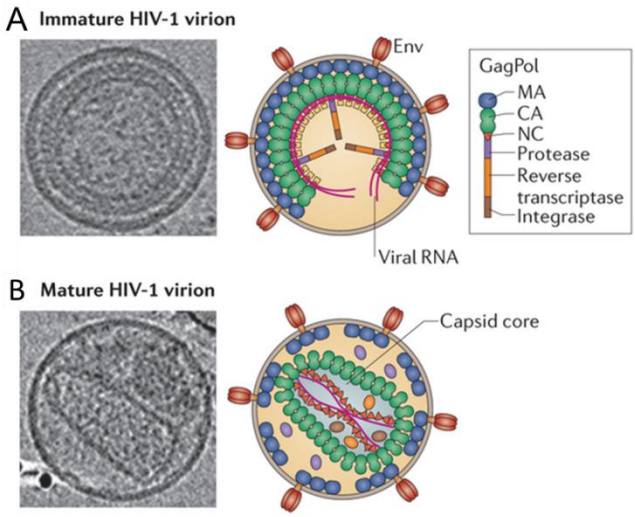


Figure 1.1.5. HIV-1 immature (A) and mature (B) virions. Representative central sections from the electron cryotomograms together with the schematic illustrations of an immature and a mature HIV-1 particles. Adapted from Freed 2015.

IN is a 32 kDa protein, encoded at the 3' end of *pol* gene, composed of three structurally and functionally different domains: (1-51) dimeric N-terminal domain containing an HH-CC Zn-binding motif (Cai, Zheng et al. 1997); (52-212) catalytic core domain containing endonuclease and polynucleotidyl transferase sites; (220-288) C-terminal domain responsible for DNA binding and oligomerization during the integration process (Zheng, Jenkins et al. 1996; Chen, Krucinski et al. 2000; Li, Xuan et al. 2015). IN catalyzes the 3′-processing step (cleavage of two nucleotides from the two 3′ ends of the double-stranded viral DNA) and the strand transfer reaction resulting in the insertion of the proviral DNA into the host cell DNA.

Env gene encodes for a polyprotein precursor $Pr160^{Env}$, whose cleavage product is the Env trimer. It is composed of three copies of non-covalently associated heterodimers of gp120 surface protein (SU) that interact with cellular receptors, and gp41 transmembrane protein (TM) necessary for fusion between the viral and cellular membranes.

The proviral DNA also encodes regulatory (Tat and Rev) and accessory (Nef, Vif, Vpr, Vpu) proteins responsible for viral pathology (Frankel and Young 1998; Malim and Emerman 2008).

1.1.4. Virus replication cycle

The HIV-1 viral life cycle can be divided into two main phases: from entry to proviral DNA integration (early) phase and post-integration (late) phase.

The early phase of the HIV-1 infection is initiated by binding of the viral envelope glycoproteins to the CD4 receptor and chemokine co-receptor 5 (CCR5) or C-X-C chemokine receptor 4 (CXCR4) on the host cell surface (Figure 1.1.6). Following viral attachment, fusion between the viral and host cell membranes occurs and results in a release of the HIV-1 conical capsid into the cytoplasm of the infected cell. There, it is believed to be present as a large ribonucleoprotein structure called the reverse transcription complex (RTC), where reverse transcription occurs resulting in synthesis of the proviral DNA (Mougel, Houzet et al. 2009). The RTC is thought to contain gRNA coated with NC and other components such as MA, CA, Vpr as well as viral enzymes RT and IN. The RTC enrolled by the capsid core uses the microtubule network of the host cell to traffic towards the nucleus. At this stage the capsid core disassembly may occur, though its exact timing and location in the cell are debated (Hulme, Perez et al. 2011; Ambrose and Aiken 2014). Then, the proviral DNA crosses the nuclear envelope and enters the nucleus in the form of the pre-integration complex (PIC). The PIC enters into the nucleus through the nuclear pore complex (NPC) which is composed by multiple copies of the nucleoporins embedded in the double layer of the nuclear membrane (Di Nunzio 2013). The viral partners involved in the nuclear entry are currently under investigation but the HIV-1 CA was evoked as one of the main viral components in this step (Campbell and Hope 2015). Following nuclear import, the proviral double-stranded DNA is inserted into the host cell chromosome.

The integration process is catalyzed by IN and consists of two steps. First, IN binds to a short sequence at each end of the LTR of the proviral DNA and catalyzes the endonucleotide cleavage called 3′-processing reaction leading to a removal of two nucleotides from each of the 3′ ends of LTR. This process is conducted by IN in dimeric form and takes place in the cytoplasm of the infected cells (Guiot, Carayon et al. 2006). Next, the processed DNA is used for the strand transfer reaction leading to the covalent insertion of the viral DNA into the host genome. The strand transfer is a series of phosphodiester transesterification reactions catalyzed by integrase, in which the 3′ OH-groups of the processed viral DNA ends attack a pair of phosphodiester bonds in the target DNA. The strand transfer occurs in the nuclei and is believed to require IN in its tetrameric form (Li, Mizuuchi et al. 2006). The integrated

provirus is responsible for the expression of viral proteins and gRNA necessary for the synthesis of the new infectious virions.

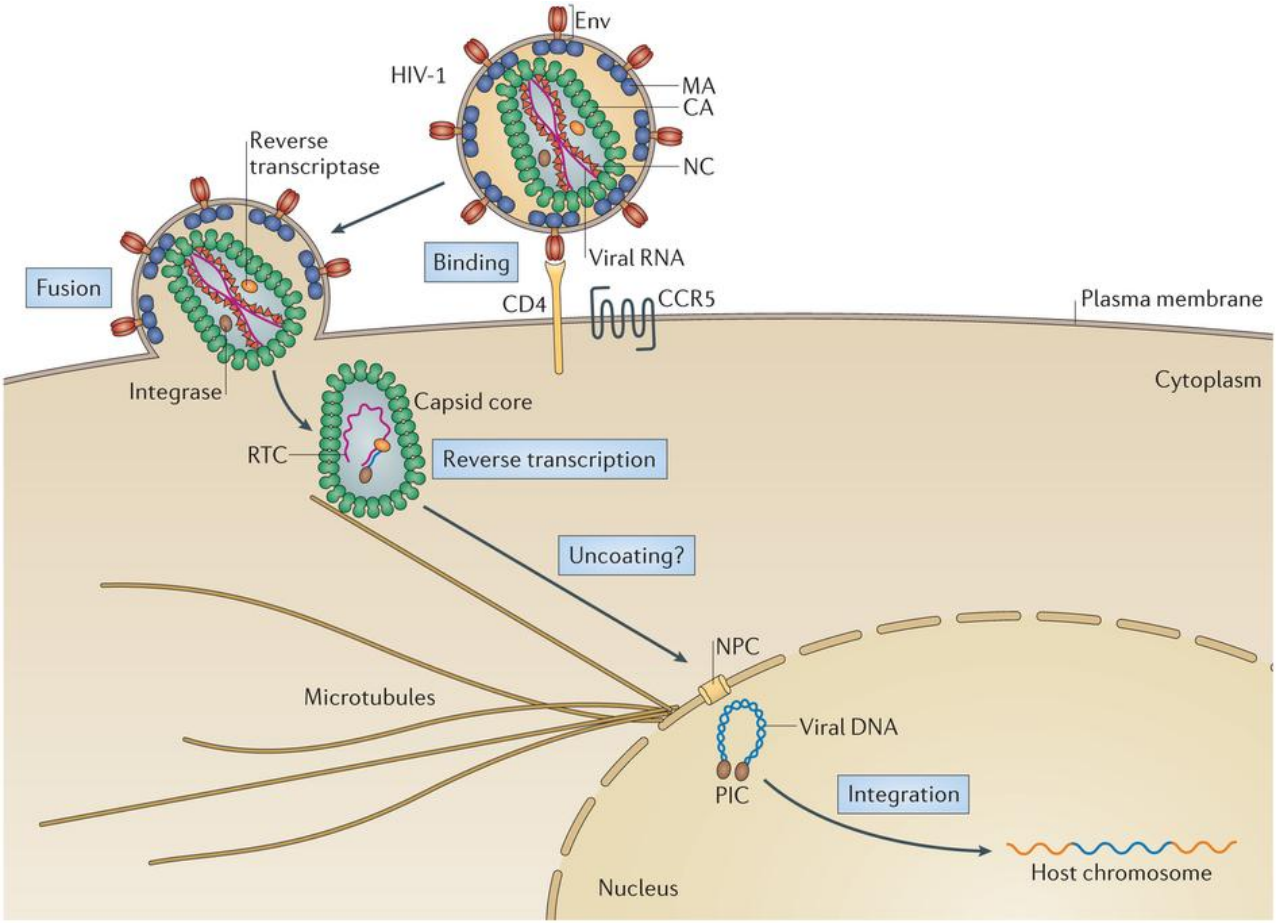


Figure 1.1.6. Early phase of the HIV-1 replication cycle. It includes binding of the viral envelope to cell receptors, fusion of the viral and cellular membranes, reverse transcription of the viral genome, nuclear import and integration of the proviral DNA. Env, envelope glycoprotein; MA, matrix protein; CA, capsid protein ; NC, nucleocapsid protein (Campbell and Hope 2015).

During the late phase of HIV-1 replication (Figure 1.1.7), viral RNAs are synthesized and exported from the nucleus to the cytoplasm where the synthesis of viral proteins occurs. Translation of viral RNAs produces Gag and GagPol polyprotein precursors, Env glycoproteins, as well as regulatory and accessory viral proteins.

Assembly of the HIV-1 particles is initiated by the trafficking of Gag to the site of assembly – the inner leaflet of plasma membrane. During this trafficking, molecules of Gag interact with viral RNA acting as a scaffold for Gag oligomerization (Chen, Nikolaitchik et al. 2009; Briggs and Krausslich 2011; Lu, Heng et al. 2011; Kuzembayeva, Dilley et al. 2014).

The process of Gag oligomerization is a series of homotypic interactions involving CA, MA and SP1 regions (Hill, Worthylake et al. 1996; Datta, Temeselew et al. 2011; de Rocquigny, El Meshri et al. 2014; Datta, Clark et al. 2016). Recently, it has been established that Gag-Gag oligomerization takes place in the cytoplasm and that NC domain plays a key role in this process (El Meshri, Dujardin et al. 2015). As deletion of NC strongly reduces binding of Gag to NAs, its role is likely related to its ability to form with NAs a scaffold for further Gag oligomerization. This scaffolding results in close packaging of Gag proteins that will next bind the plasma membrane. When oligomers progressively assemble, they likely grow in size during the trafficking from the cytoplasm toward the membrane. Other investigations suggest that a relatively small number of Gag molecules may contact gRNA in the cytosol, then move to the membrane, where several thousand additional Gag molecules localize (Jouvenet, Simon et al. 2009; Kutluay and Bieniasz 2010; Lu, Heng et al. 2011).

At the plasma membrane Gag – gRNA oligomers bind to the negatively-charged phosphatidylinositol-(4,5)-biphosphate (PIP_2) phospholipids of the inner leaflet of plasma membrane through its MA domain. This binding is ensured specifically by the myristic acid moiety of Gag that stabilizes Gag-membrane association (Ono, Ablan et al. 2004; Saad, Miller et al. 2006). Following their localization at the plasma membrane, Gag, full-length viral RNA and GagPol assemble into immature viral particles. Before packaging into virions, gRNA undergoes the process of dimerization. Previously, it has been hypothesized that the two copies of viral RNA dimerize in the cytoplasm, then the dimers in complex with Gag move to plasma membrane (D'Souza and Summers 2005; Jouvenet, Simon et al. 2009; Moore, Nikolaitchik et al. 2009; Jouvenet, Laine et al. 2011; Lu, Heng et al. 2011; Keane, Heng et al. 2015). Recently, it has been suggested that gRNA dimerizes at the plasma membrane and Gag plays a major role in this process (Chen, Rahman et al. 2016).

After incorporation of the Env glycoproteins into the assembling particles, the p6 domain of Gag recruits the cellular ESCRT machinery required for particle release. The PTAP domain of p6 binds to the tumor susceptibility gene 101 (GST101) of the ESCRT-I multiprotein complex. The LYPX domain of p6 binds to the ALG2-interacting protein X (ALIX), while the ESCRT-III and vacuolar protein sorting 4 (VPS4) complexes also participate to the virion release.

The immature particles are produced by budding during which a cleavage of Gag and GagPol polyproteins by viral PR occurs (Briggs, Simon et al. 2004; Pettit, Lindquist et al.

2005; Mougel, Houzet et al. 2009). Activation of PR may occur already during the assembly step, as evidenced by the presence of mature CA, MA and NC proteins in the cytoplasmic extracts of infected cells (Mougel, Houzet et al. 2009) and proofs of reverse transcription in budding virions (Trono 1992; Chamontin, Rassam et al. 2015).

Following budding, new virions are released and undergo the maturation step, where CA spontaneously assembles into a cone, which contains the HIV-1 genome, NC, viral replicative enzymes and several accessory proteins. The maturation process leads to the formation of an infectious virus.

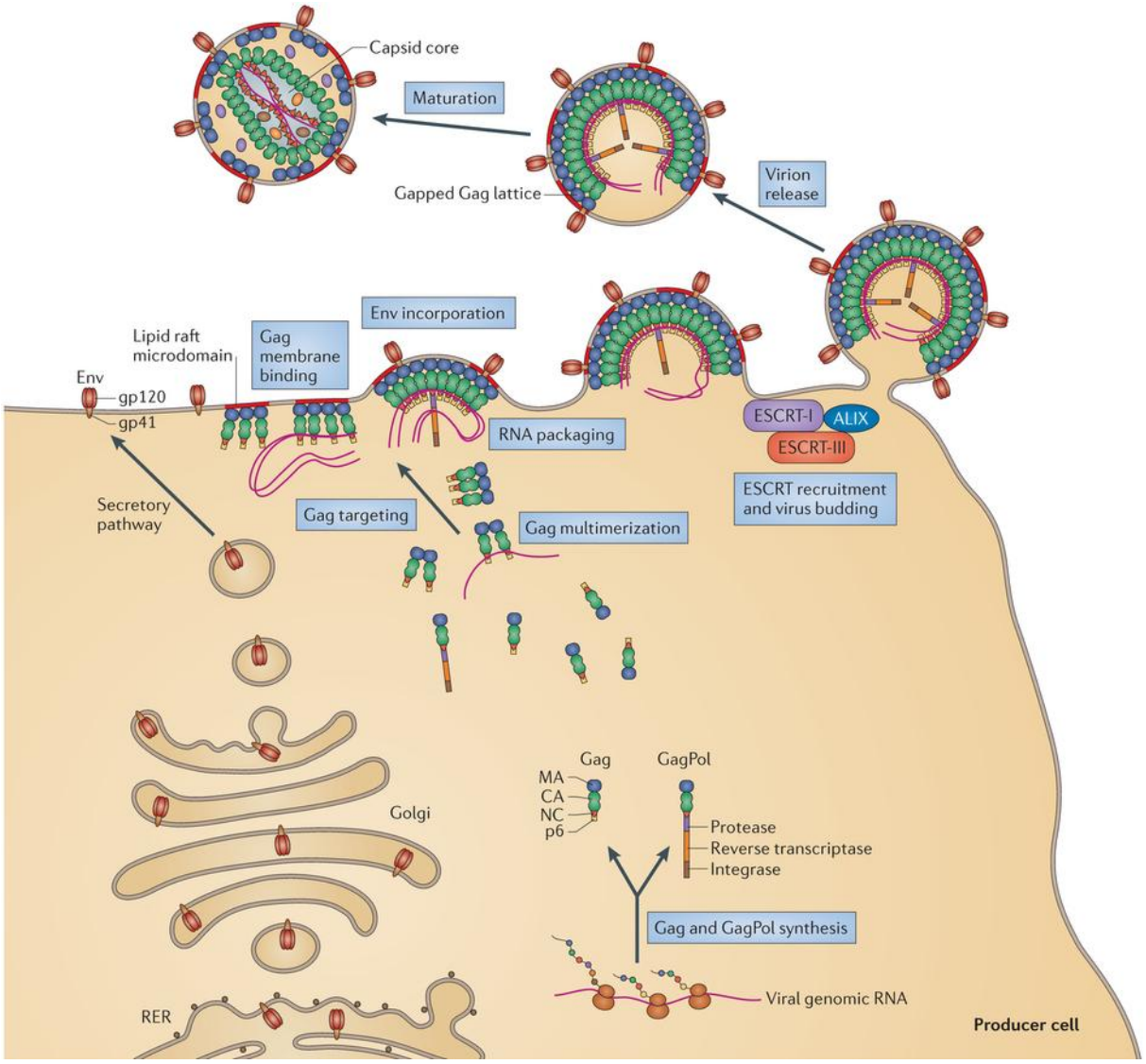


Figure 1.1.7. Late phase of the HIV-1 virus life cycle. It includes synthesis of viral RNAs, their export from the nucleus to the cytoplasm; translation of viral RNAs and production of Gag (that contains matrix MA, capsid CA, nucleocapsid NC and p6 domains), GagPol (that contains MA, CA, NC, PR, RT and IN domains), envelope Env glycoproteins (gp120 and gp41), regulatory and accessory

proteins; trafficking of Gag, GagPol and Env to plasma membrane; assembly of Gag and GagPol at the plasma membrane; encapsidation of viral RNA; incorporation of viral Env glycoproteins; ESCRT recruitment and budding of the new virions; release and virus particle maturation (Freed 2015).

1.1.5. Antiretroviral therapies

The extension of knowledge about the HIV-1 has led to an enormous progress in the development of antiretroviral drugs. Historically the first anti-HIV molecule was azidothymidine (AZT) (Figure 1.1.8), which was developed on the basis of the characterization of the viral RT activity (Broder, Yarchoan et al. 1985). It is a thymidine nucleoside analog acting as a RT inhibitor. Cellular kinases add three phosphate groups to its sugar moiety in order to convert AZT into an active metabolite zidovudine triphosphate. This active metabolite acts as a competitive substrate to dATP, dGTP, dCTP or dTTP and leads to termination of the chain elongation during the synthesis of proviral DNA. AZT was also a first therapeutic approach to prevent mother-to-child transmission of HIV.

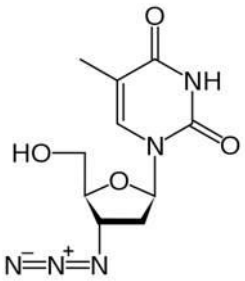
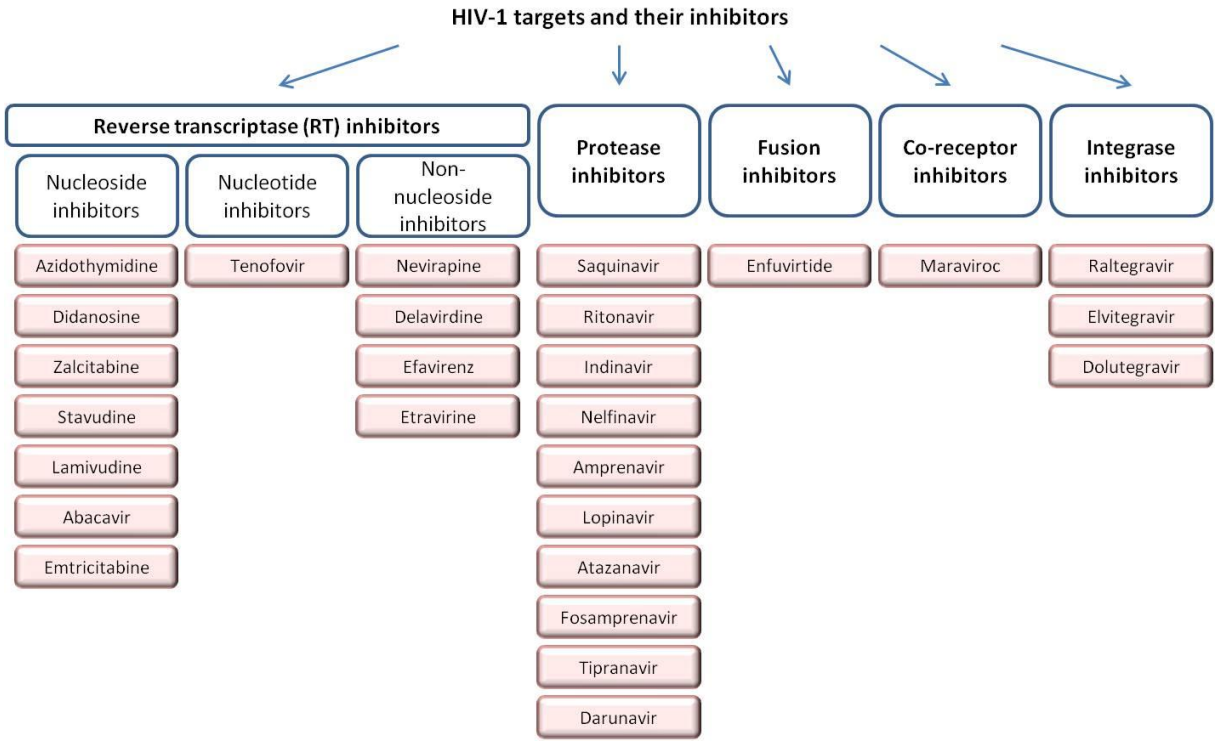


Figure 1.1.8. Structure of azidothymidine (AZT), RT nucleoside inhibitor.

Due to virus mutations, resistance to the monotherapy was quickly observed. This led to the development in the early 90s of the highly active antiretroviral therapy (HAART) based on a combination of several antiviral drugs that target viral enzymes RT, PR and more recently IN. The HAART in the clinic is mostly based on RT and PR inhibitors and in some instances on IN inhibitors. Around 25 anti-HIV drugs have been approved (De Clercq 2010) by the US food and drug administration (FDA). Based on the target, these compounds can be divided into 7 categories: (1) NRTIs (nucleoside reverse transcriptase inhibitors); (2) NtRTIs (nucleotide reverse transcriptase inhibitors); (3) NNRTIs (non-nucleoside reverse transcriptase inhibitors); (4) PIs (protease inhibitors); (5) FIs (fusion inhibitors); (6) CRIs (co-receptor inhibitors); and (7) INIs (integrase inhibitors) (Scheme 1.1.1).



Scheme 1.1.1. Approved HIV-1 inhibitors.

Over the decade, HAART has gradually evolved from drug regiments with more than 20 pills daily in 1996, to 3 pills daily in 2003, to 2 pills daily in 2004 (Truvada®), and finally to 1 pill daily in 2006 (Atripla®) (De Clercq 2007). Atripla is the first anti-HIV drug that contains three active ingredients belonging to three different classes of the HIV inhibitors (NRTIs, NtRTIs and NNRTIs). The INIs represent the most recent advance in the search for effective and selective anti-HIV agents. Combination of several anti-HIV drugs has drastically altered AIDS from an almost uniformly fatal disease to a chronic manageable one, but however does not cure the HIV infection.

Despite the immense benefits of the HAART treatment, many issues still need to be improved. Life-long treatment is associated with emerging drug-resistance, metabolic disorders and cancers. Currently many areas of the HIV investigation require fundamental research. Further insight into the very early stages of HIV infection, the establishment of viral reservoirs and the immune responses is crucial for the development of novel therapeutic and vaccine strategies.

Part 2. Nucleocapsid protein as a potential target for anti-HIV therapy

1.2.1. Structural characteristics of the nucleocapsid protein (NC)

The nucleocapsid protein is a 55 amino acid peptide resulting from the PR-mediated cleavage of the HIV-1 Gag polyprotein. During virus maturation, the initial PR cleavage liberates NCp15, a polyprotein containing NCp7, SP2 and p6 (Shehu-Xhilaga, Kraeusslich et al. 2001) (Figure 1.2.1). The subsequent cleavage results in NCp9 (71 amino acids) containing NCp7 and SP2. In fully mature viral particles, NCp7 (55 amino acids) and NCp9 forms are found in freshly collected highly infectious viral particles (Henderson, Bowers et al. 1992; Thomas and Gorelick 2008; Muriaux and Darlix 2010).

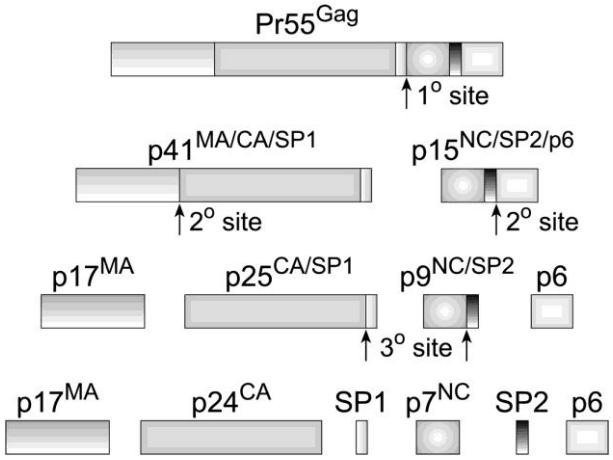


Figure1. 2.1. Proteolytic processing of HIV-1 Gag by PR. During proteolysis, NC exists in two intermediates: NCp15 and NCp9 further cleaved into the final form NCp7 (Thomas and Gorelick 2008).

NCp7 (NC) is a structural protein of HIV-1 which coats gRNA in its dimeric form. One of the functions of NC is to protect gRNA from nucleases through its NA binding and condensing properties (Tanchou, Gabus et al. 1995; Krishnamoorthy, Roques et al. 2003; Darlix, Godet et al. 2011). NC is characterized by two highly conserved CCHC zinc fingers (ZFs) connected by a short linker rich in basic residues and two unfolded N- and C-terminal parts (Figure 1.2.2A). The CCHC motifs and the basic linker are invariant in diverse HIV-1 subtypes and drug-resistant viruses (Figure 1.2.3). Sequence variations are mainly observed in residues of the N-terminal domain while many residues of the ZFs, and notably the Phe16 in ZF1 and Trp37 in ZF2, important for NA recognition and chaperone properties of the HIV-1 NC are highly invariant (Darlix, Godet et al. 2011).

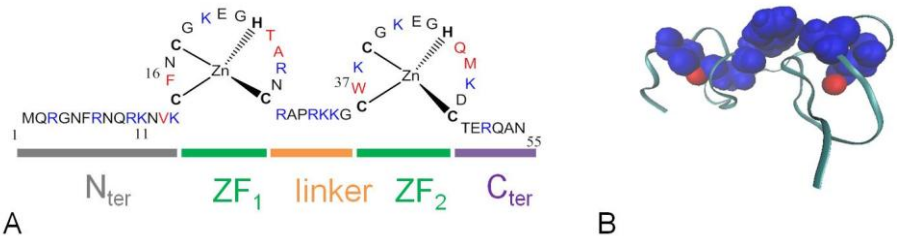


Figure 1.2.2. Sequence and 3D structure of NC protein of HIV-1. (A) Sequence of the HIV-1 nucleocapsid protein and its structural components. Positively charged amino acids are in blue, amino acids of the hydrophobic plateau are in red, CCHC motifs are in bold. (B) 3D structure of the hydrophobic plateau at the top of NC zinc finger domain, zinc ions are in red.

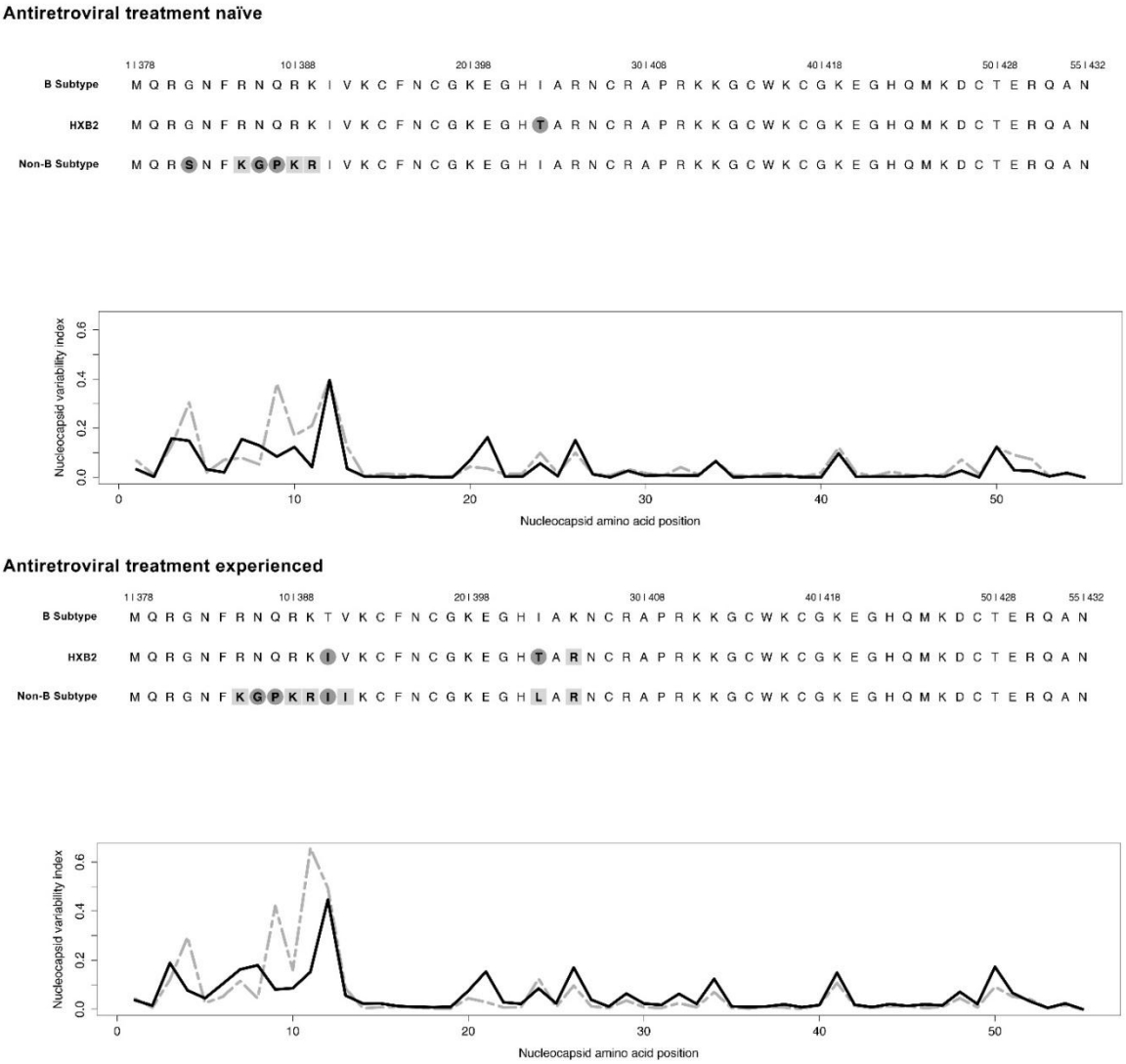


Figure 1.2.3. Sequence of NC amino acids in different HIV-1 subtypes as well as in viral isolates obtained from antiretroviral naïve (top panel) and treated (bottom panel) individuals. Top panel: NC sequences from B (594 sequences) and non-B subtypes (4938 sequences) as well as the HXB2

molecular clone, which is often considered as a representative subtype B. Bottom panel: NC sequences from B (7351 sequences), non-B (14286 sequences), and HXB2 subtypes. Grey circles on amino acids indicate non conservative amino acids, whereas grey blocks indicate conservative amino acids. Nucleocapsid variability index graphs show the variability at each position of NC (higher value corresponds to higher amino acid variability). Black lines represent the sequences of the B subtype, whereas the grey dashed lines are the non-B subtypes (Mori, Kovalenko et al. 2015).

The central ZF domain of the NC protein is folded into a tight structure, while flanking N- and C-terminal parts are flexible and independent from the central part. Functionally, the basic N-terminal domain is mainly responsible for the NA aggregation activity of NC (Stoylov, Vuilleumier et al. 1997). In many biophysical studies, the NC(11-55) sequence, lacking the N-terminal part, is used in order to limit aggregation of the NAs. Each ZF coordinates one zinc ion through three Cys and a His residue with high affinity of $10^{13} – 10^{14}$ M^{-1} (Berg 1986; Mely, Cornille et al. 1991; Mely, de Rocquigny et al. 1996). The two zinc fingers are spatially close and show similar folding patterns (Morellet, Jullian et al. 1992; Summers, Henderson et al. 1992; Morellet, de Rocquigny et al. 1994). However, they are not functionally equivalent (Fisher, Rein et al. 1998; Guo, Wu et al. 2000; Guo, Wu et al. 2002; Fisher, Fivash et al. 2006; Zargarian, Tisne et al. 2014), and the substitution of either of the ZFs by an identical counterpart leads to a loss of the NC functions (Gorelick, Chabot et al. 1993). The 7 amino acid-long linker RAPRKKG located downstream of the ZF_1 is responsible for the proximity of the ZFs and for their transient globular structure (Mely, Jullian et al. 1994; Morellet, de Rocquigny et al. 1994; Ottmann, Gabus et al. 1995; Lee, De Guzman et al. 1998). This relative orientation is additionally stabilized by the hydrophobic and aromatic residues of the proximal (Val13, Phe16, Thr24, Ala25) and distal (Trp37, Gln45, Met46) ZFs forming a hydrophobic plateau (Figure 1.2.2B). This hydrophobic plateau is of key importance for NC functions, since single point mutations in this plateau induce a loss of viral infectivity (Demene, Dong et al. 1994; Wu, Mitra et al. 2013). Similarly, mutations of the zinc-binding CCHC residues also lead to a loss of viral infectivity through preventing the appropriate folding of the ZFs and, thus, hindering the formation of the hydrophobic plateau (Aldovini and Young 1990; Gorelick, Nigida et al. 1990; Dorfman, Luban et al. 1993; Stote, Kellenberger et al. 2004). Folded ZFs and the hydrophobic plateau at their top play a critical role in recognition, destabilization, binding and dynamic rearrangements of NAs (Tisne, Roques et al. 2003; Levin, Guo et al. 2005; Cruceanu, Urbaneja et al. 2006; Darlix, Garrido et al. 2007; Thomas and Gorelick 2008; Levin, Mitra et al. 2010; Darlix, Godet et al. 2011).

1.2.2. Binding of NC to nucleic acids

The role of NC in the HIV-1 virus life cycle is mainly associated with interactions with NAs. NC can bind in a non-specific or specific manner to nearly any NA sequence of 5-8 nt in length (Darlix, Godet et al. 2011). However, the affinity of such binding may vary by several orders of magnitude and depends on the nature, sequence and folding of the NAs. NC protein binds to both RNA and DNA sequences with a preference for single-stranded oligonucleotides (ODNs) (Mirambeau, Lyonnais et al. 2006; Darlix, Garrido et al. 2007). The affinity of NC for RNAs follows the order: retroviral RNA > mRNA > rRNA> poly (rA) (Karpel, Henderson et al. 1987; Coffin 1995). The relative affinity of NC for retroviral RNA, DNA and ODNs follows the order: retroviral RNA > ssDNA, dsDNA > ODNs (Darlix, Lapadat-Tapolsky et al. 1995). Non-specific and low-affinity binding of NC to NAs is the dominant binding mode that occurs through electrostatic interactions between the positively-charged NC and negatively-charged NAs. At saturating concentrations, NAs are coated by NC molecules resulting in their protection from the nuclease degradation (Lapadat-Tapolsky, de Rocquigny et al. 1993; Krishnamoorthy, Roques et al. 2003).

Studies performed with RNA and DNA ODNs reveal that NC binds with high affinity to sequences containing unpaired guanine residues like TG, GXG and TXG (Vuilleumier, Bombarda et al. 1999; Avilov, Piemont et al. 2008; Avilov, Godet et al. 2009). This efficient binding is ensured by the interactions of the hydrophobic plateau of NC with the nucleic bases and NA phosphate backbone (de Guzman, Wu et al. 1998; Morellet, Déméné et al. 1998; Amarasinghe, De Guzman et al. 2000; Bourbigot, Ramalanjaona et al. 2008; Spriggs, Garyu et al. 2008; Bazzi, Zargarian et al. 2011). Within the hydrophobic plateau, the highly conserved aromatic Trp37 residue plays a critical role through its π-stacking interactions with the nucleobases, notably with guanosines as was shown by 3D structures of NC/ODN complexes (de Guzman, Wu et al. 1998; Amarasinghe, De Guzman et al. 2000; Bourbigot, Ramalanjaona et al. 2008; Bazzi, Zargarian et al. 2011). Careful examination of these 3D structures revealed that each ZF is specialized in one function. ZF_2 is responsible for binding of the accessible and flexible guanines, while ZF_1 either binds a second G (in case of GXG sequence) or destabilizes the stem through its hydrophobic residues (in case of TG or TXG sequences) (Zargarian, Tisne et al. 2014) (Figure 1.2.4).

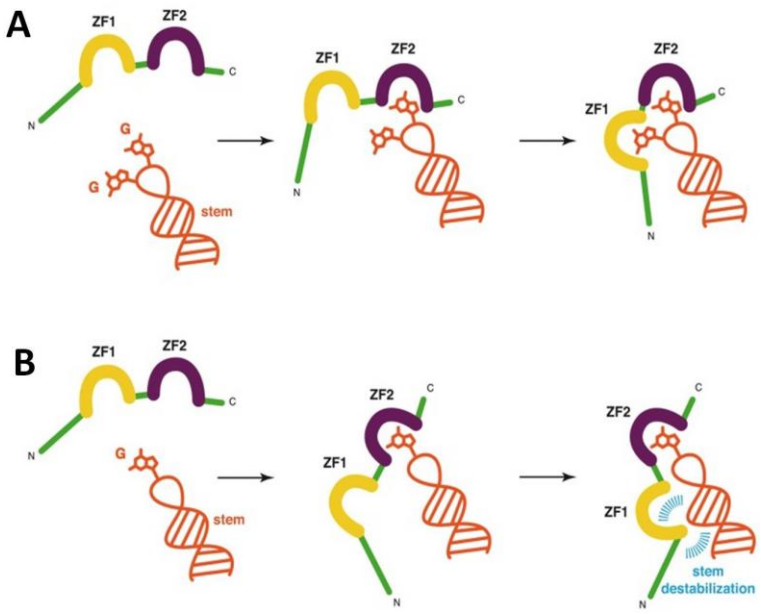


Figure 1.2.4. Schematic representation of the task specialization of each ZF upon binding to NAs. The NA fragment is in orange, NC zinc fingers are in yellow (ZF_1) and purple (ZF_2), and the rest of the protein is in green. (A) Presence of two accessible guanines. ZF_2 binds first to an accessible guanine, while ZF_1 binds the remaining guanine. The stem is not destabilized. (B) Presence of only one accessible guanine. ZF_2 binds first to this accessible guanine, while ZF_1 remains free to contact the stem via its hydrophobic platform, leading to the destabilization of the stem (in blue). Adapted from Zargarian, Tisne et al. 2014.

In addition, it was shown that NC binds to DNA and RNA sequences with opposite polarities (Bazzi, Zargarian et al. 2011). NC complexes with DNA sequences like ACGCC pentanucleotide (Morellet, Déméné et al. 1998) and ΔP(–)PBS (Bourbigot, Ramalanjaona et al. 2008), and RNA sequences like SL2 (Amarasinghe, De Guzman et al. 2000) and SL3 (de Guzman, Wu et al. 1998) show two distinct modes of NC binding. In complexes with DNAs, ZF_2 interacts with a G residue of the ODN sequence while ZF_1 contacts the residue (C or T) upstream to G (Figure 1.2.5). In complexes with RNAs, ZF_1 and ZF_2 interact with two G residues separated by a third residue (GXG). The Trp37 residue of ZF_2 stacks with the G residue located at the 3′ end of the GXG sequence. This different binding polarity is likely due the sugar moieties. The riboses possess an additional –OH group at the C2′ position as compared to deoxyriboses, which creates additional hydrophobic contacts with NC. The different binding polarities of NC to RNAs and DNAs may play a role during reverse transcription and NA recombination, since a similar behavior has been observed for a family of proteins that possess an oligosaccharide/ODN binding fold and are involved in DNA

recombination (Yang, Jeffrey et al. 2002; Bochkarev and Bochkareva 2004; Bazzi, Zargarian et al. 2011).

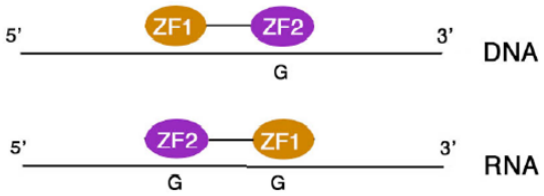


Figure 1.2.5. Schematic illustration of the two modes of NC binding to DNA and RNA sequences.

The specific binding properties of NC play a critical role in gRNA recognition and the assembly step of the viral life cycle. NC as a domain of the Gag polyprotein recognizes the ψ encapsidation signal of gRNA among a large excess of non-viral mRNAs and spliced viral mRNAs (Aldovini and Young 1990; Cimarelli and Darlix 2002; Muriaux, Darlix et al. 2004; Muriaux and Darlix 2010). This specific recognition is required for the preferential incorporation of the viral gRNA in virions (Berkowitz and Goff 1994; Clever, Sassetti et al. 1995; McBride and Panganiban 1997; de Guzman, Wu et al. 1998; Amarasinghe, de Guzman et al. 2000; Amarasinghe, Zhou et al. 2001; Mark-Danieli, Laham et al. 2005). The ψ encapsidation signal (Figure 1.1.3) is a ~120 nucleotide-long sequence composed of four stem-loops (SL1-SL4) located near the 5′ end of the genome (Lever, Gottlinger et al. 1989; Aldovini and Young 1990; Clavel and Orenstein 1990). Among these four stem-loops, SL3 is the most highly conserved sequence of the ψ RNA site (Hayashi, Ueno et al. 1993). SL2 and SL3 are thought to be the primary determinants for the interaction with NC as they show high binding affinities to NC (de Guzman, Wu et al. 1998; Amarasinghe, De Guzman et al. 2000). According to the 3D structures of NC/SL2 and NC/SL3 complexes, the protein binds preferentially to the GXG motif of SL2 and SL3 loops. The residues of the hydrophobic plateau, notably Val13, Phe16, Thr24 (Ile24), and Ala25 interact with the 10-UG-11 residues of the 8-GGUG-11 loop of SL2, and with the 8-AG-9 residues of the 6-GGAG-9 loop of SL3 (Figure 1.2.6). The G9 residue in SL2 and the G7 residue in SL3 are involved in stacking and hydrophobic interactions with the Trp37, Gln45 and Met46 residues of the hydrophobic plateau. This mode of binding corresponds to the situation with two accessible guanines where ZF_2 binds firstly to an accessible guanine, and ZF_1 binds to the remaining guanine (Figure 1.2.4A). The stems of SL2/SL3 were not shown to be destabilized. The binding constants of NC(1-55) to SL2 and SL3 were found to be similar (K_d = 110 (±50) nM and 170 (±65) nM, respectively). This high affinity contributes to their selective recognition by NC

during assembly stage of HIV-1 life cycle. Moreover, the interaction of NC with multiple stem-loop packaging elements further illustrates the ability of NC to bind various NA targets (Darlix, Godet et al. 2011; Godet, Kenfack et al. 2013).

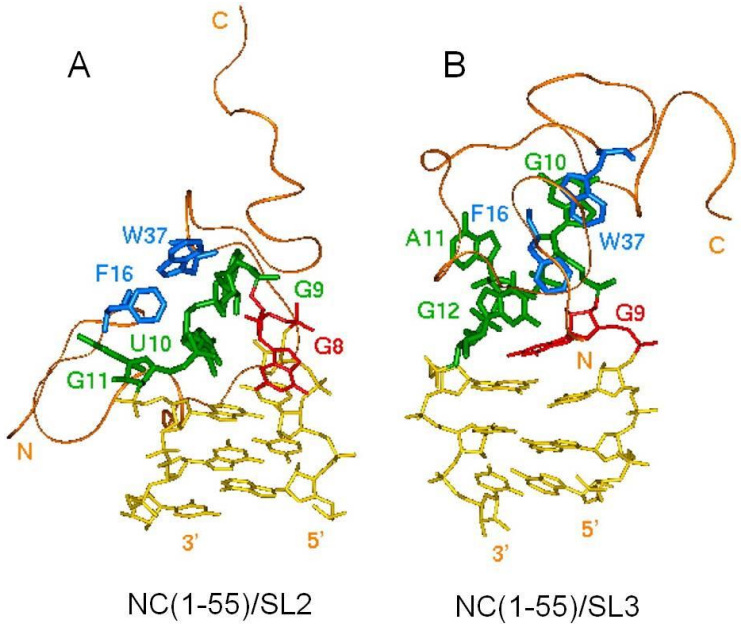


Figure. 1.2.6. Ribbon representation of the (A) NC(1-55)/SL2 complex and (B) NC(1-55)/SL3 complex. Phe16 and Trp37 are in blue. The ODNs are in yellow. Nucleobases involved in the interaction are in green: 9-GUG-11 for SL2 and 10-GAG-12 for SL3. The remaining bases from the loops are in red. Adapted from Bourbigot, Ramalanjaona et al. 2008.

1.2.3. Chaperone properties of NC

The NA molecules (mainly RNAs) can adopt a large number of conformations while only one is functional. The inactive conformations of NAs can be very stable and persistent. This key biological issue appears to be solved by a class of proteins called NA chaperones (Herschlag 1995). NA chaperones are non-specific NA binding proteins that rescue NAs trapped in unproductive folding states (Cristofari and Darlix 2002; Schroeder, Barta et al. 2004; Zuniga, Sola et al. 2009). NC is a member of this abundant class of proteins due to its ability to promote the rearrangement of NAs into their most stable conformations (Rein, Henderson et al. 1998; Godet and Mely 2010; Levin, Mitra et al. 2010). The chaperone activity of NC relies on three main components. The first one consists in the destabilization of the NA secondary structure mainly driven by the hydrophobic plateau at the top of the folded ZFs of NC (Bernacchi, Stoylov et al. 2002; Azoulay, Clamme et al. 2003; Egele, Schaub et al. 2004; Beltz, Clauss et al. 2005; Cosa, Zeng et al. 2006; Levin, Mitra et al. 2010; Wu, Rouzina et al. 2010). The destabilizing activity of NC strongly depends on the NA stability and

structure. The presence of destabilizing motifs along the NA sequence like bulges, mismatches, internal loops and terminal base pairs favors the destabilization and likely ensures a selective recognition of a limited number of NA sequences by NC during the viral life cycle (Godet and Mely 2010). The destabilization is further accompanied by an exposure and restricted mobility of the bases of the NA sequences bound to NC (Avilov, Piemont et al. 2008; Bourbigot, Ramalanjaona et al. 2008; Godet, Ramalanjaona et al. 2011; Godet, Kenfack et al. 2013). This is thought to be critical for the annealing of the complementary sequences. The second component of the NC chaperone activity consists in the promotion of the annealing of complementary sequences (Darlix, Vincent et al. 1993; You and McHenry 1994; Hargittai, Gorelick et al. 2004; Godet, de Rocquigny et al. 2006; Liu, Zeng et al. 2007; Ramalanjaona, de Rocquigny et al. 2007; Vo, Barany et al. 2009). Here, the N-terminal domain of NC rich in basic residues possesses the NA aggregation ability and is, thus, actively involved in the annealing process. The hydrophobic plateau and the two ZFs also play a role in the annealing process through activation of specific pathways (like loop-loop kissing complexes). These specific pathways are required to promote faithfully, timely and specifically the hybridization of the two complementary sequences (Godet, Ramalanjaona et al. 2011; Godet, Kenfack et al. 2013).

The third component of the NC chaperone activity consists in the fast binding and dissociation from NAs. Indirect evidence of this property has been demonstrated using single molecule stretching experiments to explore the chaperone properties of NC (Cruceanu, Gorelick et al. 2006). The efficiency in strand annealing was limited when the retroviral NC performed slow on/off binding kinetics, though it may still be capable to bind NAs with high affinity, destabilize duplexes and aggregate NAs. For example, HIV-1 NC exhibits fast on/off kinetics and is, thus, one of the most efficient NC chaperones in comparison with NCs from the Rous sarcoma virus, Moloney murine leukemia virus or HTLV-1 (Stewart-Maynard, Cruceanu et al. 2008). In contrast, the HTLV-1 NC performs poor chaperone properties, despite its strong duplex destabilizing activity, mainly due to very slow NA dissociation kinetics. The on/off binding kinetics rate is found dependent on the ZFs of NC, since mutations in the ZF domains lead to a decrease in the protein dissociation rate (Cruceanu, Gorelick et al. 2006).

Taken together, in the HIV-1 virus life cycle the chaperone activity of NC is indispensable during viral replication and assembly steps. In the present work we focused mainly on the viral replication step, so that this process will be discussed in details below.

1.2.4. Role of NC in HIV-1 reverse transcription

Reverse transcription process

The reverse transcription step of the HIV-1 is a complex process resulting in synthesis of a double-stranded linear proviral DNA from the single-stranded viral RNA (Arts and Wainberg 1996; Telesnitsky and Goff 1997). This process is controlled by RT enzyme and involves two obligatory strand transfers (Basu, Song et al. 2008). Schematically, it may be divided into the following steps (Figure 1.2.7).

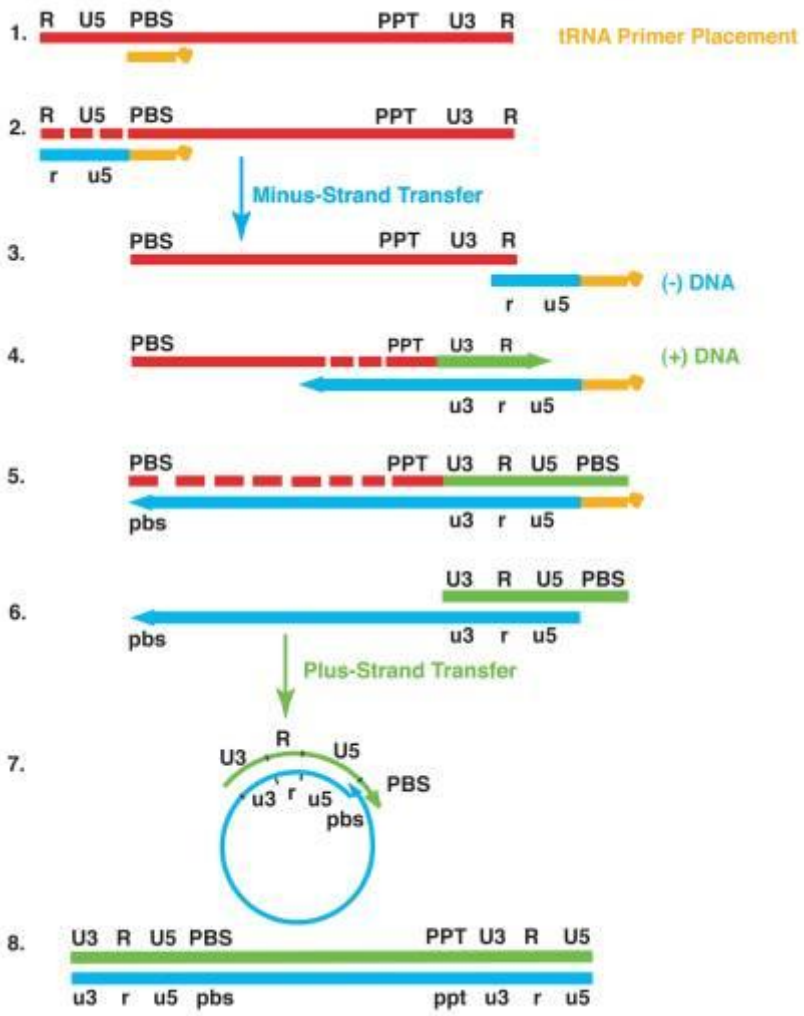


Figure 1.2.7. Schematic representation of HIV-1 reverse transcription process (Levin, Mitra et al. 2010).

Step 1. Initiation of reverse transcription occurs when 18-nt of the cellular $tRNA^{Lys3}$ primer (orange line) (Figure 1.2.7) hybridize to the 18-nt sequence PBS (red solid line) located at the 5′ end of the genome. This reaction is called the *tRNA primer placement*.

Step 2. RT enzyme promotes extension of the primer in order to synthesize the minus strong-stop DNA [(−)ssDNA] (blue solid line) that contains copies of the unique 5′ genomic sequence (U5) and repeat R regions. As the primer is extended, the RNase H activity of RT cleaves gRNA (fragments represented by short red solid lines) that has already been read.

Step 3. To continue the minus-strand, the (−)ssDNA is transferred to the 3′ end of gRNA where annealing of the complementary R sequences at the 3′ ends of gRNA and (−)ssDNA occurs. This step is called *minus-strand transfer*.

Step 4. Following minus-strand transfer, elongation of (−)ssDNA proceeds to reach the PBS sequence of the 5′ end, while RNase H degradation continue. The polypurine tracts PPT and cPPT located near the U3 are resistant to RNase H activity and initiate the synthesis of the plus-strand DNA [(+)ssDNA] towards the 3′ end of gRNA using newly synthesized (−)ssDNA as a template.

Steps 5 and 6. When the PBS region of the plus strand is reconstituted, the formed sequence is called (+)ss DNA. RNase H removes the tRNA and PPT primers from minus- and plus-strand DNAs, respectively. A *plus-strand transfer* is required to complete the synthesis of viral DNA.

Step 7. Annealing of the two complementary PBS sequences at 3′ ends of (+)ssDNA and minus-strand DNA results in a circular DNA intermediate.

Step 8. Complete minus- and plus-strand DNAs are elongated into a linear double-stranded DNA with an LTR at each end. This proviral DNA is used for further integration step.

The NC protein assists RT during the whole reverse transcription process (Levin, Guo et al. 2005; Thomas and Gorelick 2008; Levin, Mitra et al. 2010; Darlix, Godet et al. 2011). It directs annealing of the $tRNA^{Lys3}$ primer with the PBS, chaperones both strand transfers, ensures fidelity of the plus-strand DNA priming at two polypurine tracts (Hergott, Mitra et al. 2013), and facilitates removal of $tRNA^{Lys3}$ primer from the 5′ end of the minus strand DNA (Levin 2010). Also NC assists RT in the fidelity of vDNA synthesis by providing nucleotide excision repair (Bampi, Bibillo et al. 2006).

tRNA^{Lys3} – PBS interaction

The HIV-1 reverse transcription initiation consisting in base-pairing of tRNA^{Lys3} to PBS RNA requires NC chaperone activity (Levin, Mitra et al. 2010; Darlix, Godet et al. 2011; Sleiman, Goldschmidt et al. 2012). NMR studies revealed that both basic and ZF domains are involved in the formation of the tRNA-RNA complex (Tisne, Roques et al. 2004; Barraud, Gaudin et al. 2007). The role of basic residues consists in destabilization of a few base pairs of the tRNA, while the ZFs affect the ternary interactions within the tRNA molecule. As a precondition for reverse transcription, tRNA primer, viral RNA template and RT form a productive ternary ribonucleoprotein complex under the action of NC.

Interestingly, exploration of the tRNA^{Lys3} specificity as a primer of HIV-1 identified an 8-nucleotide sequence (nucleotides 123-130, Figure 1.2.8) in the U5 region of viral RNA complementary to 48-55 nucleotides of tRNA^{Lys} primer (Beerens, Groot et al. 2001; Beerens and Berkhout 2002). This sequence of the viral RNA is called the primer activation signal (PAS) and is involved in annealing with the so-called anti-PAS region of tRNA^{Lys3}. Using single- molecule FRET assays it was shown that this interaction appears to be dynamic and stimulated by NC (Beerens, Jepsen et al. 2013). Therefore, the PAS/anti-PAS interaction is an additional illustration of NC chaperone activity (Sleiman, Barraud et al. 2013).

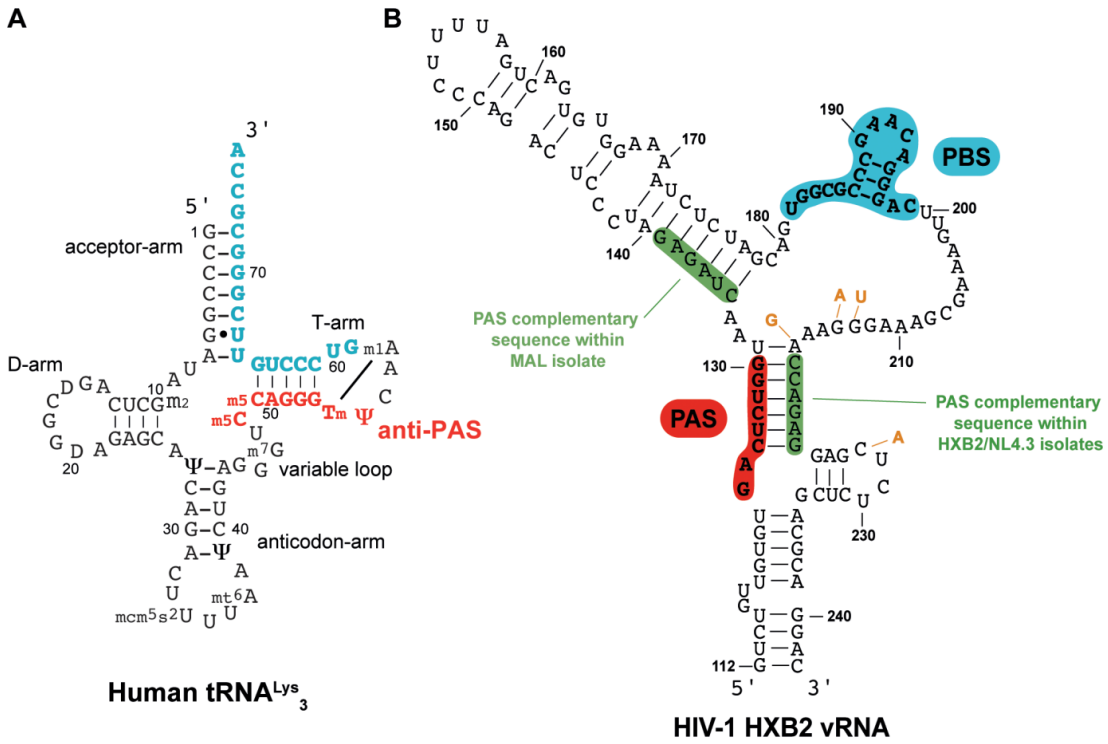


Figure 1.2.8. Secondary structures of (A) the human tRNA^{Lys3}, in blue is the sequence complementary to the PBS and in red is the anti-PAS sequence; (B) PBS domain within the HIV-1 HXB2 isolate, in

blue is the PBS sequence, in red is the PAS sequence, in green are the sequences complementary to the PAS sequence, the nucleotides that are different between HXB2 and NL4.3 are indicated in orange. Adapted from Sleiman, Barraud et al. 2013.

Minus-strand transfer

The minus-strand transfer is mediated by base pairing of the R region at the 3′ end of gRNA with the complementary r region at the 3′ end of (−)ssDNA. The R region consists of the transactivation response element TAR and a portion of the polyadenylation signal (polyA) (Feng and Holland 1988; Berkhout, Klaver et al. 1995), while the r region contains the complementary cTAR and cpolyA (Berkhout, Vastenhouw et al. 2001; Bernacchi, Stoylov et al. 2002). TAR and cTAR sequences (Figure 1.2.9) are imperfect stem-loops with short double-stranded segments separated by bulges, mismatches and internal loops (Baudin, Marquet et al. 1993). Their annealing in the absence of NC is a very slow process (You and McHenry 1994; Godet, de Rocquigny et al. 2006; Vo, Barany et al. 2006; Vo, Barany et al. 2009) likely due to the high stability of the stem-loops.

Addition of NC accelerates the rate of TAR/cTAR hybridization up to 3000-fold (You and McHenry 1994; Guo, Wu et al. 2000; Vo, Barany et al. 2006; Vo, Barany et al. 2009), as a result of its NA chaperone properties. NC binds TAR and cTAR stem-loops and destabilizes their secondary structure. It was found that the destabilization is much more pronounced with cTAR than with TAR, due to the higher stability of the TAR RNA structure (Bernacchi, Stoylov et al. 2002). Upon binding to cTAR, NC destabilizes the base pairs of the lower part of the stem, which results in the melting of the lower half of the stem (Bernacchi, Stoylov et al. 2002; Azoulay, Clamme et al. 2003; Cosa, Harbron et al. 2004; Liu, Cosa et al. 2005; Cosa, Zeng et al. 2006; Godet, Kenfack et al. 2013). The ZFs and hydrophobic plateau of NC were reported to be responsible for this cTAR destabilization (Beltz, Clauss et al. 2005). Substitution of several cTAR residues with the fluorescent adenosine analog 2-aminopurine revealed that G_{10} and G_{50} residues (Lai strain) directly interact with NC (Figure 1.2.9), likely through a stacking interaction with its Trp37 residue (Godet, Kenfack et al. 2013). This interaction with NC results in the solvent-exposure of the nucleosides of the lower half of cTAR stem, making them competent for further annealing with the complementary TAR RNA sequence.

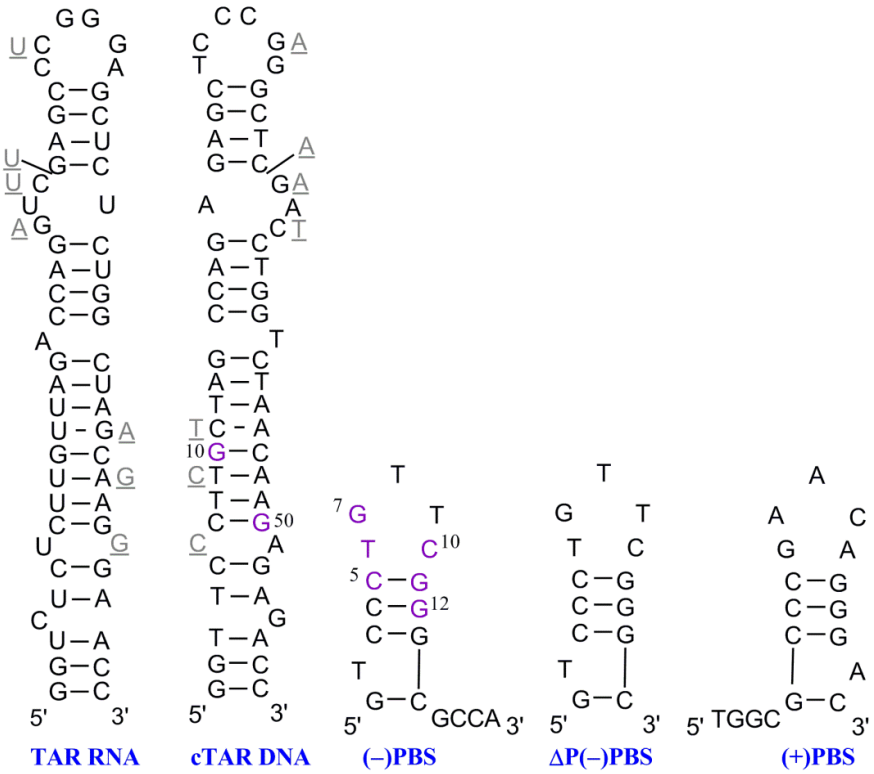


Figure 1.2.9. Structures of TAR, cTAR, (–)PBS, (+)PBS, and ΔP(–)PBS sequences. TAR RNA and cTAR DNA sequences are shown with variations observed in the HIV-1 MAL strain (black letters) or Lai strain (grey letters). Their secondary structures were predicted from the structure of TAR (Baudin, Marquet et al. 1993) using the *m*-fold program. In violet: G_{10} and G_{50} residues of cTAR, and the two preferential binding sites 5-CTG-7 and 10-CGG-12 of (–)PBS involved in NC(11-55) interaction.

Using the 55 nt TAR/cTAR sequences from the Mal strain (Godet, de Rocquigny et al. 2006), or the 59 nt sequences from the Lai strain (Vo, Barany et al. 2009), it was further demonstrated that the annealing reaction follows a second-order kinetics. The mechanism of reaction involves two steps. The first bimolecular step leads to the formation of an intermediate complex that is further converted into the final extended duplex (Figure 1.2.10). Depending on the NA substrates and solution conditions, the annealing reaction can proceed through a loop-loop interaction pathway (Godet, de Rocquigny et al. 2006; Berkhout, Vastenhouw et al. 2001; Kanevsky, Chaminade et al. 2005; Vo, Barany et al. 2006) or a zipper pathway through the 3′/5′ ends (Godet, de Rocquigny et al. 2006; Vo, Barany et al. 2009; Liu, Zeng et al. 2007; Zeng, Liu et al. 2007; Chen, Maskri et al. 2015). NC is able to switch the annealing mechanism from the kissing to zipper pathway depending on the NC/ODN ratio. Indeed, in the absence or at low NC/ODN ratio, the hybridization proceeds through a loop-loop intermediate, while at high ratios, the two sequences hybridize through their stem ends. Thus, the mechanism of TAR/cTAR annealing depends on the local stability

of the NA sequences and the NC destabilization activity, and appears to be both sequence and protein-concentration dependent.

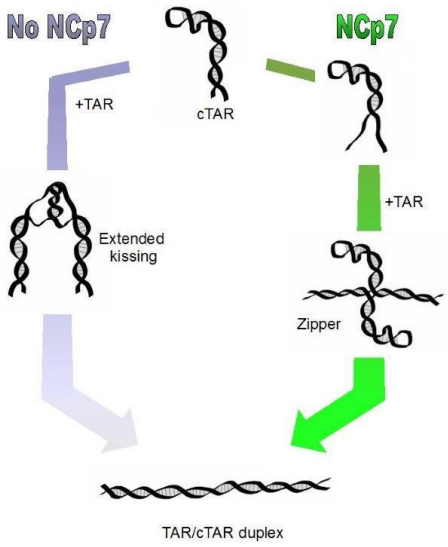


Figure 1.2.10. Schematic representation of the TAR/cTAR annealing pathways in the absence and in the presence of NC. In the absence of NC or at low NC/ODN ratios, the annealing reaction is mainly driven through fast formation of the kissing intermediate complex, which is further converted into the extended duplex. NC destabilizes cTAR by melting the lower part of the stem. Then, the annealing reaction is promoted through the stem ends, through a zipper pathway (Godet and Mely 2010).

Plus-strand transfer

The synthesis of the full-length copy of the viral (+)DNA requires the plus-strand transfer mediated by annealing of the complementary 18 nt (–)PBS and (+)PBS sequences at the 3′ ends of the minus-strand DNA and (+)ssDNA, respectively. The (–)PBS and (+)PBS sequences are folded into short stable stem-loops with terminal single-stranded overhangs (Figure 1.2.9). Similarly to minus-strand transfer, this process is chaperoned by NC. In the absence of NC, the bases of (–)PBS and (+)PBS loops are oriented towards the interior of the loop, which is not favorable for annealing. Therefore, the hybridization of (–)PBS and (+)PBS proceeds spontaneously mainly through the flexible single-stranded overhangs, and possibly through the bulged bases at the bottom of the stems (Figure 1.2.11) (Egele, Schaub et al. 2004; Ramalanjaona, de Rocquigny et al. 2007; Godet, Ramalanjaona et al. 2011).

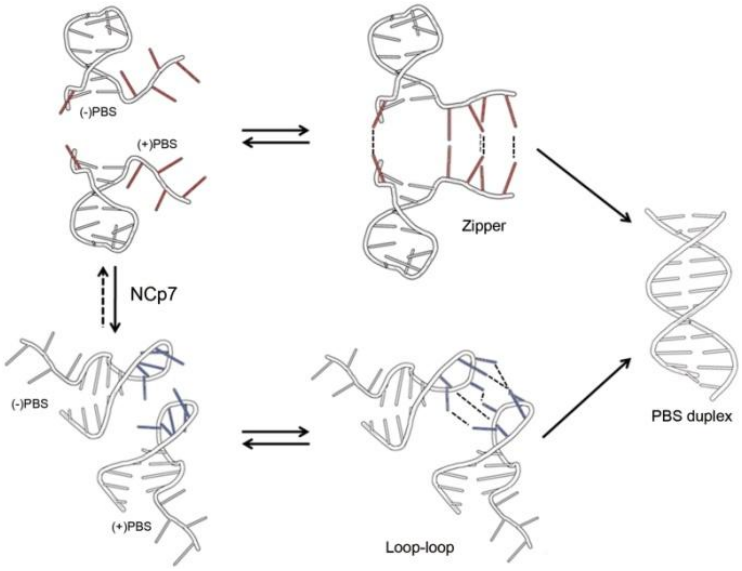


Figure 1.2.11. Proposed pathways for (–)/(+)PBS annealing in the absence (upper) and in the presence (bottom) of NC. In the absence of NC, the annealing is nucleated through the single-stranded overhangs and possibly, through the exposed bulge at the bottom of the stem, presenting a zipper mechanism. In the presence of NC, the loop is stretched and bases are exposed to the solvent, favoring the loop-loop kissing mechanism and thus, formation of the extended duplex (Godet, Ramalanjaona et al. 2011).

NC binds with rather high affinity at the level of the loop and the overhang of the PBS sequence (Egele, Schaub et al. 2004). NMR experiments further evidenced the structural changes induced by NC on ΔP(–)PBS DNA, a truncated version of (–)PBS lacking the 3′-overhang. The 5-CTG-7 region of ΔP(–)PBS (Figure 1.2.12) was identified as the preferential binding site of NC with an affinity differing by a factor of 5 to 10 in comparison to the 10-CCG-12 sequence, the second binding site (K_b = 1.2 (±)0.4×10^7 M^{-1} and 2.5 (±0.5)×10^6 M^{-1}, respectively) (Johnson, Turner et al. 2000; Bourbigot, Ramalanjaona et al. 2008). According to obtained 3D structures, binding of NC induces an exposure of the T6 and G7 residues to the exterior of the loop (Figure 1.2.12), as well as changes in the orientation of C10 residue from almost stacked with G11 in ΔP(–)PBS free state to almost perpendicular to G11 in its bound form. NC also makes T8 and T9 residues more accessible to interactions with (+)PBS. The stacking interactions of T6 and G7 residues with the aromatic Phe16 and Trp37 residues of NC, respectively, were found to restrict the mobility of these bases and stretch the entire loop of (–)PBS. These specific interactions stabilize and optimize the (–)PBS conformation for further annealing with the complementary (+)PBS. Exposure of the bases and stretching of the loop switches the annealing pathway from a zipper to a loop-loop kissing mechanism

(Egele, Schaub et al. 2004; Ramalanjaona, de Rocquigny et al. 2007; Godet, Ramalanjaona et al. 2011) (Figure 1.2.11). Determination of the kinetic rate parameters of NC-mediated (–)/(+)PBS annealing shows that in the absence of NC, the hybridization occurs with a rather slow rate, while NC increased the rate of the reaction by about two orders of magnitude through a strong activation of the loop-loop interactions (Ramalanjaona, de Rocquigny et al. 2007).

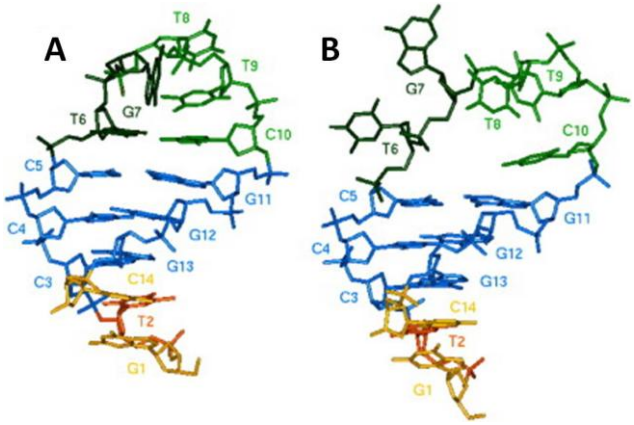


Figure 1.2.12. Comparison of 3D structures of ΔP(–)PBS sequence (A) in its free from and (B) in complex with NC(12-55) (Bourbigot, Ramalanjaona et al. 2008).

Experiments with NC mutants confirm that the hydrophobic plateau at the top of the folded ZFs is responsible for the rearrangement of the NC loop and the loop-loop pathway. The annealing activity is thought to be mainly governed by the N-terminal domain of NC due to its aggregating crowding properties (Stoylov, Vuilleumier et al. 1997; Le Cam, Coulaud et al. 1998; Williams, Rouzina et al. 2001).

1.2.5. Role of NC in HIV-1 assembly

Selection of gRNA and its packaging

During virus assembly, NC acts as a domain of the Gag polyprotein precursor (GagNC) with preserved chaperone activity (Rein 2010). GagNC plays a critical role in the specific recognition and packaging of gRNA, and participates in Gag assembly (Adamson and Freed 2007; Freed 2015). GagNC specifically interacts with gRNAs among a large excess of monomeric RNAs, spliced viral mRNAs and cellular mRNAs in order to ensure the non-spliced gRNA selection and packaging. These specific interactions of GagNC with gRNA are mediated by the ZFs of NC and its hydrophobic plateau that bind directly conserved RNA

packaging signals (ψ sites) located near the 5′ end of viral RNA. Therefore, GagNC regulates both the efficiency and specificity of the gRNA packaging.

Gag-Gag multimerization

The NC domain of Gag plays also a critical role in Gag-Gag oligomerization and assembly since deletion of the NC domain or its two ZFs largely alters both Gag oligomerization and its accumulation at the plasma membrane (El Meshri, Dujardin et al. 2015). The exact role of NC in this process is not fully established. Through its ability to interact with NAs, the NC domain of Gag may tether Gag molecules on the RNA, so that NC bound to RNA plays a role of a scaffold for assembling Gag molecules (Burniston, Cimarelli et al. 1999; Cimarelli, Sandin et al. 2000; Ott, Coren et al. 2009; El Meshri, Dujardin et al. 2015). This scaffolding also allows the Gag proteins to pack closely together and efficiently bind to the plasma membrane.

According to another scenario, protein-protein or protein-membrane contacts may mainly direct Gag assembly, since deletion of NC does not disrupt the formation of the virus like particles (VLP) that, however, lacked gRNA (Zhang, Qian et al. 1998; O'Carroll, Soheilian et al. 2013).

Interaction of Gag with the plasma membrane

The deletion of GagNC or the substitution of its basic residues with neutral ones were reported to partially attenuate the fixation of Gag to the plasma membrane (Grigorov, Decimo et al. 2007; Hogue, Hoppe et al. 2009). These results suggest that in addition to the MA domain, the NC domain can also contribute to binding of Gag to the inner leaflet of plasma membrane. In this context, it was shown that NC, free or bound to an ODN, binds with high affinity ($\sim 10^7$ M^{-1}) to negatively charged model membranes (Kempf, Postupalenko et al. 2015). In contrast to MA, NC did not exhibit any preference for PIP_2 lipids. Importantly, a decrease in the percentage of negatively-charged lipids induced a decrease in the number of NC binding sites but preserved a high level of binding affinity to these sites. This indicates that NC and NC-ODN complexes can recruit negatively-charged lipids to ensure optimal binding of Gag to lipid membranes.

RNA dimerization

GagNC is critical also for the formation of a stable dimer complex between the two molecules of gRNA (Feng, Campbell et al. 1999; Kafaie, Song et al. 2008; Jalalirad and Laughrea 2010). Viral dimerization is involved in several steps of HIV-1 replication such as encapsidation (Berkhout and van Wamel 1996; Houzet, Paillart et al. 2007), recombination fueling virus variability (Balakrishnan, Fay et al. 2001; Chin, Rhodes et al. 2005) and reverse transcription (Paillart, Berthoux et al. 1996; Shen, Jette et al. 2000). The 35-nucleotides-long SL1 sequence of the 5'-UTR of gRNA is a highly conserved sequence that adopts a stem-loop structure with a hexanucleotide self-complementary loop, usually GUGCAC (subtype A, Mal isolate) or GCGCGC (subtype B, Lai isolate) together with 3′ and 5′ flanking purine nucleotides (Figure 1.2.13) (Theilleux-Delalande, Girard et al. 2000). These self-complementary sequences of SL1 promote genome dimerization by forming loop-loop RNA kissing complexes (Muriaux, Fosse et al. 1996; Paillart, Skripkin et al. 1996; Laughrea, Jette et al. 1997). Experiments *in vitro* showed that this dimerization may be facilitated at high temperature (55 °C) or in the presence of NC (Muriaux, Fosse et al. 1996; Laughrea, Jette et al. 1997; Muriaux, de Rocquigny et al. 1996; Takahashi, Baba et al. 2000; Theilleux-Delalande, Girard et al. 2000). NC facilitates RNA dimerization via its NA chaperone activity that results in the formation of a more thermodynamically stable RNA extended duplex.

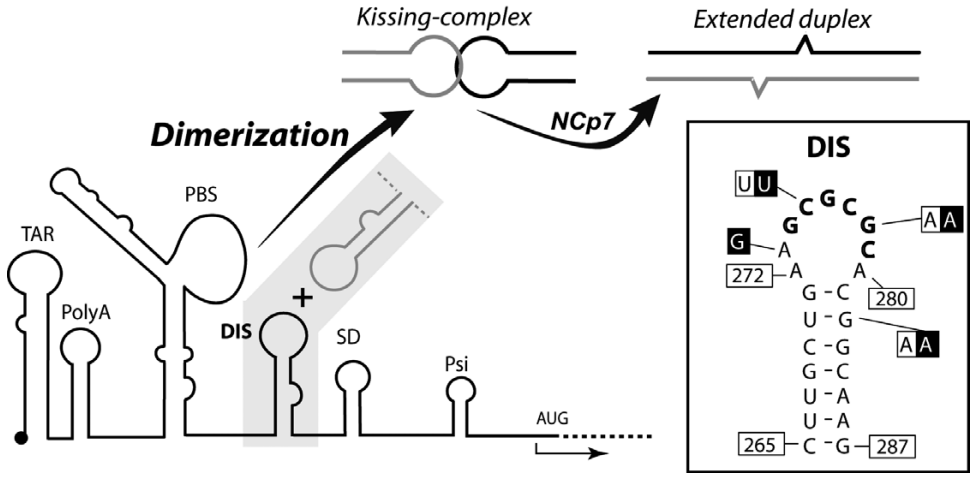


Figure 1.2.13. Localization of the dimerization initiation site (DIS) on the 5′-UTR of the HIV-1 gRNA and mechanism of dimerization. In the inset, the sequence of SL1 stem-loop from the subtype B is shown with variations observed in subtypes A (black boxes) and F (white boxes) (Blond, Ennifar et al. 2014).

Interaction of NC with host cell proteins

Several cellular proteins were described to interact with GagNC during viral assembly and budding. The human protein Staufen 1 binds gRNA via GagNC and may be involved in the gRNA selection and packaging (Mouland, Mercier et al. 2000; Chatel-Chaix, Clement et al. 2004; Chatel-Chaix, Boulay et al. 2008). Nucleolin, an RNA chaperone protein, was found in cytoplasm of the infected cells and was suggested to participate in the nucleoplasmatic transport of the gRNA to the translation site. Additionally, nucleolin may play a role during viral particle release favoring the budding of virions (Ueno, Tokunaga et al. 2004; Gao, Li et al. 2014). The cellular ATP-binding protein ABCE1 was found in association with Gag early after Gag translation and at the assembly sites of the particle at the plasma membrane. It was demonstrated that the NC domain of Gag is critical for interaction with ABCE1, whereas other domains of Gag could be deleted without hindering this association (Lingappa, Dooher et al. 2006). Likely the role of ABCE1 is to facilitate the HIV-1 capsid formation (Zimmerman, Klein et al. 2002), though it dissociates from Gag during virus production and is not incorporated into virions (Dooher, Schneider et al. 2007).

HIV-1 recruits the ESCRT cell membrane fission machinery to promote virus budding and release. Two motifs of the p6 region of Gag, PTAP and LYPX, recruit the ESCRT-associated proteins TSG101 and ALIX. The NC domain of Gag assists p6 binding to ALIX (Popov, Popova et al. 2008; Dussupt, Javid et al. 2009). Two structural domains of ALIX, namely the N-terminal Bro1 and central V-shaped domain interact with NC and p6 domains of Gag, respectively. A detailed analysis of the Bro1-NC binding reveals an implication of NC basic residues in this interaction suggesting an involvement of RNA in this complex as a bridge (Sette, Dussupt et al. 2012). Similarly to ALIX, GagNC was found to interact with TSG101 since mutating of GagNC resulted in a release of not RNA- but DNA-containing virions (Chamontin, Rassam et al. 2015).

Actin and actin-binding proteins were also found incorporated in the HIV-1 particles. Thus, filamentous actin has been suggested to bind to the GagNC domain (Liu, Dai et al. 1999; Wilk, Gowen et al. 1999). However, recent studies claim that the reported NC – actin interactions do not appear to play a role in HIV-1 assembly since disruption of the filamentous actin network does not influence the HIV-1 assembly rates (Stauffer, Rahman et al. 2014). These observations do not exclude a role of actin in the intracellular trafficking or

release of viral particles, but the NC domain and its association with actin was shown to be not needed for these processes.

The NC region of HIV-1 Gag is also required for efficient incorporation of APOBEC3G (A3G) and APOBEC3F (A3F) cell-encoded cytidine deaminases that restrict the replication of HIV-1 virions lacking the viral protein Vif (Viral Infectivity Factor). In the absence of Vif, A3G is packaged into virions during virus assembly. A3G is found in the viral core and is thought to be associated with ribonucleoprotein complexes in the target cells (Goila-Gaur, Khan et al. 2007). Following initiation of reverse transcription, A3G deaminates the (–)ssDNA leading to G to A hypermutation in the (+)ssDNA. Additionally, A3G was reported to interact with IN and inhibit the integration process (Luo, Wang et al. 2007). An effect on the viral DNA synthesis and integration results in inactivation of HIV-1. Vif interacts with A3G and A3F and neutralizes the cellular anti-HIV defense by diverse mechanisms including their degradation or translational inhibition (Jager, Kim et al. 2012). Several groups reported that packaging of A3G into HIV-1 particles involves its interaction with GagNC domain since A3G and A3F are not packaged into virus-like particles with Gag lacking the NC domain (Luo, Liu et al. 2004; Schafer, Bogerd et al. 2004; Apolonia, Schulz et al. 2015). This interaction is RNA-dependent, meaning that RNA may be acting as a bridge or platform for A3G-NC complexes. Interestingly, while RNA binding is clearly required for APOBEC3 proteins, viral gRNA is not essential for A3G packaging (Apolonia, Schulz et al. 2015). The N-terminal region of NC and its linker were found critical for A3G – GagNC interaction (Luo, Liu et al. 2004).

Part 3. Fluorescent amino acid analogs as markers for peptide-nucleic acid interactions

1.3.1. Introduction to fluorescence

Much research in recent years has been focused on the study of biological objects using the fluorescence phenomenon. Its ultimate sensitivity, possibility to study complex structures and functions of living organisms in real time and space together with its non-destructive character make this technique a powerful versatile tool for biological and diagnostic applications.

Fluorescence is an optical phenomenon corresponding to the emission of light by a molecule as a result of the absorption of photons (UV and visible light) (Valeur 2004). Fluorescence is part of a more general phenomenon called luminescence. Luminescence is divided into different sub-classes according to the excitation source. Their types are listed in Table 1.3.1.

Table 1.3.1. Different types of luminescence

Types of luminescence	Excitation source
Photoluminescence (fluorescence and phosphorescence)	Photons
Radioluminescence	Ionizing radiation
Cathodoluminescence	Accelerated electrons
Electroluminescence	Electric field
Chemiluminescence	Chemical reaction
Bioluminescence	Enzymatic reaction
Triboluminescence	Deformation
Sonoluminescence	Ultrasound
Thermoluminescence	Temperature increase

Fluorescence and phosphorescence are two forms of photoluminescence that differ by the nature of the excited state. Fluorescence and phosphorescence emissions occur from the excited singlet state and excited triplet state, respectively.

Classically, the process of fluorescence is illustrated by the Jablonski diagram (Lakowicz 2006) shown on Figure 1.3.1A. The process is initiated by the absorption of a photon by a fluorophore being in its singlet electronic ground state (S_0). Following light absorption, the fluorophore is excited to any of the vibrational levels of S_1 or S_2. The transitions between the ground and excited states occur in about 10^{-15} s, a time too short for a significant displacement of the nuclear configuration of a molecule (the set of bond distances and bond angles), as stated by the *Franck-Condon* principle. Then the molecule relaxes quickly to the first singlet excited state (S_1), through a process called vibrational relaxation (or additionally through *internal conversion* if the excited level is superior to S_1). The average time that a fluorophore spends in the excited state before emitting a photon is about 10^{-10}- 10^{-8} s, and is called the *fluorescence lifetime*. Return to the ground state can be accompanied by a photon emission (fluorescence) or can occur through a non-radiative pathway. Non-radiative quenching of the fluorophore can be due to several possible processes, including molecular collisions, photoinduced electron transfer, bond rotation or vibration (Doose, Neuweiler et al. 2009). The excited state can also undergo an *intersystem crossing* (ISC) to the triplet excited state (T_1) and subsequent relaxation by either photon emission (phosphorescence) or a non-radiative pathway. Another important pathway from the excited state is the Förster resonance energy transfer (FRET) to an acceptor molecule. This process is distance-dependent and used to measure the proximity of labeled entities (Sapsford, Berti et al. 2006).

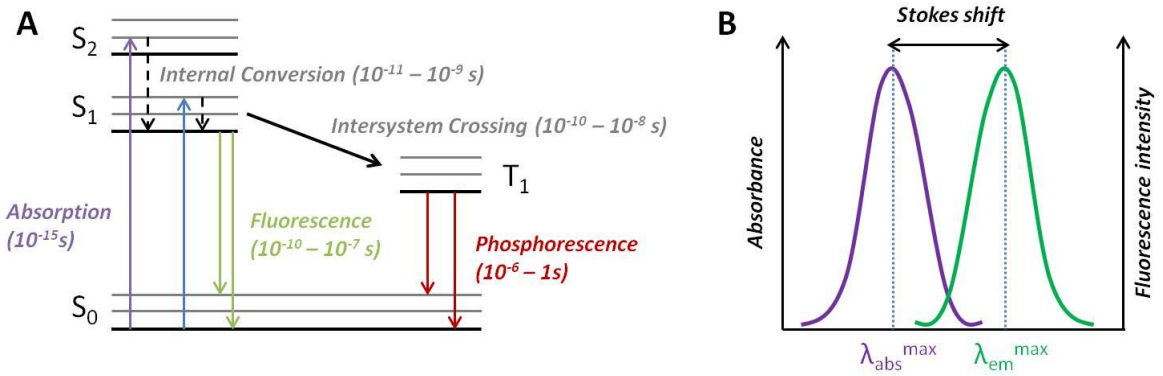


Figure 1.3.1. Jablonski diagram (A) and schematic view of absorption and emission spectra (B).

According to the Jablonski diagram, the energy of absorption is higher than that of the emission. Thus, emission typically occurs at lower energies and is shifted to longer wavelengths with respect to absorption (Figure 1.3.1 B). The difference between absorption and emission maxima is called the *Stokes shift*. The fluorescence efficiency of the fluorophore

is defined by its *fluorescence quantum yield* (QY) that is the ratio of the number of emitted photons to the number of absorbed photons (Eq. 1.3.1).

$$QY = \frac{N_{emitted}}{N_{absorbed}} \quad (1.3.1)$$

The *fluorescence lifetime* τ is one more important characteristic of the fluorophore that is defined by the radiative (k_r) and non-radiative (k_{nr}) rate constants:

$$\tau = \frac{1}{k_r + k_{nr}} \quad (1.3.2),$$

The non-radiative rate constant is the sum of rate constants for vibrational relaxation, k_{ic}, intersystem crossing, k_{isc}, and if there is quenching due to interactions with other molecules $k_q[q]$, where [q] is the concentration of quenchers (Festy, Ameer-Beg et al. 2007):

$$k_{nr} = k_{ic} + k_{isc} + k_q[q] \quad (1.3.3)$$

The fluorescence lifetime is independent on the excitation intensity, fluorophore concentration (if it is sufficiently low), photobleaching, or scattering.

The positions of the absorption and emission maxima, the Stokes shift, fluorescence quantum yield, and fluorescence lifetime are key parameters of any fluorescent molecule.

Development and applications of fluorescence-based methods have largely benefited from the development of a large number of fluorescent dyes that exhibit varied photophysical properties.

1.3.2. Fluorescent probes

Fluorescent molecules are usually characterized by the families to which they belong. One classification divides the fluorescent probes into 2 categories: intrinsic or endogenous occurring naturally in the biomolecules of interest, and extrinsic probes that are synthetically incorporated. Based on their structure, the fluorescent probes can be also divided into small organic molecules, fluorescent proteins and fluorescent quantum dots. Their different molecular and electronic structures lead to a wide range of photophysical properties such as molar absorption, quantum yield, Stokes shift, fluorescence lifetime, and ability to respond to the environment.

Small organic molecules represent the largest group of fluorophores. They generally correspond to organic compounds with extended π-conjugated bonds. This group subdivides into several families of fluorophores that are the derivatives of one core structure. All families of dyes are characterized by close absorption/emission profiles and show associated advantages and disadvantages depending on the application (Sapsford, Berti et al. 2006). The first well-characterized fluorophore was the natural product *quinine* (Figure 1.3.2). Excited by ultraviolet light, quinine in aqueous solution emits near 450 nm (Herschel 1845). Interestingly, quinine also stimulated the development of the first spectrofluorimeters in 1950s that promoted the use of fluorescence spectroscopy as an analytical technique (Lakowicz 2006). Quinine sulfate is still frequently used as a reference fluorophore for QY determination (Melhuish 1961).

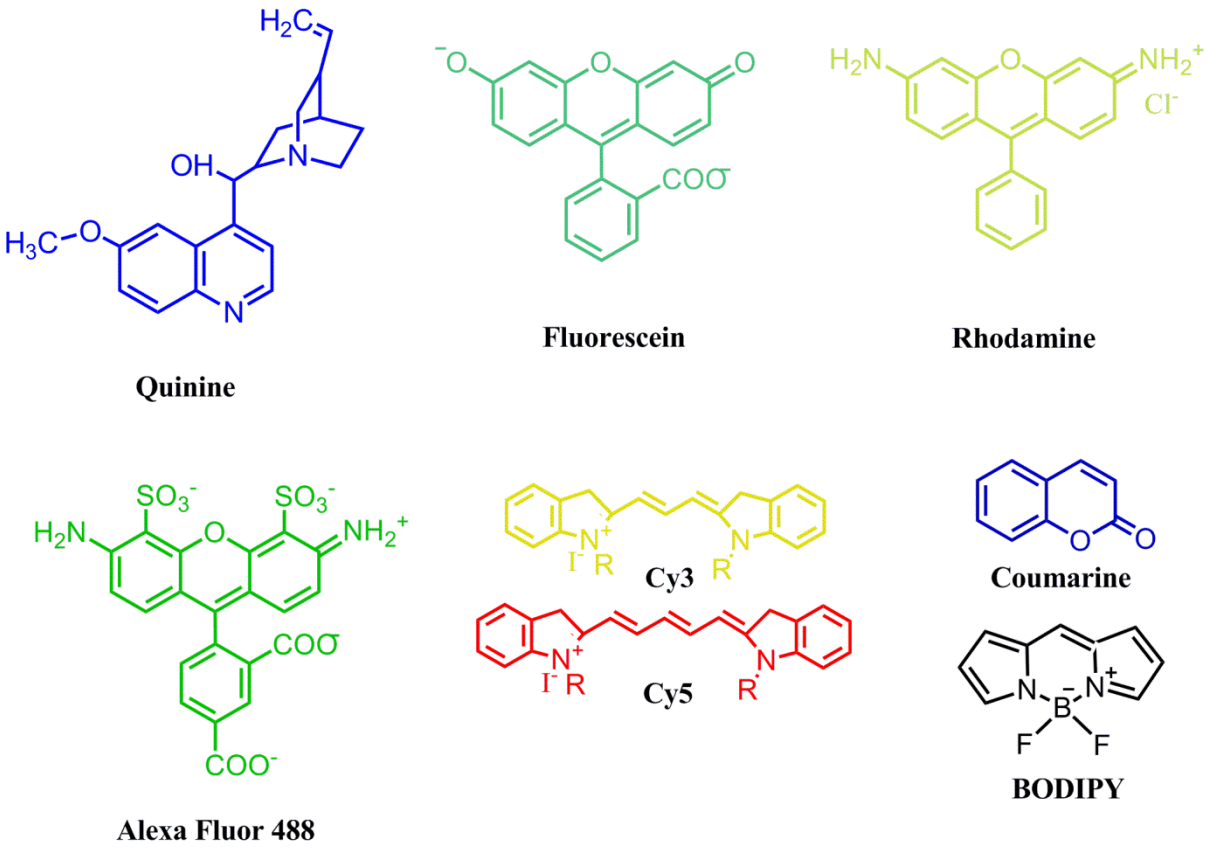


Figure 1.3.2. Core structures of small organic fluorophores.

Fluorescein is considered as the most popular fluorescent label. It was synthesized by Bayer in 1871, then in 1958 Riggs and colleagues reported the synthesis of fluorescein isothiocyanate (FITC), its improved derivative. The popularity of fluorescein dyes is determined by their high QY, solubility and conjugation susceptibility. However, they show poor photostability, pH sensitivity and self-quenching at high degrees of substitution

(Sapsford, Berti et al. 2006; Lavis and Raines 2008). Fluorescein dyes are widely used for flow cytometry, fluorescence microscopy and immunoassays.

Rhodamine dyes compose another group of classical organic fluorophores. Their key characteristics include low pH sensitivity and tunable spectral properties. Rhodamine B and 6G are used as laser dyes, while tetramethylrhodamine TMR, carboxytetramethylrhodamine TAMRA and other derivatives are suitable for monitoring binding events using FRET, anisotropy measurements, fluorescence correlation spectroscopy, flow cytometry and immunoassays.

Coumarin is a natural fluorophore with blue-green emission. The coumarin family represents a broad class of fluorophores (Lavis and Raines 2008). These fluorophores exhibit environmental sensitivity, notably to polarity and viscosity. They are applied mostly as laser dyes and in ion detection studies.

BODIPY (boron difluoride dipyrromethane) is an increasingly popular fluorophore that was described by Treibs and Kreuzer in 1968, and constitutes the core of numerous derivatives (Loudet and Burgess 2007; Ulrich, Ziessel et al. 2008; Kowada, Maeda et al. 2015). BODIPYs exhibit large extinction coefficients, high QYs and a wide range of fluorescence emission wavelengths. Derivatization of the core molecule allows mimicking the emission characteristics of traditional dyes like fluorescein, tetramethylrhodamine, etc. Moreover, the lipophilic character of BODIPY allows its incorporation into non-polar systems (Karolin, Johansson et al. 1994).

Alexa Fluor series are rhodamine derivatives that form a family of highly photostable probes with a wide range of excitation and emission wavelengths. They are applied in single-molecule studies, fluorescence-based assays and biosensing (Geng, Uknalis et al. 2006).

Cyanine dyes were firstly synthesized over a century ago and form a family of polymethine fluorophores with long-wavelength emission above 550 nm. Cyanine dyes are characterized by small Stokes shift, high photostability, large absorption coefficients and good QY. Variation of the R-groups in their structures (Figure 1.3.2) produces a myriad of fluorophores applied in bioimaging and single molecule experiments.

The second class of fluorophores – **fluorescent proteins** – is a class of proteins originally found in jellyfish and corals. The discovery of green fluorescent protein (GFP) in 1962 by Shimomura and co-workers (Shimomura, Johnson et al. 1962) opened a new page in

fluorescent probes and was awarded the Nobel Prize in chemistry in 2008. GFP was isolated from the bioluminescent jellyfish *Aequorea victoria* and later its crystal structure was resolved (Ormo, Cubitt et al. 1996). GFP is composed of a highly fluorescent chromophore surrounded by constrained and protective β-sheets of the protein (Figure 1.3.3) (Day and Davidson 2009). This chromophore is encoded by the primary amino acid sequence that forms upon cyclization and oxidation of the sequence –Ser65-Tyr66-Gly67. In the final folded structure, the chromophore is wrapped by a polypeptide barrel which enhances its fluorescence quantum yield. Interestingly, the chromophore forms spontaneously, without cofactors or enzymes. This key characteristic allows expression of the fluorescent protein in living cells. Introducing mutations into the amino acid sequence or modifying the chemical structure of the chromophore generated a large set of fluorescent proteins with various chemical and spectral properties (Nienhaus and Nienhaus 2014; Walker, Lukyanov et al. 2015). However, fluorescent proteins show limitations due to their molecular weight which is close to 30 kDa and thus comparable or even larger than the one of the proteins of interest. This may perturb the protein activity, functions and interactions or alter their subcellular localization. Additionally, prolonged excitation induces photobleaching of these fluorophores (Zhang, Zheng et al. 2015).

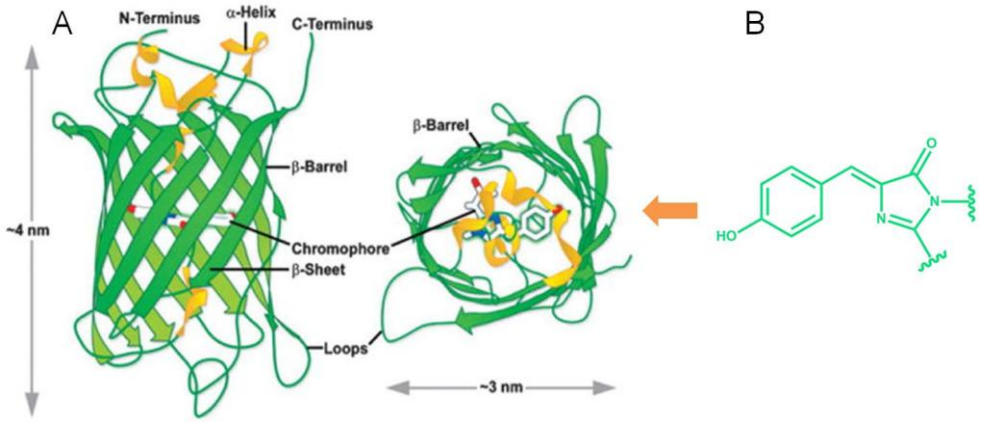


Figure 1.3.3. (A) β-barrel architecture and approximate dimensions of GFP. Drawing based on Protein Data Bank ID: 1w7s. Adapted from Day and Davidson 2009. (B) The GFP chromophore structure.

The third class – **fluorescent quantum dots (QDs)** – are semiconductor nanocrystals which have found a wide range of applications in biological imaging. QDs are composed of a core with diameter 1-10 nm and a shell that can be functionalized to increase their solubility, or be conjugated with biomolecules of interest (Figure 1.3.4). Light absorption by QDs leads to an electron-hole recombination which gives rise to luminescence emission (Jameson 2014;

Wegner and Hildebrandt 2015). Elements of several groups like III-V, II-IV, or IV-VI have been used to fabricate QDs, but the most popular are the cadmium-based QDs. QDs show excellent photophysical characteristics like high molar extinction coefficients (10^5-10^7 M^{-1} cm^{-1}), high QY and photostability. Interestingly, their absorption and emission wavelengths can be tuned by the size of the particle. However, the major problem of the QDs is their toxicity, so a continuous research is conducted towards more biocompatible materials, like silicon QDs or fluorescent nanodiamonds (Montalti, Cantelli et al. 2015).

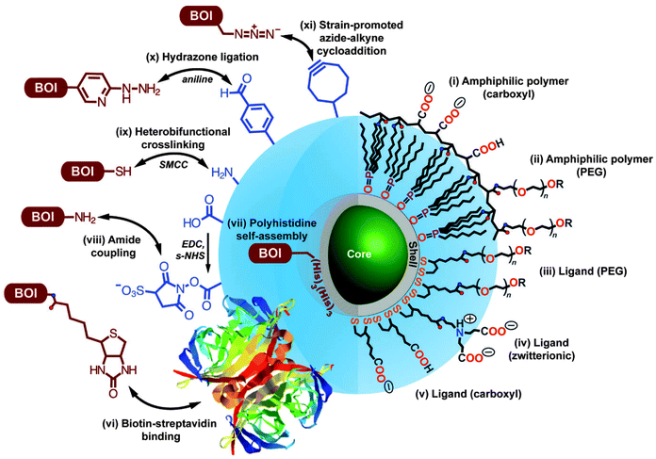


Figure 1.3.4. Structure, bioconjugation possibilities (left side, BOI = biomolecule of interest) and surface coating (right side) strategies for QDs. Two surface coating strategies are presented: encapsulation with amphiphilic polymers (i, ii) and cap exchange with hydrophilic ligands exploiting the thiol-affinity of the ZnS shell of the QD (iii–v) (Petryayeva, Algar et al. 2013).

1.3.3. Fluorescence approaches to study protein – nucleic acid interactions

Various fluorescence-based techniques have been developed to solve questions in analytical chemistry, protein biochemistry, biophysics, molecular cell biology, biotechnology and medicine. These techniques are grounded on several fundamental photophysical principles like Förster resonance energy transfer (FRET) (Jares-Erijman and Jovin 2003), fluorescence anisotropy (Yengo and Berger 2010), fluorescence correlation spectroscopy (Elson 2011), fluorescence polarization (Jameson and Ross 2010) etc. Advanced instrumentation has allowed fluorescence-based techniques to be extended for a wide range of applications like high-throughput screening assays (Inglese, Johnson et al. 2007), drug discovery or super-resolution imaging in cells (Ha and Tinnefeld 2012).

In this work, we focused on the binding of proteins to NAs. Several fluorescence-based approaches are usually applied for this purpose. The most *classical approaches* are based on

changes in the emission spectra, anisotropy and fluorescence lifetimes of tryptophan residues used as an intrinsic protein probe, or extrinsic probes introduced in the protein or NA sequence. *The FRET approach* is also largely used. FRET is a non-radiative energy transfer process between a donor D (usually fluorophore) in the excited state and an acceptor molecule A in the ground state located at distance r from the donor. FRET requires that the acceptor absorbs photons at the emission wavelength of the donor, and that the distance between the two molecules should be less than 10 nm (Figure 1.3.5A) (Lakowicz 2006; Sapsford, Berti et al. 2006). Figure 1.3.5A shows the absorption (Abs) and emission (Em) spectra of one of the classical FRET pairs: fluorescein as donor and rhodamine as acceptor (Sapsford, Berti et al. 2006). Fluorescein can be efficiently excited at 480 nm and emits at around 520 nm. A spectral overlap between fluorescein emission and rhodamine absorption is observed at 500–600 nm. The Förster distance R_0 for this pair is 55 Å. Thus, in appropriate conditions ($0.5R_0 < r < 1.5R_0$), excitation of fluorescein results in significant FRET to the rhodamine.

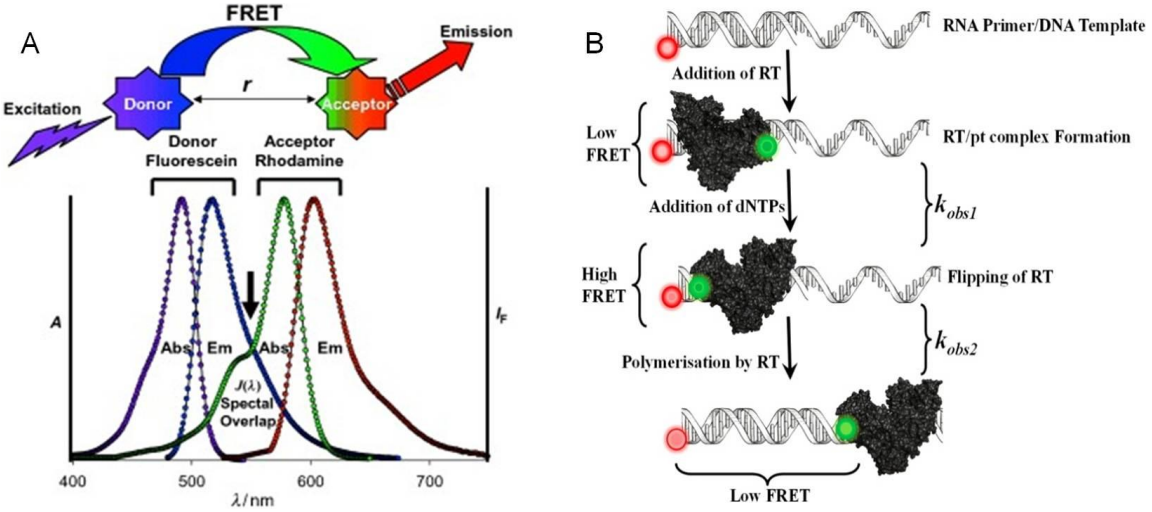


Figure 1.3.5. Schematic representation of (A) FRET between fluorescein and rhodamine molecules and their corresponding absorption and emission spectra (Sapsford, Berti et al. 2006); (B) FRET-bases assay for monitoring flipping and polymerase activity of reverse transcriptase (Sharma, Przybilla et al. 2015). The DNA template was labeled at 3′ end with TMR acting as a FRET donor, RT was labeled close to its RNase H domain with Alexa488 acting as a FRET acceptor. Firstly, due to the RNA nature of the primer, RT orients in its RNase H conformation on the duplex, showing low FRET between TMR and Alexa488. Addition of dNTPs leads to flipping of RT from RNase H to polymerase orientation, inducing an increase in FRET. Further dislocation of RT as a result of its polymerase activity is accompanied by a FRET decrease due to elongation of the distance between the two fluorophores.

One example of application of FRET for monitoring protein – NA interaction is the real time FRET-based assay for direct measurement of the binding orientation and polymerase activity of the HIV-1 reverse transcriptase (RT) (Sharma, Przybilla et al. 2015), which is responsible for the conversion of the single-stranded viral RNA into double-stranded DNA. It was reported that depending on the nature of the primer, RT adopts opposite binding orientations on primer/template duplexes (Abbondanzieri, Bokinsky et al. 2008). On duplexes with DNA, the primer RT adopts its polymerase conformation (DNA polymerase domain is oriented toward the 3′-OH of primer) to direct the DNA synthesis, while on the duplexes with RNA primer, RT adopts its RNase H conformation (RNase H domain is oriented toward 3′-OH of primer) in order to perform its RNase H activity. Such flipping from polymerase-competent to RNase H-competent orientation was found to be altered by either adding cognate nucleotides (decrease in the flipping rate) or non-nucleoside RT inhibitors (increase in the flipping rate). Based on these observations, a continuous FRET-based assay was designed to visualize both the change in RT binding orientation and the primer elongation step (Figure 1.3.5B). RT was labeled in the p51 subunit close to the RNase H domain by Alexa 488 used as a FRET donor, while the DNA template was labeled at its 3′ end by TMR, used as a FRET acceptor. As initially RT binds RNA/DNA primer/substrate in its RNase H conformation, a large distance between Alexa 488 and TMR is expected resulting in a low FRET. Addition of cognate deoxynucleotide triphosphates (dNTPs) leads to flipping of RT from RNase H to polymerase orientation and result in a closer proximity of the two dyes and, thus, higher FRET. Further primer extension leads to a progressive increase of the distance between the two probes and, thus, a progressive FRET decrease.

Based on the FRET approach, *molecular beacons* were designed and widely applied in ODN hybridization and peptide – NA interactions in order to monitor DNA polymerase reaction (Summerer and Marx 2002), enzymatic cleavage of single-stranded DNA (Li, Fang et al. 2000), or annealing reactions occurring during the HIV-1 reverse transcription (Godet, de Rocquigny et al. 2006; Ramalanjaona, de Rocquigny et al. 2007). This technique exploits a NA sequence that is labeled with fluorophore at one end and a non-fluorescent quencher at the other end (Figure 1.3.6) (Li, Fang et al. 2000; Tyagi and Kramer 1996). The sequence forms a hairpin loop that brings fluorophore and a quencher close enough (< 1 nm) so that quenching occurs through excitonic coupling (Bernacchi, Piémont et al. 2003). Upon destabilization or hybridization, the two labels are separated, so that a distant-dependent FRET will result in a partial restoration the emission of the fluorophore.

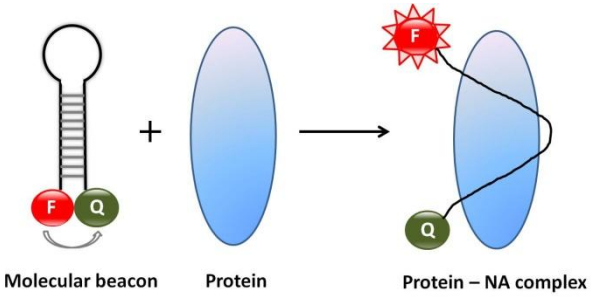


Figure 1.3.6. Schematic representation of the use of a molecular beacon to monitor binding with a protein. F is the fluorophore and Q is the quencher.

The main limitation of the FRET-based approach is that it requires double-labeling, which can be tedious and may perturb the binding process (Klymchenko and Mely 2013). Single fluorescence labeling approaches are simpler and less perturbing. The simplest and most straightforward approach to monitor protein/NA interactions is to use the natural aromatic amino acids of the proteins, as intrinsic fluorescence probes.

1.3.4. Natural fluorescent amino acids

The fluorescence properties of the aromatic amino acids tryptophan, tyrosine and phenylalanine (Figure 1.3.7) were firstly described by Weber (Teale and Weber 1957). Phenylalanine is not commonly used as a fluorescent reporter due to its low extinction coefficient coupled with low QY (Ross and Jameson 2008). Its brightness is far below those of tyrosine and tryptophan. Tyrosine may be used as an intrinsic fluorescent probe, but only in proteins that do not contain tryptophan. In aqueous solution, tyrosine has an extinction coefficient ε^{275} = 1400 M^{-1} cm^{-1} (Fasman 1989), a QY = 0.14 (Chen 1967) and a fluorescence lifetime around 3.3-3.8 ns (Ferreira, Stella et al. 1994). However, in proteins the QY of tyrosine strongly drops. Tryptophan is the most emissive amino acid exhibiting ε^{279} = 5580 M^{-1} cm^{-1} (Fasman 1989) and QY = 0.14 (Chen 1967; Eisinger and Navon 1969), thus about 3.6-fold brighter than tyrosine. Its fluorescence lifetime in proteins ranges from several picoseconds to nearly 10 ns (Beechem and Brand 1985; Ross and Jameson 2008). Moreover, the tryptophan emission maximum and QY are sensitive to polarity due to a large dipole moment of the molecule in the excited state (Pierce and Boxer 1995). This environment sensitivity has been used to monitor a variety of processes, including protein dynamics, folding, and ligand binding (Beechem and Brand 1985; Engelborghs 2003; Royer 2006). Being apolar, tryptophan is often located in the hydrophobic interior of proteins (Royer 2006). Binding of a ligand or unfolding of the protein is accompanied by the exposure of tryptophan

to bulk water, which results in a red-shift of its emission maximum and changes of its QY (Sinkeldam, Greco et al. 2010). Tryptophan can be also used as an intrinsic FRET donor in pair with acceptor-labeled ligand for real-time detection of binding events (Sapsford, Berti et al. 2006; Ghisaidoobe and Chung 2014).

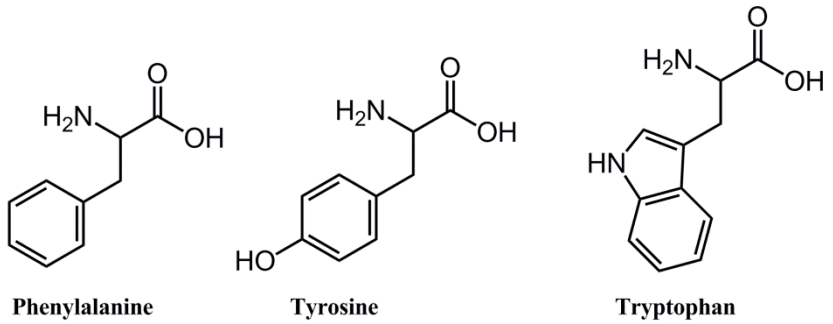


Figure 1.3.7. Structure of the fluorescent natural amino acids.

However, the relative abundance of tryptophan in proteins and its UV excitation strongly limit its application as a fluorescent probe in complex systems. This stimulated the development of **synthetic amino acid analogs** with improved photophysical properties (Sinkeldam, Greco et al. 2010). The general approach for the design of synthetic amino acid analogs is based on mimicking the tryptophan molecule or attaching an established fluorophore to the side chain of a non-emissive amino acid such as alanine or lysine. Amino acid analogs with environment-sensitive fluorophores are of special interest since they can detect changes in the probe microenvironment that commonly accompany biomolecular interactions.

1.3.5. Environment-sensitive probes

Environment-sensitive fluorophores exhibit changes in their absorption/emission spectra in response to a change in their environment such as pH, ionic strength, specific ions (Ca^{2+}, Cl^-), O_2 saturation, solvation or polarity (Sapsford, Berti et al. 2006). Conjugation of these probes to relevant biomolecules provides valuable information about their functions and interactions in the context of living systems. Noticeably, pH and ion-sensitive dyes are not always considered as environment sensitive, as their response is associated with changes in their chemical structures (protonation/deprotonation, formation of metal complexes) (Klymchenko and Mely 2013).

Since specific protein/ligand recognition and interactions strongly depend on hydrophobic contacts, *polarity-sensitive (solvatochromic) fluorescent* probes are of particular interest. These probes are composed of an electron-donor group (for example, amino groups), a spacer, and an electron acceptor (for example, carbonyl group), forming a typical intramolecular charge transfer system (Yang, Cao et al. 2014). In response to polarity changes, these fluorophores exhibit changes in their emission maxima, fluorescence lifetime or quantum yield. Often, all three parameters are affected simultaneously, highlighting the dynamic range of the fluorophore response. In order to understand how the solvatochromic dyes work, one need to understand what is the nature of *solvatochromism*. Molecules in their excited state, before returning to their ground state, interact with the surrounding solvent molecules depending on their polarity. These excited-state solute-solvent interactions result in changes in the spectral position and shape of the fluorescence emission bands as well as in the fluorescence QY and lifetimes of the fluorophores. The solvent dependence of the maxima of the fluorescence emission spectra is commonly called solvatochromism (Lakowicz 2006), that can be schematically illustrated on figure 1.3.8.

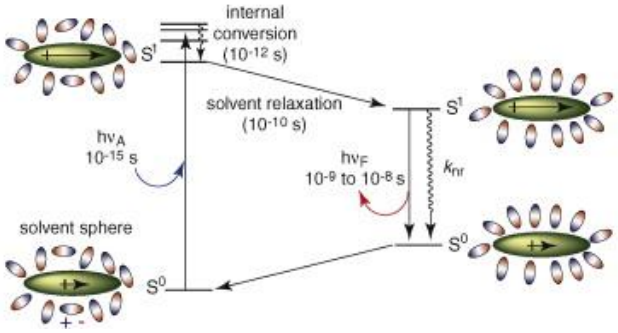


Figure 1.3.8. Principle of the solvatochromism phenomenon (Loving, Sainlos et al. 2010).

In the ground state S_0, a fluorophore is surrounded by solvent molecules and its dipole moment is μ_g. Absorption of a photon results in the excitation of the fluorophore to the S_1 level accompanied by a change in its dipole moment (μ_e) because of an intramolecular charge transfer from the electron-donor group to the electron acceptor group (Klymchenko and Mely 2013). Consequently, the molecules of solvent need to reorient their dipoles to optimize their interaction with the new dipole of the fluorophore, resulting in a more highly ordered arrangement. This reorientation of the solvent molecules corresponds to solvent relaxation. The net effect of this relaxation is to decrease the energy of the excited singlet state, which results in a decrease of the energy and thus to a shift to the red of the emitted photon. The degree of solvent relaxation increases with the solvent polarity (Loving, Sainlos et al. 2010).

In addition, protic solvents (with –O-H, -N-H_2, -N=H groups) can also interact with the dye through H-bonding that further decreases the energy of the excited S_1 state. Thus, both dipole-dipole and H-bonding interactions in polar solvents induce a red-shift of the fluorophore emission (Klymchenko and Mely 2013).

The characteristics of the most frequently used solvatochromic probes and their corresponding amino acid analogs are briefly presented below.

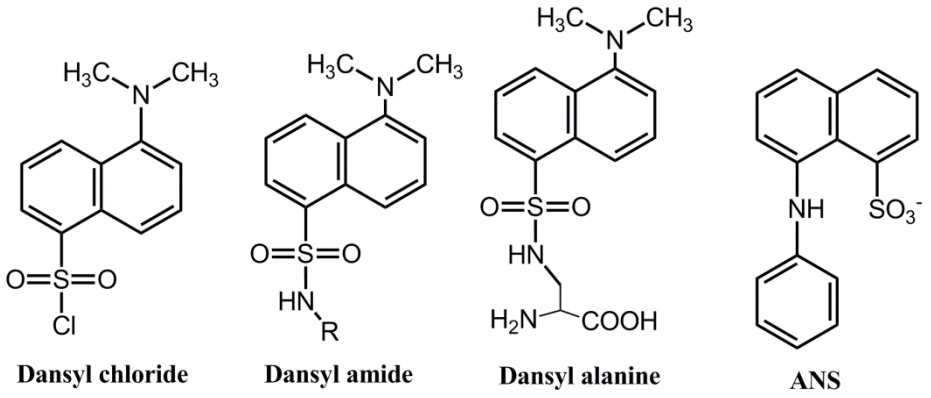


Figure 3.9. Structures of selected aminonaphtalene-sulfonic acid derivatives.

Dansyl chloride (1,5-dimethylaminonaphtalene sulfonyl chloride) (Figure 1.3.9) and its derivatives were among the first characterized solvatochromic fluorescent dyes with excitation at 350 nm and emission around 520 nm (Lakowicz 2006). They are still applied for protein labeling and membrane imaging (Goretta, Kinoshita et al. 2012; Alaimo 2013). Moreover, the amino acid analog **dansyl alanine** was genetically expressed and used to monitor protein unfolding (Summerer, Chen et al. 2006).

ANS (1,8-anilinonaphtalene sulfonate) firstly reported by Slavik (Slavik 1982) is another solvatochromic fluorescent probe for protein and lipid membrane labeling. ANS is weakly emissive in water, but its QY increases significantly, accompanied by a blue-shift, as the polarity of the environment decreases (Jameson 2014). This was observed upon binding to bovine serum albumin or heat-denatured proteins and applied for studies of protein folding/unfolding events and aggregation (Hawe, Sutter et al. 2008).

PRODAN (6-propionyl-2-(dimethylamino)naphthalene) (Figure 1.3.10) is another classical example of solvatochromic fluorophores (Weber and Farris 1979) because of its remarkable sensitivity to solvent polarity and its small size (Klymchenko and Mely 2013). PRODAN found its wide application in monitoring biological and model membrane

environments (Parasassi, Krasnowska et al. 1998). The alanine analog of PRODAN called DANA or Aladan (6-(2-dimethylaminonaphthoyl) alanine) (Figure 1.3.10) was synthesized in parallel by the groups of Imperiali and Jan (Cohen, McAnaney et al. 2002; Nitz, Mezo et al. 2002). DANA/Aladan probe is applied for probing protein electrostatics, peptide-protein and protein-protein interactions. For example, DANA was attached to the octapeptide involved in the phosphorylation-dependent protein-protein interactions essential for the cell cycle regulation (Vazquez, Nitz et al. 2003). Upon binding to the hydrophobic pocket of the target protein 14-3-3 it shows a strong fluorescence increase accompanied by a blue-shift of its emission maximum, highlighting the change in its local environment upon binding.

Several PRODAN derivatives were designed to overcome its UV excitation and improve its photophysical properties. For instance, the fluorene-based derivative FR0 shifted the excitation wavelength of the probe to 400 nm (Kucherak, Didier et al. 2010) showed an increased environment sensitivity due to its larger dipole moment change in the excited state. Another perspective PRODAN analog ACRYLODAN selectively labels thiol moieties in proteins (Prendergast, Meyer et al. 1983). This fluorophore shows enhanced QY when bound to the hydrophobic domains of biological macromolecules.

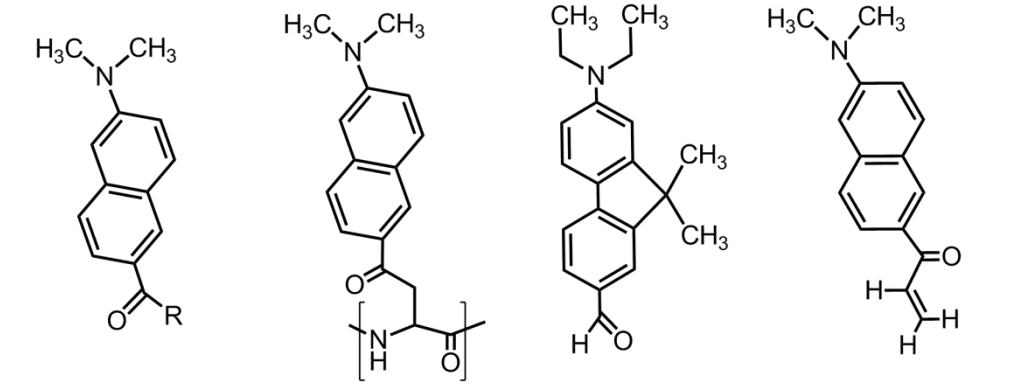


Figure 1.3.10. Structure of PRODAN and its derivatives.

LAURDAN is a member of the 2,6-naphtalene family and a PRODAN derivative. In LAURDAN, the propionic acid tail is substituted by the lauric acid tail that increases its lipophilicity and therefore improves its insertion into membranes. Both excitation and emission spectra of LAURDAN can sense the physical state of the lipid bilayer, so they are applied to follow alterations of the lipid membrane (Parasassi, Krasnowska et al. 1998).

The 4-aminophtalimide family of fluorophores (Figure 1.3.11) is represented by a series of derivatives with different spectroscopic properties. Their maximum emission wavelength increases with the extension of the conjugation moiety. **4-DMAP** (4-*N*,*N*-dimethylaminophtalimide), similarly to PRODAN, absorbs in the UV region, has a low molar extinction coefficient and exhibits a strong polarity sensitivity (Soujanya, Fessenden et al. 1996; Saroja, Soujanya et al. 1998). Like ANS, 4-DMAP is strongly quenched in polar solvents like water and becomes highly emissive in non-polar media. This turn-on effect is useful for monitoring changes in the probe microenvironment, though can be a limitation when the label is exposed to water at all steps of the interaction. **6-DMN** (6-*N*,*N*-dimethylamino-2,3-naphtalimide) is an improved environment-sensitive derivative of 4-DMAP that absorbs in UV region, but emits in the 500-600 nm range (Vazquez, Blanco et al. 2005). **4-DMN** (4-*N*,*N*-dimethylamino-1,8-naphtalimide) molecule was designed to minimize the damaging effects of UV excitation. It can be excited by visible light at around 400-450 nm. The key advantages of this probe are its higher chemical stability, small size and characteristic switch-like spectroscopic properties (Loving and Imperiali 2008).

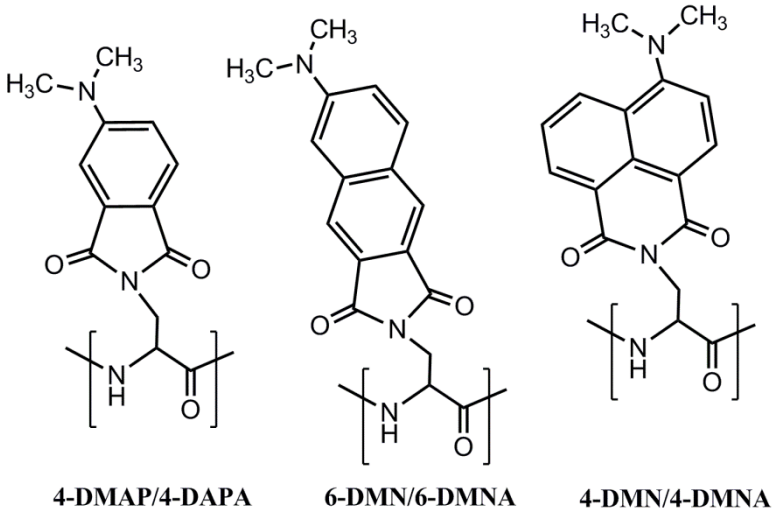


Figure 1.3.11. Structures of 4-aminonaphtalene dyes and their amino acid analogs.

The corresponding amino acid analogs 4-DAPA (4-(N,N-dimethylamino)-phtalimide propionic acid), 6-DMNA and 4-DMNA (Figure 1.3.11) were also synthesized (Vazquez, Rothman et al. 2004; Vazquez, Blanco et al. 2005; Loving and Imperiali 2008). All these mimics have sizes comparable to the natural aromatic amino acids allowing their incorporation into peptides and proteins without significant perturbation. Moreover, they exhibit the switch-like emission properties characteristic of the dimethylaminophthalimide

series and are applied in the study of protein – protein interactions (Loving and Imperiali 2008; Krueger and Imperiali 2013).

To summarize, both solvatochromic and fluorescence characteristics of the dye should be taken into account when choosing an appropriate fluorescent probe. Current development of new environment-sensitive fluorophores is directed towards:

- Improved photophysical properties and sensitivity in highly polar media;
- Shift in the excitation wavelengths towards the visible region in order to decrease the photodamage and autofluorescence;
- Increased photostability and brightness.

One more important characteristics of the solvatochromic dye is to be easily synthesized as a Fmoc building block in order to be introduced into proteins in the form of an amino acid analog. The development of the solid-phase peptide synthesis in 1963 (Merrifield 1963) facilitated incorporation of amino acid analogs into proteins.

1.3.6. Ratiometric environment-sensitive probes based on 3-hydroxychromones

An alternative type of environment-sensitive probes is the ratiometric probes that possess a dual-band emission. The ratio between their two bands depends only on the microenvironment of the probe and not on the concentration of the dye or the instrumental setting. This is an essential advantage of ratiometric probes over conventional one-band probes. One of the most wide-spread ratiometric environment-sensitive probes are the 3-hydroxychromones (3HCs) (Figure 1.3.12), that are oxygen-containing heterocycles found notably in natural flavonoids. Flavonoids are abundant in common plant-based foods and beverages, like onions, berries, tea, and red wine (Sengupta, Reilly et al. 2015). As 3HCs are a part of normal human diet, they are low toxic (Harborne 1988). In 1979, Sengupta and Kasha observed yellow to yellow-green fluorescence in several flavonoids, such as kaempferol, quercetin and myricetin (Sengupta and Kasha 1979). They studied in detail 3-hydroxyflavone (3HF) as a skeletal precursor of flavonols, the most commonly occurring class of flavonoids.

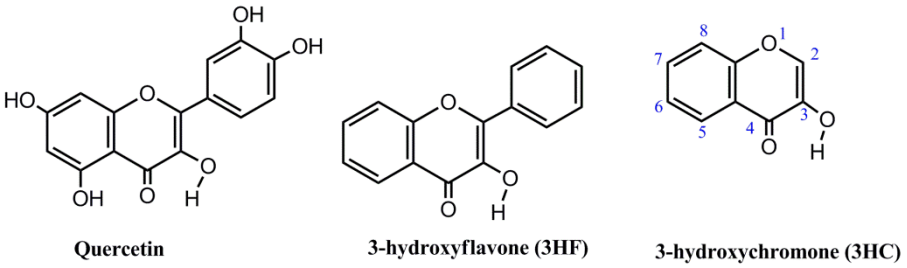


Figure 1.3.12. Structures of the natural flavonoid quercetin, 3-hydroxyflavone and 3-hydroxychromone.

They reported that 3HFs and 3HCs (as a larger family) show two well-resolved emission bands with a wavelength difference about 100 nm. This dual emission is the result of the formation of two excited state tautomers (N* and T*) due to an excited-stated intramolecular proton transfer reaction (ESIPT) (Figure 1.3.13) (Sengupta and Kasha 1979; Demchenko, Tang et al. 2013).

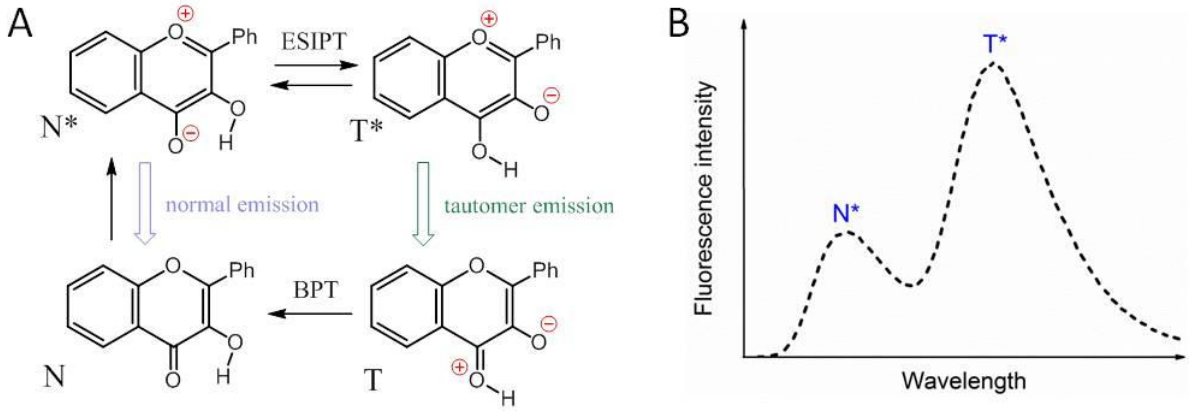


Figure 1.3.13. (A) Excited-state intramolecular proton transfer (ESIPT) reaction in 3HCs. BPT corresponds to back proton transfer, N* and T* represent normal and tautomeric emissive forms, respectively. (B) Representative fluorescence spectrum of a 3HC derivative.

The proton transfer reaction occurs between the proximate proton donor (3-hydroxyl group) and acceptor (oxygen of the 4-carbonyl) groups connected by a hydrogen bond in the same photoexcited molecule (Kasha 1986; Barbara, Walsh et al. 1989; Tomin, Demchenko et al. 2015). The ESIPT reaction transforms the normal excited state into the tautomer excited state. The mechanism of the ESIPT reaction is related to changes in the acidity of the proton donor 3-OH group and basicity of the proton acceptor 4-carbonyl group in the excited state. In the excited state, the electronic charges are redistributed resulting in a strongly acidic 3-OH

group and strongly basic carbonyl group. This favors the proton transfer and gives a tautomer form T*.

The ESIPT reaction may follow reversible or irreversible mechanisms (Tomin, Demchenko et al. 2015). However, the definition of the mechanism is a point of issue since the reaction may be complicated by ground and excited-state heterogeneities and the presence of side reactions. Generally, it is considered that the N* form is the only initially excited species and it transforms to the T* form with a kinetic rate constant k_+ together with radiative and non-radiative conversion to the ground state N with kinetic constants k_R^{N*} and k_{NR}^{N*}, respectively (Figure 1.3.14). Resulting from the ESIPT reaction, the T* species are characterized by reverse transition to the N* state with a rate constant $k_–$ and by the radiative and non-radiative conversion to the ground state T with rate constants k_R^{T*} and k_{NR}^{T*}, respectively. Thus, the kinetic description of the ESIPT reaction consists of six rate constants and the interplay of these constants defines the time-dependent decay of emission of both forms.

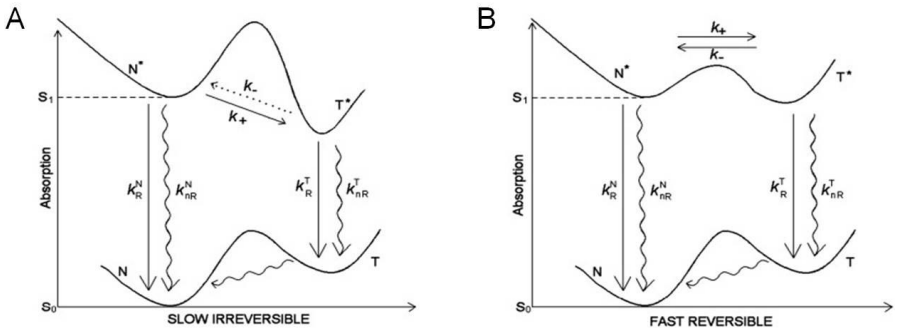


Figure 1.3.14. Schematic representation of two possible mechanisms of ESIPT reaction. (A) A slow irreversible reaction is observed when $k_– \rightarrow 0$ or $k_+ > k_–$. (B) In a fast reversible reaction, both forward and reverse reaction rates (expressed by k_+ and $k_–$) are fast. Adapted from Tomin, Oncul et al. 2007.

Depending on the nature of the N*↔T* transition, two distinct mechanisms can be observed (Demchenko, Tang et al. 2013; Tomin, Demchenko et al. 2015). In a first case, the relative populations of N* and T* forms are mainly regulated by the **kinetic factors**. In this case the energy of the excited tautomer state is much lower than the energy of the normal state (Figure 1.3.14A). This makes the reverse $k_–$ rate slow and largely shifts the equilibrium of the ESIPT reaction toward the T* form. However, the rate of this reaction may be slow due to some energy barrier. In a second case, the relative excited state distribution is **controlled thermodynamically**. In this case, the energies of the N* and T* forms are close, the k_+ and $k_–$

rate constants are similar and the thermodynamic equilibrium is reached very fast (Figure 1.3.14B). As a result, the two emission bands are of comparable intensities (Chou, Martinez et al. 1993; Swinney and Kelley 1993). Their intensity ratio is determined by the Boltzmann distribution of the excited-state species between the two forms (Klymchenko and Demchenko 2003).

Both the ESIPT mechanisms and spectroscopic properties of each 3HC derivative vary largely on their structure. The kinetic control was found to govern the parent 3HF structure, while the thermodynamic mechanism is observed for 3HF analogs with strong electron-donor substituent at position 2. Two representative examples were selected to illustrate the spectroscopic properties of 3HC: 4′-(diethylamino)-3-hydroxyflavone and 2-(2-furyl)-3-hydroxychromone (Figure 1.3.15).

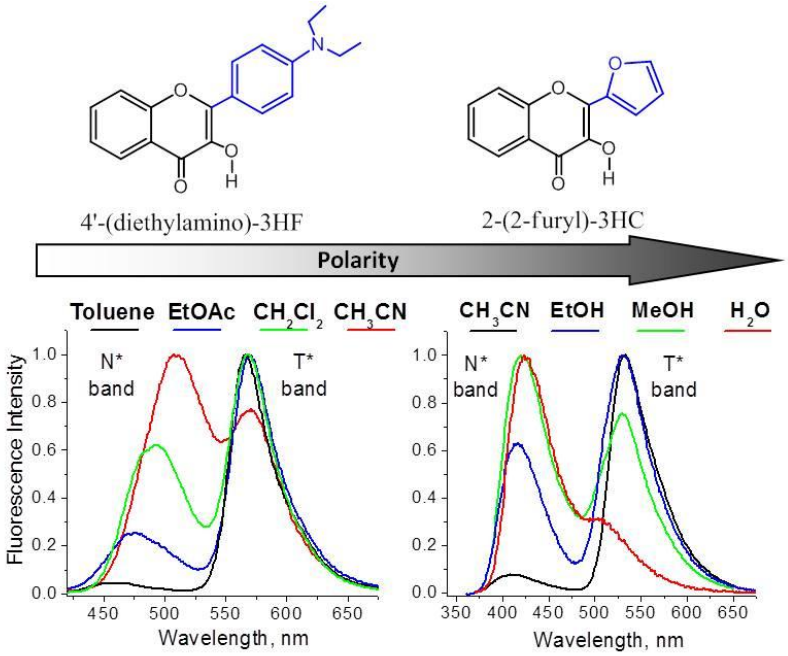


Figure 1.3.15. Structures of the selected 3HC dyes and their solvent-dependent emission spectra. Adapted from Klymchenko 2012.

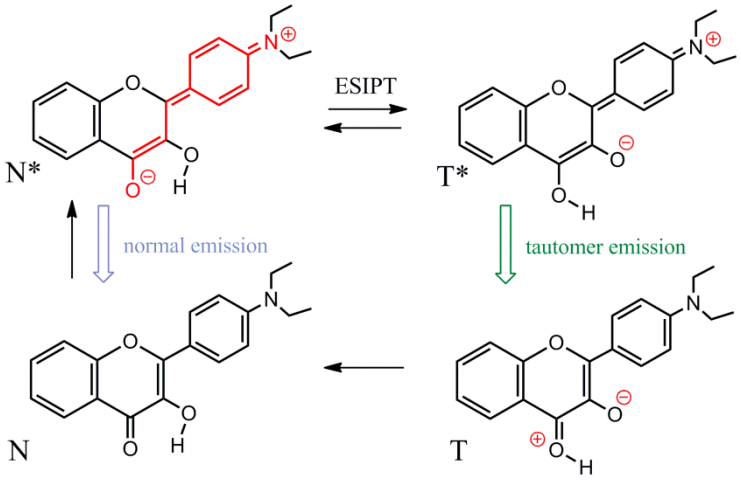


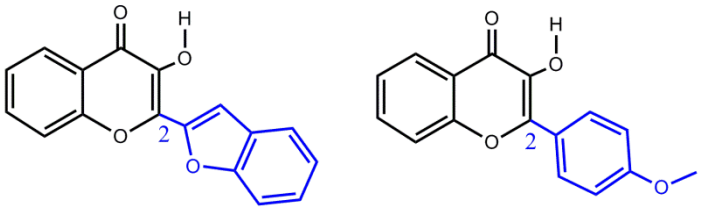
Figure 1.3.16. Photophysical cycle of 4′-(diethylamino)-3HF probe. Adapted from Klymchenko and Demchenko 2003.

4′-(diethylamino)-3HF dye exhibits a large dipole moment in the N* excited state due to an electron transfer from the dialkylamino group to the chromone moiety (Klymchenko and Demchenko 2003) (Figure 1.3.16). In contrast, the T* product exhibits a smaller charge separation and, thus, a smaller dipole moment. As a consequence, the N* band of the 4′-(diethylamino)-3HF emission spectra shows a significant red-shift upon an increase of the solvent polarity, while T* band maximum remains insensitive. Moreover, the red shift of the N* band is accompanied by an increase of the fluorescence intensity of the N* band as a result of the larger stabilization of the N* state as compared to the T* state (Shynkar, Mely et al. 2003). The fluorescence spectra of 4′-(diethylamino)-3HF performed in solvents with varying polarity indicated that the intensity ratio of the two bands, called N*/T*, is an important indicator of the solvent polarity (Figure 1.3.15).

Interestingly, 4′-(diethylamino)-3HF probe shows dual emission only in low polar and polar aprotic solvents, while in polar protic solvents the ESIPT is inhibited so that the T* band emission is not observed (Klymchenko and Demchenko 2003). Thus, 4′-(diethylamino)-3HF probe is suitable for sensing biological environment of low polarity and hydration, such as in lipid bilayers.

Other 3HC derivatives demonstrate improved photophysical characteristics in water and other polar protic solvents. In contrast to 2-dialkylamino-containing 3HC, derivatives that contain an aryl group in position 2 show a dual emission in polar protic solvents and significant fluorescence QY in water. This is the case of 2-(2-furyl)-3HC dye that was designed to probe highly polar media (Klymchenko, Pivovarenko et al. 2003; Klymchenko

and Demchenko 2004; Enander, Choulier et al. 2008; Das, Klymchenko et al. 2009) (Figure 1.3.15). This dye presents a relatively low dipole moment in its N* state due to the weaker electron-donor properties of the 2-furanyl group in comparison with the 4′-diethylaminophenyl group, so that the charge transfer character of its N* state is much less pronounced. Therefore, the N* band of 2-(2-furyl)-3HC emission spectra shows only a slight red-shift upon increase of the solvent polarity, but a significant N*/T* ratio change upon going from ethanol to water (Figure 1.3.15) (Klymchenko and Demchenko 2004). Since the N* and T* bands of 2-(2-furyl)-3HC emission spectra remain well separated even in highly polar solvents, it may be beneficially used for different molecular sensor applications such as for peptides and NAs studies.



2-benzofuryl-3HC **4'-methoxy-3HF**

Figure 1.3.17. Structures of 2-aryl-substituted derivatives of 3HC.

Other examples of 3HC probes with high sensitivity to polar protic environments are 2-benzofuryl-3HC and 4′-methoxy-3HF dyes (Figure 1.3.17) (Klymchenko and Demchenko 2004; Zamotaiev, Postupalenko et al. 2011). In comparison with 2-(2-furyl)-3HC, both probes in polar protic solvents show a superior QY. However, the studies of 2-benzofuryl-3HC in water were limited due to its low solubility (Klymchenko and Demchenko 2004). In contrast, 4′-methoxy-3HF dye displays higher sensitivity to solvent polarity and H-bond donor ability, allowing significantly improved detection of peptide interactions with NAs (Zamotaiev, Postupalenko et al. 2011; Sholokh, Zamotaiev et al. 2015). Moreover, 4′-methoxy-3HF exhibits higher pKa values of its 3-OH group, which avoids the interference of the anionic form, one of the limitations of the 2-(2-furyl)-3HC probe.

Based on the highly prominent properties of the 2-furyl-3-HC and 4’-methoxy-3HF, their amino acid analogs 3HCaa and M3HFaa, respectively, (Figure 1.3.18) were synthesized for sensing the environment in highly polar media (Strizhak, Postupalenko et al. 2012; Sholokh, Zamotaiev et al, 2015). By introducing these amino acids into different positions of the viral nucleocapsid peptide NC(11-55), the interactions of this peptide with ODNs were

monitored. Due to their small size and aromatic character, 3HCaa and M3HFaa were able to mimic the highly conserved Trp37 residue, critical for the biological activity of this peptide.

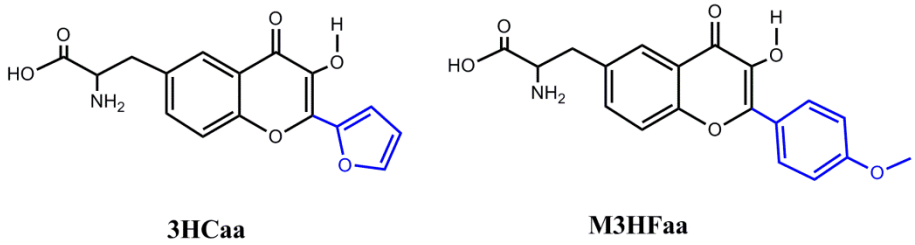
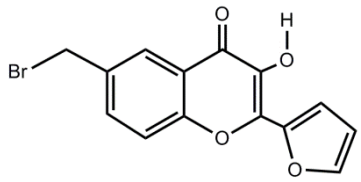


Figure 1.3.18. Structures of the 3-HC-based amino acid analogs.

1.3.7. Applications of 3HC probes and 3HC-based fluorescent amino acid analogs for monitoring biomolecular interactions

1. *Structure, conformation and function of the proteins.* 3HC probes can be incorporated site-selectively in proteins in order to study the protein conformational changes. For instance, the alpha-antitrypsin protein was labeled by 6-bromomethyl-2-(2-furanyl)-3-hydroxychromone (BMFC) (Figure 1.3.19) to investigate its conformation changes on interaction with proteinase enzymes such as the pancreatic elastase (Boudier, Klymchenko et al. 2009). This enzyme binds the reactive center loop of the alpha-antitrypsin resulting in a formation of a covalently bound intermediate complex that further rearranges into a stable inhibitory complex. The conformational changes were indirectly monitored through the labeling of a unique free Cys-232 residue at the C-terminus of the alpha-antitrypsin by the BMFC probe, a thiol-reactive derivative of 2-(2-furyl)-3HC (Figure 1.3.19). Upon interaction of the BMFC-labeled alpha-antitrypsin with pancreatic elastase the fluorescence changes of the probe (increase of the N*/T* ratio and decrease of the QY) suggested an exposure of the label at Cys-232 site to the bulk water. Further stopped-flow kinetic measurements of the BMFC-labeled protein – enzyme association allowed observing the well-studied fast formation of the intermediate complex and revealed an additional slow step likely corresponding to the protein-enzyme reorganization. Therefore, the BMFC probe appears as a convenient tool to study protein conformational changes.



BMFC

Figure 1.3.19. Structure of 6-bromomethyl-2-(2-furanyl)-3-hydroxychromone (BMFC).

2. *Binding events*. Binding of peptides with ligands generally decrease solvent exposure of the peptide binding sites that can be sensed by environment-sensitive amino acid analogs. Importantly, the sensitivity of the probe to its microenvironment should cover highly-polar environments (aqueous solutions) and it is desirable that the probe remains fluorescent when the peptide forms a complex with its target.

Peptide−NAs interactions can be illustrated by the site-selective labeling of the HIV-1 nucleocapsid peptide NC(11-55) by the 2-(2-furyl)-3-hydroxychromone (3HCaa) amino acid analog (Figure 1.3.18) (Strizhak, Postupalenko et al. 2012). This aromatic 3HCaa was used to substitute Tpr37 and Ala30 residues of the NC(11-55) peptide. Interactions of the peptide with its target ODNs, namely ΔP(–)PBS DNA, SL2 and SL3 RNAs, were explored using the 3HCaa. Based on the N*/T* ratio it was shown that 3HCaa is partially screened from water by the NC(11-55) peptide folding. Interactions with ODNs of the Trp37-substituted NC(11-55) led to a decrease in the N*/T* ratio and QY, and a red-shift of the T* band maximum (Figure 1.3.20) suggesting a drop of the local hydration of the 3HCaa probe when bound to ODNs. These changes were rationalized using the 3D structures of the NC/ODN complexes and were assigned to the stacking interactions of the Trp37 residue with the ODN nucleic bases. In contrast, limited changes were observed for the Ala30-labeled NC(11-55) in complexes with the ODNs (Figure 1.3.20), due to the large distance between the Ala30 amino acid and the nucleic bases in NC-ODN complexes. The obtained results show that the fluorescence response of the amino acid analog correlates with the proximity to the interacting ODN bases. Moreover, the changes in the N*/T* ratio or the QY of 3HCaa were used to measure the binding constant of NC−SL3 association that matched closely the constant obtained with the native peptide. This indicates that the fluorescent amino acid analog may be also used for the quantification of peptide−ODN interactions.

In comparison with Trp, 3HCaa possesses red-shifted absorption and emission, preserves its fluorescence on peptide binding to NAs, and presents higher environment sensitivity in highly polar media.

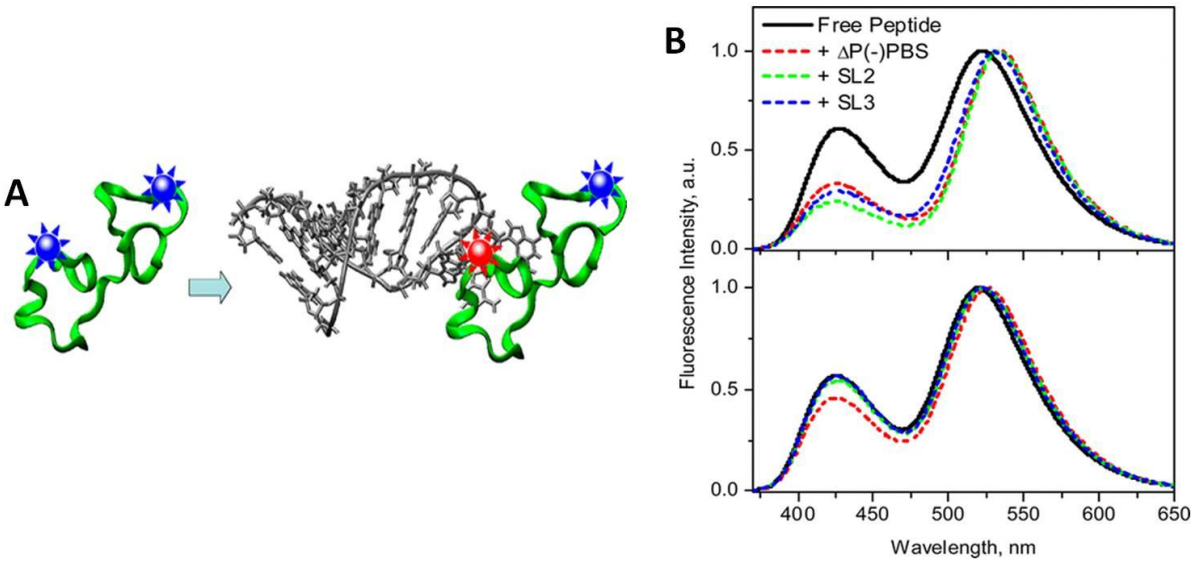


Figure 1.3.20. Monitoring peptide – ODN interaction using an environment-sensitive amino acid analog 3HCaa. (A) NC(11-55)-labeled peptide (in green) where 3HCaa replaced either Trp37 or Ala30 residues (blue stars) interacts with ODN, leading to changes in the 3HCaa emission; (B) Effect of ODN binding on the dual fluorescence of Trp37- (top) and Ala30- (bottom) substituted NC(11-55) peptides. As the target ODNs ΔP(–)PBS, SL2 and SL3 were used. Adapted from Strizhak, Postupalenko et al. 2012.

Protein – protein interactions result commonly in changes of polarity. Environment-sensitive probes were used as fluorescent biosensors for performing protein quantification assays (Enander, Choulier et al. 2008; Choulier, Shvadchak et al. 2010) (Figure 1.3.21). In these assays, the biosensor consists of a 3HC-labeled peptide containing a binding site recognised by antibody fragments. Interaction of antibody fragments with the peptide significantly changed the ratio of the 3HC emission bands due to exclusion of water molecules from the interaction site. Thus, ratiometric peptide-based biosensors can be designed for quantitative target detection.

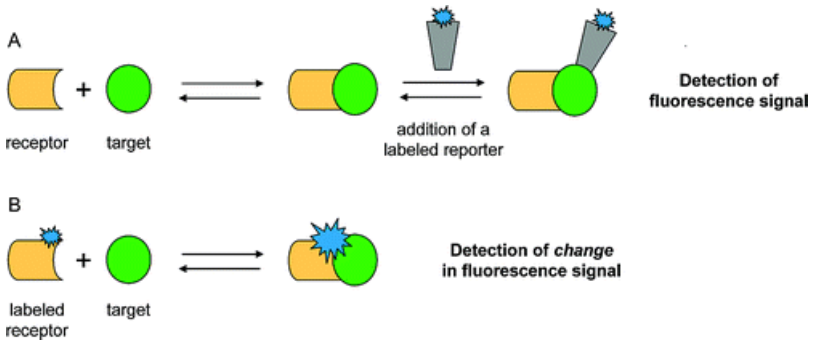


Figure 1.3.21. Comparison of the two strategies for protein quantification. (A) Conventional sandwich assay using a labeled reporter molecule for the detection. Target molecule should provide two distinct binding sites, and washing to separate free and bound molecules after each step is required. (B) Proposed molecular biosensor where the recognition and signaling events are integrated by the attachment of one or more fluorescent probe to the receptor. The interaction is detected by a change in the fluorescence signal. This assay requires a fluorophore with environment-sensitive properties to change its signal when the target molecule binds. Adapted from Enander, Choulier et al. 2008.

Peptide – lipid membranes interactions can be illustrated with the example of melittin, a cationic membrane-binding peptide of 26 amino acids, corresponding to the main toxic component of European honey bee venom (Postupalenko, Zamotaiev et al. 2013). Depending on the membrane properties and lipid/protein ratio, melittin can form pores with transmembrane perpendicular or parallel orientation to membrane surface. Using 4′-dimethylamino-3-HC amino acid analog, the insertion depth and orientation of melittin was determined in model and cellular membranes (Figure 1.3.22) (Postupalenko, Shvadchak et al. 2011; Postupalenko, Zamotaiev et al. 2013). Melittin labeled at several positions (Leu9 and Trp19) was added to model lipid bilayers and the N*/T* intensity ratio was determined from the emission spectra of the peptide-lipid complexes. The N*/T* ratio values varied linearly with the dielectric constant and thus, correlated with the environment polarity of the label. Using the dielectric model of the lipid bilayer, the depth of insertion of the probe into membrane could thus be determined. This depth of insertion was further confirmed using the parallax method (Postupalenko, Shvadchak et al. 2011). In cells, addition of the labeled melittin induced a disruption of the plasma membrane and formation of vesicles. Furthermore, melittin orientation in cell membranes could be further determined (Figure 1.3.22B). Altogether, the data show that melittin's α-helix binds parallel to the bilayer surface and does not adopt a transmembrane orientation in both model and cellular membranes.

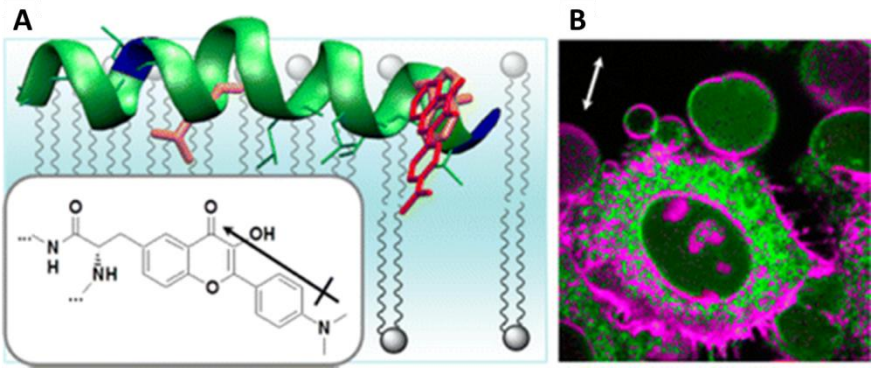


Figure 1.3.22. Melittin labeling, orientation and insertion in lipid membranes. (A) 3D structure of melittin oriented parallel to the lipid bilayer. In red leucine-9 and tryptophan-19 residues are shown. 4′-dimethylamino-3-HC amino acid analog is overlaid with Trp19 for size comparison. Inset: structure of 4′-dimethylamino-3-HC amino acid analog, black arrow shows the direction of the transition dipole moment of the fluorophore. (B) Two-photon fluorescence microscopy ratiometric image of HeLa cells in the presence of the labeled melittin is shown. The ratiometric image was obtained by recording two images in the blue and red channels and calculating I_{blue}/I_{red} intensity ratio. White arrow indicates the orientation of the laser light polarization (Postupalenko, Zamotaiev et al. 2013).

Part 4. Fluorescent nucleic acid analogs: diversity, properties and applications

1.4.1. Introduction to nucleic acids

Nucleic acids (NAs) are biological polymers coding the genetic information. There are two types of NAs: DNA and RNA. They are composed of four nucleic bases that are planar aromatic heterocycle molecules. Nucleic bases are divided into two groups: purines – adenine and guanine, and pyrimidines – cytosine, thymine in DNA and uracil in RNA. Their major tautomeric forms are shown on Figure 1.4.1. Attaching a sugar to the base forms a nucleoside. Nucleosides are divided into ribonucleosides (R = OH, Figure 1.4.1) when attached to a ribose, and deoxyribonucleosides (R = H) when attached to a 2'-deoxyribose. The pyrimidines are linked at N1 to the sugar; the purines are linked at N9.

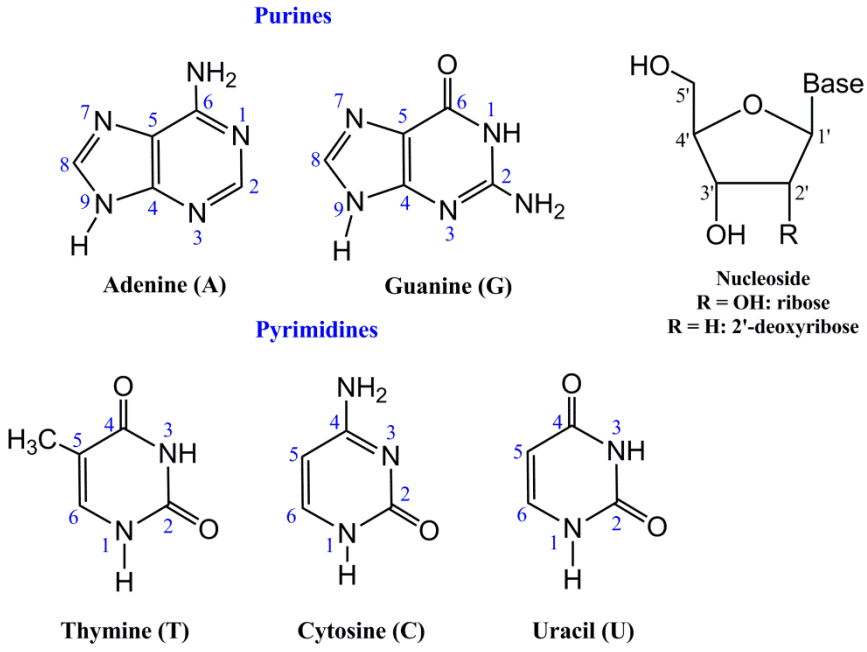


Figure 1.4.1. DNA and RNA building blocks.

Adding a phosphate group to a nucleoside forms a nucleotide. The linear sequence of nucleotides linked together by phosphodiester bonds forms the primary structure of NAs. An example of DNA primary structure is presented in figure 1.4.2A. The sequence of the bases is written in the 5′→ 3′ direction. The secondary structure of NAs results from a set of interactions between nucleobases. The secondary structures of DNAs and RNAs show significant differences: DNAs are mostly double-stranded, while RNAs are mainly single-

stranded with complex base-pairing interactions. In a double helix, the nucleobases are bound by hydrogen bonds to form adenine: thymine/uracil (A•T or A•U) and guanine : cytosine (G•C) Watson-Crick base pairs. The A•T/U base pair has two hydrogen bonds in comparison to three in the G•C base pair (Figure 1.4.2B). The distances between the C1′ atoms of sugars on opposite strands are almost identical for A•T/U and G•C base pairs, so they can form a regular helix for any sequence (Donohue and Trueblood 1960; Bloomfield 1999). The heterocyclic bases are found in *anti* conformation, where the hydrogen bonding groups are directed away from the sugar moiety. This allows for the negatively charged phosphate groups to be projected toward the polar media, while the aromatic bases are oriented toward each other creating a hydrophobic cavity. The anti-parallel helix forms two grooves: a major one and a minor one. The minor grooves occur where the backbones are close together and are thus, partly shielded from the solvent. Thus, most interactions occur on the major grooves, which are much more exposed (Bloomfield 1999).

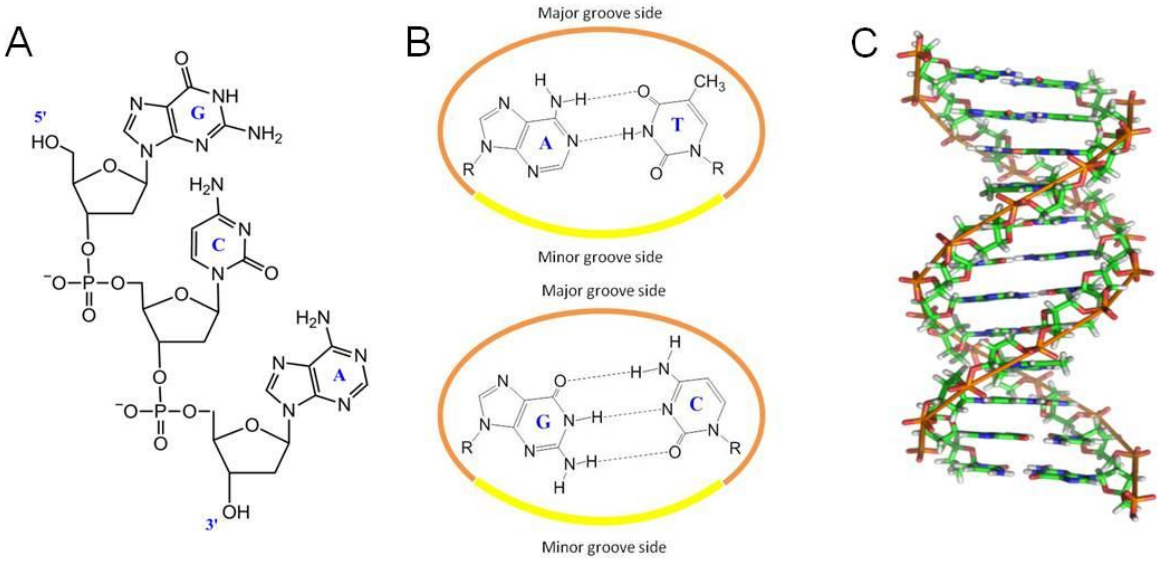


Figure 1.4.2. (A) GCA DNA primary structure, (B) Watson-Crick base pairs, and (C) B-DNA double strand conformation.

When the nucleobases are in their energetically dominant neutral tautomeric form, the purine-pyrimidine base pairs G•T/U and A•C typically do not adopt a Watson-Crick like geometry but rather form G•T, G•U or A^+•C wobble base pairs that deviate from the Watson-Crick geometry (Figure 1.4.3). There are other rarely observed purine-pyrimidine base pairs that undergo energetically unfavorable base ionization or tautomerization to form A•T or G•C^+ Hoogsteen base pairs where the purine base rotates ~180° about the N-glycosidic bond to change its *anti* conformation to a *syn* one (Hoogsteen 1963; Nikolova, Kim et al. 2011;

Kimsey and Al-Hashimi 2014). Similarly to the Watson-Crick base pairs, two H-bonds are preserved in A•T Hoogsteen base pair. In contrast, the formation of the G•C$^+$ Hoogsteen base pair requires protonation of N3 of cytosine and one H-bond is lost. Both base pairs require compression of the C1′-C1′ distance by about 2.5Å. Recent studies show that these base pairs are ubiquitous but transient and in low abundance, stabilized through intermolecular interactions or chemical modifications.

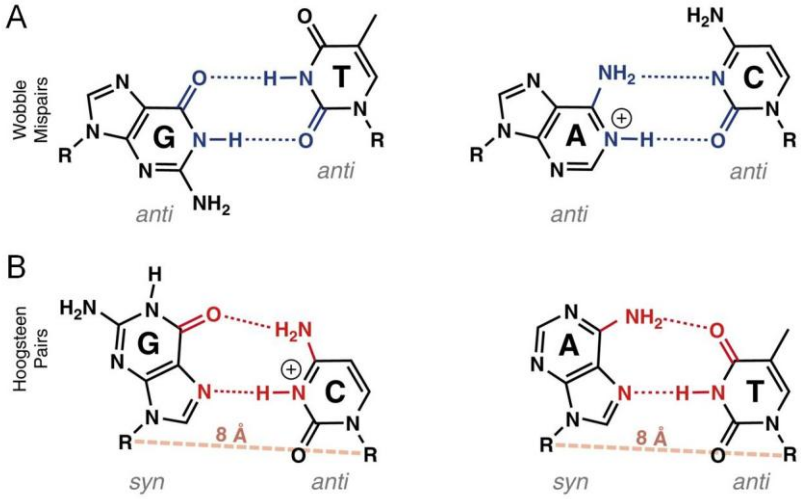


Figure 1.4.3. Low and high energy purine-pyrimidine base pairing. (A) Energetically favorable neutral G•T and ionized A$^+$•C wobble base pairs (hydrogen bonds in blue); (B) high energy G•C$^+$ and A•T Hoogsteen base pairs (hydrogen bonds in red), where purines are shown in the *syn* conformation. The average C1′-C1′ distance is shown in dashed red line. Adapted from Kimsey and Al-Hashimi 2014.

The tertiary structure of NAs describes their three-dimensional arrangement, based on geometric and steric constraints. The three major double-stranded DNA conformations are A, B and Z (Zimmerman 1982; Jovin, Soumpasis et al. 1987; Hartmann and Lavery 1996). Both A and B-DNA are right-handed double helical structures. Their interconversion is governed by the hydration conditions. B-DNA is the common DNA structure stable in high hydration conditions (Figure 1.4.2C), but is converted to A-DNA when the water access is limited (Franklin and Gosling 1953; Wahl and Sundaralingam 1997). In contrast to A and B, Z-DNA is a left-handed helical conformation with peculiar zig-zag disposition of the sugar-phosphate backbone. This structure is generally adapted by sequences of alternating purines and pyrimidines in the presence of high salt concentrations (Boon and Barton 2003).

1.4.2. Emissive nucleoside analogs and their classification

Natural NA building blocks are practically non-emissive in neutral aqueous solutions. This stimulated an extensive search for emissive nucleoside analogs. Since purine and pyrimidine heterocycles are susceptible for diverse modifications, one can find a large diversity of synthetic fluorescent nucleoside analogs in the literature (Wilson and Kool 2006). Their structure, properties and applications vary largely. Fluorescent nucleoside analogs can be site-selectively incorporated into NA strands by an enzymatic polymerization reaction or by chemical synthesis. As a result, their location and geometry are well-defined, and their signal reports on the structure, dynamics and biomolecular interactions of the NAs (Hawkins 2001; Wilhelmsson 2010; Dai and Kool 2011; St-Pierre, McCluskey et al. 2014); Wilson and Kool 2006; Sinkeldam, Greco et al. 2010).

An "ideal" fluorescent nucleoside analog should present the following characteristics (Tor 2009):

1) High structural similarity to the native nucleic base in order to faithfully mimic its size and shape, as well as its hybridization and recognition properties;
2) Red-shifted absorption and emission bands;
3) High quantum efficiency;
4) Sensitivity of its photophysical parameters to changes in the microenvironment.

NA analogs are often classified based on their structural characteristics. They can be grouped into isomorphic base analogs, polycyclic aromatic hydrocarbons, pteridines, nucleosides containing expanded nucleobases, nucleosides containing extended nucleobases, and universal base analogs.

Isomorphic base analogs

Isomorphic base analogs are one of the most applied sub-classes of NA analogs. They are based on aromatic heterocycles closely mimicking the corresponding natural nucleic bases. They preserve the overall dimensions and Watson-Crick base pairing of the natural nucleobases in order to induce minimal distortion when incorporated into ODNs. As with natural nucleobases, the NA backbone arranges the fluorescent analogs in a predictable location and orientation within the double helix. The design and synthesis of such molecules is always challenging since desirable photophysical properties are usually associated with

significant structural changes and extension of the conjugation (Sinkeldam, Greco et al. 2010).

The earliest known and most extensively used isomorphic nucleoside analog is *2-aminopurine* (Ward, Reich et al. 1969) (Figure 1.4.4). The high usefulness and still strong popularity of 2Ap rely on its ability to form Watson-Crick base pairs with T/U (Figure 1.4.4), its high quantum yield (0.68 in water), its isolated absorption band (303 nm), its strong environment sensitivity and availability (Ward, Reich et al. 1969; Guest 1991; Rachofsky, Osman et al. 2001; Jean and Hall 2002). However, incorporation of 2Ap into single and especially double stranded ODNs results in a strong decrease of its fluorescence quantum yield (Figure 1.4.5) (Guest 1991; Law, Eritja et al. 1996; Stivers 1998; Avilov, Piemont et al. 2008; Avilov, Godet et al. 2009) with appearance of large amounts of dark species (Larsen, Van Stokkum et al. 2001; Godet, Ramalanjaona et al. 2011; Godet, Kenfack et al. 2013). This strong quenching of 2Ap results mainly from a dynamic quenching by its neighbor residues due to a charge transfer process (Kelley and Barton 1999; Wan, Fiebig et al. 2000; Fiebig, Wan et al. 2002; O'Neill and Barton 2002; Larsen, Van Stokkum et al. 2001) or relaxation into a low-lying non-emissive electronic state (Jean and Hall 2001).

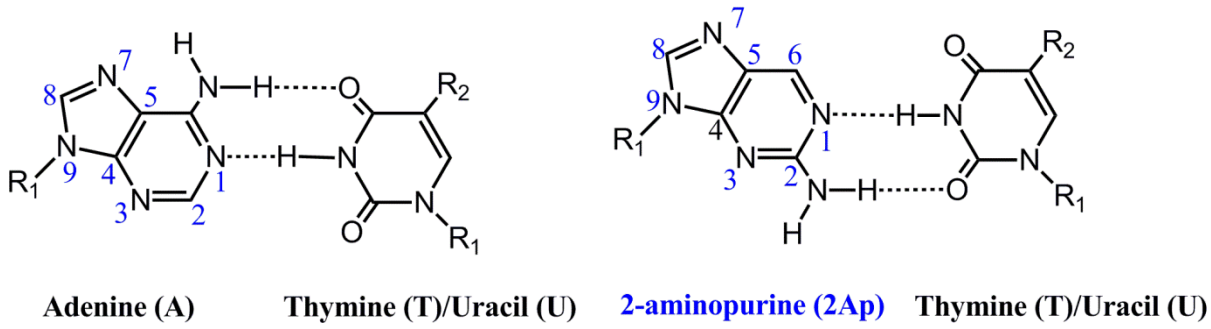


Figure 1.4.4. Hydrogen bonding of A-T(U) and 2Ap-T(U), where R_1 is 2′-deoxyribose or ribose, R_2 is CH_3 (T) or H (U).

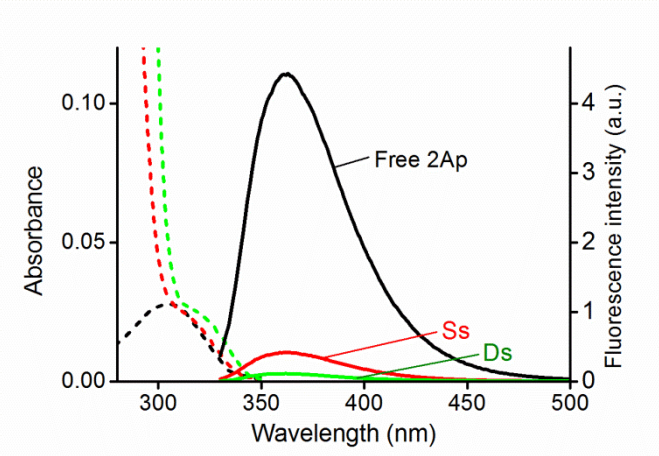


Figure 1.4.5. Absorption (dashed line) and emission (solid line) spectra of d2Ap in its free form in water (black), incorporated in a model single-stranded DNA (red) and in a model double-stranded DNA (green).

Moreover, 2Ap has a tendency to mispair with cytosine at much higher frequency than A, thus being mutagenic (Sowers, Fazakerley et al. 1986). Due to the difference in the amino group position between A and 2Ap, one of the hydrogen bonds with the opposite T is moved from the major to minor groove of the duplex on replacing A by 2Ap. This was shown to alter the dynamics and lower the stability of the duplex, by considerably reducing the lifetime of the 2Ap-T base pair and its immediate adjacent base pairs as compared to the corresponding duplex with the canonical A-T base pair (Dallmann, Dehmel et al. 2010).

The properties of 2Ap were successfully used in numerous applications, such as monitoring RNA/DNA conformational dynamics and heterogeneity (Hall 2009; Zhao and Xia 2009; Sakakibara, Abeysirigunawardena et al. 2012), DNA-protein interactions (Ward, Reich et al. 1969; Bandwar and Patel 2001; Goel, Kumar et al. 2013; Kuznetsova, Kuznetsov et al. 2014), dynamics and conformations of damaged DNA (Rachofsky, Alexander Ross et al. 2001; El-Yazbi and Loppnow 2012), riboswitch structural dynamics (Ilgu, Fulton et al. 2014; St-Pierre, McCluskey et al. 2014), mismatched base pair recognition (Guest 1991; Sharma, Doucette et al. 2013), local melting (Nordlund, Xu et al. 1993; Lee, Waters et al. 2012) and quadruplex formation (Johnson, Okyere et al. 2013; Gray, Trent et al. 2014).

Numerous other fluorescent nucleoside analogues were designed (review (Sinkeldam, Greco et al. 2010)) to replace either A (Ben Gaied, Glasser et al. 2005; Kenfack, Piémont et al. 2008), C (Liu and Martin 2001; Berry, Jung et al. 2004), U (Xie, Dix et al. 2009) or T (Greco and Tor 2005; Greco and Tor 2007; Srivatsan and Tor 2007).

In the case of **fluorescent G analogues**, *3-methyl-isoxanthopterin* (3-MI) *and 6-methyl-isoxanthopterin* (6-MI) (Charubala, Maurinsh et al. 1997; Driscoll, Hawkins et al. 1997; Widom, Rappoport et al. 2013) (Figure 1.4.6) were the first emissive guanosine analogs, belonging to the family of pteridines. Pteridines is a class of naturally occurring highly emissive bicyclic heterocycles, structurally similar to purines, developed mainly by the group of Hawkins (Hawkins 2001; Hawkins 2008). 3-MI and 6-MI are both highly fluorescent (quantum yields of 0.88 and 0.70 in TRIS buffer, respectively), absorb at around 340 nm and emit at 430 nm (Hawkins 2001). Their fluorescence emission is quenched upon incorporation into single- and double-stranded ODNs. As for 2Ap, the extent of their fluorescence quenching constitutes a measure of the degree of stacking between the inserted base analog and its flanking bases. 3-MI and 6-MI fluorescent analogs found their application mainly in systems where their fluorescent signal increased due to the weakening of base-stacking interaction. 3-MI was applied in a real-time assay for monitoring the HIV-1 integrase 3′-processing reaction. By incorporating the probe at the cleavage site of the double-stranded DNA, an increase of the fluorescence signal was observed when the integrase cleaved the 3′-dinucleotide with 3-MI (Hawkins, Pfleiderer et al. 1995). 6-MI was applied for the detection of bulges in DNA strands and studying DNA conformational changes (Datta, Johnson et al. 2012).

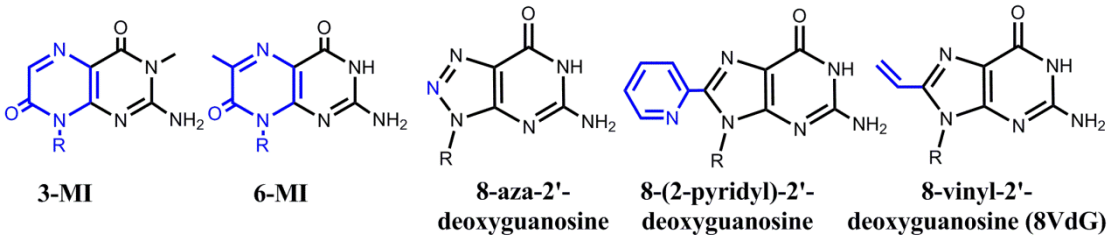


Figure 1.4.6. Fluorescent G analogs (R = ribose or 2′-deoxyribose).

Next, *8-azaguanosine* was developed (Seela, Jiang et al. 2009) (Figure 1.4.6), a perfect shape mimic of G, forming a strong base pair with C and displaying pH-dependent fluorescence (pK_a = 8.4). Its anionic form is ~3.5-5-fold brighter than the neutral one. Incorporation of 8-azaguanine into single-stranded DNA induces a 10-fold decrease of its QY, that is comparable to that of 2Ap. In double strands the most important quenching of 8-azaguanine was observed in matched duplexes, while it was less quenched in mismatches. A change in the pK_a values upon transition from single-stranded to double-stranded ODNs and different emission signal in duplexes allow 8-azaguanine to be applied for the detection of the

base pair microenvironment, ionization states of nucleobases in NAs or single nucleotide polymorphism.

Recently, the fluorescent analogs *8-(2-pyridyl)-2 '-deoxyguanosine (2PyG)* (Dumas and Luedtke 2010) and *8-vinyl-2'-deoxyguanosine (8VdG)* (Nadler, Strohmeier et al. 2011) (Figure 1.4.6) were synthesized and applied for sensing G-quadruplexes. 2PyG exhibits a much higher QY in acetonitrile (~0.71) than in water (~0.02) and shows an environmental sensitivity. Noticeably, the probe is minimally disruptive to G-quadruplex folding and exhibit relatively high QY when base-stacked with other dG residues (0.04-0.15). Therefore, it was applied for sensing G-quadruplex folding and energy transfer.

8VdG preserves the Watson-Crick H-bonding and the overall geometry of its parent molecule. The fluorescence QY of the free probe is about 0.72, but is 9-12-fold quenched upon incorporation into single-stranded DNA, and undergoes even more pronounced drop of its QY in duplexes. Neighboring dG and dT nucleobases led to the highest quenching of 8VdG fluorescence, while adjacent dC and dA were less influential. Interestingly, 8VdG was shown to be capable of adopting both the nucleotide *syn* and *anti* conformations required for different quadruplex structures. Thus, the probe was applied to differentiate between topologies of G-quadruplexes. Moreover, 8VdG was shown to be processed by DNA polymerases to create fluorescent DNA, being useful for investigating DNA hybridization and topology (Holzberger, Strohmeier et al. 2012).

Thieno[3,4-d]-guanosine (thG) was designed as a highly isomorphic fluorescent analog of G within the emissive ribonucleoside alphabet derived from thieno[3,4-*d*]pyrimidine (Figure 1.4.7) (Shin, Sinkeldam et al. 2011).

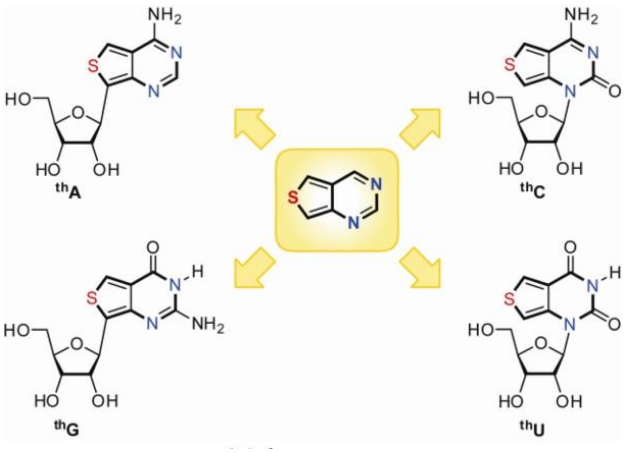


Figure 1.4.7. Isomorphic fluorescent nucleoside analogs derived from a thieno[3,4-*d*]pyrimidine core (Shin, Sinkeldam et al. 2011).

Importantly, thG was found to faithfully replace G with minimal perturbation and exhibit interesting photophysical properties, being endowed with good quantum yield, emission in the visible spectrum and environment sensitive spectral properties. Theoretical calculations showed that thG can form a Watson-Crick base pair with complementary C with similar H-bonding energy as for the natural G-C base pair (Samanta and Pati 2013). Moreover, thGTP has been also used as a GTP surrogate in T7 RNA transcription reactions (McCoy, Shin et al. 2014). After enzymatic incorporation of thGTP, the modified nucleotide could initiate and maintain transcription reactions, leading to the formation of fully substituted and highly emissive RNA transcripts with thG replacing all guanosine residues. The modified substrate was effectively cleaved by the natural RNA enzyme, confirming the isomorphic character of the thG nucleotide and its ability to maintain the proper folding of the substrate.

Next, *thieno[3,4-d]-deoxyguanosine* (dthG), the DNA form of the thG analog was synthesized and characterized (Park, Otomo et al. 2014; Sholokh, Sharma et al. 2015). This compound exhibited spectroscopic properties similar to thG and was successfully applied for monitoring B-Z DNA transition and sensing specific interactions with peptides (Park, Otomo et al. 2014).

A next generation of isomorphic nucleoside analogs that present even higher structural and electronic similarity to the natural nucleic bases has been recently reported (Figure 1.4.8) (Rovira, Fin et al. 2015). The new alphabet was synthesized using an isothiazolo[4,3-*d*]pyrimidine core that contains a basic nitrogen corresponding to N7 of the purine core. This modification allows improving the functionality of the NA analogs since many biomolecular

interactions of purine nucleosides/nucleotides rely on the basicity and coordinating ability of this residue.

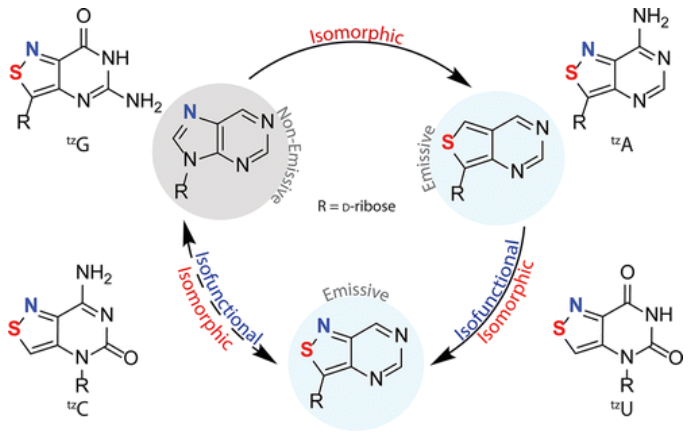


Figure 1.4.8. Isomorphic and isofunctional fluorescent nucleoside analogs derived from isothiazolo[4,3-*d*]pyrimidine core, R = ribose (Rovira, Fin et al. 2015).

The photophysical properties of the ^{tz}-analogs were found to be similar to the th-analogs. They present similar absorption and emission maxima and environment sensitive properties, however, significantly lower QY values in water (2-40-fold decrease). The utility of the ^{tz}-analogs was demonstrated on the example of the enzymatic deamination of adenosine to inosine, when adenosine deaminase was able to deaminate ^{tz}A as effectively as adenosine, the native substrate.

Chromophoric base analogs

The chromophoric base analogs replace the natural nucleobases with established fluorophores. These analogs typically lack the Watson-Crick hydrogen bonding groups, so they do not participate in base pairing. However, they can substitute any of the natural nucleobase and thus, are considered as universal nucleoside analogs.

Polycyclic aromatic hydrocarbons (PAH) compose a large family of chromophoric base analogs with isolated absorption bands and high quantum yields (Wilson and Kool 2006; Sinkeldam, Greco et al. 2010). Pyrene and perylene-based nucleosides (Figure 1.4.9) show emission properties similar to their parent fluorophores (emission maxima at 375 and 443 nm, QY of 0.12 and 0.88 in methanol for pyrene and perylene analogs, respectively) (Ren, Schweitzer et al. 1996; Gao, Strassler et al. 2002). Since PAH analogs are significantly less polar than natural nucleoside structures, they favor stacking interactions and were mostly

applied for studying non-covalent interactions of ODNs (i.e. host-guest interactions) (Schweitzer and Kool 1994).

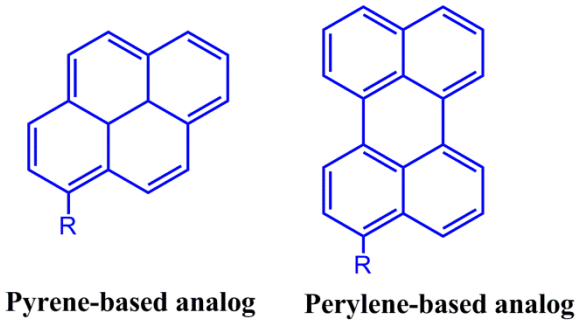


Figure 1.4.9. Examples of polycyclic aromatic hydrocarbons.

Another example is a nucleoside surrogate bearing a 2-thienyl-3-hydroxychromone fluorophore as a nucleobase attached directly to the sugar (Figure 1.4.10) (Dziuba, Postupalenko et al. 2012). It was shown that this analog can replace any natural nucleobase in single- and double stranded DNAs with minimal alteration of their thermodynamic stability. In comparison with existing nucleoside analogs, this probe demonstrates all advantages of the ratiometric detection, with the ratio of the two emission bands being sensitive to the polarity and hydration changes of the environment. This analog was applied to monitor the local conformational changes of model single-stranded ODNs upon interaction with the HIV-1 nucleocapsid protein NC (Dziuba, Postupalenko et al. 2012).

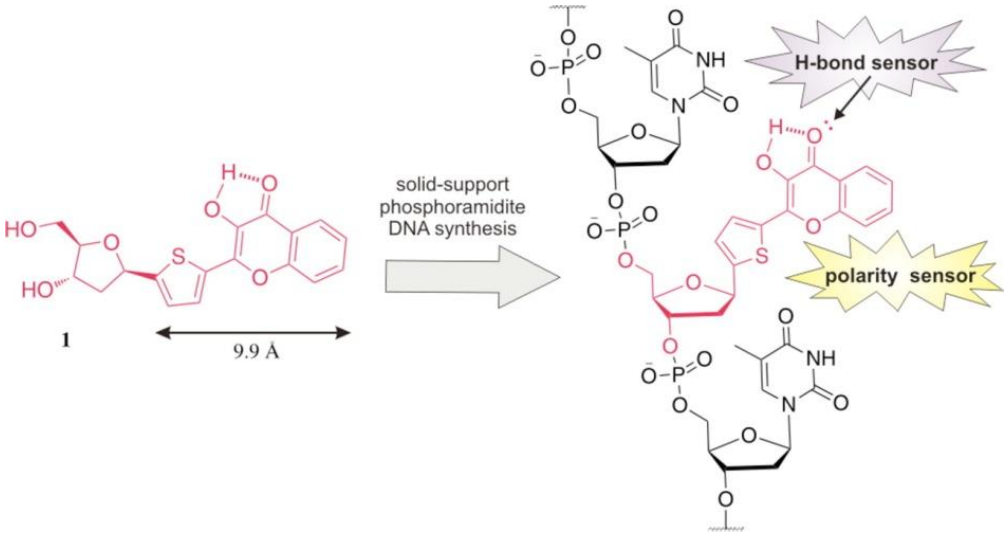


Figure 1.4.10. Hydrogen-bonding- and polarity-sensitive nucleoside 1 containing 2-thienyl-3-hydroxychromone as a base surrogate and its incorporation into DNA (Dziuba, Postupalenko et al. 2012).

Nucleosides with expanded nucleobases

Fusing additional aromatic rings to purine and pyrimidine moieties extends the conjugation of the natural nucleobases and generates diverse DNA-like systems with expanded diameters. By addition of a benzene ring, a bicyclic purine is converted into a three-ring analog, and a monocyclic pyrimidine into a bicyclic structure. One example of these compounds is *benzoadenosine* (Figure 1.4.11), firstly synthesized by Leonard and co-workers in 1970s to sense the active sites of ATP-dependent enzymes, but was not studied in DNA (Leonard, Morrice et al. 1975; Scopes, Barrio et al. 1977). Later the research group of E. Kool has designed a complete set of benzo-homologated ribo- and deoxyribonucleosides (xRNA/xDNA) (Figure 1.4.11) (Liu, Gao et al. 2004; Krueger and Kool 2008).

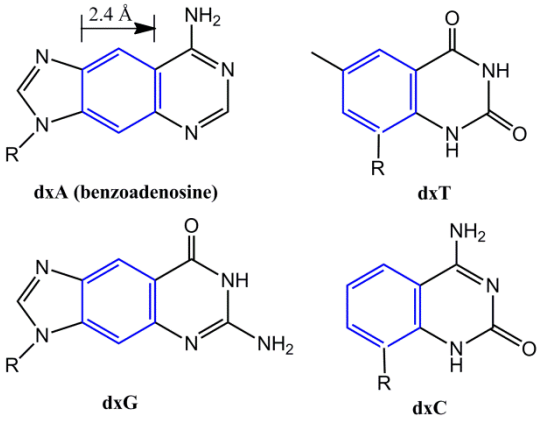


Figure 1.4.11. Size-expanded xDNA nucleosides, R = ribose or 2′-deoxyribose.

Insertion of a benzene ring into a natural nucleoside moiety leads to a 2.4 Å expansion of the base. Nevertheless, the lengths and angular shifts in the hydrogen-bonding faces of the bases are preserved, thus enabling the formation of a regular helical conformation when paired (Winnacker and Kool 2013). Using these expanded nucleobases, various DNA helixes were designed where a benzo-homologated base forms a base pair with its complementary natural base, yielding a fully expanded duplex composed of eight genetic letters instead of four. Such DNA-like molecules show useful properties, like high binding affinity, high helix stability and fluorescence. Because of the enlarged bases, xDNA stack much more strongly that their natural counterparts, resulting in an increased stability of the duplexes. Free xDNA bases are brightly fluorescent (QY in methanol 0.30-0.52), with emission maxima at 380-410 nm. When incorporated into single-stranded DNA the xC and xT bases are quenched by about 2-fold, while xA and xG by about 5-fold. Further fluorescence quenching in the duplexes is highly dependent on the opposite base inducing either small or 2-4-fold decrease in xDNA

base emission (Krueger and Kool 2008). The potential of benzo-expanded DNA to encode and transfer the genetic information was explored. It was found that polymerases selected the correctly paired partner for each xDNA base but with lower efficiency than with the natural bases (Lu, Krueger et al. 2010). This artificial genetic system without the constraints of purine-pyrimidine geometry provides more flexibility in molecular designs, interesting for addressing bioanalytical questions and potentially useful for the development of new biotechnological tools.

Nucleosides containing extended nucleobases

Finally, extended NA analogs were made of natural purines and pyrimidines linked to fluorescent moieties via flexible or rigid linkers. *Pyrene-based* base-discriminating fluorescent (BDF) probes reported by the group of Saito are an example of such compounds (Figure 1.4.12). In these systems, the natural nucleoside is coupled via a short tether to an emissive pyrene (Okamoto, Tainaka et al. 2005; Okamoto, Tainaka et al. 2006) working as an antenna. When properly paired, pyrene is located outside the duplex and is highly fluorescent (Figure 1.4.12). In the mismatch, it is believed that the pyrene moiety instead of the nucleobase is inserted into the duplex, and is thus highly quenched by the neighboring bases (Okamoto, Kanatani et al. 2004). A homogeneous single-nucleotide polymorphism typing assay has been also developed based on this principle (Okamoto, Tainaka et al. 2006).

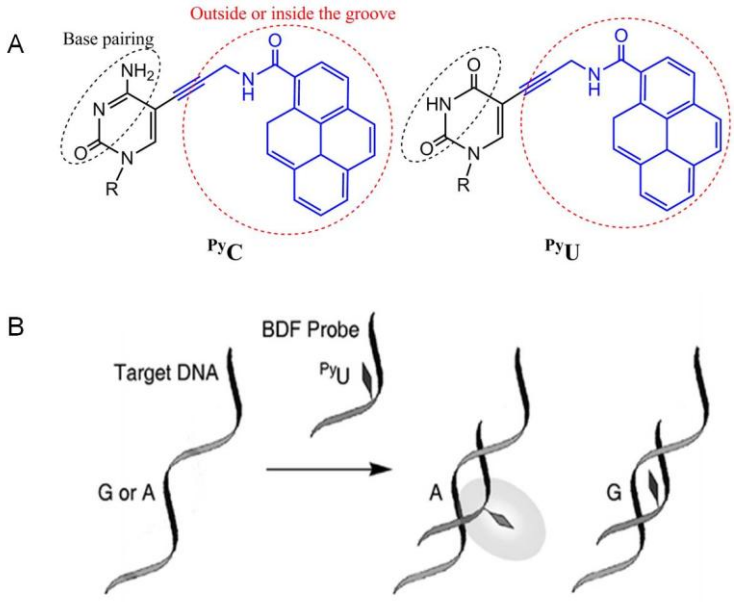
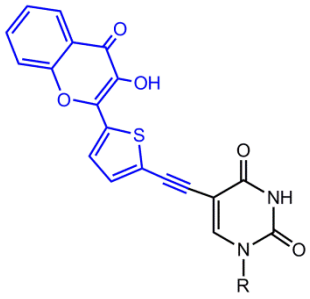


Figure 1.4.12. Base discrimination by fluorescent nucleoside analogs. (A) Structures of pyrene-based nucleoside analogs containing extended nucleobases. (B) Schematic illustration of a single nucleotide

polymorphism typing assay principle using ^{Py}U probe. The probe shows a strong fluorescence only when its opposite base is the complementary one (adenosine). Adapted from Okamoto, Tainaka et al. 2006.

Another interesting example of a fluorescent extended nucleobase analog is the *TCeU* probe composed of a 2-thienyl-3-hydroxychromone fluorophore and an uracil aromatic fragment (Figure 1.4.13) developed in the group of A. Burger (Dziuba, Karpenko et al. 2014). The two aromatic electron-donor and electron-acceptor moieties are assembled through an electron-conducting acetylenic tether resulting in advanced fluorescent properties. This probe exhibits red-shifted absorption and emission, strong brightness and environment sensitivity. Moreover, it shows dual emission corresponding to the ESIPT reaction, characteristic for the 3HC probes. This dye constitutes a promising building block for NA labeling.



TCeU

Figure 1.4.13. Example of the 3HC-containing expanded nucleoside analog, R = 2′-deoxyribose.

1.4.3. Applications of fluorescent nucleoside analogs

To summarize, unlike traditional fluorescence probes, which mostly report overall changes, nucleoside analogs incorporated in ODNs report local changes occurring around the site of interest at the nucleotide level. Mainly, fluorescent nucleoside analogs have found their applications in:

- Structural, conformational, and dynamical studies of NAs. Different forms, conformations, aggregated states and polymorphs of NAs can be related to their specific cellular function. Suitable fluorescent nucleosides can provide a photophysical signature of NA hybridization, folding or conformational change.
- Single nucleotide polymorphism (SNP) detection is a rather recent application of the emissive nucleobase analogs (Wilson and Kool 2006). SNPs have been linked to specific diseases or liability to particular therapeutics. As a

consequence, many efforts have been focused on the design of base-discriminating nucleobases. Various modes of detection are used including emission enhancement, quenching or shifts in emission wavelength of the fluorescent nucleosides.

- Microenvironment of the NAs. Thee polarity of NA grooves as well as base flipping, base stacking and nucleobase damage can be monitored using environment-sensitive fluorescent nucleoside analogs.
- Ligand binding. Fluorescent nucleoside analogs found wide application in reporting NA binding events. In this case, isomorphic nucleobases are of special interest due to their minimally perturbing nature for the folding and binding patterns of NA - ligand complexes.

Research objectives

One of the main research axes of the Laboratory of Biophotonics and Pharmacology is the investigation of the molecular and cellular mechanisms implied in HIV-1 infection using an interdisciplinary approach (physics, chemistry, pharmacology, biochemistry, cellular and molecular biology) in order to propose new therapeutic strategies. The highly conserved HIV-1 nucleocapsid protein NC plays several key roles in the HIV-1 replication cycle and, thus, appears a potential target for the anti-viral research. The functions of NC rely mainly on its chaperone activity which drives the rearrangement of NAs into their most stable conformations and favors the hybridization of complementary NA sequences. Although the mechanistic understanding of the chaperone activity of NC has achieved important progresses, its dynamic aspects are still not completely understood, mainly due to the lack of appropriate tools to monitor them on the molecular level.

In this context, the objective of my thesis project was to characterize on the molecular level the dynamics of the interaction of NC with NAs using the fluorescence tools developed in the laboratory or in collaboration. Aiming to follow the interaction from both sides, we used fluorescent amino acid analogs introduced into the NC protein and fluorescent nucleoside analogs introduced in the ODNs.

To label NC, a ratiometric amino acid analog from the 3-hydroxychromone family was used to replace the residues Trp37 in the distal zinc finger and Ala30 in the linker between the two zinc fingers of the peptide. The probe presents a multiparametric response to the changes in its local microenvironment, allowing us to obtain site-selective information about the NC/NA binding, notably concerning the conformations of NC in complex with DNA/RNA molecules.

In a second approach, in order to monitor the dynamics of ODNs complexed with NC, two types of fluorescent nucleoside analogs were applied: an isomorphic guanosine analog – thienodeoxyguanosine and a universal nucleoside analog – 2-thienyl-3-hydroxychromone. The advantage of isomorphic analogs is their strong resemblance with the natural nucleosides and their ability to form Watson-Crick base pairs with the complementary bases. The advantage of universal nucleoside analogs is their ability to replace any natural base, and in the case of 3HC-based nucleoside analogs, their strong sensitivity to the microenvironment. The complementary information from the two probes should allow us getting a complete

picture about the peptide – ODN and ODN− ODN interactions during the (–)/(+)PBS annealing process of the HIV-1 reverse transcription step. This understanding is fundamental to develop rational strategies in order to target the specific role of NC in this step.

This work would never be possible without our kind collaborators:

- Dr. O. Zamotaiev, Dr. V. Postupalenko, Dr. L. Richert, Dr. D. Dujardin, Dr. C. Boudier, Dr. C. Kenfack and Dr. A. Klymchenko from the Laboratory of Biophotonics and Pharmacology, Université de Strasbourg.
- Prof. Dr. V. Pivovarenko from the Department of Chemistry, Kyiv National Taras Shevchenko University, Ukraine.
- Dr. D. Dziuba, Dr. N. Barthes, Dr. B. Michel and Dr. A. Burger from the Institut de Chimie de Nice, Université de Nice Sophia Antipolis.
- Dr. D. Shin and Prof. Dr. Y. Tor from the Department of Chemistry and Biochemistry, University of California, San Diego, USA.
- Dr. R. Das from the Department of Chemistry, West Bengal State University, India.
- Dr. R. Improta from the Consiglio Nazionale delle Ricerche, Istituto di Biostrutture e Bioimmagini, Italy.
- Dr. M. Mori and Prof. Dr. M. Botta from the Dipartimento di Biotecnologie, Chimica e Farmacia, Universita degli Studi di Siena, Italy.
- Dr. K. Voltz and Dr. R. Stote from the Department of Integrative Structural Biology, IGBMC, Université de Strasbourg.

Chapter 2. Materials and Methods

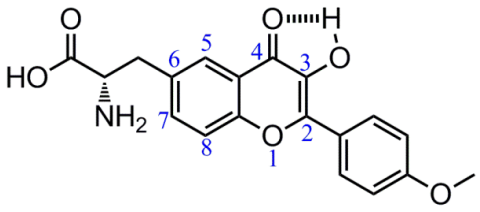
Part 1. Materials

2.1.1. Synthesis of fluorescent amino acid and nucleoside analogs

The solvents and chemicals used in this work were purchased from Sigma-Aldrich Chemical Company, Fluka, Carl Roth and Fisher Scientific, unless mentioned otherwise. For absorption and fluorescence studies, solvents were of spectroscopic grade. Mass spectra were recorded on Bruker HCT Ultra and Agilent Technologies Accurate-Mass Q-TOF LC/MS 6520 mass spectrometers.

Synthesis of the 4′-methoxy-3-hydroxyflavone amino acid analog

The fluorescent amino acid analog 4′-methoxy-3-hydroxyflavone (M3HFaa) (Figure 2.1.1) was synthesized by Dr. Oleksandr Zamotaiev in collaboration with the group of Prof. Vasyl Pivovarenko from the Kyiv National Taras Shevchenko University. The scheme and protocol of the synthesis of Fmoc-M3HFaa are described in (Sholokh, Zamotaiev et al. 2015).

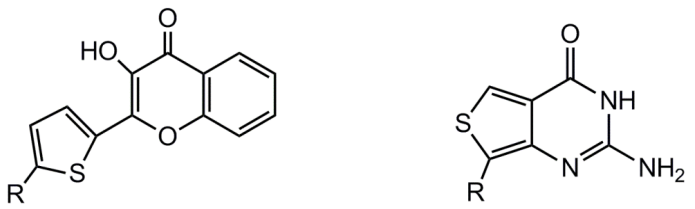


4'-methoxy-3-hydroxyflavone (M3HFaa)

Figure 2.1.1. Structure of the fluorescent amino acid analog M3HFaa.

Synthesis of the nucleic acid analogs: 2-thienyl-3-hydroxychromone and deoxythienoguanosine

The fluorescent nucleoside analogs 2-thienyl-3-hydroxychromone (3HCnt) and 2-aminothieno-[3,4-*d*]pyrimidin-4(3*H*)-one-7-*β*-D-2′-deoxyribofuranoside (deoxythienoguanosine, $d^{th}G$) (Figure 2.1.2) were synthesized by the group of Prof. Alain Burger from the Nice Sophia Antipolis University and the group of Prof. Yitzhak Tor from the University of California, San Diego, respectively. The detailed protocols of their syntheses are described in (Dziuba, Postupalenko et al. 2012; Sholokh, Sharma et al. 2015).



2-thienyl-3-hydroxychromone nucleoside (3HCnt) Deoxythienoguanosine ($d^{th}G$)

Figure 2.1.2. Fluorescent nucleoside analogs, where R is 2′-deoxyribose.

2.1.2. Peptide synthesis

The HIV-1 nucleocapsid protein, NC(11-55), lacking the N-terminal domain, and labeled NC(11-55)-W37-M3HFaa and NC(11-55)-A30-M3HFaa peptides, where M3HFaa probe substitutes Ala30 or Trp37 residues, respectively (Figure 2.1.3), were synthesized by the solid-phase peptide synthesis on 433A synthesizer (ABI, Foster City, CA) as previously described (de Rocquigny, Ficheux et al. 1991; Strizhak, Postupalenko et al. 2012).

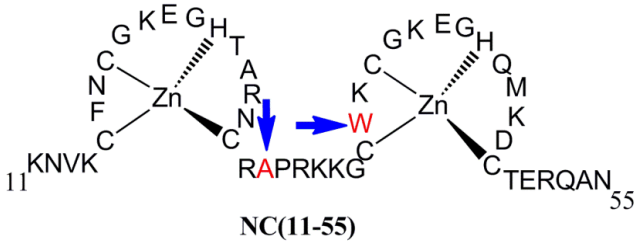


Figure 2.1.3. Structure of NC(11-55) peptide.

Since its introduction in 1963 by Bruce Merrifield, solid-phase peptide synthesis has become the most successful method for synthesizing peptides. It generally involves several steps:

1) Chain assembly: deprotection, activation, coupling;
2) Cleavage from resin and removal of side-chain protecting groups;
3) Purification;
4) Characterization.

Chain assembly usually uses two strategies based on the Fmoc and Boc groups that protect the α-amino group of an amino acid. The peptide is assembled from the C-terminal towards the N-terminal with the α-carboxyl group of the amino acid attached to a solid support. The first step in chain assembly is deprotection, or removal of the protecting group. After deprotection, the next amino acid is coupled to the deprotected amino end of the

growing peptide, forming a peptide bond. Chain assembly is thus a series of deprotections and couplings. Schematically, the chain assembly may be illustrated by Figure 2.1.4:

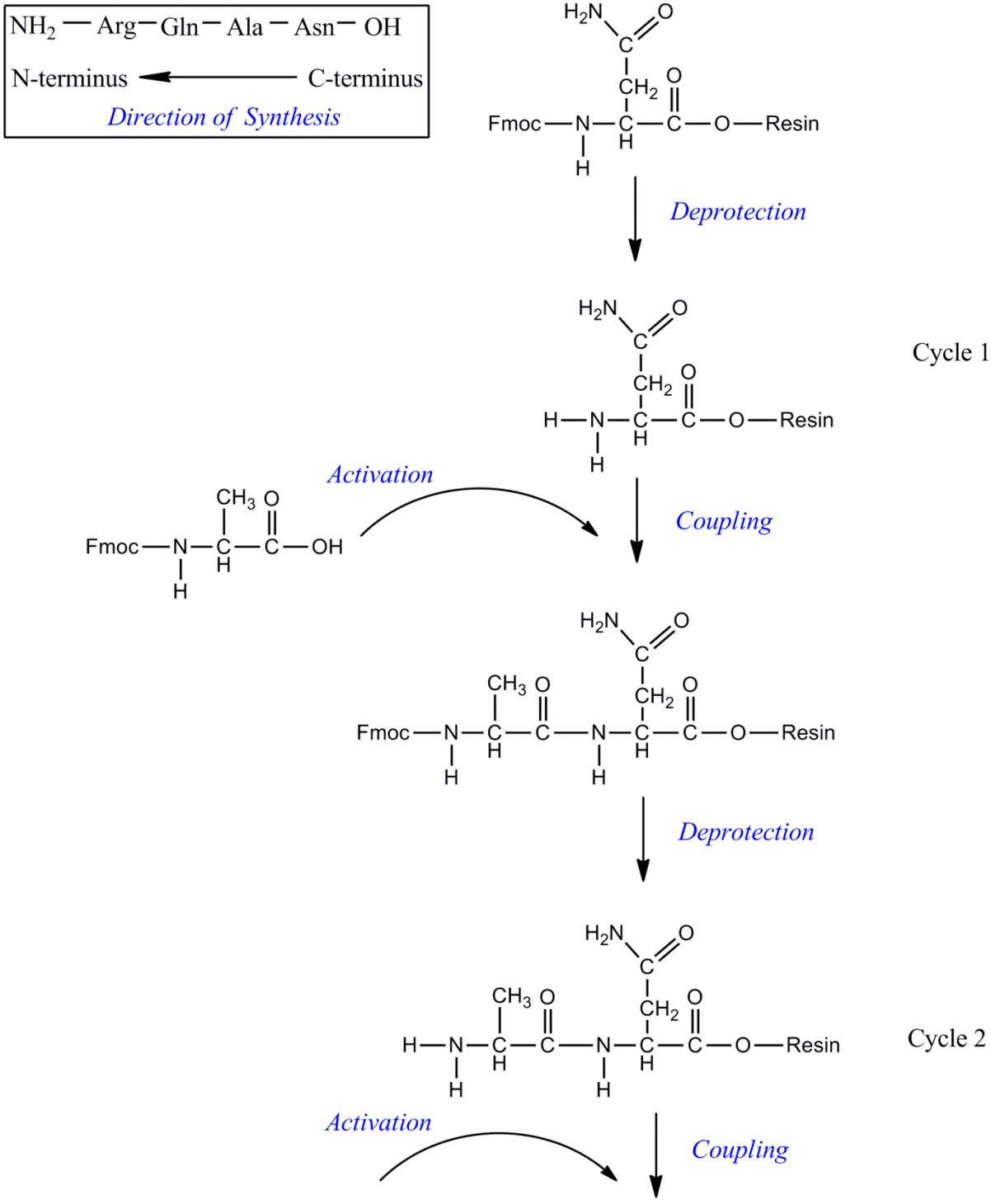


Figure 2.1.4. Scheme of the chain assembly of the solid-phase peptide synthesis.

Detailed protocol of the labeled NC(11-55) peptide synthesis

- *Synthesis*

The procedure was performed at 0.1 mmol scale using the standard 9-fluorenylmethoxycarbonyl (Fmoc)-amino acid-coupling protocol on a 433A synthesizer (ABI, Foster City, CA). LL-HMP resin (ABI, 0.44 mmol/g reactive group concentrations) was used as a solid support. Deprotection steps were performed by piperidine and automatically controlled by UV absorbance. At the end of the synthesis, the peptidylresin was isolated, and washed with NMP, methanol and dichloromethane.

- *Labeling*

At selected positions of the peptide, the fluorescent amino acid analogue M3HFaa was incorporated using the following procedure. In a flask, 2 mol equivalents (0.11 mmol) of the Fmoc-M3HFaa were dissolved in 1 mL of 1-methyl-2-pyrrolidone (NMP) and mixed with two equivalents of HBTU/HOBt coupling solution (in DMF). This mixture was immediately added to the Fmoc-deprotected peptidyl resin (0.055 mmol) swelled in 1 mL of NMP. After a few minutes of shaking, four equivalents of DIEA solution were added. Then, the reaction mixture was stirred for 2 hours at 37 °C. After this, the resin was washed by NMP several times. The labeling procedure was repeated twice (the second time, the reaction mixture was stirred ~ 20 hours at 37 °C). After this, the resin was washed with NMP.

- *Capping*

In order to covalently block unreacted amino acids, a capping procedure was performed. This procedure was done in the acetic anhydride mixture (0.5 M acetic anhydride, 0.125 M DIEA, and 0.015 M HOBt in NMP). The reaction mixture was stirred for 10-20 min at 37 °C. Then, the resin was filtrated, washed by NMP and peptide synthesis was continued on the synthesizer. At the end of the synthesis, the Fmoc-deprotected peptidyl resin 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 pure peptide were combined and lyophilized.

All purified labeled peptides were analyzed by ESI Mass Spectrometry. NC(11-55)-W37-M3HFaa peptide: calculated M = 5288.84, found M = 5288.35; NC(11-55)-A30-M3HFaa peptide: calculated M = 5403.97, found M = 5403.42.

2.1.3. Preparation of Zn-bound NC(11-55) peptides

After the synthesis and purification, powder of the lyophilized peptide was dissolved in de-ionized water (≈ 0.5 mg in 500 µL). Then, about 10 µL of this solution was used to determine the peptide concentration using an extinction coefficient of ε_{280} = 5700 $M^{-1}cm^{-1}$ for the non-labeled peptide and ε_{360} = 23000 $M^{-1}cm^{-1}$ for the M3HFaa-labeled NC(11-55). Next, 2.0 molar equivalents of $ZnSO_4$ was added to the peptide and pH was raised to its final value by adding buffer. Adjustment of the pH to the neutral was done only at the end of the preparation to prevent oxidation of the cysteines.

2.1.4. Peptide activity tests

The ability of the labeled peptides to destabilize the secondary structure of ODNs was tested using cTAR DNA (DNA hairpin of 55 nucleotides), the complementary sequence of the transactivation response element, involved in minus strand DNA transfer during reverse transcription. To this end, we used a cTAR sequence labeled at its 3' and 5' ends by 5(6)-carboxytetramethylrhodamine (TMR) and 4-(4′-methylaminophenylazo) benzoic acid (Dabcyl), respectively. In the absence of NC, cTAR is mainly in a non-fluorescent closed form where the TMR and Dabcyl labels are close together, giving excitonic coupling (Bernacchi, Piémont et al. 2003) (Figure 2.1.5). The destabilization was measured through the fluorescence intensity increase of TMR in the labeled cTAR as a function of NC(11-55) concentration (Bernacchi, Stoylov et al. 2002; Beltz, Azoulay et al. 2003). The ability of the labeled peptides to promote the annealing of cTAR with its complementary dTAR DNA sequence was compared to that of the unlabeled NC peptide. The annealing kinetics were performed under the pseudo-first-order conditions by using unlabeled dTAR at a concentration which was 30-fold higher than the concentration of cTAR labeled with carboxytetramethylrhodamine (TMR) at its 5′ end and with 5/6-carboxyfluorescein (Fl) at its 3′ end (Godet, de Rocquigny et al. 2006). Excitation and emission wavelengths were 480 and 520 nm, respectively, for monitoring the Fl fluorescence. Peptides were added to each reactant separately at a peptide : ODN ratio of 3:1, and then, the reaction was initiated by mixing the peptide-coated ODNs together.

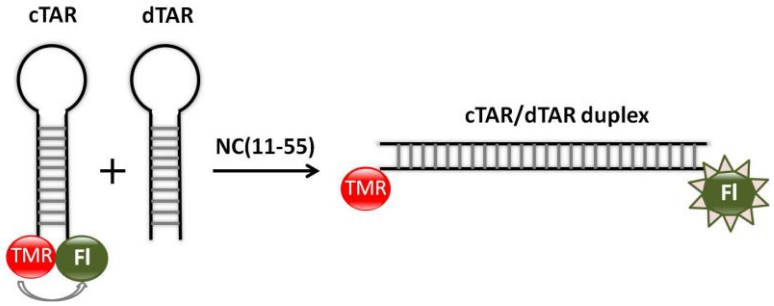
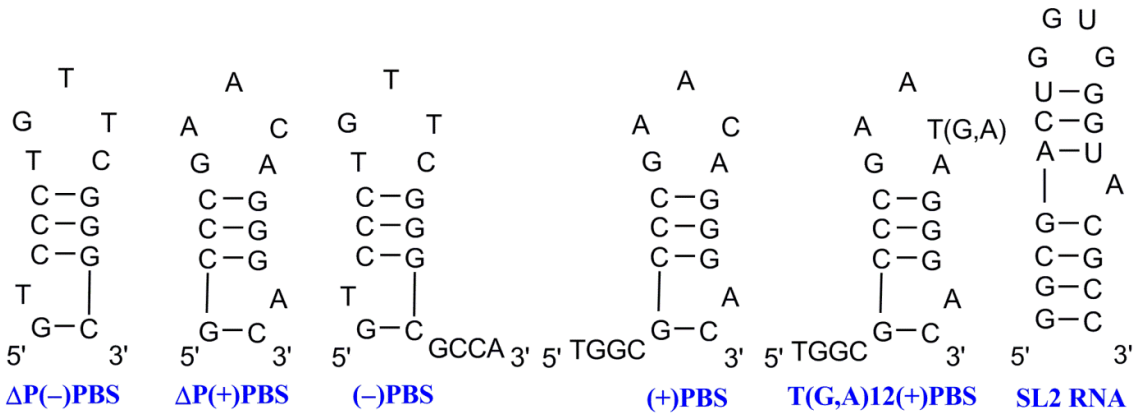


Figure 2.1.5. Schematic presentation of the activity test.

2.1.5. Oligonucleotide sequences used in this work

Non-labeled ODNs ΔP(–)PBS DNA, ΔP(+)PBS DNA, (–)PBS DNA, (+)PBS DNA, T12(+)PBS DNA, G12(+)PBS, A12(+)PBS, SL2 RNA, TAR RNA, cTAR DNA and dTAR DNA, calf thymus double-stranded DNA (CT-DNA), and labeled d2Ap7(–)PBS DNA were synthesized and purified by IBA GmbH Nucleic Acids Product Supply (Germany) (Figure 2.1.6). Their stock solutions were prepared in de-ionized water, and concentrations were calculated from their absorbance using the molar extinction coefficients at 260 nm provided by the supplier.



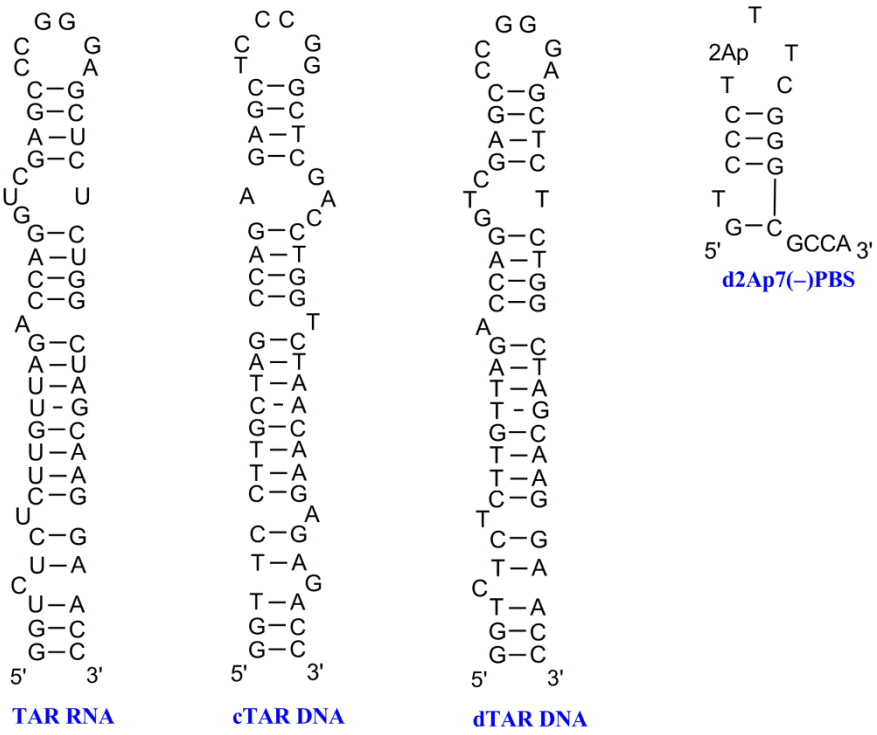


Figure 2.1.6. Oligonucleotide sequences used in this thesis.

2.1.6. Synthesis of the labeled 3HCnt-(–)PBS DNA and d^{th}G7(–)PBS DNA

Syntheses of the 3HCnt- and d^{th}G7-labeled ΔP(–)PBS and (–)PBS DNA sequences were performed using the solid-phase ODN synthesis in the group of Prof. Alain Burger and the group of Prof. Yitzhak Tor, respectively. The 3HCnt nucleobase was selectively introduced at positions 2, 7, 8, 9 and 10 in the ΔP(–)PBS sequences. For kinetic measurements, 3HCnt was introduced at position 9 of the full-length (–)PBS DNA. The d^{th}G nucleobase was introduced at position 7 of (–)PBS DNA (Figure 2.1.7). The descriptions of the ODNs syntheses are presented in (Sholokh, Sharma et al. 2015) and (Sholokh, Grytsyk et al., in preparation).

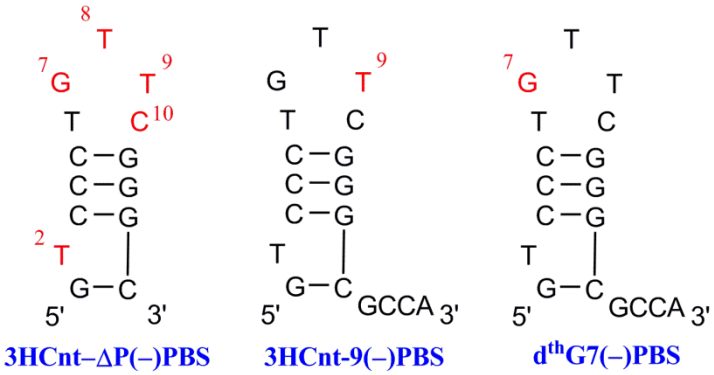


Figure 2.1.7. ODN sequences labeled with fluorescent nucleoside analogs, where the positions of labeling are marked in red.

After synthesis, lyophilized labeled ODNs were dissolved in deionized water and their concentrations were determined by absorption measurements using their molar extinction coefficients at 260 nm.

2.1.7. Preparation of the double-stranded ODNs

The double stranded ODNs were prepared by mixing two complementary sequences (labeled and non-labeled) 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. Their equimolar concentrations (1 µM) were used for the thermal denaturation experiments, and 1:3 molar ratio of labeled:nonlabeled sequence for all other experiments.

2.1.8. Lipids

Zwitterionic dioleoylphosphatidylcholine (DOPC) and negatively-charged dioleoylphosphatidylserine (DOPS) were purchased from Sigma-Aldrich. The concentrations of phospholipid stock solutions in chloroform were determined by dry weight. Lipid membrane models composed of DOPC or DOPS were used to characterize NC(11-55)-lipid interactions. These phospholipids consist of a hydrophilic head, a charged phosphate group, and a hydrophobic tail of long saturated or unsaturated hydrocarbon chain (Figure 2.1.8).

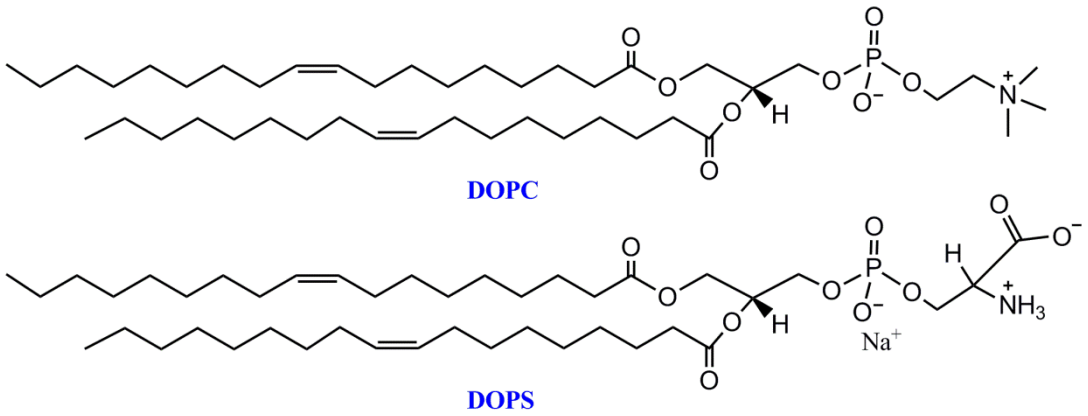


Figure 2.1.8. Lipid structures used in this thesis.

Preparation of large unilamellar vesicles (LUVs)

To prepare liposomes of pure or mixed lipids, the lipids were first dissolved in an organic solvent (chloroform or chloroform/methanol mixtures). The organic solvent was then removed by rotary evaporation, yielding a thin lipid film on the wall of a round bottom flask. The lipid film was thoroughly dried to remove the residual organic solvent by continuing evaporation with additional heating during 30 min. Hydration of the dry lipid film was done by adding 10 mM phosphate buffer, 30 mM NaCl, pH 6.5 or 7.0 to the flask for approximately 30 min. Typically, lipid suspensions were prepared to obtain a 0.3 mM final concentration, although higher concentrations could be achieved. After hydration, the solution was thoroughly vortexed during 2 min, giving rise to a suspension of multilamellar vesicles with a heterogeneous size distribution.

Once a suspension has been obtained, the vesicles were downsized by extrusion. Lipid extrusion is a technique in which a lipid suspension is forced through a polycarbonate filter with a defined pore size to yield vesicles having a diameter near the pore size. An extruder (LipexBiomembranesInc) with polycarbonate filters of calibrated pores (Nucleopore) was used. Prior to extrusion through the final pore size, the lipid suspension was first downsized by passing through a large pore size (0.2 μm) filter 7 times. Then the suspension was passed through a filter with the final pore size (0.1 μm) 10 times. This final extrusion through a filter with 0.1 μm pores yields large unilamellar vesicles (LUV) with a mean diameter of 110-120 nm (according to dynamic light scattering technique). This method ensures a homogeneous size distribution of the final suspension. All preparation steps including hydration, vortexing and extrusion were done at a room temperature.

2.1.9. Experiments with free nucleosides

Adenosine, guanosine, cytidine and thymidine were purchased from Sigma-Aldrich. To determine the sensitivity of M3HFaa spectroscopic properties to nucleosides, 2 mL of aqueous stock solution of free nucleoside (35 mM) was mixed with 0.5 mL of M3HFaa in buffer (12 µM M3HFaa, 10 mM phosphate, 60 mM NaCl, 1.5 M sucrose). To avoid guanosine precipitation, the spectroscopic properties of the sample were recorded immediately after heating and cooling down to room temperature. Measurements were repeated 3 times for each nucleoside.

2.1.10. Preparation of M3HFaa in a polystyrene film

Polystyrene (MW = 4 kDa) was from Sigma-Aldrich (Figure 2.1.9). 0.3 g of polystyrene was dissolved in 0.5 mL of a 3 µM M3HFaa solution in ethyl acetate. To evaporate the solvent, the resulting viscous solution was placed onto a glass coverslip. Fluorescence measurements were performed directly on the solid polystyrene film at the surface of the coverslip.

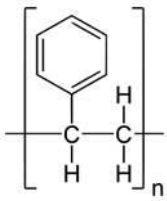


Figure 2.1.9. Structure of polystyrene.

Part 2. Physical measurements and procedures

2.2.1. UV/visible absorption and steady-state fluorescence measurements

Absorption measurements are performed using spectrophotometer. Typical spectrophotometer is based on the double-beam principle, when the light absorbed by one solution is compared to that absorbed by a blank solution (Figure 2.2.1).

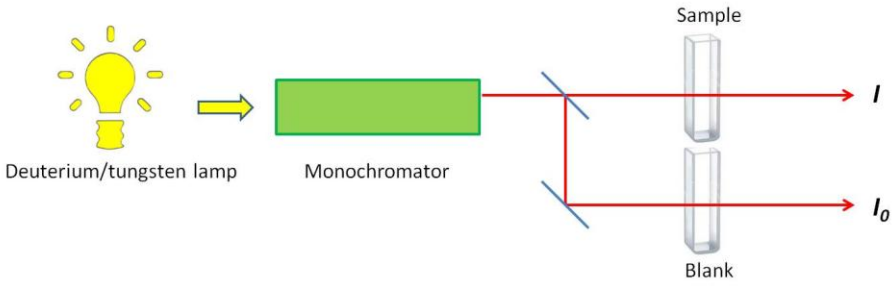


Figure 2.2.1. Scheme of the spectrophotometer instrument.

Quantification of the absorbance is usually performed using the Beer-Lambert law:

$$A = log_{10}\left(\frac{I_0}{I}\right) = \varepsilon cl, \quad (2.2.1)$$

where A is the absorbance of light through a material of pathlength 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.

A typical instrumental block diagram for molecular fluorescence is shown in Figure . In contrast to instruments for absorption spectroscopy, the optical paths for the source and detector are usually positioned at an angle of 90°.

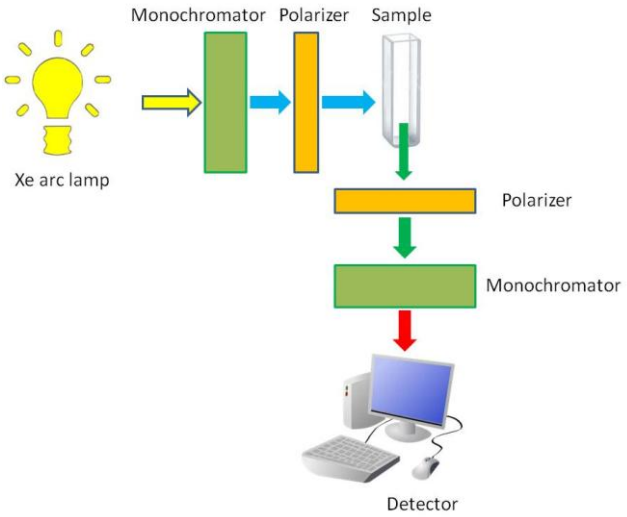


Figure 2.2.2. Scheme of the spectrofluorimeter.

Absorption spectra were recorded on a Cary 400 or 4000 UV-visible spectrophotometer (Varian). 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 yields were calculated using quinine sulfate in 0.5 M sulfuric acid

(QY = 0.546) (Melhuish 1961) or free 2Ap deoxyriboside (QY = 0.68) (Ward, Reich et al. 1969) as references.

Fluorescence quantum yield, that is the number of photons emitted divided by the number of photons absorbed, was calculated using the equation:

$$QY_x = QY_r \frac{I_x}{I_r}\frac{A_r}{A_x}\left(\frac{n_x}{n_r}\right)^2, \quad (2.2.2)$$

where QY_x is the quantum yield of the sample, QY_r is the known quantum yield of the reference dye, I_x and I_r are their respective integral fluorescence intensities, A_x and A_r are the absorbance of the sample and the reference at the excitation wavelength, respectively, n_x and n_r are the refractive indices of the respective solvents of the sample and the reference.

Photodegradation experiments

To check the photostability, the solutions of $d^{th}G$ 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 experiments were performed in dark quartz micro-cuvettes with 50 µL total volume. The samples were illuminated using Xenon lamp of a FluoroMax 4 spectrofluorimeter, excitation and emission slits were of 8 nm. The emission signal was detected as a function of time.

2.2.2. Thermal denaturation experiments

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 7.0. 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.

2.2.3. Competition experiments

To determine the relative affinity of the labeled sequences comparatively to the non-labeled sequences for NC(11-55), competition experiments were performed. To this end, 2 µM of labeled 3HCnt-ΔP(−)PBS was mixed with 2 µM of NC(11-55), and then with 2 µM of the non-labeled ΔP(−)PBS. The distribution 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 or another sequence will result in deviations from this value. Experiments were performed in 10 mM phosphate buffer, 30 mM NaCl, pH 6.5. Excitation wavelength was 360 nm.

2.2.4. Quenching experiments

A number of chemical and physical processes induce the reduction in fluorescence intensity in a phenomenon called quenching. Several mechanisms lead to a loss of emission from the fluorophore including intersystem crossing or the heavy atom effect, electron exchange or the photoinduced electron tranfer. In principle, external quenching may provide information on the accessibility of fluorophore to solvent. Commonly the quenching processes are divided into static and dynamic quenching. In the present work we mainly focused on the dynamic or collisional quenching that occurs when the excited fluorophore contacts with an atom or a molecule that can facilitate non-radiative transitions to the ground state. Common quenchers include O_2, I^-, Br^-, Cl^- and acrylamide (more complete list of quenchers is reviewed in (Lakowicz 2006). Potassium iodide (KI) was used for the quenching experiments wih $d^{th}G$ and d2Ap. Free labels and labeled ODNs were titrated by the aliquots of the concentrated aqueous solution of KI. $Na_2S_2O_3$ 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:

$$\frac{F_0}{F} = 1 + K_{SV}[Q] = 1 + k_q\tau_0[Q] \text{ , (2.2.3)}$$

where F and F_0 are the fluorescence intensities observed in the presence and absence of quencher, respectively, K_{SV} is the Stern-Volmer quenching constant, [Q] is the quencher

concentration, k_q is the diffusion-controlled quenching rate, and τ_0 is the lifetime in the absence of the quencher.

2.2.5. *Time-resolved fluorescence spectroscopic measurements*

Frequently macromolecules exist in more than one conformation, especially when bound to a ligand. However, this molecular information is lost when performing the steady-state measurements because of the time-averaging process. For quantitative interpretations of the fluorophore conformational states, quenching mechanisms, anisotropy, or the donor-acceptor transitions, time-resolved fluorescence is usually used.

The fluorescence lifetime τ is the average time a fluorophore remains in the excited state after excitation, and it is defined as $\tau = \frac{1}{k_r + k_{nr}}$, (2.2.4) where k_r is the radiative rate constant, and k_{nr} is the non-radiative rate constant (Lakowicz 2006). After excitation, the molecule loses its energy either by emitting a fluorescence photon or by dissipation energy in the form of heat. N excited fluorophores will leave the excited state according to the following rate equation:

$$dN = (k_r + k_{nr})N(t)dt \text{ , (2.2.5)}$$

where t is time. Integration of equation 5, using equation 4 and taking into account the proportionality of the excited state population N(t) to the fluorescence intensity I(t) results in:

$$I(t) = I_0 e^{-t/\tau} \text{ , (2.2.6)}$$

where I_0 is the fluorescence intensity at t = 0. The decay of the fluorescence intensity follows an exponential equation, and the key parameter characterizing this fluorescence decay is τ. Fluorescence lifetime is an absolute value, independent on the excitation intensity, fluorophore concentration (if it is sufficiently low), photobleaching (if there are no fluorescent photoproducts), light pathlength or scattering (Festy, Ameer-Beg et al. 2007).

Nanosecond time-resolved measurements typically require complex and expensive instrumentation. There are two dominant methods for the measurements: time-domain or frequency-domain (Lakowicz 2006). In the time-domain approach the sample is excited with a pulse of light (Figure 2.2.3).

The width of the pulse is made as short as possible, and is preferably much shorter than the decay time τ of the sample. The time-dependent intensity is measured following the excitation pulse, and the decay time τ is calculated from the slope of a plot of log $I(t)$ versus t, or from the time at which the intensity decreases to $1/e$ of the intensity at $t = 0$. The intensity decays are often measured through a polarizer oriented at 54.7° from the vertical z-axis. This

condition is used to avoid the effects of rotational diffusion and/or anisotropy on the intensity decay.

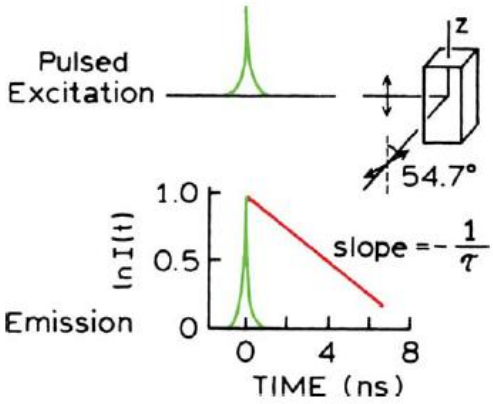


Figure 2.2.3. Time-domain lifetime measurements (Lakowicz 2006).

Time-resolved fluorescence measurements were performed with the time-correlated single-photon counting (TCSPC) technique (Lakowicz 2006). This technique consists in exciting fluorophores under the conditions when no more than a single photon could be detected (less than 5 fluorescence photons detected per 100 excitation pulses). A time interval between the initial pulse and the detection of each single photon is the fluorescence lifetime. 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) (Figure 2.2.4) (Godet, Kenfack et al. 2013). The fluorescence emission was collected 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 $<\tau>$ was calculated from the individual fluorescence lifetimes (τ_i) and their relative amplitudes (α_i) according to $<\tau> = \sum \alpha_i \times \tau_i$ (2.2.7). The population of dark species (α_0) was calculated by: $\alpha_0 = 1 - \tau_{free}/(\tau_{ODN} \times R_m)$, (2.2.8) 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 $\alpha_{ic} = \alpha_i \times (1 - \alpha_0)$ (2.2.9). Time-resolved intensity data were fitted using the maximum entropy method (Pulse 5 software) (Brochon 1994). This method presents no *a priori* assumption that is a distinct advantage over the traditional iterative reconvolution

technique based on a statistical method of least squares, where the results may depend on the choice of the initial conditions. The maximum entropy method assumes that all possible lifetimes in a given lifetime range have equal probability. Thus, an initial assumed distribution plot of amplitudes versus lifetimes is flat. After the analysis is completed, this flat distribution is transformed into a structured distribution that adequately fits the experimental decay data. In all cases, the χ^2 values were close to 1, indicating an optimal fit.

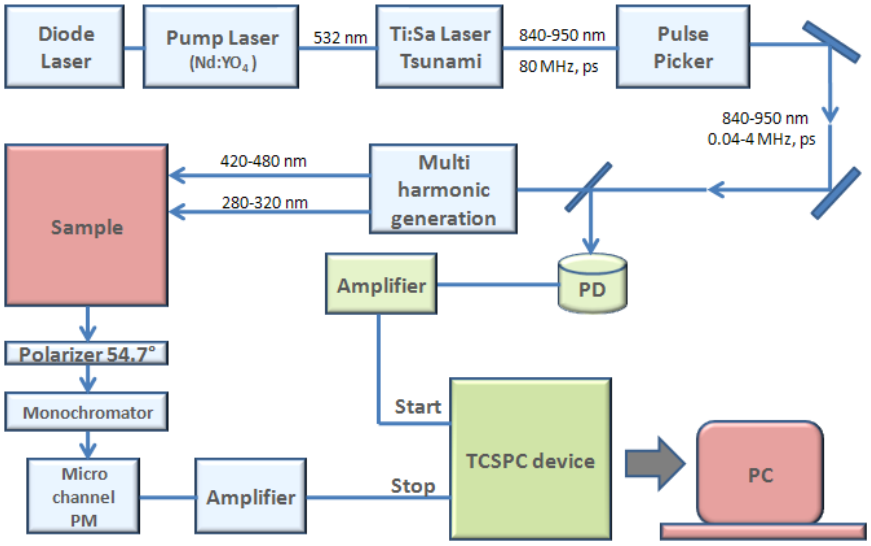


Figure 2.2.4. Scheme of the TCSPC setup.

2.2.6. Time-resolved fluorescence anisotropy measurements

Fluorescence anisotropy measures the orientation of the fluorophore with respect to a particular axis, usually the laboratory vertical axis, at the moment of the fluorescence emission. This parameter reports on the movement of the fluorophore during its excited state lifetime and provides information about the size, shape of the biomolecules and the rigidity of their environment. Anisotropy is described by the equation:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad , (2.2.10)$$

where $I_{\parallel}$ and $I_{\perp}$ is the observed parallel (vertical) and perpendicular (horizontal) intensities, respectively.

Anisotropy method is usually applied to study protein interactions, like its aggregation state, protein-ligand interactions, protease activity and antigen-antibody binding (Jameson 2014).While equilibrium steady-state measurements reflect an average of conformations of

fluorophore, to separate local mobility of the fluorophore from the overall rotation of the protein, time domain measurements are used. Time-resolved anisotropy provides information on changes in fluorophore orientation as a function of time.

Technically, the sample is illuminated by a short pulse of vertically polarized light and the decay over time of both the vertical and horizontal components of the emission are recorded. The anisotropy function is then plotted versus time and analyzed by:

$$r(t) = \frac{I_{\parallel}(t) - G \times I_{\perp}(t)}{I_{\parallel}(t) + 2G \times I_{\perp}(t)} = r_0 \sum_i \beta_i \times \exp\left(-\frac{t}{\theta_i}\right), \quad (2.2.11)$$

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.

2.2.7. Fluorescence lifetime imaging microscopy (FLIM)

One important application of the time-resolved fluorescence is cellular imaging fluorescence microscopy. Using the local intensity of the fluorophore in the fluorescence microscopy may be problematic due to unknown local concentration of the fluorophore and its change during the measurement because of wash out or photobleaching. In contrast, fluorescence lifetime is independent on the probe concentration and allows creating the lifetime images, where the color contrast is based on the lifetime in each region of the sample. Fluorescence lifetime imaging microscopy (FLIM) has become an accessible and increasingly applied tool in the cell biology. FLIM can be used as an imaging technique in confocal microscopy, two-photon excitation microscopy and multiphoton tomography (Lakowicz 2006; Festy, Ameer-Beg et al. 2007). In the present work, two-photon excitation microscopy was used. The phenomenon of two-photon excitation arises from the simultaneous absorption of two photons in a single quantitized event. Since the energy of a photon is inversely proportional to its wavelength, the two absorbed photons must have a wavelength about twice that required for one-photon excitation. For example, a fluorophore that normally absorbs ultraviolet light (≈ λ = 400 nm) can also be excited by two photons of near-infrared light (≈ λ = 800 nm) if both reach the fluorophore at the same time.

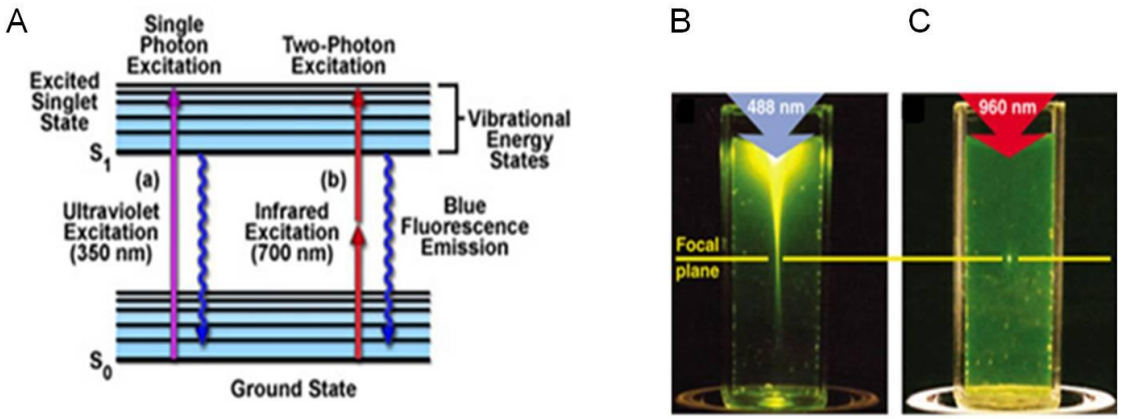


Figure 2.2.5. A) two-photon Jablonsky Energy Diagram;
B) cone of fluorescence resulting from a single-photon excitation
C) fluorescent volume resulting from a multi-photon excitation

The de-excitation of a molecule excited by two photons is similar to that following single-photon excitation (Figure 2.2.5A). However, unlike single-photon excitation, excitation at two photons is a nonlinear phenomenon: the probability of exciting a molecule is very high in a small volume (less than femtoliter), and then decreases very rapidly. The observation volume is thus centered on the focal point (Figure 2.2.5B). To obtain two photon excitation, a pulsed femtosecond laser is generally used. Two-photon microscopy has the benefit of using low energy photons (infrared region), which reduces their impact on biological tissues. Moreover, the excitation volume is localized around the focal point, limiting the photobleaching at a single focal plane, and increasing the observation clearness since out of focus fluorescence is substantially reduced.

Principle of FLIM

Fluorescence microscopy experiments were performed on a home-built two-photon laser scanning setup based on an Olympus IX70 inverted microscope with an Olympus 60x 1.2NA water immersion objective (Figure 2.2.6) (Azoulay, Clamme et al. 2003). Two-photon excitation was provided by a Ti-sapphire (Tsunami, Spectra Physics) or an Insight DeepSee (Spectra Physics) laser. Imaging was carried out using two fast galvo-mirrors operating in the descanned fluorescence collection mode. Photons were detected with an avalanche photodiode (APD SPCM-AQR-14-FC, Perkin-Elmer) which was connected to a time-correlated single photon-counting (TCSPC) module (SPC830, Becker and Hickl) operating in reversed start-stop mode. Typical acquisition time was 5 s with an excitation power around 2.5 mW (760 nm) at the sample. Subsequent data analysis using the software SPC Image

(Becker and Hickl) allowed us extracting the fluorescence lifetimes from the decays, and their visualization in a FLIM image, using an arbitrary color scale.

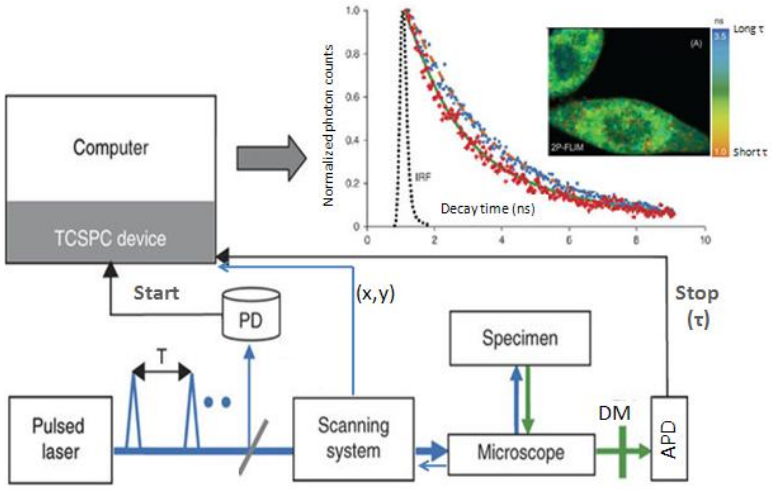


Figure 2.2.6. Scheme of the multiphoton excitation TCSPC FLIM setup.

2.2.8. Kinetic measurements

The stopped-flow experiments with W37-M3HFaa-labeled NC(11-55) were performed in second-order conditions where 75nM W37-M3HFaa was mixed with 75 nM ΔP(–)PBS DNA in 10mM phosphate buffer, 30 mM NaCl (pH = 6.5). Measurements were performed using a stopped-flow instrument SFM-3 Bio-Logic, which allows rapid mixing of 80 µL of each reagent (dead time 2.8 ms). Excitation wavelength was 365 nm and the fluorescence emission was detected at 424 nm through an interferential filter (Kodak) or above 500 nm through a cutoff filter (Kodak).

To determine the kinetic parameters of the (–)/(+)PBS annealing reaction the kinetic measurements were performed under pseudo-first-order condition by mixing 50 or 100 nM labeled 3HCnt-9(–)PBS DNA with a 10- to 100-fold molar excess of non-labeled (+)PBS DNA. The experiments were performed either in the absence or in the presence of NC(11-55) added at a 3:1 ratio to each complementary ODN prior to the start of the annealing reaction. Measurements were performed either by using a FluoroMax 4 spectrofluorimeter (Horiba) or a stopped-flow instrument SFM-3 Bio-Logic. With the spectrofluorimeter, excitation and emission wavelengths were 360 and 540 nm, respectively. For the stopped-flow instrument, excitation wavelength was 365 nm and the emission signal was detected above 500 nm through a cutoff filter (Kodak).

The stopped-flow technique is a kinetic method of analysis designed to rapidly mix samples and reagents when using reactions with very fast kinetics. Sample and reagents are loaded into two syringes (Schematic view of the instrument on Figure 2.2.7), and precisely measured volumes are dispensed by the action of a syringe drive. 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, 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.

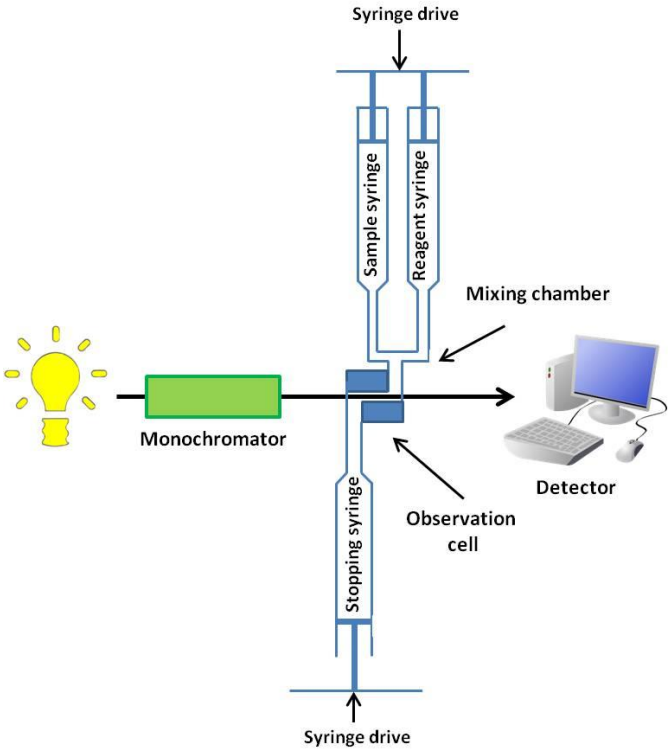


Figure 2.2.7. Schematic view of the stopped-flow instrument.

Fitting of the kinetic curves

Second-order kinetics of binding of M3HFaa-labeled NC(11-55) with ΔP(–)PBS

The obtained curves of W37-M3HFaa mixed with 75 nM ΔP(–)PBS were fitted according to equations:

$I(t) = (I_0 - I_f)/(1+k(I_0 - I_f)(t - t_0)) + \text{offset}$, (2.2.12) for the N* band emission

$I(t) = (I_f - I_0) - ((I_f - I_0)/(1+k(I_f - I_0)(t - t_0))) + \text{offset}$, (2.2.13) for the T* band emission

where I_0 and I_f are the initial and final fluorescence intensities, k is the bimolecular kinetic rate constant, t_0 is the dead time, the offset is the difference between the I_f (or I_0) and the baseline signal of buffer.

These equations were derived for the second-order reactions of type II, when two different species A (W37-M3HFaa) and B (ΔP(–)PBS) are mixed at the same initial concentrations that remain equal throughout the reaction. If A_0 represents the initial number of moles of the reacting species A and B, and x the number of moles of each reactant in a time interval t, then the rate of formation of product k_2 can be described by the following mechanism and rate equation:

$$A + B \rightarrow C \qquad (2.2.14)$$

Amount at time t = 0: $[A]_0 + [A]_0 \quad 0$

Amount at time t = t: $([A]_0 - x)\ ([A]_0 - x)\ x$

$$\frac{dx}{dt} = k_2([A]_0 - x)^2 \qquad (2.2.15)$$

$$\frac{dx}{([A]_0 - x)^2} = k_2 dt \qquad (2.2.16) \text{ or}$$

$$-\int \frac{d([A]_0 - x)}{([A]_0 - x)^2} = k_2 \int dt \qquad (2.2.17)$$

$$\frac{1}{([A]_0 - x)} + I = k_2 t \qquad (2.2.18)$$

The integration constant I is evaluated from the limiting condition that at t = 0, x = 0. Equation (18) then reduces to $I = -1/[A]_0$. Substituting for I back into (2.2.18) yields:

$$\frac{1}{([A]_0 - x)} - \frac{1}{[A]_0} = \frac{x}{[A]_0([A]_0 - x)} = k_2 t \qquad (2.2.19) \text{ or}$$

$$k_2 = \frac{1}{t}\frac{x}{[A]_0([A]_0 - x)} \qquad (2.2.20)$$

If x is replaced by $([A]_0 - [A])$, equation (2.2.20) takes on the form

$$[A] = \left[\frac{[A]_0}{1 + k_2[A]_0 t}\right] + constant$$

Pseudo-first-order kinetics of 3HCnt-labeled (–)PBS with (+)PBS annealing reaction

The kinetic constants of 3HCnt-labeled (–)PBS with (+)PBS annealing reaction were calculated using analytical and numerical approaches. By *analytical* approach the kinetic curves were fitted using the following mono- and bi-exponential functions:

$$I(t) = I_f - \left(I_f - I_0\right) exp\left(-k_{obs1}(t - t_0)\right), \quad (2.2.21)$$

$$I(t) = I_f - \left(I_f - I_0\right)(a\, exp\left(-k_{obs1}(t - t_0)\right) + (1 - a) exp\,(-k_{obs2}(t - t_0)), \quad (2.2.22)$$

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

The kinetic rate constants of both components were then determined from the dependence of $k_{obs1,2}$ on the (+)PBS concentration according to:

$$k_{obs1} = k_1[(+)PBS] + k_{-1}, \quad (2.2.23)$$

$$k_{obs2} = (k_f K_m[(+)PBS])/(1 + K_m[(+)PBS]) + k_{-f}, \quad (2.2.24)$$

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.

The Dynafit software (Kuzmic 1996; Kuzmic 2009) 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. For the Dynafit calculations the data files reporting the fluorescence intensity changes of the 3HCnt-labeled (–)PBS as a function of time were used. These data files were used with a script that includes the mechanism of the reaction, approximate constants taken from the analytical calculations, initial concentrations of 3HCnt-9(–)PBS and (+)PBS, and the responses of the fluorescent products that are their fluorescence intensities divided by the concentration of 3HCnt-9(–)PBS.

A typical script for the 3HCnt-9(–)/(+)PBS system in the absence of NC is presented below:

```
[task]
 task  = fit
 data  = progress
[mechanism]
   A + B <===> AB   :   k1  k-1
   AB <===> C  :   kf  k-f
[constants]
  k1   = 3000 ?
  k-1   = 0.0027 ?
  kf =  0.00036 ?
  k-f = 0.000001 ?
[concentrations]
  A = 50e-9
 [progress]
  directory      C:\Users\msholokh\Desktop\...
  extension   txt
  file        500nM
  concentration   A = 50e-9 , B = 500e-9
  responses  A = 3.58e+11 , AB = 1.79e+12 ? , C = 1.79e+12 ?
  file        1000nM
  concentration   A = 50e-9 , B = 1000e-9
  responses  A = 3.49e+11 , AB = 1.74e+12 ? , C = 1.74e+12 ?
  file        2000nM
  concentration   A = 50e-9 , B = 2000e-9
  responses  A = 4.44e+11 , AB = 1.83e+12 ? , C = 1.83e+12 ?
  file        3000nM
  concentration   A = 50e-9 , B = 3000e-9
  responses  A = 5.14e+11 , AB = 1.87e+12 ? , C = 1.87e+12 ?
  file        4000nM
  concentration   A = 50e-9 , B = 4000e-9
  responses  A = 6.17e+11 , AB = 1.97e+12 ? , C = 1.97e+12 ?
  file        5000nM
  concentration   A = 50e-9 , B = 5000e-9
  responses  A = 6.59e+11 , AB = 1.98e+12 ? , C = 1.98e+12 ?
[output]
  directory   ./FP/out_rtpt
[end]
```

2.2.9. Cell cultures

HeLa cells were cultured in Dulbecco's modified Eagle medium (D-MEM, high glucose, Gibco-Invitrogen) supplemented with 10% (v/v) fetal bovine serum (FBS, Lonza), 1% antibiotic solution (penicillin-streptomycin, Gibco-Invitrogen) in a humidified incubator with 5% CO_2, 95% air atmosphere at 37 °C. Cells plated on a 75 cm^2 flask at a density of 5-10×10^4 cells/mL were harvested at 80% confluence with trypsin–EDTA (Sigma) and seeded on a glass coverslip at a density of 2.5-5×10^4 cells/coverslip, 48 hours before the microscopy measurements. Finally, cells on the cover slip were mounted on the microscope holder and covered with D-MEM solution.

2.2.10. Microinjection

Microinjection refers to the process of using a glass micro-needle to insert substances, such as NAs, proteins or fluorescent organic molecules into a single living cell. It is a simple mechanical process in which a needle of 1 µm in diameter penetrates inside the cytoplasm and/or the nucleus. The desired contents are then injected into the desired sub-cellular compartment and the needle is removed (Figure 2.2.8). Microinjection was performed on cells seeded on a gridded cover slip (Cellocate; Eppendorf), which allows locating cells. Before loading into the micro-needle, the solution of 200 µM labeled peptide was centrifuged for 20 min at 4 °C. Microinjection was performed using a micromanipulator (model 5171; Eppendorf) coupled to a microscope (model DMIRBE; Leica) at 40× magnification. The positions of the injected cells were noted to find them back after placing the cover slip under the FLIM microscope.

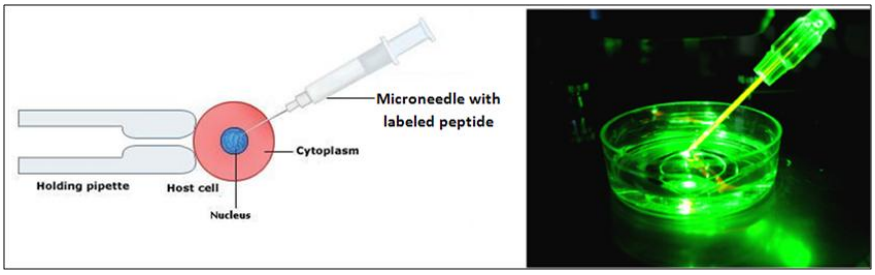


Figure 2.2.8. Scheme of the microinjection process.

2.2.11. Deconvolution of the spectra

Deconvolution of the fluorescence spectra into N* and T* bands was performed using the Siano software kindly provided by Prof. A.O. Doroshenko (Kharkiv, Ukraine), as previously described (Doroshenko, Sychevskaya et al. 2002). This deconvolution is based on an iterative nonlinear least-squares method, where the individual emission bands were approximated by a log-normal function accounting for several parameters: maximal amplitude, I_{max}, spectral maximum position, ν_{max}, and position of half-maximum amplitudes, ν_1 and ν_2, for the blue and red parts of the band, respectively. These parameters determine the shape parameters of the log-normal function, namely the full width at the half-maximum, FWHM = $\nu_1-\nu_2$, and the band asymmetry, P = $(\nu_1-\nu_{max})/(\nu_{max}-\nu_2)$. All parameters were allowed to vary in the iteration process, with the exception of the FWHM parameter of the T*

band which was fixed in cases where the intensity of the T* band was low. In these last cases, the FWHM value of the T* band was fixed at 2600 cm^{-1} based on the data from binary solvent mixtures. The resulting integral fluorescence intensities, I_{N*}/I_{T*}, of the separated N* and T* bands were used for calculation of the ratio of the two bands.

Software used

For many analytical purposes Origin software (academic license for LBP-UMR7213) has been used. Chemical structures and schemes have been prepared with ChemBioDraw software (personal academic license). Other software was mentioned otherwise wherever required.

Chapter 3. Results and Discussion

Part 1. Interaction of NC with its molecular partners using fluorescent amino acid analogs based on 3-hydroxychromones

Due to their ultimate sensitivity, fluorescence-based methods are ideal for monitoring protein conformational changes and molecular interactions. In this respect, non-natural amino acids bearing environment-sensitive fluorophores are of special interest since they can be inserted at well-defined positions of the protein. The main requirements for these amino acid analogs are their small sizes (in order to not perturb the folding and function of the labeled protein), high brightness and strong sensitivity to the environment. The amino acid analogs based on 3-hydroxychromones (3HC) meet these requirements and were chosen to study the interaction of NC with its biological targets.

The synthesis of the L-amino acid M3HFaa (Figure 1A, scheme S1, publication 1) containing the 4′-methoxy-3-hydroxyflavone fluorophore was performed by O. Zamotaiev. This compound was synthesized to replace a first-generation amino acid analog bearing 2-furyl-3HC fluorophore (3HCaa) that showed good ability to replace tryptophan, but low fluorescence QY as well as ground-state deprotonation of its 3-OH group that complicated the data analysis and its application for cellular studies. M3HFaa overcomes these problems and additionally exhibits substantially higher sensitivity to water, allowing significantly improved detection of peptide/ligand interactions. Therefore, we synthesized two labeled NC(11-55) peptides, namely A30-M3HFaa and W37-M3HFaa, where the fluorescent amino acid M3HFaa replaced the Ala30 and Trp37 residues located in the linker and in the second zinc finger of NC(11-55), respectively (Figure 1B, publication 1). The W37-M3HFaa peptide showed fully preserved chaperone activity, nearly identical to those of the non-labeled NC(11-55) (Figures S2 and S3, publication 1), indicating that M3HFaa maintains the NC(11-55) peptide folding and can functionally substitute the key Tpr37 residue. The A30-M3HFaa peptide, however, showed possible perturbation of the NC(11-55) linker that may impact the relative orientation of the zinc fingers and formation of the hydrophobic plateau.

In buffer, the two labeled peptides showed distinct dual-emission fluorescence spectra indicating that the probe is shielded from water in case of W37-M3HFaa while more water-exposed in A30-M3HFaa. The fluorescence QY of W37-M3HFaa is about 2-fold higher than that of A30-M3HFaa (Figure 3, Table 1, publication 1). Interactions of W37-M3HFaa with SL2 RNA or ΔP(−)PBS DNA stem-loops induced a significant red-shift of the T* emission

band and a dramatic drop of the N*/T* ratio of M3HFaa, suggesting a stacking interaction of M3HFaa with the nucleobases in both complexes, that correlates with the 3D structures of the complexes solved by NMR. For A30-M3HFaa, the changes of QY and N*/T* ratio were moderate on interaction with SL2 and more pronounced with ΔP(–)PBS. This was rationalized by the different distances between the A30 residue and the nucleobases evidenced by NMR in these two complexes. Therefore, our data indicate that M3HFaa appropriately substitutes Trp residue and likely preserves the interaction mode of the peptide with its target nucleic acid sequences. The interactions of the labeled peptides with double-stranded calf thymus CT-DNA and the non-specific RNA binding sequence TAR showed that M3HFaa at position 37 stacks with the nucleobases similarly in all complexes, while at position 30 less conserved interactions were observed. This suggests that NC binds differently to the non-specific binding sites of TAR or CT-DNA in comparison with its specific binding sites in SL2 or ΔP(–)PBS. Finally, binding of the M3HFaa-labeled NC(11-55) to DOPC/DOPS lipid membranes was also explored and showed a strong increase in the QY values and a decrease in the N*/T* ratio for both peptides. This indicates that both residues experience comparable poorly hydrated environments due to their insertion into the lipid bilayer, which is significantly less polar and water accessible than ODNs.

To further characterize the photophysical properties of the labeled peptides, we performed time-resolved fluorescence experiments. Firstly, we characterized the free M3HFaa in buffer and organic solvents to explore the effect of polarity on its fluorescence lifetimes (Table S1, publication 1). The fluorescence decays were monitored both for the N* band (monoexponential decay with a fast τ_1^{N*} component) and the T* band (biexponential decay with a fast rise component τ_1^{T*} and a longer component τ_2^{T*}). The fast decay component τ_1^{N*} agrees well with the rise component τ_1^{T*}, in line with an ESIPT reaction (N* → T*) and provides information about the ESIPT kinetics. Interestingly, the τ_2^{T*} lifetime of M3HFaa was found to decrease with an increase of polarity, as reported previously for 2-furyl-3HC and 2-thienyl-3HC.

The fluorescence decays of the T* band for both labeled NC(11-55) peptides were characterized by a short rise component τ_1^{T*} and two longer components τ_2^{T*} and τ_3^{T*} (Table 2, publication 1). The values of the longer components, likely associated with two different orientations of the fluorophore in the peptide, were dependent on the labeling position and on the nature of the ligand bound to the peptide. Binding to the ODNs induced a significant

increase of the long-lived lifetime of the W37-M3HFaa, likely associated with a stacking of the M3HFaa residue with the nucleobases. This stacking interaction likely restricts the relative motion of the two aromatic moieties of M3HFaa, favoring a more planar and hence a more emissive conformation of the probe. This also explains the low accessibility of the probe to water in the peptide/ODN complexes. The attribution of the long-lived lifetime component to stacking interactions was further substantiated by inclusion of the M3HFaa into a polystyrene matrix that mimics highly rigid environment with possible π-stacking interactions of the probe with the polystyrene aromatic rings. The high τ_3^{T*} lifetime value of M3HFaa in polystyrene matrix clearly confirmed our hypothesis. Moreover, through model experiments of M3HFaa in the presence of highly concentrated free nucleosides (A, C, G, or T) (Figure 5, publication 1), the decay component τ_2^{T*} of M3HFaa was assigned to peptide conformations where the M3HFaa residue interacts transiently with the ODN bases. For the A30-M3HFaa peptide bound to ODNs, the decay of the T* form was largely dominated by the τ_2^{T*} component, close to the lifetime of M3HFaa in polar solvents. This indicated that the M3HFaa residue is largely exposed to water, in agreement with NMR data for the corresponding Ala30 residue. In the presence of lipids, both peptides exhibited similar lifetime values and amplitudes. By analogy with the peptide/ODN complexes, the long-lived lifetime τ_3^{T*} was attributed to a highly emissive flat conformation of the probe resulting from its insertion into the lipids, while the shorter lived lifetime τ_2^{T*} may be related to a more shallow insertion of the probe into the lipids which induces a less restricted conformation of the dye. Taken together, our data show that the fluorescence lifetimes of M3HFaa in NC(11-55) are highly sensitive to the nature of the interacting partners and the position of the probe in the peptide. They provide thus information about the conformations of M3HFaa in complexes with nucleic acids and lipids.

Finally, we microinjected the W37- and A30-M3HFaa peptides in HeLa cells to identify the interacting partners of NC(11-55) in cells using two-photon FLIM microscopy. The mean lifetimes for both peptides in their free form and complexed with various ligands were recorded using the FLIM setup (Figure 7, publication 1) and compared with the lifetime values obtained in cells (Figure 8, publication 1). Both peptides were found to accumulate in the cytoplasm and spherical nuclear compartments likely corresponding to nucleoli. This distribution is fully in line with that observed with over-expressed eGFP-labeled NC(11-55), showing that the protein can diffuse all over the cell and cross the nuclear membrane. The observed lifetime values were consistent with binding of both labeled peptides to RNAs. Since the characteristic lifetimes for the free or lipid-bound peptides were not observed, this

indicates that only minimal amounts of NC(11-55) were in their free form or bound to lipid membranes in cells.

As a conclusion, the newly-synthesized M3HF-based amino acid shows a strong sensitivity of its dual emission to the microenvironment and constitutes a nearly perfect Trp substitute. Analysis of its dual emission, QY and lifetimes show that interaction of the peptide with ODNs or lipid bilayers induces dramatic changes in the water accessibility or in rigidity of the probe. Moreover, FLIM analysis of the microinjected peptides allows us identifying RNAs as the main binding partner of NC in nucleoli and cytoplasm. This new amino acid analog based on ESIPT dye constitutes thus an efficient tool for detection and characterization of peptide interactions in solution and cells.

As a next step, we monitored the kinetics of the structural rearrangement of NC(11-55) upon binding to nucleic acids. For this, we performed stopped-flow kinetics measurements with labeled W37-M3HFaa in the presence of ΔP(−)PBS DNA. The peptide and ODN were mixed at 1:1 ratio in second-order conditions. The emission of the N* or T* bands was followed, showing a decrease (Figure 3.1.1) and an increase (data not shown) of the fluorescence signal, respectively.

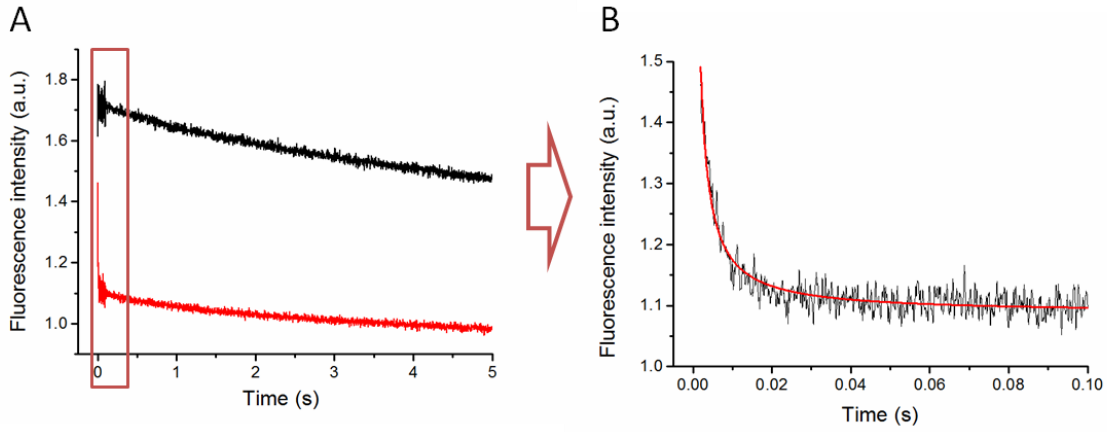
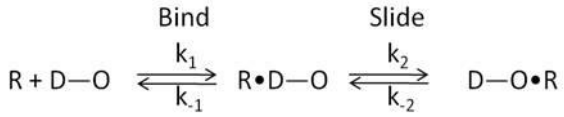


Figure 3.1.1. Representative stopped-flow kinetic traces of 75 nM W37-M3HFaa (A) in the absence (black line) and in the presence (red) of 75 nM ΔP(−)PBS DNA. (B) Fitting of the fluorescence signal of the first 0.1s of formation of the W37-M3HFaa/ ΔP(−)PBS complex (red line). Experiments were performed in 10 mM phosphate, 30 mM NaCl (pH = 6.5). Excitation wavelength was 365 nm and the fluorescence emission was detected at 424 nm through an interferential filter (Kodak).

In the absence of ΔP(−)PBS, the fluorescence signal of the labeled W37-M3HFaa (Figure 3.1.1 A black line) showed a continuous decrease, likely related to the photobleaching

of the 3HC probe. In the presence of ΔP(−)PBS, an about 2-fold fast decrease of the fluorescence of M3HFaa was observed, followed by a slower continuous decrease due to probe photobleaching. This photobleaching was somewhat less pronounced than in the absence of ΔP(−)PBS, probably due to the screening of the probe in the peptide-NA complex.

These obtained kinetic curves (Fig. 3.1.1B) were fitted according to eq. 2.2.21 in Materials and Methods, given a $1.2(\pm 0.3)\times 10^{10}$ $M^{-1}s^{-1}$ value for the bimolecular association constant. A similar value was obtained when the increase in T* band emission was monitored. This value may not be very accurate due to the very short time scale of the fluorescence decrease. Nevertheless, this value is 10-100 fold higher than expected for a diffusion-limited reaction of two macromolecules ($\sim 10^8$ $M^{-1}s^{-1}$). Similar high values were observed with the sequence-specific *lac* repressor protein/DNA binding system where a bimolecular rate constant of about 5×10^{10} $M^{-1}s^{-1}$ was described (Riggs et al, 1970). Von Hippel, Berg and co-workers proposed that such fast binding is observed when the protein binds DNA molecules on non-specific binding sites and then slides along the DNA molecule to reach the preferential binding site (Berg, Winter et al. 1981; Winter, Berg et al. 1981). Such two-step mechanism was called "bind-and-slide" and is depicted on Scheme 3.1.1.



Scheme 3.1.1. "Bind-and-slide" mechanism of protein/DNA binding. Protein (R) binds DNA (D) in a non-specific manner, then slides to the specific binding site (O) in order to form a specific complex. k_n are the kinetic rate constants.

The extremely fast kinetics observed with the W37-M3HFaa/ΔP(−)PBS system suggests that the observed changes in the M3HFaa fluorescence occur simultaneously to binding. No further slower changes in the probe signal could be evidenced when the protein binds its specific binding site. This suggests that either the conformational changes evidenced by NMR at this site occur immediately on binding or that they are spectroscopically silent.

Publication 1. Fluorescent amino acid undergoing excited state intramolecular proton transfer for site-specific probing and imaging of peptide interactions.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Fluorescent Amino Acid Undergoing Excited State Intramolecular Proton Transfer for Site-Specific Probing and Imaging of Peptide Interactions

Marianna Sholokh,[†,‡] Oleksandr M. Zamotaiev,[‡] Ranjan Das,[§] Viktoriia Y. Postupalenko,[†] Ludovic Richert,[†] Denis Dujardin,[†] Olga A. Zaporozhets,[‡] Vasyl G. Pivovarenko,[‡] Andrey S. Klymchenko,[†] and Yves Mély*[,†]

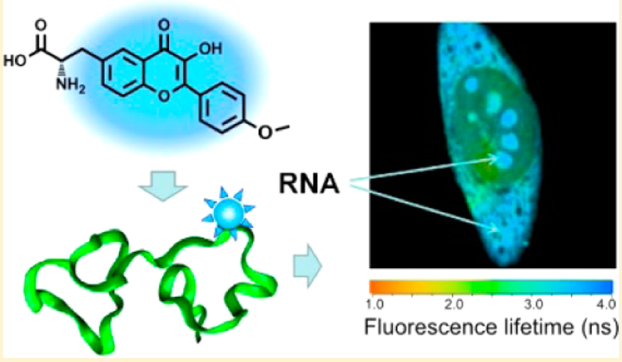
[†]Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74, Route du Rhin, 67401 Illkirch Cedex, France

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

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

Supporting Information

ABSTRACT: Fluorescent amino acids bearing environment-sensitive fluorophores are highly valuable tools for site-selective probing of peptide/ligand interactions. Herein, we synthesized a fluorescent L-amino acid bearing the 4′-methoxy-3-hydroxyflavone fluorophore (M3HFaa) that shows dual emission, as a result of an excited state intramolecular proton transfer (ESIPT). The dual emission of M3HFaa was found to be substantially more sensitive to hydration as compared to previous analogues. By replacing the Ala30 and Trp37 residues of a HIV-1 nucleocapsid peptide, M3HFaa was observed to preserve the peptide structure and functions. Interaction of the labeled peptides with nucleic acids and lipid vesicles produced a strong switch in their dual emission, favoring the emission of the ESIPT product. This switch was associated with the appearance of long-lived fluorescence lifetimes for the ESIPT product, as a consequence of the rigid environment in the complexes that restricted the relative motions of the M3HFaa aromatic moieties. The strongest restriction and thus the longest fluorescence lifetimes were observed at position 37 in complexes with nucleic acids, where the probe likely stacks with the nucleobases. Based on the dependence of the lifetime values on the nature of the ligand and the labeled position, two-photon fluorescence lifetime imaging was used to identify the binding partners of the labeled peptides microinjected into living cells. Thus, M3HFaa appears as a sensitive tool for monitoring site selectively peptide interactions in solution and living cells.



INTRODUCTION

Environment-sensitive probes become an essential tool for monitoring site specifically the interaction of biomolecules and studying their conformational changes.[1−4] In the case of peptides and proteins, environment-sensitive dyes are of particular value when they are inserted at well-defined positions. In particular, fluorescent amino acids bearing environment-sensitive fluorophores are of high interest for peptide research and their development is currently in progress.[5−12] However, strict requirements need to be fulfilled for these derivatives in order to use them in biological applications. At first, they should be of small size, similar to the aromatic amino acid tryptophan. Second, they should exhibit high brightness and strong sensitivity to the environment. There are a few examples in literature that fit these requirements, such as amino acids bearing Prodan,[6,7] 4-dimethylaminonaphthalimide (4-DMAP) and its analogues,[8−10] and 3-hydroxychromones (3HCs).[11,12] The first two bear conventional push−pull fluorophores, which change their emission color and/or intensity on changes in the environment polarity. Thus, Prodan is a unique molecule showing strong positive solvatochromism in fluorescence,[13,14] so that its emission shifts to the red on increase in solvent polarity and H-bond donor ability. A Prodan-derived amino acid was successfully applied to study the interaction of S-peptide with ribonuclease S[7] and δ-opioid receptor with antagonists[15] as well as to estimate the local dielectric constant of the B1 domain of the staphylococcal protein G.[6] 4-DMAP and its analogues with

Special Issue: Photoinduced Proton Transfer in Chemistry and Biology Symposium

Received: August 29, 2014
Revised: October 9, 2014
Published: October 13, 2014

extended conjugation also show positive solvatochromism, but their most important property is their strong fluorescence quenching in aqueous media, which is used for intensiometric detection.[8−10] A fluorescent amino acid based on 4-DMAP incorporated into octapeptides was used to monitor their binding to 14−3−3bp protein,[8] while its improved analogues (e.g., 6-dimethylaminonaphthalimide) were applied to investigate SH2 phosphotyrosine binding domains,[9] major histocompatibility complexes (MHC) at the cell surface,[16] and peptide−calmodulin interactions.[10] An alternative approach is to utilize fluorophores based on 3-hydroxyflavones (3HFs) and 3HCs as a larger family that show dual emission, as a result of the formation of two excited-state tautomers (N* and T*) due to an excited state intramolecular proton transfer (ESIPT).[17,18] As ESIPT can be easily modulated by dipolar and H-bonding interactions, the dual emission of 3HC dyes depends on local polarity and hydration with exquisite sensitivity.[1,19−27] This allows for instance distinguishing single from double stranded DNA,[28] detecting the interaction of peptides with nucleic acids,[11,29,30] proteins,[21,31] and lipid membranes,[12,32] and monitoring conformational changes of proteins.[33] Recently, we reported on 3HCaa, the first amino acid undergoing ESIPT, that bears 2-furyl-3-hydroxychromone,[11] which was particularly suited for sensing the environment in highly polar media.[19−21,34] By introducing this amino acid at different positions, we could map the interaction of a viral peptide with oligonucleotides (ODNs).[11] Moreover, due to its small size, this amino acid was able to perfectly replace a highly conserved tryptophan residue, critical for the biological activity of this peptide. However, the limitations of the 2-furyl-3-hydroxychromone are its relatively low fluorescence quantum yield and ground state deprotonation of its 3-OH group, which complicated the data analysis and do not allow its use in the cellular context. The 4′-methoxy-3-hydroxyflavone fluorophore overcame these problems and in addition showed much higher sensitivity to the environment hydration, allowing significantly improved detection of peptide interactions with nucleic acids.[30] Therefore, preparation of a fluorescent amino acid based on 4′-methoxy-3-hydroxyflavone and its validation in peptides became an urgent task.

The NC(11−55) peptide corresponding to the zinc finger domain of the nucleocapsid protein (NC) of the Human Immunodeficiency Virus, type 1 (HIV-1), is an interesting model for evaluation of new fluorescent amino acids, as it can interact with a variety of biological targets[35] and can easily be prepared by solid phase peptide synthesis.[36] This peptide binds notably to a large range of nucleic acid sequences[37−45] mainly through a hydrophobic platform at the top of its folded fingers.[38,40,42] This platform also plays a key role in the nucleic acid chaperone properties of NC, which enables directing the rearrangement of ODNs into their most stable conformation.[46−48] Within this platform, the highly conserved Trp37 residue of the distal zinc finger plays a crucial role, through its π-stacking with the nucleobases.[38,40,49,50] During the viral cycle, this protein plays important roles in both the cytosol and the nucleus of the host cell. Therefore, labeling of NC by fluorescent amino acids should allow site specific monitoring of its interactions with biological targets in solution as well as in living cells.

In the present work, we synthesized a fluorescent amino acid bearing the ESIPT fluorophore 4′-methoxy-3-hydroxyflavone and incorporated it into two positions of the NC(11−55) peptide. We showed that interactions of the labeled peptides with different ligands can produce a nearly complete switch in their two emission bands. The new amino acid showed much higher sensitivity to interactions compared to the previous analogue. Moreover, significant changes were observed in the fluorescence lifetime values of the T* emission, which allows identification by two-photon fluorescence lifetime microscopy of the intracellular binding partners of the microinjected peptide.

MATERIALS AND METHODS

All solvents and chemicals were purchased from Sigma-Aldrich Chemical Co. For absorption and fluorescence studies, the solvents were of spectroscopic grade. ΔP(−)PBS DNA, SL2 RNA, TAR RNA, cTAR DNA, and calf thymus double stranded DNA (CT-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. Dioleoylphosphatidylcholine (DOPC), dioleoylphosphatidylserine (DOPS), polystyrene (MW = 4 kDa), adenosine, guanosine, cytidine, and thymidine were from Sigma-Aldrich. The concentration of phospholipid stock solutions in chloroform was determined by dry weight.

Proton NMR spectra were recorded on a 300 MHz Bruker spectrometer. Mass spectra were recorded on Bruker HCT Ultra and Agilent Technologies Accurate-Mass Q-TOF LC/MS 6520 mass spectrometers.

Synthesis of Fmoc-M3HFaa. The scheme and protocol of the synthesis of Fmoc-M3HFaa are described in the Supporting Information.

Peptide Synthesis. The labeled NC(11−55) peptides were synthesized by solid phase peptide synthesis[36] as detailed in the Supporting Information. The obtained peptides were characterized by ESI-MS. NC(11−55)-W37-M3HFaa peptide: calculated M = 5288.84, found M = 5288.35; NC(11−55)-A30-M3HFaa peptide: calculated M = 5403.97, found M = 5403.42.

Preparation of Zn-Bound Peptides. Lyophilized labeled peptides were dissolved in water (≈ 0.5 mg in 500 μL). Then, about 10 μL of this solution were used to determine the peptide concentration using an extinction coefficient of 23 000 M^{-1} cm^{-1} at 360 nm. Next, 2.0 mol equiv of $ZnSO_4$ were added to the peptide and the pH was raised to its final value, by adding buffer. This last step was done only at the end to prevent peptide oxidation. Noticeably, a large excess of Zn^{2+} ions should be avoided since this ion could affect the M3HFaa fluorescence.

Peptide Activity Tests. The ability of the labeled peptides to destabilize the secondary structure of ODNs[47,51] and to promote the annealing of cTAR with its complementary dTAR DNA sequence[52] was tested as described in the Supporting Information.

Experiments with Free Nucleosides. To determine the sensitivity of M3HFaa spectroscopic properties to nucleosides, 2 mL of aqueous stock solution of free nucleoside (35 mM) were mixed with 0.5 mL of M3HFaa in buffer (12 μM M3HFaa, 10 mM phosphate, 60 mM NaCl, 1.5 M sucrose). To avoid guanosine precipitation, the spectroscopic properties of the sample were recorded immediately after heating and cooling down to room temperature. Measurements were repeated 3 times for each nucleoside.

Preparation of M3HFaa in a Polystyrene Film. 0.3 g of polystyrene (M = 4 kDa) was dissolved in 0.5 mL of a 3 μM M3HFaa solution in ethyl acetate. To evaporate the solvent, the

resulting viscous solution was placed onto a glass coverslip. Fluorescence measurements were performed directly on the solid polystyrene film at the surface of the coverslip.

Preparation of Large Unilamellar Vesicles (LUVs). LUVs were obtained by extruding a suspension of multilamellar vesicles with a Lipex Biomembranes extruder and polycarbonate filters of calibrated pores (Nucleopore). The size of the filters was first 0.2 μm (7 passages) and then 0.1 μm (10 passages). This protocol leads to LUVs with a mean diameter of 110−120 nm according to dynamic light scattering measurements.

Spectroscopic Measurements. Unless otherwise indicated, experiments were performed in 10 mM phosphate buffer, pH 7.0, 30 mM NaCl, at 20 °C. Absorption spectra were recorded on a Cary 4000 UV−visible spectrophotometer (Varian). Fluorescence spectra were recorded on a FluoroMax3 spectrofluorimeter (Jobin Yvon) equipped with a thermostated cell compartment. Fluorescence spectra were corrected for Raman scattering and lamp fluctuations. Quantum yields were calculated using quinine sulfate in 0.5 M sulfuric acid (quantum yield is 0.546) as a reference.[53] The excitation wavelength was 360 nm for M3HFaa.

Deconvolution of Spectra. Deconvolution of the fluorescence spectra into N* and T* bands was performed using the Siano software kindly provided by Prof. A.O. Doroshenko (Kharkiv, Ukraine), as previously described.[54] This deconvolution is based on an iterative nonlinear least-squares method, where the individual emission bands were approximated by a log-normal function accounting for several parameters: maximal amplitude, I_{max}, spectral maximum position, ν_{max}, and position of half-maximum amplitudes, ν_1 and ν_2, for the blue and red parts of the band, respectively. These parameters determine the shape parameters of the log-normal function, namely the full width at the half-maximum, fwhm = $\nu_1-\nu_2$, and the band asymmetry, P = $(\nu_1-\nu_{max})/(\nu_{max}-\nu_2)$. All parameters were allowed to vary in the iteration process, with the exception of the fwhm parameter of the T* band which was fixed in cases where the intensity of the T* band was low. In these last cases, the fwhm value of the T* band was fixed at 2600 cm^{-1} based on the data from binary solvent mixtures. The resulting integral fluorescence intensities, I_{N*}/I_{T*}, of the separated N* and T* bands were used for calculation of the ratio of the two bands.

Time-Resolved Spectroscopy. Time-resolved fluorescence measurements were performed using the time-correlated single-photon counting technique (TCSPC)[55] as described in the Supporting Information.

Fluorescence Lifetime Imaging Microscopy. Fluorescence microscopy experiments were performed on a home-built two-photon laser scanning setup based on an Olympus IX70 inverted microscope with an Olympus 60× 1.2NA water immersion objective.[56] Two-photon excitation was provided by a Ti-sapphire (Tsunami, Spectra Physics) or an Insight DeepSee (Spectra Physics) laser. Imaging was carried out using two fast galvo-mirrors operating in the descanned fluorescence collection mode. Photons were detected with an avalanche photodiode (APD SPCM-AQR-14-FC, PerkinElmer) which was connected to a time-correlated single photon-counting (TCSPC) module (SPC830, Becker and Hickl) operating in reversed start−stop mode. The typical acquisition time was 5 s with an excitation power around 2.5 mW (760 nm) at the sample. Subsequent data analysis using the software SPC Image (Becker and Hickl) allowed us to extract the fluorescence lifetimes from the decays, and their visualization in a FLIM image, using an arbitrary color scale.

Cell Cultures. HeLa cells were cultured in Dulbecco's modified Eagle medium (D-MEM, high glucose, Gibco-invitrogen) supplemented with 10% (v/v) fetal bovine serum (FBS, Lonza), 1% antibiotic solution (penicillin-streptomycin, Gibco-invitrogen) in a humidified incubator with 5% CO_2, 95% air atmosphere at 37 °C. Cells plated on a 75 cm^2 flask at a density of (5−10) × 10^4 cells/mL were harvested at 80% confluence with trypsin−EDTA (Sigma) and seeded on a glass coverslip at a density of (2.5−5) × 10^4 cells/coverslip, 48 h before the microscopy measurements. Finally, cells on the coverslip were mounted on the microscope holder and covered with D-MEM solution.

Microinjection. Microinjection was performed on cells seeded on a gridded coverslip (Cellocate; Eppendorf), which allows locating cells. Before loading into the microneedle, the solution of 200 μM labeled peptide was centrifuged for 20 min at 4 °C. Microinjection was performed using a micromanipulator (model 5171; Eppendorf) coupled to a microscope (model DMIRBE; Leica) at 40× magnification. The positions of the injected cells were noted to find them after placing the coverslip under the FLIM microscope.

RESULTS AND DISCUSSION

The L-amino acid M3HFaa bearing the 4′-methoxy-3-hydroxyflavone fluorophore was prepared by using a previously described synthesis strategy.[11] In this procedure (Scheme S1), L-tyrosine was C-acetylated to the corresponding *o*-hydroxyacetophenone **1** using a Fries reaction.[57,58] Then, the amino function was protected with the Boc group, and the product **2** was condensed with 4-methoxybenzaldehyde in the presence of alkali. An Algar−Flynn−Oyamada reaction was then used to convert the obtained chalcone **3** into a 3HF derivative (Boc-M3HFaa, **4**). Then, the desired amino acid M3HFaa **5** (Figure 1A) was obtained by removing the Boc group. Finally, this compound was reacted with the Fmoc group to obtain product **6**, which was used in solid-phase peptide synthesis. We checked that the basic conditions used in the synthesis do not alter the optical purity of the obtained amino acid.[11,12]

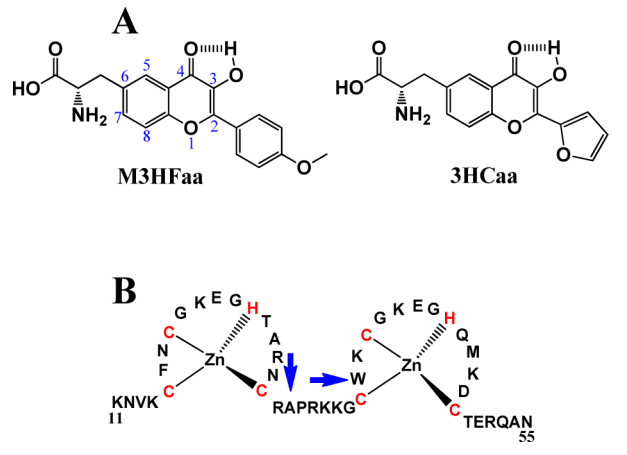


Figure 1. Amino acid and peptide structures. (A) Structure of the new fluorescent amino acid (M3HFaa) and the previously synthesized 3HCaa.[11] (B) Sequence of the NC(11−55) peptide. The arrows show the two positions where the natural amino acids were replaced by the M3HFaa.

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

The spectroscopic properties of the new fluorescent amino acid were characterized in different solvents (Figure S1, Table S1). The absorption spectrum displayed a single band at 355−360 nm which varies mildly in different solvents. The position of the dual emission bands was found to be moderately sensitive to solvent polarity, whereas the ratio of the two emission bands (N*/T*) dramatically increased from 0.04 to 4.80 with an increase in solvent polarity and hydrogen bond donor ability (Table S1). Moreover, the values of the fluorescence quantum yield showed only slight variations within the polar solvents but increased by 2-fold in the nonpolar 1,4-dioxane. The absorption and emission maxima of M3HFaa were close to those reported previously for the F6C fluorophore,[30] indicating that the substitution at position 6 only moderately affects the spectroscopic properties of the fluorophore.

Previously, we showed that 3HC derivatives bearing a 2-furyl or 4-methoxyphenyl group at the 2-position are highly sensitive to water, which enabled their use for determining local hydration.[11,59] Therefore, we investigated the fluorescence properties of M3HFaa in water/*N*,*N*-dimethylformamide (DMF) mixtures (Figure 2) to determine whether this probe

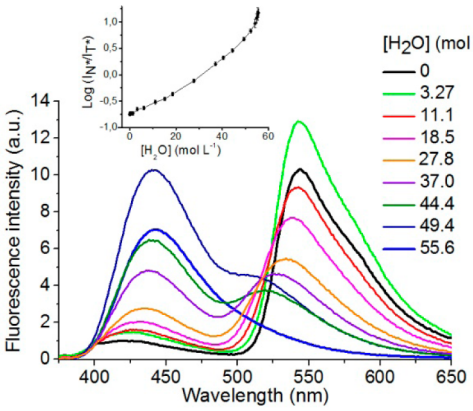


Figure 2. Fluorescence spectra of M3HFaa in water−DMF mixtures. Inset: plot of Log(I_{N*}/I_{T*}) as a function of water concentration in water−DMF mixtures. Excitation wavelength was 360 nm. Concentration of M3HFaa was 3 μM.

exhibits the same properties. As expected, the dual emission of M3HFaa was also found to depend strongly on the water concentration in DMF, showing an increase of the relative intensity of its N* band in more hydrated media. This increase is likely a consequence of the inhibition of the ESIPT reaction by H-bonding of the N* form with water molecules.[1,19,27] The N* and T* bands were then deconvoluted from the spectra, and the integral fluorescence intensities, I_{N*}/I_{T*}, of the separated N* and T* bands were plotted as a function of the water concentration (Figure 2, inset). Much stronger changes were observed compared to the amino acid 3HCaa,[11] in agreement with the previously reported superiority of 2-(4-methoxyphenyl)- over 2-(2-furyl)-3HCs,[30,59] indicating that M3HFaa exhibits a substantially higher sensitivity to water than the first generation 3HCaa fluorescent amino acid analogue. Noticeably, the observed deviation from the linearity, mainly at high water concentrations, could be explained by preferential solvation of the dye with DMF molecules in water, commonly observed with fluorescent dyes in solvent mixtures.[60,61]

Using compound **6**, we then synthesized two labeled NC(11−55) peptides, A30-M3HFaa and W37-M3HFaa, where the fluorescent amino acid replaced the Ala30 and Trp37 residues located in the linker and the distal zinc finger, respectively (Figure 1B). The labeled peptides were found to be highly pure and of the expected mass, underlining that the Fmoc-M3HFaa derivative was compatible with the standard protocols of peptide synthesis and purification. In order to check whether the M3HFaa may affect the peptide structure and activity, we compared the chaperone activity of the labeled peptides with that of the native peptide. To this end, we characterized the ability of the peptides to destabilize the cTAR sequence and promote its annealing with its complementary dTAR sequence. Both activities are critically dependent on the hydrophobic platform at the top of the folded fingers[44] and notably on the Trp37 residue, since its replacement by a nonaromatic residue leads to profound alterations in the nucleic acid binding and chaperone properties[39,44,46,62] as well as a complete loss of HIV-1 infectivity.[63] Remarkably, the W37-M3HFaa peptide showed nucleic acid destabilization and annealing activities nearly identical to those of the native peptide (Figures S2 and S3). The fully preserved chaperone activity of W37-M3HFaa indicates that, similarly to 3HCaa,[11] M3HFaa does not affect the peptide folding and can functionally substitute the key Trp37 residue. This is likely because M3HFaa bears a relatively small aromatic fluorophore, so that it can mimic the stacking interactions of the Trp37 indole moiety with nucleobases.[28,29,38−40,42,64] While A30-M3HFaa efficiently promoted the cTAR/dTAR annealing, it was twice less active than the wild-type NC(11−55) in its destabilization activity. Though Ala30 is not critical for the chaperone activity of the peptide, its substitution by the more bulky M3HFaa amino acid may perturb the relative orientation of the two zinc fingers, and thus the proper formation of the hydrophobic plateau.

In contrast to the free M3HFaa in buffer, the labeled peptides (Figure 3) showed distinct dual emission bands with a much lower N*/T* ratio (Table 1), which indicates a partial screening of the M3HFaa fluorophore from water, leading to a more efficient ESIPT reaction. Remarkably, this ratio is smaller for the W37-M3HFaa peptide compared to A30-M3HFaa, suggesting a more efficient screening of the probe from bulk water in the former mutant. Moreover, similarly to 3HCaa-labeled peptides,[11,30] we observe a significant blue shift of the position of the T* band in comparison to the free amino acid in alcohols or aprotic solvents. This blue-shifted position is a hallmark of extensive exposure to water for this class of 3HC-based amino acid analogues, which was explained by the strong H-bonding of the tautomer form with the solvent.[65] Therefore, though the exposure of M3HFaa to water in both peptides is less than that for the free amino acid, it is still very high. Noticeably, while the fluorescence quantum yield of A30-M3HFaa is comparable to that in polar solvents, the quantum yield of the W37-M3HFaa is two times higher. Thus, in its somewhat less solvent-exposed environment, the M3HFaa at position 37 may undergo specific interactions with its neighboring amino acids that can increase its emission.

In a next step, we made a comparative investigation of the changes in the emission spectra of the M3HFaa- and 3HCaa-labeled peptides on interaction with ODNs. As representative examples of RNA and DNA target sequences of NC, we used the SL2 and ΔP(−)PBS stem-loop sequences (Figure S4), respectively. For both sequences, the 3D structure of their

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

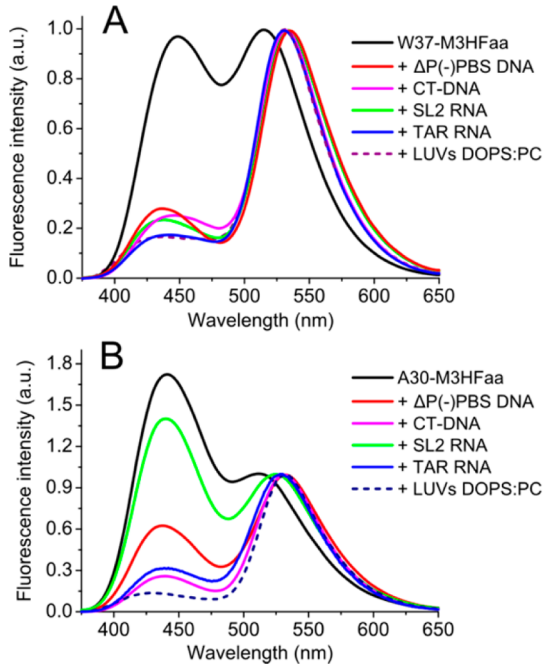


Figure 3. Effect of ODN interaction on the fluorescence spectra of the M3HFaa-labeled NC(11−55) peptides. Normalized fluorescence spectra of the W37-M3HFaa (A) and A30-M3HFaa (B) peptides. Concentrations of peptides, and ΔP(−)PBS and SL2 ODNs were 1 μM; TAR RNA, 0.25 μM; CT-DNA, 40 μM; and LUVs, 400 μM. Measurements were done in 10 mM phosphate buffer, 30 mM NaCl, pH 7.0. Excitation wavelength was 360 nm.

complexes with NC was solved by NMR,[40,42] so that the observed changes in the emission spectra could be correlated with the structure. For W37-M3HFaa, a large red shift of the T* emission band and a dramatic decrease of the N*/T* ratio were observed on interaction with the two ODNs (Figure 3). These changes are fully in agreement with the stacking interaction of the W37 residue with the nucleobases in both complexes[40,42] and, thus, confirm that M3HFaa appropriately substitutes the Trp residues. In the case of A30-M3HFaa, only moderate changes in the position of the T* band and of the N*/T* ratio were observed on interaction with SL2, which corroborates with the distant location of A30 with respect to the nucleobases in the SL2/NC complex.[42] In contrast, larger changes in both of these parameters were observed on interaction with ΔP(−)PBS, where A30 is more closely located to the nucleobases in the complex.[40] Thus, for both positions, the M3HFaa residue likely preserves the mode of interaction of the peptide with its ODNs targets. Importantly, both the shift of the T* band and the decrease in the N*/T* ratio were significantly larger for the M3HFaa-labeled peptide compared to the 3HCaa-labeled peptide (Figure 4), which agrees well with the stronger environment sensitivity of M3HFaa compared to 3HCaa. Moreover, the quantum yield of the M3HFaa-labeled peptides in complexes with both ODNs was two to three times higher than that for the 3HCaa-labeled peptides. Thus, M3HFaa is a superior dye to 3HCaa with respect to brightness and sensitivity.

To further characterize the spectroscopic changes of A30-M3HFaa and W37-M3HFaa on interaction with NC targets, we investigated the binding of the two peptides to calf thymus DNA (CT-DNA) and TAR RNA (Figure S4), as models of a double-stranded DNA and nonspecific RNA binding sequence, respectively. Interestingly, similar values of the N*/T* ratio were observed for W37-M3HFaa with all ODNs (Table 1), suggesting that the M3HFaa residue at position 37 stacks similarly with the nucleic bases in all of the complexes. A much larger range of values of the N*/T* ratio was observed for M3HFaa at position 30, indicating a less conserved interaction at this position. The large decrease in the N*/T* ratio for binding of the A30-M3HFaa peptide to TAR RNA or CT-DNA compared to ΔP(−)PBS or SL2 RNA clearly indicates a closer interaction of the M3HFaa residue with the ODNs in the former complexes. Thus, it suggests that NC(11−55) binds differently to the nonspecific binding sites of TAR RNA and CT-DNA compared to its specific binding site in SL2 RNA and ΔP(−)PBS, respectively.

Finally, binding of the labeled peptides to lipid membranes was investigated using DOPS/DOPC large unilamellar vesicles (LUVs) with a lipid ratio of 32:68 (Figure 3, Table 1) that were shown to bind NC(11−55) with good affinity (data not

Table 1. Spectroscopic Properties of the M3HFaa-Labeled NC(11−55) Peptides in Complexes with ODNs[a]

	ligand	λ_{ABS} (nm)	λ_{N^*} (nm)	λ_{T^*} (nm)	N*/T*	QY (%)
M3HFaa, buffer	−	355	442	520	4.80	7.8
M3HFaa, methanol	−	357	434	535	0.95	6.7
W37-M3HFaa	−	360	451	515	1.04	16.2
	ΔP(−) PBS DNA	368	441	534	0.27	14.6
	CT-DNA	365	445	532	0.25	15.0
	SL2 RNA	364	438	532	0.21	14.2
	TAR RNA	367	443	532	0.18	15.2
	DOPS/PC (32:68)	360	442	532	0.16	25.1
A30-M3HFaa	−	360	441	512	1.60	5.1
	ΔP(−) PBS DNA	365	439	531	0.58	4.8
	CT-DNA	360	438	531	0.26	8.0
	SL2 RNA	360	441	524	1.26	5.0
	TAR RNA	365	439	530	0.30	8.2
	DOPS/PC (32:68)	360	428	532	0.14	22.9

[a] λ_{ABS}, λ_{N^*} and 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 $n \geq 2$ experiments. Standard deviations are ∼5% for N*/T* and QY. Excitation wavelength was 360 nm. Measurements were done in 10 mM phosphate buffer, 30 mM NaCl, pH 7.0.

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

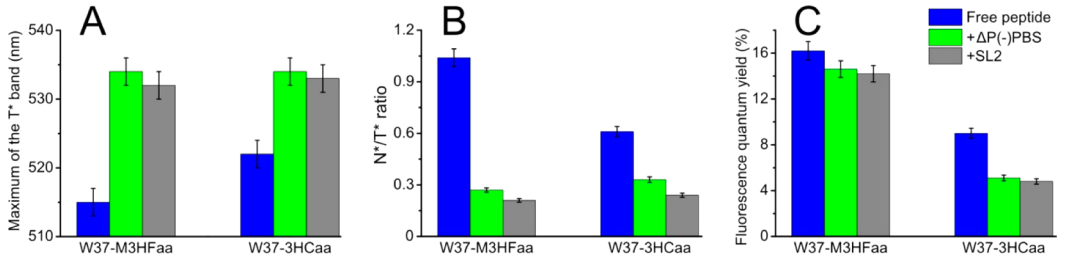


Figure 4. Comparison of λ_{T^*} (A), N*/T* ratio (B), and fluorescence quantum yield (C) values for the NC(11−55) peptides labeled at position 37 by M3HFaa or 3HCaa.

Table 2. Lifetime Values of the M3HFaa Label and M3HFaa-Labeled NC(11−55) Peptides in Complexes with ODNs[a]

	Ligand/solvent	τ_1^{T*} (ns)	a_1^{T*}	τ_2^{T*} (ns)	a_2^{T*}	τ_3^{T*} (ns)	a_3^{T*}	$\langle\tau\rangle$ (ns)
M3HFaa	Methanol	0.135	−1	0.41	1	–	–	0.41
	Polystyrene	ND	ND	2.84	0.32	5.80	0.68	4.56
	Adenosine	0.133	−1	0.74	0.70	2.15	0.30	1.16
W37-M3HFaa	–	0.081	−1	0.75	0.65	3.15	0.35	1.59
	+ΔP(−)PBS DNA	0.050	−1	1.93	0.47	6.07	0.53	4.12
	+CT-DNA	0.056	−1	2.00	0.40	6.27	0.60	4.56
	+SL2 RNA	0.068	−1	2.06	0.38	6.14	0.62	4.59
	+TAR RNA	0.051	−1	2.24	0.43	6.31	0.57	4.56
	+DOPS/PC (32:68)	0.062	−1	2.53	0.57	5.17	0.43	3.66
A30-M3HFaa	–	0.056	−1	0.32	0.89	1.55	0.11	0.44
	+ΔP(−)PBS DNA	0.037	−1	0.39	0.75	3.38	0.25	1.14
	+CT-DNA	0.036	−1	0.92	0.52	5.23	0.48	2.99
	+SL2 RNA	0.051	−1	0.34	0.92	3.00	0.08	0.55
	+TAR RNA	0.037	−1	0.75	0.58	4.91	0.42	2.50
	+DOPS/PC (32:68)	0.051	−1	2.58	0.41	5.22	0.59	4.14

[a] τ_i are the fluorescence lifetimes, and α_i their amplitudes. $\langle\tau\rangle$ is the mean fluorescence lifetime, calculated by using only the lifetimes associated with positive amplitudes according to $\langle\tau\rangle = \sum\alpha_i \tau_i$. Excitation and emission wavelengths were 315 and 550 nm, respectively. The reported data are the mean of $n \geq 2$ experiments. Standard deviations are <15% for lifetimes and amplitudes. ND, nondetermined because of strong scattering of the light by the polystyrene film.

shown). For both A30-M3HFaa and W37-M3HFaa peptides, we observed a strong decrease in the N*/T* ratio on binding to LUVs, indicating that both residues experienced comparable poorly hydrated environments due to insertion of the probes into the lipid bilayer. Interestingly, the interaction with lipid membranes was also accompanied by a strong increase in the quantum yield for both peptides. This increase (from 16.2% to 25.1% for W37-M3HFaa and from 5.1% to 22.9% for A30-M3HFaa) may be partly attributed to the nonpolar environment in the membrane, since a significant increase in the quantum yield was also observed when the free M3HFaa residue was transferred from water to dioxane (Table S1). Both the changes in the N*/T* ratio and in quantum yield clearly confirm the recently demonstrated ability of the positively charged NC(11−55) peptide to effectively bind the negatively charged lipid membranes (data not shown) and further indicate that the amino acids at positions 30 and 37 in the complexes with LUVs experienced a significantly less polar and less water accessible environment than in the complexes with ODNs.

To gain further information on the photophysical properties of the labeled NC peptides in complexes with their targets, time-resolved fluorescence measurements were first performed using the time-correlated single photon counting (TCSPC) technique with the free label in buffer, methanol, ethanol, *N*,*N*-dimethylformamide, and 1,4-dioxane to investigate the effect of polarity (Table S1). The fluorescence decays were monitored both for the N* form (emission wavelength 430 nm) and for the T* form (emission wavelength 550 nm). In these solvents, the time-resolved decays monitored at the N* emission band were monoexponential with a fast decay component (τ_1). For the T* emission band, the time-resolved decays were biexponential with a fast rise component (τ_1^{T*}) and a longer decay component (τ_2^{T*}). In addition, the fast decay component associated with the monexponential decay of the N* form agrees quite well with the rise component of the T* form, suggesting its formation through an ESIPT (N* → T*) reaction. The magnitude of the fast decay component of the N* form increases with polarity from 42 ps in DMF to 294 ps in water. Furthermore, the τ_2^{T*} lifetime of the T* form was found to decrease with an increase in polarity (from 1.63 ns in dioxane to 410 ps in methanol). The results observed are close to those reported earlier for (2-furyl)-3-hydroxychromone,[19] the fluorophore moiety of 3HCaa, and for 2-thienyl-3-hydroxychromone,[66] suggesting that, similar to these two compounds, M3HFaa may also undergo an irreversible ESIPT reaction. This explains that only the ESIPT associated decay time τ_1^{N*} is observed for the N* form. As for the peptides and their complexes, a similar irreversible ESIPT reaction occurs (data not shown); we only reported the decay parameters of the T* form which yielded more detailed information on the environment of the labeled peptides (Table 2).

The fluorescence decays of the T* form for both the A30-M3HFaa and the W37-M3HFaa peptides were characterized by a short rise and two longer decay components. The magnitude of the rise component τ_1^{T*} was comparable in both peptides and is poorly dependent on the nature of the ligand bound to

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

the peptides, thus indicating that the ESIPT reaction contributes moderately to the ratio of intensities of the dual emission bands. In sharp contrast, the values of the longer decay components τ_2^{T}* and τ_3^{T}* were strongly dependent on the position of the fluorophore label in the peptide, as well as on the nature of the ligand bound to the peptide. These two fluorescence lifetimes are likely associated with two different orientations of the fluorophore in the peptide. Both, the similarity in the value of τ_2^{T}* (0.32 ns) of the free A30-M3HFaa peptide to that in the polar protic solvents and its very high amplitude (89%) are fully consistent with the high water accessibility of M3HFaa at this position suggesting that this lifetime may correspond to a peptide conformation where the M3HFaa label is largely exposed to the solvent. In contrast, the τ_2^{T}* value (0.75 ns) of W37-M3HFaa and the τ_3^{T}* values of both A30-M3HFaa (1.55 ns) and W37-M3HFaa (3.15 ns) peptides were much higher than the M3HFaa lifetimes in polar solvents. These rather high values of τ_2^{T}* and τ_3^{T}* cannot solely be explained by changes in the polarity of the probe environment, but they are likely the consequence of interactions of the M3HFaa fluorophore with the peptide backbone or neighboring side chains of the peptide. As discussed below for the complexes with ODNs, these lifetimes may originate from the restriction of the relative motions of the two aromatic moieties of the M3HFaa label by the peptide backbone or neighboring amino acids, favoring a more emissive conformation of the probe.

Interestingly, addition of the ODNs was found to induce a significant increase of both τ_2^{T}* and τ_3^{T}* values for the W37-M3HFaa peptide (Table 2). As the major mode of interaction of Trp37 in peptide/ODN complexes relies on stacking interactions with the nucleobases,[38,40,42,50] the long-lived decay components (τ_3^{T}* > 6.0 ns) associated with high amplitudes (≥53%) observed for W37-M3HFaa may be associated with the M3HFaa residues stacked with the nucleobases. These interactions likely restrict the relative motions of the two aromatic moieties of M3HFaa, favoring a more planar and hence a more emissive conformation of the molecule. This behavior is fully reminiscent of that of DNA intercalating agents, which become highly fluorescent when they are intercalated in their planar conformation within the DNA base pairs.[67−70] Moreover, this stacking interaction of the planar emissive conformation of M3HFaa also explains well the low accessibility of the fluorescent amino acid to water molecules in all of the complexes of W37-M3HFaa with the ODNs (Table 1). To further validate the attribution of these long-lived decay components to stacked conformations, we included the M3HFaa probe in a polystyrene matrix where the M3HFaa probe is expected to be present in a highly rigid environment with possible π-stacking interactions with the polystyrene aromatic rings. In line with our expectations, the decay of the T* form of the probe in this matrix was shown to exhibit a 5.8 ns component with a major contribution (~70%) which clearly confirmed our conjecture. Moreover, due to the low water accessibility of the M3HFaa probe in all the W37-M3HFaa/ODN complexes, the relatively shorter decay components (τ_2^{T}*) with values ranging from 1.9 to 2.2 ns may tentatively be attributed to peptide conformations where the M3HFaa residue interacts transiently with the ODNs bases or backbone. To validate this hypothesis, we tried to mimic the behavior of M3HFaa in the peptide/ODN complexes, by adding M3HFaa to a highly concentrated solution (28 mM) of a free nucleoside (A, C, G, or T). In these highly concentrated solutions, M3HFaa is thought to collide efficiently with the nucleosides and form transient complexes that may restrict the relative motions of the two aromatic rings of M3HFaa and enhance its emission. Interestingly, a distinct increase in intensity of the T* band was observed in the emission spectrum of M3HFaa with the following order: C < T < A (Figure 5). This order may be related to the relative stacking

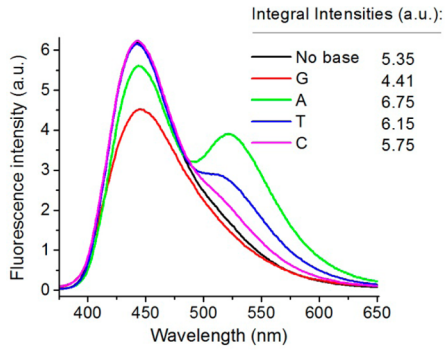


Figure 5. Fluorescence spectra of M3HFaa probe in the presence of 28 mM of free nucleosides. Probe concentration was 2.4 μM. Measurements were done in 10 mM phosphate buffer, 0.3 M sucrose, pH 7.0. Excitation wavelength was 360 nm.

abilities of these nucleobases which increase with the same order.[71] As a consequence of its strong stacking ability, the A nucleobase may efficiently increase the T* band emission of M3HFaa, both by screening M3HFaa from the bulk water molecules and by restricting the relative motion of the two aromatic rings of M3HFaa. In the presence of A, we observed a 2.15 ns lifetime that matches well with the values of the τ_2^{T}* component in the complexes, which lends support to our hypothesis. Noticeably, in agreement with the steady-state data (Table 1), the similar values of the time-resolved decay parameters for the complexes of W37-M3HFaa with all tested ODNs confirmed that the M3HFaa residue at position 37 stacks in a similar manner with the nucleobases in these complexes, in line with the key role of Trp37 in the binding of NC to nucleic acids.[38,40,42,49,50] Interestingly, in contrast to the other three nucleosides, G induced a small decrease in the emission of the M3HFaa probe, indicating that G could quench the M3HF emission, likely through an electron transfer mechanism.[72]

For the A30-M3HFaa peptide in complex with SL2, we observe that the decay of the T* form is largely dominated (92%) by a 0.34 ns component, close to the lifetime of M3HFaa in polar solvents. This is in agreement with the NMR data showing that the Ala30 residue is quite far from SL2 in the complex and is largely exposed to water.[42] The minor 3.0 ns component may likely be ascribed to a different conformation of the peptide in interaction with SL2 or to an interaction of the peptide with a different binding site of SL2. For the complex with ΔP(−)PBS, the short-lived lifetime (0.39 ns) is also the major component (75%), indicating that the M3HFaa residue is mainly exposed to water and not in contact with the ΔP(−)PBS. Moreover, the increase in the value of the long-lived lifetime (3.38 ns) and its amplitude (25%) for ΔP(−)PBS compared to SL2 corroborates with the closer proximity of the Ala30 residue to ΔP(−)PBS compared to SL2 in the complexes,[40,42] as well as with the existence of alternate binding sites on ΔP(−)PBS.[40,46] Finally, a further increase in

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

the value of the long-lived lifetime and its amplitude together with an increase in the τ_2^{T*} value were observed for CT-DNA and TAR, indicating even closer interactions of the M3HFaa residue at position 30 with the ODNs in these complexes. Therefore, the time-resolved fluorescence data confirmed that NC(11−55) binds differently to the nonspecific binding sites of CT-DNA and TAR, compared to the specific binding sites of ΔP(−)PBS and SL2.

In the presence of DOPS/DOPC LUVs, the decay of both the labeled peptides exhibits similar lifetime values ($\tau_2^{T*} \approx 2.5$ ns and $\tau_3^{T*} \approx 5.2$ ns) and amplitudes. By analogy with the complexes with ODNs, the long-lived lifetime may be attributed to a highly emissive flattened conformation of the dye that results from the insertion of the M3HFaa probe into the lipids. Moreover, the short-lived lifetime (τ_2^{T*}) may be attributed to a less deep insertion of the dye into the lipids which induces a less restricted conformation of the dye. Since the amplitude of τ_2^{T*} for each peptide correlates well with the τ_3^{T*} amplitude of the other peptide, it is suggested that the peptide adopts two binding modes which differs by the relative insertions of the residues at positions 30 and 37.

It should be noted that the fluorescence lifetime data correlate reasonably well with the variation in the dual emission bands of the probe. Thus, from the plot of the N*/T* ratio vs mean lifetime of the T* form (Figure 6), it is observed that

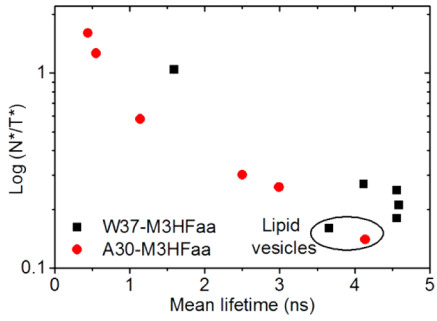


Figure 6. Correlation of the logarithm of the N*/T* intensity ratio vs the mean lifetime of the labeled peptides and their complexes with ODNs and lipid vesicles (LUVs).

higher lifetimes correspond to lower values of the ratio. Moreover, it may further be noted that, within the same labeled mutant, the data correlate even better, with the only deviation for W37-M3HFaa in lipid vesicles. This correlation results from the changes in the hydration and rigidity (microviscosity) of the local environment of M3HFaa. Indeed, the interaction of the peptide with its partners leads to two correlated processes, namely exclusion of the water molecules from the site of interaction as well as formation of a confined rigid environment. The observed deviation for W37-M3HFaa in lipids is likely due to the different nature of the local environment of the complexes with lipids and nucleic acids. Indeed, the probe is efficiently intercalated within the nucleic acid bases that ensures very high rigidity and low hydration, whereas in the case of lipids the environment is expected to be even less hydrated (lower N*/T* ratio) but less rigid (lower lifetime). The absence of this deviation for A30-M3HFaa confirms that intercalation of its probe within the nucleic acid bases is weak, while the mode of insertion into the lipid membranes is very similar to that for W37-M3HFaa.

Taken together, our data show that, in addition to the N*/T* intensity ratio, the fluorescence lifetimes of M3HFaa in labeled NC(11−55) peptides are highly sensitive to the nature of their interacting partners and to the position of the M3HFaa probe in the peptide. As a consequence, the nature of the intracellular interacting partners of NC could in principle be identified by imaging the lifetime distribution of the microinjected labeled W37-M3HFaa and A30-M3HFaa peptides in cells by two-photon FLIM microscopy since, in contrast to the fluorescence intensity, the fluorescence lifetime is an absolute parameter which is independent of the label concentration. In contrast to the TCSPC measurements in solution which allow measuring a large number of photons, FLIM images are collected with a limited number of photons per pixel, giving thus poor statistics which preclude analyzing the decays with more than one or two lifetime components. Moreover, our FLIM setup uses an avalanche photodiode detector with a broad response function, which does not allow recovering lifetimes below 400 ps. Therefore, as a first step to properly analyze the FLIM images of microinjected cells, we repeated the lifetime measurements with the labeled peptides and their complexes using the FLIM setup to obtain reference lifetime data with this setup. The obtained decays were analyzed by a two component analysis which provided a good fit for all solutions. The data were expressed in the form of the mean lifetimes in Figure 7. The mean lifetimes determined by FLIM

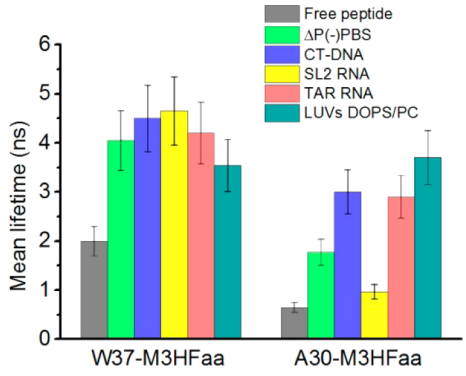


Figure 7. FLIM data of solutions of W37-M3HFaa and A30-M3HFaa, either as free peptides or in complexes with ODNs or LUVs. Two-photon excitation wavelength was at 760 nm. Emission filter was 500−540 nm.

were in reasonable agreement with those calculated from TCSPC measurements (Table 2), except for the free A30-M3HFaa peptide and its complexes with ΔP(−)PBS and TAR, as a consequence of the limited time resolution of the FLIM detector that does not allow the short-lived lifetimes of these species to be properly recovered. According to Figure 7, free peptides should be easy to recognize since they provide short-lived lifetimes at both 30 and 37 positions (0.65 and 2.0 ns, respectively). Moreover, discrimination between the peptides bound to nucleic acids and lipid membranes should also be possible, since the mean lifetimes with lipid membranes differed somewhat from those with nucleic acids for both peptides.

Two photon FLIM images of HeLa cells microinjected with 100 μM W37-M3HFaa and A30-M3HFaa are given in Figure 8A and B, respectively. For each pixel, the recorded intensity decay was analyzed by a biexponential fit. Then, the lifetime image was built using the mean lifetimes and an arbitrary color scale where shorter lifetimes are shown with warm colors and longer lifetimes with cold colors. Both peptides were found distributed all over the cell, but with much higher

dx.doi.org/10.1021/jp508748e | J. Phys. Chem. B 2015, 119, 2585−2595

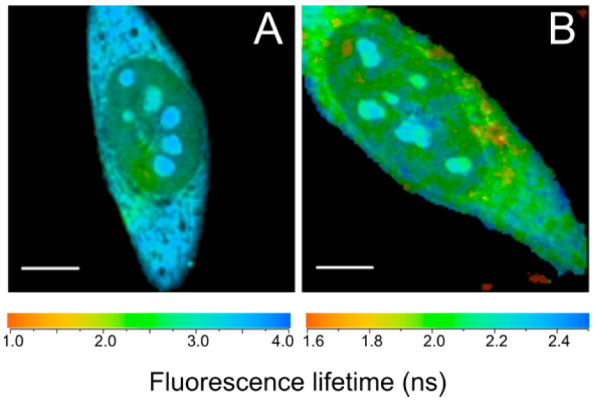


Figure 8. Fluorescence lifetime imaging of Hela cells microinjected with 100 μM of W37-M3HFaa (A) and A30-M3HFaa (B) peptides after 10 min of incubation. The color of each pixel represents the lifetime value, while the pixel intensity corresponds to the total fluorescence intensity. Two-photon excitation wavelength was at 760 nm. Emission filter was 500−540 nm. Scale bar corresponds to 10 μm.

concentrations in the cytoplasm and spherical nuclear compartments likely corresponding to nucleoli than in the nucleoplasm. This distribution is fully in line with that observed with overexpressed eGFP-labeled NC (Anton et al., submitted), showing that NC(11−55) can diffuse all over the cells and can easily cross the nuclear membrane, and that the distribution of the peptide is independent of the label. Though NC(11−55) is distributed all over the cells, different colors could be observed, suggesting that NC may interact with different partners depending on its location. For both peptides, similar colors and thus lifetimes were observed in the nucleoli and cytoplasm. Indeed, lifetimes of ∼4.0 ns and ∼2.4 ns were respectively observed for W37-M3HFaa and A30-M3HFaa in these two compartments. These lifetimes are consistent with a binding of both NC peptides to nucleic acids, in line with previous data demonstrating that NC binds mainly to rRNA in ribosomes (Anton et al., submitted). Interestingly, since no lifetimes lower than 2.5 and 1.7 ns were observed for W37-M3HFaa and A30-M3HFaa, respectively, only marginal amounts of free peptides may be present in the cells. A similar conclusion was reached with eGFP-NC, since diffusion measurements clearly indicated that the protein diffuses only in the form of large complexes in cells (Anton et al., submitted). Finally, the absence of lifetimes >2.8 ns in the case of the A30-M3HFaa peptide further suggests that only limited amounts of NC peptides may be bound to lipid membranes. Taken together, the FLIM data obtained with both M3HFaa-labeled peptides confirm that NC binds mainly to RNA in the nucleoli and cytoplasm.

CONCLUSION

A new L-amino acid bearing an ESIPT fluorophore based on 3-hydroxyflavone was synthesized. This new amino acid showed much stronger sensitivity of its dual emission to hydration compared to a previous analogue. Using a peptide corresponding to the zinc finger domain of the HIV-1 nucleocapsid protein, the fluorescent amino acid was found to constitute a nearly perfect Trp substitute. The interaction of labeled peptides with nucleic acids and lipid vesicles produced a strong switch in their dual emission, so that the emission of the ESIPT product was favored. These spectral changes were associated with the appearance of long-lived fluorescence lifetimes, indicating that, within the complexes, the probe environment was significantly more rigid. The highest rigidity of the surrounding environment was observed in complexes with nucleic acids, where the probe was probably stacked with the nucleobases. Thus, both the decrease in water exposure and the increase of rigidity are characteristic changes in the environment of aromatic amino acid residues, when peptides interact with nucleic acids and lipid membranes. Based on the differences in lifetimes for the complexes of the labeled peptides with their binding partners, two-photon fluorescence lifetime imaging of the labeled peptides microinjected into cells confirmed that the nucleocapsid peptides bind mainly to RNAs in the nucleoli and cytoplasm. Taken together, our data validate this new amino acid based on an ESIPT dye as an efficient tool for detection and characterization of peptide interactions in solutions and in living cells. This new tool may be also highly useful in establishing sensitive screening assays to identify compounds able to inhibit the binding of labeled peptides to their ligands.

ASSOCIATED CONTENT

Supporting Information

The description of the syntheses, activity tests of the labeled peptides, spectroscopic and time-resolved characterization of the fluorescent amino acid in solvents, and the sequences of the ODNs used in this study are available in the Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Telephone: 3338368854263. E-mail: yves.mely@unistra.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.S. was supported by an AMI-GDR 2588 scholarship and a fellowship from the Ministère de la Recherche. O.M.Z. was supported by the ARCUS Program. V.G.P. was supported by Université de Strasbourg. This work was supported by the European Project THINPAD "Targeting the HIV-1 Nucleocapsid Protein to fight Antiretroviral Drug Resistance" (FP7−Grant Agreement 601969) and the French-Ukrainian Dnipro program.

REFERENCES

(1) Demchenko, A. P.; Mely, Y.; Duportail, G.; Klymchenko, A. S. Monitoring Biophysical Properties of Lipid Membranes by Environment-Sensitive Fluorescent Probes. *Biophys. J.* **2009**, *96*, 3461−3470.

(2) Klymchenko, A. S.; Mely, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Prog. Mol. Biol. Transl. Sci.* **2013**, *113*, 35−58.

(3) Yang, Z.; Cao, J.; He, Y.; Yang, J. H.; Kim, T.; Peng, X.; Kim, J. S. Macro-/Micro-Environment-Sensitive Chemosensing and Biological Imaging. *Chem. Soc. Rev.* **2014**, *43*, 4563−4601.

(4) Loving, G. S.; Sainlos, M.; Imperiali, B. Monitoring Protein Interactions and Dynamics with Solvatochromic Fluorophores. *Trends Biotechnol.* **2010**, *28*, 73−83.

(5) Krueger, A. T.; Imperiali, B. Fluorescent Amino Acids: Modular Building Blocks for the Assembly of New Tools for Chemical Biology. *ChemBioChem* **2013**, *14*, 788−799.

(6) Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. Probing Protein Electrostatics with a Synthetic Fluorescent Amino Acid. *Science* **2002**, *296*, 1700−1703.

(7) Nitz, M.; Mezo, A. R.; Ali, M. H.; Imperiali, B. Enantioselective Synthesis and Application of the Highly Fluorescent and Environment-Sensitive Amino Acid 6-(2-Dimethylaminonaphthoyl) Alanine (DANA). *Chem. Commun.* **2002**, *17*, 1912−1913.
(8) Vazquez, M. E.; Rothman, D. M.; Imperiali, B. A New Environment-Sensitive Fluorescent Amino Acid for Fmoc-Based Solid Phase Peptide Synthesis. *Org. Biomol. Chem.* **2004**, *2*, 1965−1966.
(9) Vazquez, M. E.; Blanco, J. B.; Imperiali, B. Photophysics and Biological Applications of the Environment-Sensitive Fluorophore 6-*N*,*N*-Dimethylamino-2,3-naphthalimide. *J. Am. Chem. Soc.* **2005**, *127*, 1300−1306.
(10) Loving, G.; Imperiali, B. A Versatile Amino Acid Analogue of the Solvatochromic Fluorophore 4-*N*,*N*-Dimethylamino-1,8-Naphthalimide: A Powerful Tool for the Study of Dynamic Protein Interactions. *J. Am. Chem. Soc.* **2008**, *130*, 13630−13638.
(11) Strizhak, A. V.; Postupalenko, V. Y.; Shvadchak, V. V.; Morellet, N.; Guittet, E.; Pivovarenko, V. G.; Klymchenko, A. S.; Mély, Y. Two-Color Fluorescent L-Amino Acid Mimic of Tryptophan for Probing Peptide-Nucleic Acid Complexes. *Bioconjugate Chem.* **2012**, *23*, 2434−2443.
(12) Postupalenko, V. Y.; Zamotaiev, O. M.; Shvadchak, V. V.; Strizhak, A. V.; Pivovarenko, V. G.; Klymchenko, A. S.; Mely, Y. Dual-Fluorescence L-Amino Acid Reports Insertion and Orientation of Melittin Peptide in Cell Membranes. *Bioconjugate Chem.* **2013**, *24*, 1998−2007.
(13) Weber, G.; Farris, F. J. Synthesis and Spectral Properties of a Hydrophobic Fluorescent Probe: 6-Propionyl-2-(dimethylamino)-naphthalene. *Biochemistry* **1979**, *18*, 3075−3078.
(14) Parasassi, T.; Krasnowska, E. K.; Bagatolli, L.; Gratton, E. Laurdan and Prodan as Polarity-Sensitive Fluorescent Membrane Probes. *J. Fluoresc.* **1998**, *8*, 365−373.
(15) Chen, H.; Chung, N. N.; Lemieux, C.; Zelent, B.; Vanderkooi, J. M.; Gryczynski, I.; Wilkes, B. C.; Schiller, P. W. [Aladan^3]TIPP: A Fluorescent δ-Opioid Antagonist with High δ-Receptor Binding Affinity and δ Selectivity. *Biopolymers* **2005**, *80*, 325−331.
(16) Venkatraman, P.; Nguyen, T. T.; Sainlos, M.; Bilsel, O.; Chitta, S.; Imperiali, B.; Stern, L. J. Fluorogenic Probes for Monitoring Peptide Binding to Class Ii Mhc Proteins in Living Cells. *Nat. Chem. Biol.* **2007**, *3*, 222−228.
(17) Sengupta, P. K.; Kasha, M. Excited State Proton-Transfer Spectroscopy of 3-Hytdroxyflavone and Quercetin. *Chem. Phys. Lett.* **1979**, *68*, 382−385.
(18) Demchenko, A. P.; Tang, K. C.; Chou, P. T. Excited-State Proton Coupled Charge Transfer Modulated by Molecular Structure and Media Polarization. *Chem. Soc. Rev.* **2013**, *42*, 1379−1408.
(19) Das, R.; Klymchenko, A. S.; Duportail, G.; Mély, Y. Unusually Slow Proton Transfer Dynamics of a 3-Hydroxychromone Dye in Protic Solvents. *Photochem. Photobiol. Sci.* **2009**, *8*, 1583−1589.
(20) Shynkar, V. V.; Klymchenko, A. S.; Piemont, E.; Demchenko, A. P.; Mely, Y. Dynamics of Intermolecular Hydrogen Bonds in the Excited States of 4′-Dialkylamino-3-hydroxyflavones. On the Pathway to an Ideal Fluorescent Hydrogen Bonding Sensor. *J. Phys. Chem. A* **2004**, *108*, 8151−8159.
(21) Enander, K.; Choulier, L.; Olsson, A. L.; Yushchenko, D. A.; Kanmert, D.; Klymchenko, A. S.; Demchenko, A. P.; Mély, Y.; Altschuh, D. A Peptide-Based, Ratiometric Biosensor Construct for Direct Fluorescence Detection of a Protein Analyte. *Bioconjugate Chem.* **2008**, *19*, 1864−1870.
(22) Chou, P.-T.; Martinez, M. L.; Clements, H. Reversal of Excitation Behavior of Proton-Transfer vs Charge-Transfer by Dielectric Perturbation of Electronic Manifolds. *J. Phys. Chem.* **1993**, *97*, 2618−2622.
(23) Ormson, S. M.; Brown, R. G.; Vollmer, F.; Rettig, W. Switching between Charge- and Proton-Transfer Emission in the Excited State of a Substituted 3-Hydroxyflavone. *J. Photochem. Photobiol., A* **1994**, *81*, 65−72.
(24) Swinney, T. C.; Kelley, D. F. Proton Transfer Dynamics in Substituted 3-Hydroxyflavones: Solvent Polarization Effects. *J. Chem. Phys.* **1993**, *99*, 211−221.
(25) Klymchenko, A. S.; Demchenko, A. P. Multiparametric Probing of Intermolecular Interactions with Fluorescent Dye Exhibiting Excited State Intramolecular Proton Transfer. *Phys. Chem. Chem. Phys.* **2003**, *5*, 461−468.
(26) Strandjord, A. J. G.; Barbara, P. F. Proton-Transfer Kinetics of 3-Hydroxyflavone: Solvent Effects. *J. Phys. Chem.* **1985**, *89*, 2355−2361.
(27) McMorrow, D.; Kasha, M. Intramolecular Excited-State Proton Transfer in 3-Hydroxyflavone. Hydrogen-Bonding Solvent Perturbations. *J. Phys. Chem.* **1984**, *88*, 2235−2243.
(28) Klymchenko, A. S.; Shvadchak, V. V.; Yushchenko, D. A.; Jain, N.; Mély, Y. Excited-State Intramolecular Proton Transfer Distinguishes Microenvironments in Single- and Double-Stranded DNA. *J. Phys. Chem. B* **2008**, *112*, 12050−12055.
(29) Shvadchak, V. V.; Klymchenko, A. S.; De Rocquigny, H.; Mély, Y. Sensing Peptide - Oligonucleotide Interactions by a Two-Color Fluorescence Label: Application to the HIV-1 Nucleocapsid Protein. *Nucleic Acids Res.* **2009**, *37*, e25.
(30) Zamotaiev, O. M.; Postupalenko, V. Y.; Shvadchak, V. V.; Pivovarenko, V. G.; Klymchenko, A. S.; Mély, Y. Improved Hydration-Sensitive Dual-Fluorescence Labels for Monitoring Peptide-Nucleic Acid Interactions. *Bioconjugate Chem.* **2011**, *22*, 101−107.
(31) Choulier, L.; Shvadchak, V. V.; Naidoo, A.; Klymchenko, A. S.; Mély, Y.; Altschuh, D. A Peptide-Based Fluorescent Ratiometric Sensor for Quantitative Detection of Proteins. *Anal. Biochem.* **2010**, *401*, 188−195.
(32) Postupalenko, V. Y.; Shvadchak, V. V.; Duportail, G.; Pivovarenko, V. G.; Klymchenko, A. S.; Mély, Y. Monitoring Membrane Binding and Insertion of Peptides by Two-Color Fluorescent Label. *Biochim. Biophys. Acta* **2011**, *1808*, 424−432.
(33) Boudier, C.; Klymchenko, A. S.; Mely, Y.; Follenius-Wund, A. Local Environment Perturbations in Alpha$_1$-Antitrypsin Monitored by a Ratiometric Fluorescent Label. *Photochem. Photobiol. Sci.* **2009**, *8*, 814−821.
(34) Yushchenko, D. A.; Fauerbach, J. A.; Thirunavukkuarasu, S.; Jares-Erijman, E. A.; Jovin, T. M. Fluorescent Ratiometric Mfc Probe Sensitive to Early Stages of α-Synuclein Aggregation. *J. Am. Chem. Soc.* **2010**, *132*, 7860−7861.
(35) Darlix, J. L.; Godet, J.; Ivanyi-Nagy, R.; Fossé, P.; Mauffret, O.; Mély, Y. Flexible Nature and Specific Functions of the HIV-1 Nucleocapsid Protein. *J. Mol. Biol.* **2011**, *410*, 565−581.
(36) De Rocquigny, H.; Ficheux, D.; Gabus, C.; Fournie-Zaluski, M. C.; Darlix, J. L.; Roques, B. P. First Large Scale Chemical Synthesis of the 72 Amino Acid HIV-1 Nucleocapsid Protein NCp7 in an Active Form. *Biochem. Biophys. Res. Commun.* **1991**, *180*, 1010−1018.
(37) Clever, J.; Sassetti, C.; Parslow, T. G. RNA Secondary Structure and Binding Sites for Gag Gene Products in the 5′ Packaging Signal of Human Immunodeficiency Virus Type 1. *J. Virol.* **1995**, *69*, 2101−2109.
(38) de Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F. Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 Psi-RNA Recognition Element. *Science* **1998**, *279*, 384−388.
(39) Vuilleumier, C.; Bombarda, E.; Morellet, N.; Gerard, D.; Roques, B. P.; Mely, Y. Nucleic Acid Sequence Discrimination by the HIV-1 Nucleocapsid Protein NCp7: A Fluorescence Study. *Biochemistry* **1999**, *38*, 16816−16825.
(40) Bourbigot, S.; Ramalanjaona, N.; Boudier, C.; Salgado, G. F. J.; Roques, B. P.; Mely, Y.; Bouaziz, S.; Morellet, N. How the HIV-1 Nucleocapsid Protein Binds and Destabilises the (−)Primer Binding Site During Reverse Transcription. *J. Mol. Biol.* **2008**, *383*, 1112−1128.
(41) Shubsda, M. F.; Paoletti, A. C.; Hudson, B. S.; Borer, P. N. Affinities of Packaging Domain Loops in HIV-1 RNA for the Nucleocapsid Protein. *Biochemistry* **2002**, *41*, 5276−5282.
(42) Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F. NMR Structure of the HIV-1 Nucleocapsid Protein Bound to Stem-Loop SL2 of the Psi-

RNA Packaging Signal. Implications for Genome Recognition. *J. Mol. Biol.* **2000**, *301*, 491−511.
(43) Fisher, R. J.; Rein, A.; Fivash, M.; Urbaneja, M. A.; Casas-Finet, J. R.; Medaglia, M.; Henderson, L. E. Sequence-Specific Binding of Human Immunodeficiency Virus Type 1 Nucleocapsid Protein to Short Oligonucleotides. *J. Virol.* **1998**, *72*, 1902−1909.
(44) Beltz, H.; Clauss, C.; Piemont, E.; Ficheux, D.; Gorelick, R. J.; Roques, B.; Gabus, C.; Darlix, J. L.; de Rocquigny, H.; Mely, Y. Structural Determinants of HIV-1 Nucleocapsid Protein for cTAR DNA Binding and Destabilization, and Correlation with Inhibition of Self-Primed DNA Synthesis. *J. Mol. Biol.* **2005**, *348*, 1113−1126.
(45) Mely, Y.; de Rocquigny, H.; Sorinasjimeno, M.; Keith, G.; Roques, B. P.; Marquet, R.; Gerard, D. Binding of the HIV-1 Nucleocapsid Protein to the Primer tRNA(3)(Lys), in-Vitro, Is Essentially Not Specific. *J. Biol. Chem.* **1995**, *270*, 1650−1656.
(46) Godet, J.; Ramalanjaona, N.; Sharma, K. K.; Richert, L.; De Rocquigny, H.; Darlix, J. L.; Duportail, G.; Mély, Y. Specific Implications of the HIV-1 Nucleocapsid Zinc Fingers in the Annealing of the Primer Binding Site Complementary Sequences During the Obligatory Plus Strand Transfer. *Nucleic Acids Res.* **2011**, *39*, 6633−6645.
(47) Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mely, Y. HIV-1 Nucleocapsid Protein Activates Transient Melting of Least Stable Parts of the Secondary Structure of TAR and Its Complementary Sequence. *J. Mol. Biol.* **2002**, *317*, 385−399.
(48) Post, K.; Kankia, B.; Gopalakrishnan, S.; Yang, V.; Cramer, E.; Saladores, P.; Gorelick, R. J.; Guo, J. H.; Musier-Forsyth, K.; Levin, J. G. Fidelity of Plus-Strand Priming Requires the Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein. *Nucleic Acids Res.* **2009**, *37*, 1755−1766.
(49) Morellet, N.; Déméné, H.; Teilleux, V.; Huynh-Dinh, T.; De Rocquigny, H.; Fournié-Zaluski, M. C.; Roques, B. P. Structure of the Complex between the HIV-1 Nucleocapsid Protein NCp7 and the Single-Stranded Pentanucleotide d(ACGCC). *J. Mol. Biol.* **1998**, *283*, 419−434.
(50) Bazzi, A.; Zargarian, L.; Chaminade, F.; Boudier, C.; De Rocquigny, H.; René, B.; Mély, Y.; Fossé, P.; Mauffret, O. Structural Insights into the cTAR DNA Recognition by the HIV-1 Nucleocapsid Protein: Role of Sugar Deoxyriboses in the Binding Polarity of NC. *Nucleic Acids Res.* **2011**, *39*, 3903−3916.
(51) Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mely, Y. Impact of the Terminal Bulges of HIV-1 cTAR DNA on Its Stability and the Destabilizing Activity of the Nucleocapsid Protein NCp7. *J. Mol. Biol.* **2003**, *328*, 95−108.
(52) Godet, J.; De Rocquigny, H.; Raja, C.; Glasser, N.; Ficheux, D.; Darlix, J. L.; Mely, Y. During the Early Phase of HIV-1 DNA Synthesis, Nucleocapsid Protein Directs Hybridization of the TAR Complementary Sequences Via the Ends of Their Double-Stranded Stem. *J. Mol. Biol.* **2006**, *356*, 1180−1192.
(53) Melhuish, W. H. Quantum Efficiencies of Fluorescence of Organic Substances: Effect of Solvent and Concentration of the Fluorescent Solute. *J. Phys. Chem.* **1961**, *65*, 229−235.
(54) Doroshenko, A. O.; Sychevskaya, L. B.; Grygorovych, A. V.; Pivovarenko, V. G. Fluorescence Probing of Cell Membranes with Azacrown Substituted Ketocyanine Dyes. *J. Fluoresc.* **2002**, *12*, 455−464.
(55) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Kluwer Academic, New York, 2006.
(56) Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mély, Y. Destabilization of the HIV-1 Complementary Sequence of TAR by the Nucleocapsid Protein through Activation of Conformational Fluctuations. *J. Mol. Biol.* **2003**, *326*, 691−700.
(57) Hufford, C. D.; Oguntimein, B. O.; Shoolery, J. N. Angoluvarin, an Antimicrobial Dihydrochalcone from Uvaria angolensis. *J. Org. Chem.* **1987**, *52*, 5286−5288.
(58) Chen, C.; Zhu, Y. F.; Wilcoxen, K. An Improved Synthesis of Selectively Protected L-Dopa Derivatives from L-Tyrosine. *J. Org. Chem.* **2000**, *65*, 2574−2576.
(59) Pivovarenko, V. G.; Zamotaiev, O. M.; Shvadchak, V. V.; Postupalenko, V. Y.; Klymchenko, A. S.; Mély, Y. Quantification of Local Hydration at the Surface of Biomolecules Using Dual-Fluorescence Labels. *J. Phys. Chem. A* **2012**, *116*, 3103−3109.
(60) Królicki, R.; Jarzęba, W.; Mostafavi, M.; Lampre, I. Preferential Solvation of Coumarin 153 - the Role of Hydrogen Bonding. *J. Phys. Chem. A* **2002**, *106*, 1708−1713.
(61) Silva, M. A. D. R.; Da Silva, D. C.; Machado, V. G.; Longhinotti, E.; Frescura, V. L. A. Preferential Solvation of a Hydrophobic Probe in Binary Mixtures Comprised of a Nonprotic and a Hydroxylic Solvent: A View of Solute−Solvent and Solvent−Solvent Interactions. *J. Phys. Chem. A* **2002**, *106*, 8820−8826.
(62) Avilov, S. V.; Piemont, E.; Shvadchak, V.; de Rocquigny, H.; Mely, Y. Probing Dynamics of HIV-1 Nucleocapsid Protein/Target Hexanucleotide Complexes by 2-Aminopurine. *Nucleic Acids Res.* **2008**, *36*, 885−896.
(63) Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Gottlinger, H. G. Mapping of Functionally Important Residues of a Cysteine-Histidine Box in the Human Immunodeficiency Virus Type 1 Nucleocapsid Protein. *J. Virol.* **1993**, *67*, 6159−6169.
(64) Dziuba, D.; Postupalenko, V. Y.; Spadafora, M.; Klymchenko, A. S.; Guérineau, V.; Mély, Y.; Benhida, R.; Burger, A. A Universal Nucleoside with Strong Two-Band Switchable Fluorescence and Sensitivity to Environment for Investigating DNA Interactions. *J. Am. Chem. Soc.* **2012**, *134*, 10209−10213.
(65) Kenfack, C. A.; Klymchenko, A. S.; Duportail, G.; Burger, A.; Mély, Y. Ab Initio Study of the Solvent H-Bonding Effect on ESIPT Reaction and Electronic Transitions of 3-Hydroxychromone Derivatives. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8910−8918.
(66) Dziuba, D.; Karpenko, I. A.; Barthes, N. P. F.; Michel, B. Y.; Klymchenko, A. S.; Benhida, R.; Demchenko, A. P.; Mély, Y.; Burger, A. Rational Design of a Solvatochromic Fluorescent Uracil Analogue with a Dual-Band Ratiometric Response Based on 3-Hydroxychromone. *Chem.—Eur. J.* **2014**, *20*, 1998−2009.
(67) Lee, L. G.; Chen, C. H.; Chiu, L. A. Thiazole Orange: A New Dye for Reticulocyte Analysis. *Cytometry* **1986**, *7*, 508−517.
(68) Köhler, O.; Seitz, O. Thiazole Orange as Fluorescent Universal Base in Peptide Nucleic Acids. *Chem. Commun.* **2003**, *9*, 2938−2939.
(69) Wenge, U.; Wagenknecht, H. A. Synthetic GFP Chromophore and Control of Excited-State Proton Transfer in DNA: An Alternative Concept for Fluorescent DNA Labels with Large Apparent Stokes' Shifts. *Synthesis* **2011**, *3*, 502−508.
(70) Spielmann, H. P.; Wemmer, D. E.; Jacobsen, J. P. Solution Structure of a DNA Complex with the Fluorescent Bis-Intercalator TOTO Determined by NMR Spectroscopy. *Biochemistry* **1995**, *34*, 8542−8553.
(71) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T. Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J. Am. Chem. Soc.* **2000**, *122*, 2213−2222.
(72) Kelley, S. O.; Barton, J. K. Electron Transfer between Bases in Double Helical DNA. *Science* **1999**, *283*, 375−381.

Supporting Information

Fluorescent Amino Acid Undergoing Excited State Intramolecular Proton Transfer for Site-specific Probing and Imaging of Peptide Interactions

*Marianna Sholokh[§†], Oleksandr M. Zamotaiev[†], Ranjan Das[#], Viktoriia Y. Postupalenko[§], Ludovic Richert[§], Denis Dujardin[§], Olga A. Zaporozhets[†], Vasyl G. Pivovarenko[†], Andrey S. Klymchenko[§], Yves Mély[§]**

[§]Laboratoire de Biophotonique et Pharmacologie, UMR 7213 CNRS, Université de Strasbourg, Faculté de Pharmacie, 74, Route du Rhin, 67401 ILLKIRCH Cedex, France.

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

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

Synthesis of Fmoc-M3HFaa. The synthesis of the compounds 1 and 2 (Scheme S1) was performed as described previously.[1-2]

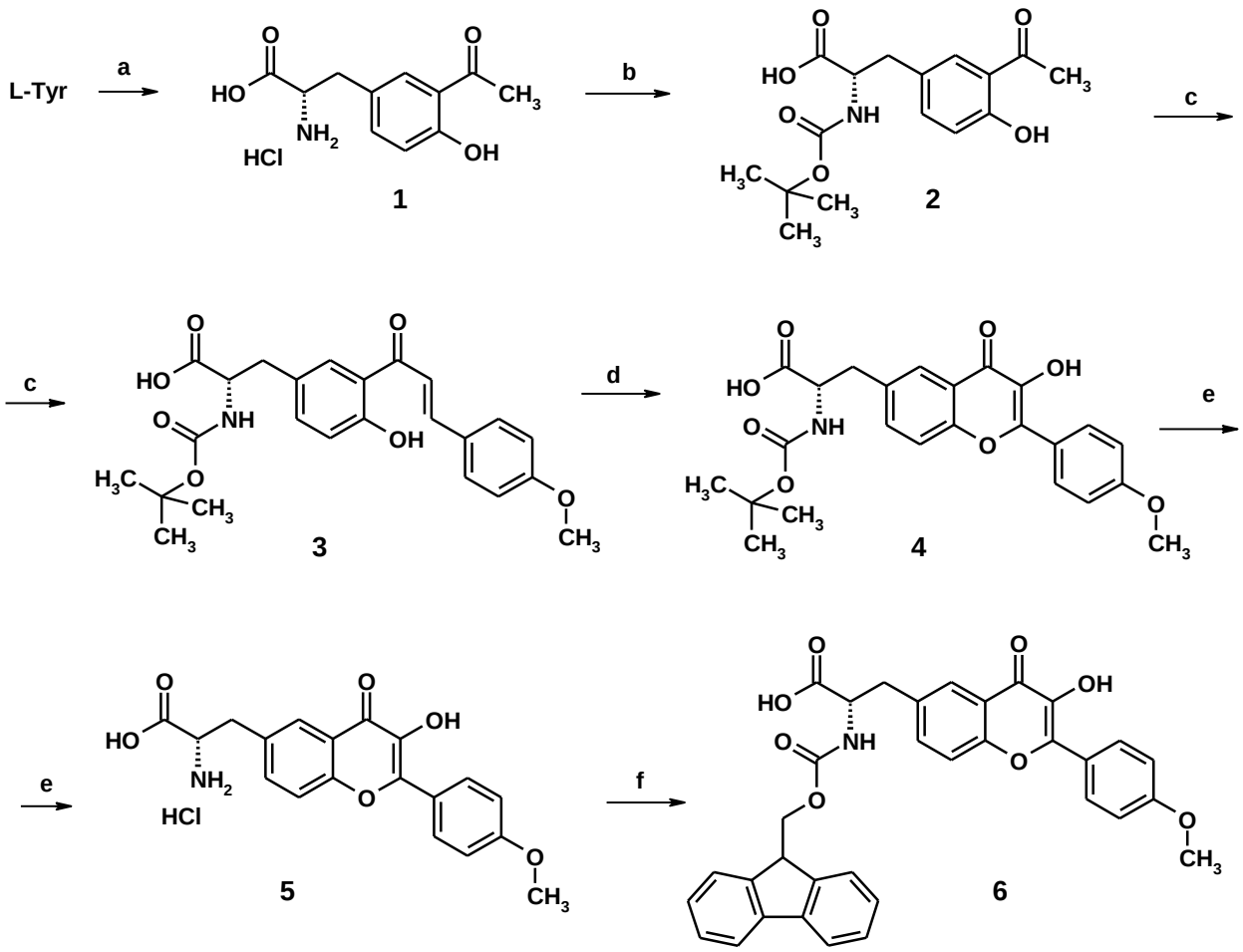
N-(tert-butoxycarbonyl)-3-[(2E)-3-(2-methoxyphenyl)prop-2-enoyl]-L-tyrosine (3). 5.00 g (15.48 mmol) of 3-acetyl-N-Boc-L-tyrosine (2) were dissolved in 37 mL of ethanol upon stirring under Ar-atmosphere. Then, degassed solution of sodium hydroxide (9.25 g in 28 mL of water) was slowly added to the reaction mixture in cold water bath. After that, 3.25 mL of 4-methoxybenzaldehyde (26.81 mmol, 1.73 eq.) were added and the mixture was stirred for 16 h at room temperature under Ar. The reaction was monitored by TLC (CH_2Cl_2: MeOH, 7: 3 v/v). For that, small aliquots were carefully acidified with HCl and extracted with ethyl acetate. Then, the mixture was carefully acidified to pH 5 with 10% HCl, diluted to 180 mL with water and extracted 3 times with ethyl acetate. The obtained solution was washed with water, dried with Na_2SO_4 and the solvent was evaporated *in vacuo*. The obtained oil was crystallized upon standing overnight under petroleum ether. Compound was recrystallized from toluene/petroleum ether mixture. Yield 5.94 g (87%) as an yellow powder. 1H-NMR (300MHz, CDCl3): δ 1.41 (s, 9H, Boc), 3.05 - 3.25 (m, 2H, CH2), 3.87 (s, 3H, OCH3), 4.65 (m, 1H, CH), 6.93 (m, 1H, ArH), 6.95 (d, 2H, J=9Hz, ArH), 7.28 (m, 1H, ArH), 7.52 (d, 1H, J=15Hz, CH), 7.65 (d, 2H, J=9Hz, ArH), 7.73 (s, 1H, ArH), 7.91 (d, 1H, J=15Hz, CH), 12.87 (br. s, 1H, COOH). Mass spectroscopy, ion polarity – negative, m/z: 440.0 $[M–H]^–$.

N-(tert-butoxycarbonyl)-3-[2-(4-methoxyphenyl)-3-methyl-4-oxo-4H-chromen-6-yl]-L-alanine (Boc-M3HFaa, 4). 5.00 g (11.32 mmol) of chalcone (3) was dissolved in 50 mL of ethanol, cooled in ice bath and 47mL of 1 M solution of sodium hydroxide was added under stirring. After that, 2.85 mL of 30% hydrogen peroxide was added and the mixture was stirred in ice bath for approximately 6 h under Ar-atmosphere. The reaction was monitored by TLC (CH_2Cl_2:MeOH (8:2 v/v). After completing the reaction, the mixture was diluted to 300 mL with water, acidified with HCl to pH 5 and the obtained yellow precipitate was filtered and washed with water. It was purified by crystallization from methanol/water

(4:1) mixture. Yield 3.25 g (63%) of yellow crystals. 1H-NMR (300MHz, CDCl3): δ 1.48 (s, 9H, Boc), 3.39 (m, 2H, CH2), 3.88 (s, 3H, OCH3), 4.79 (m, 1H, CH), 7.01 (d, 2H, J=9Hz, ArH), 7.51 (m, 2H, ArH), 8.15 (s, 1H, ArH), 8.20 (d, 2H, J=9Hz, ArH). Mass spectroscopy, ion polarity – negative, m/z: 454.0 $[M–H]^-$; ion polarity – positive, m/z: 456.0 $[M+H]^+$.

3-[2-(4-methoxyphenyl)-3-methyl-4-oxo-4H-chromen-6-yl]-L-alanine hydrochloride (M3HFaa hydrochloride, 5). 4 g (8.78 mmol) of chromone (4) was dissolved in 51 mL of dioxane and then, 51 mL of conc. HCl was added dropwise to the reaction mixture cooled in an ice water bath. After 10 min, the bath was removed and the mixture was stirred for 2-3 h under Ar-atmosphere. Then, the obtained yellowish-white fuzzy precipitate was filtered, washed with water twice and dried *in vacuo*. Yield 1.98 g (57.6 %). 1H-NMR (300MHz, DMSO-d6): δ 3.21 (m, 2H, CH2), 3.84 (s, 3H, OCH3), 3.93 (m, 1H, CH), 7.10 (d, 2H, J=9Hz, ArH), 7.67 (m, 2H, ArH), 7.98 (s, 1H, ArH), 8.17 (d, 2H, J=9Hz, ArH), 9.11 (br. s, 2N NH2). Mass spectroscopy, ion polarity – negative, m/z: 354.0 $[M–H]^-$ (base form); ion polarity – positive, m/z: 356.0 $[M+H]^+$ (base form).

N-[(9H-fluoren-9-ylmethoxy)carbonyl]-3-[2-(4-methoxyphenyl)-3-methyl-4-oxo-4H-chromen-6-yl]-L-alanine (Fmoc-M3HFaa, 6). 1.1 g (2.81 mmol) of chlorohydrate of amino acid (5) was dissolved in a solution of 950 mg (11.3 mmol) of sodium bicarbonate in 38 mL of water and then 55 mL of DMF was added. After that, 950 mg (1 eq.) of FmocOSu was added and the mixture was stirred for 24 h under Ar-atmosphere. The mixture was then diluted with 38 mL of water and the precipitate of sodium salt of the crude product was filtered. It was purified by dissolution of the sodium salt in hot water (about 100 mL per 1 g) and further acidified with HCl. Yield 1.33 g (82%), yellow powder. 1H-NMR (300MHz, DMSO-d6): δ 3.22 (m, 2H, CH2), 3.85 (s, 3H, OCH3), 4.17 (m, 2H, CH2), 4.28 (m, 1H, CH), 7.12 (d, 2H, J=9Hz, ArH), 7.25 (m, 2H, J=9Hz, ArH), 7.36 (m, 3H, ArH), 7.58 (m, 2H, ArH), 7.68 (m, 2H, ArH), 7.79 (s, 1H, ArH), 7.84 (d, 2H, J=9Hz, ArH), 8.03 (br. s, 1H, OH), 8.19 (d, 2H, J=9Hz, ArH), 9.41 (br. s, 1H, NH), 12.80 (br. s, 1H, COOH). Mass spectroscopy, ion polarity – positive, m/z: 578.0 $[M+H]^+$.



a) AcCl, $AlCl_3$, $PhNO_2$; b) Boc_2O; c) 4-methoxybenzaldehyde, NaOH, $EtOH/H_2O$; d) H_2O_2, NaOH, EtOH ; e) HCl, dioxane; f) FmocOSu, $NaHCO_3$, DMF/H_2O.

Scheme S1. Synthesis of Fmoc-M3HFaa aminoacid for solid-phase peptide synthesis.

Peptide synthesis. The labeled NC(11-55) peptides were synthesized by solid phase peptide synthesis on a 433A synthesizer (ABI, Foster City, CA), as previously described.[3] The synthesis was performed at a 0.1 mmol scale using the standard fluorenylmethoxycarbonyl (Fmoc)-amino acid-coupling protocol. LL-HMP resin (ABI, 0.44 mmol/g reactive group concentrations) was used as a solid support. Deprotection steps were performed by piperidine and automatically controlled by UV absorbance. At selected positions of the peptide, the fluorescent amino acid analogue was incorporated by the following procedure. In a

flask, 2 mol equivalents (0.11 mmol) of the Fmoc-M3HFaa were dissolved in 1 mL of 1-methyl-2-pyrrolidone (NMP) and mixed with two eq. of HBTU/HOBt coupling solution (in DMF). This mixture was immediately added to the Fmoc-deprotected peptidylresin (0.055 mmol) swelled in 1 mL of NMP. After a few minutes of shaking, four eq. of DIEA solution were added. Then, the reaction mixture was stirred for 2 hours at 37 °C. After this, the resin was washed by NMP several times. The labeling procedure was repeated twice (the second time, the reaction mixture was stirred ~ 20 hours at 37 °C). After this, the resin was washed with NMP. In order to covalently block unreacted amino acids, a capping procedure was performed. This procedure was done in an acetic anhydride mixture (0.5 M acetic anhydride, 0.125 M DIEA, and 0.015 M HOBt in NMP). The reaction mixture was stirred for 10-20 min at 37 °C. Then, the resin was filtrated, washed by NMP and peptide synthesis was continued on the synthesizer. At the end of the synthesis, the Fmoc-deprotected peptidylresin was isolated and washed with NMP, methanol and dichloromethane. 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 pure peptide were combined and lyophilized.

Peptide activity tests. The ability of the labeled peptides to destabilize the secondary structure of ODNs was tested using cTAR DNA (a DNA hairpin of 55 nucleotides), the complementary sequence of the transactivation response element, involved in minus strand DNA transfer during reverse transcription. To this end, we used a cTAR sequence labeled at its 3' and 5' ends by 5(6)-carboxytetramethylrhodamine (TMR) and 4-(4′- methylaminophenylazo) benzoic acid (Dabcyl), respectively. In the absence of NC, cTAR is mainly in a non-fluorescent closed form where the TMR and Dabcyl labels are close together, giving excitonic coupling.[4] The destabilization was measured through the fluorescence intensity increase of TMR in the labeled cTAR as a function of NC(11-55) concentration.[5-6]

The ability of the labeled peptides to promote the annealing of cTAR with its complementary dTAR DNA sequence was compared to that of the unlabeled NC peptide. The annealing kinetics were performed under pseudo-first-order conditions by using unlabeled dTAR at a concentration which was 30-fold higher than the concentration of cTAR labeled with carboxytetramethylrhodamine (TMR) at its 5′ end and with 5/6-carboxyfluorescein (Fl) at its 3′ end.[7] Excitation and emission wavelengths were 480 and 520 nm, respectively, for monitoring the Fl fluorescence. Peptides were added to each reactant separately at a peptide:ODN ratio of 3:1, and then, the reaction was initiated by mixing the peptide-coated ODNs together.

Time-resolved spectroscopy. Time-resolved fluorescence measurements were performed using the time-correlated single-photon counting technique (TCSPC). Excitation pulses were generated by a pulse-picked frequency-tripled Ti-sapphire laser (Tsunami, Spectra Physics) pumped by a Millenia X laser (Spectra Physics).[5, 8] Excitation wavelength was set at 315 nm, with a repetition rate of 4 MHz. The fluorescence emission was collected at 430 nm and 550 nm through a polarized 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) was recorded using a polished aluminum reflector, and its full-width at half-maximum was ~50 ps. Experimentally measured fluorescence decays were deconvoluted with the instrumental response function and fitted to retrieve the most probable lifetime distribution using the maximum entropy method (Pulse 5 software).[9-10] In all cases, the $\chi 2$ values were close to 1, indicating an optimal fit.

Table S1. Spectroscopic and time-resolved parameters of M3HFaa in various solvents.[a]

	λ_{ABS} (nm)	λ_{N^*} (nm)	λ_{T^*} (nm)	N*/T*	QY (%)	$\tau_1^{N^*}$ (ns)	$\alpha_1^{N^*}$	$\tau_1^{T^*}$ (ns)	$\alpha_1^{T^*}$	$\tau_2^{T^*}$ (ns)	$\alpha_2^{T^*}$
Buffer	355	442	520	4.80	7.8	0.294	1	ND	ND	ND	ND
MeOH	357	434	535	0.95	6.7	0.159	1	0.135	-1	0.41	1
EtOH	360	429	535	0.51	7.2	0.138	1	0.100	-1	0.53	1
DMF	355	421	539	0.11	6.8	0.042	1	0.039	-1	0.81	1
1,4-dioxane	355	416	540	0.04	15.9	ND	ND	0.025	-1	1.63	1

[a]Reference for the QY: quinine sulfate in 0.5M H_2SO_4 (QY=0.546). τ_i are the fluorescence lifetimes and α_i their amplitudes. For lifetime measurements, excitation wavelength was 315 nm, while emission wavelengths were 430 nm and 550 nm, for the N* and T* bands, respectively. ND, non-determined, due to the too low emission of the corresponding band. The reported data are the mean of n ≥ 2 experiments. Standard deviations are ~ 5% for N*/T* and QY, and < 15% for lifetimes and amplitudes, respectively.

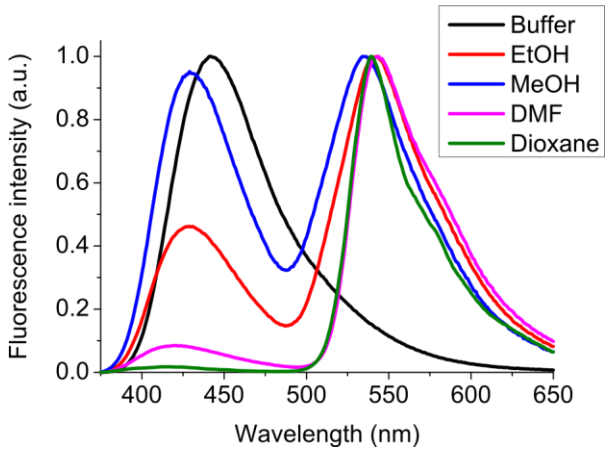


Figure S1. Fluorescence spectra of M3HFaa in different solvents. Excitation wavelength was 360 nm.

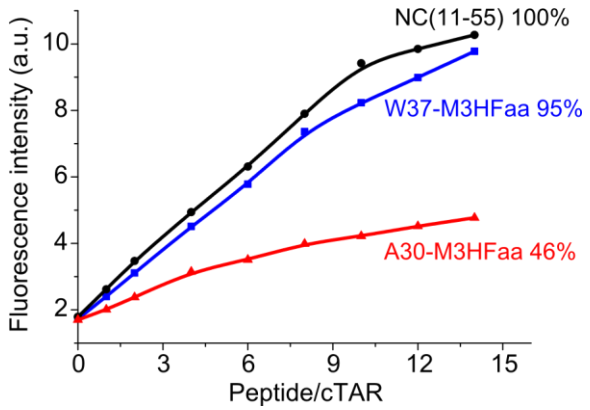


Figure S2. Nucleic acid destabilization activity of the M3HFaa-labeled peptides. Destabilization of doubly-labelled cTAR DNA by NC(11-55) (squares), W37-M3HFaa (disks) and A30-M3HFaa (triangles). Titration of 100 nM Dabcyl-5'-cTAR-3'-TMR by NC peptides was performed in 25 mM TRIS (pH 7.5), 30 mM NaCl and 0.2 mM $MgCl_2$. To monitor TMR fluorescence, excitation wavelength was set at 520 nm.

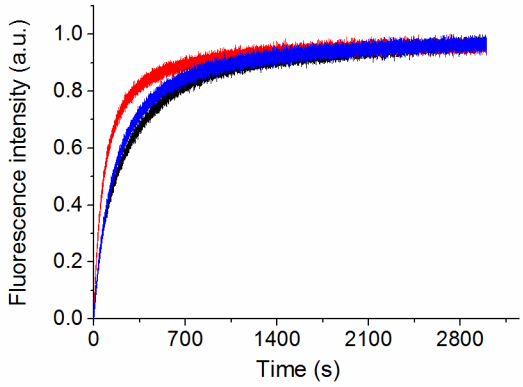


Figure S3. Nucleic acid annealing activity of the M3HFaa-labeled peptides. The annealing kinetics of 10 nM TMR-5′-cTAR-3′-Fl with 300 nM dTAR in the presence of the unlabeled NC(11-55) peptide (black) or W37-M3HFaa (blue) and A30-M3HFaa (red) peptides added at a peptide:ODN ratio of 3 :1 were monitored in 25 mM Tris (pH 7.5), 30 mM NaCl and 0.2 mM $MgCl_2$. Excitation and emission wavelengths were 480 and 520 nm, respectively.

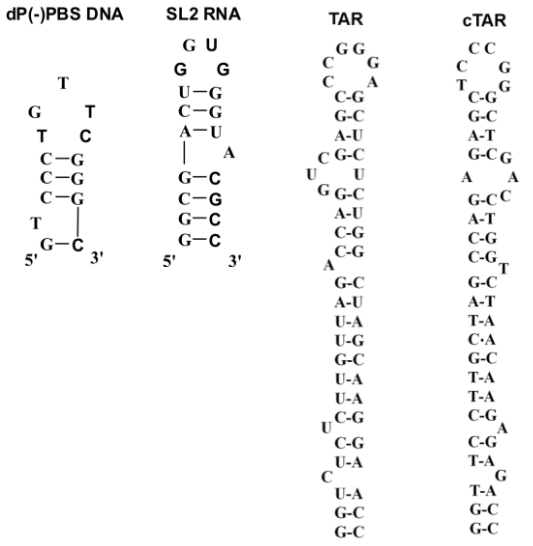


Figure S4. Sequences of the oligonucleotides used in this study.

REFERENCES

1. Hufford, C. D.; Oguntimein, B. O.; Shoolery, J. N., Angoluvarin, an Antimicrobial Dihydrochalcone from Uvaria Angolensis. *J. Org. Chem.* **1987,** *52*, 5286-5288.

2. Chen, C.; Zhu, Y. F.; Wilcoxen, K., An Improved Synthesis of Selectively Protected L-Dopa Derivatives from L-Tyrosine. *J. Org. Chem.* **2000,** *65*, 2574-2576.

3. De Rocquigny, H.; Ficheux, D.; Gabus, C.; Fournie-Zaluski, M. C.; Darlix, J. L.; Roques, B. P., First Large Scale Chemical Synthesis of the 72 Amino Acid Hiv-1 Nucleocapsid Protein Ncp7 in an Active Form. *Biochem. Biophys. Res. Comm.* **1991,** *180*, 1010-1018.

4. Bernacchi, S.; Piémont, E.; Potier, N.; Van Dorsselaer, A.; Mély, Y., Excitonic Heterodimer Formation in an Hiv-1 Oligonucleotide Labeled with a Donor-Acceptor Pair Used for Fluorescence Resonance Energy Transfer. *Biophys. J.* **2003,** *84*, 643-654.

5. Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mely, Y., Hiv-1 Nucleocapsid Protein Activates Transient Melting of Least Stable Parts of the Secondary Structure of Tar and Its Complementary Sequence. *J. Mol. Biol.* **2002,** *317*, 385-399.

6. Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mely, Y., Impact of the Terminal Bulges of Hiv-1 Ctar DNA on Its Stability and the Destabilizing Activity of the Nucleocapsid Protein Ncp7. *J. Mol. Biol.* **2003,** *328*, 95-108.

7. Godet, J.; De Rocquigny, H.; Raja, C.; Glasser, N.; Ficheux, D.; Darlix, J. L.; Mely, Y., During the Early Phase of Hiv-1 DNA Synthesis, Nucleocapsid Protein Directs Hybridization of the Tar Complementary Sequences Via the Ends of Their Double-Stranded Stem. *J. Mol. Biol.* **2006,** *356*, 1180-1192.

8. Godet, J.; Kenfack, C.; Przybilla, F.; Richert, L.; Duportail, G.; Mély, Y., Site-Selective Probing of Ctar Destabilization Highlights the Necessary Plasticity of the Hiv-1 Nucleocapsid Protein to Chaperone the First Strand Transfer. *Nucleic Acids Res.* **2013,** *41*, 5036-5048.

9. Brochon, J. C., Maximum Entropy Method of Data Analysis in Time-Resolved Spectroscopy. *Methods Enzymol.* **1994,** *240*, 262-311.

10. Livesey, A. K.; Brochon, J. C., Analyzing the Distribution of Decay Constants in Pulse-Fluorimetry Using the Maximum Entropy Method. *Biophys. J.* **1987,** *52*, 693-706.

Part 2. Dynamics of (–)/(+)PBS annealing reaction and its promotion by the HIV-1 nucleocapsid protein using fluorescent isomorphic and environment-sensitive nucleoside analogs

3.2.1. Full mechanism of (–)/(+)PBS annealing revealed by site-selective labeling of (–)PBS by thienyl-3-hydroxychromone nucleoside analog

A second approach to monitor site-selectively the spontaneous and NC-promoted (–)/(+)PBS annealing process consists in labeling one of the ODN species. Previously, the annealing mechanism was explored using either PBS sequence labeled at its 5′ and 3′ ends by the rhodamine 6G fluorophore (Rh6G) and the fluorescence quencher Dabcyl, respectively (Ramalanjaona, de Rocquigny et al. 2007), or by using one of the PBS sequences labeled in its loop by the 2-aminopurine (2Ap) nucleoside analog (Godet, Ramalanjaona et al. 2011). However, the PBS species doubly labeled by Rh6G and Dabcyl were found to be insensitive to annealing of the loop regions, while the very low brightness of 2Ap required high PBS concentrations leading to kinetics too fast to be accurately monitored. Consequently, a complete kinetics of the annealing pathways could not be provided. Therefore, to obtain a full mechanistic description of (–)/(+)PBS annealing and notably to describe the molecular events in the loop regions during the annealing reactions, we used two nucleoside analogs, namely the 2-thienyl-3-hydroxychromone-based nucleobase (3HCnt) and the isomorphic deoxythienoguanosine ($d^{th}G$) nucleobase in order to obtain complementary information.

Recently synthesized fluorescent nucleoside analog 3HCnt (Figure 1, publication 2) (in collaboration with the group of Alain Burger) acts as a universal nucleoside analog able to substitute any nucleobase. Its high brightness, ratiometric signal, and environment sensitivity allow the 3HCnt probe to provide a multiparametric response in peptide – ODN and ODN – ODN interactions. To reach our objective, we site-selectively labeled either the native (–)PBS sequence or its ΔP(–)PBS truncated version lacking the 3′ overhang (Figure 1, publication 2) in order to characterize the (–)/(+)PBS annealing kinetics, in the absence and in the presence of NC.

Firstly, to characterize the thermodynamic, NC-binding and spectroscopic properties of the 3HCnt-labeled sequences, we used the ΔP(–)PBS stem-loop sequence, whose 3D structure in both free form and bound to NC(11-55) has been solved by NMR. 3HCnt was introduced at

positions 2, 7, 8, 9 or 10 of ΔP(–)PBS to identify the most appropriate labeling position. We found that the 3HCnt base only marginally affects the stability of ΔP(–)/ΔP(+)PBS duplexes (Figure S1, Table S1, publication 2), consistent with the limited effect of the 3HCnt nucleobase on the stability of 15 base pairs model duplexes and the behavior of 3HCnt as a universal nucleobase. The absorption and fluorescence emission spectra of 3HCnt-labeled ΔP(–)PBS (Table S2, Figure 2, publication 2) indicated that the 3HCnt fluorophore in ΔP(–)PBS is hidden from the solvent through stacking interactions with the neighboring nucleobases. This conclusion was further supported by the low N*/T* intensity ratio values, indicating the low polar environment of the label in all 3HCnt-labeled sequences. Rather low QYs of about 0.4 – 4% were observed for all positions, except 7 (Table S2, publication 2), likely due to the presence of G and/or C neighbors, reported to be efficient quenchers of 3HCnt fluorescence.

Interaction of the labeled ODNs with NC(11-55) peptide was accompanied by an increase of the N*/T* ratio and the QY of the 3HCnt nucleobase at all positions (Figure 3, publication 2), confirming the ability of NC(11-55) to bind all along the (–)PBS loop as well as at the 5′ end of the stem. The increase of the N*/T* ratio shows that upon binding to NC(11-55) the exposure of the 3HCnt nucleobases to the solvent increases, in line with the NMR data, while the increase in the QY indicates a restriction in the rotation between the thienyl and 3HC moieties, so that the two aromatic rings become co-planar, similarly to the 4′-methoxy-3HF amino acid (M3HFaa) probe described above. Such restriction probably results from direct contacts between 3HCnt and the hydrophobic platform of NC(11-55), as was previously evidenced using 2Ap-labeled (–)PBS sequences.

Next, binding competition experiments were performed to determine the relative affinity of NC(11-55) for labeled ΔP(–)PBS versus non-labeled ΔP(–)PBS in order to determine whether introduction of 3HCnt modifies the binding pattern of NC(11-55) to ΔP(–)PBS. It showed that 3HCnt at positions 8 and 9 does not alter the affinity of ΔP(–)PBS to NC(11-55), while at positions 2, 7 and 10 the affinity of the labeled ΔP(–)PBS to NC(11-55) was perturbed (Figure 4, publication 2).

Finally, the emissive properties of 3HCnt were studied in transitions from the stem-loop (single-strand) to the extended duplex. The transitions were accompanied by a decrease in the N*/T* ratio and a substantial increase in the QY of all samples, except at position 10 that was almost insensitive to duplex formation. Decreased N*/T* ratios in the duplexes are in line

with the expected polarity decrease of 3HCnt microenvironment, as a consequence of shielding from water when the probe is intercalated within the double strands. Moreover, this intercalation is likely the reason of the observed increase of the QY, inducing a strong restriction of the rotation between the thienyl and 3HC moieties.

To further rationalize the observed spectroscopic properties of the 3HCnt-labeled ΔP(−)PBS, time-resolved fluorescence measurements were performed. Similarly to M3HFaa, the 3HCnt fluorophore undergoes an irreversible ESIPT reaction, where the fast decay component associated with the N* form (τ_1^{N*} = 0.10±0.01 ns in ethanol) matches the short rise component of the T* form (τ_1^{T*} = 0.08±0.01 ns). Additionally, a longer decay component associated with the T* band τ_2^{T*} (1.2±0.1 ns) is perceived. Due to the irreversibility of the ESIPT reaction, we used only the decay parameters of the T* form of 3HCnt (Table 3.2.1) to discuss the environment of the 3HCnt probe in ΔP(−)PBS. Both in its free form and when incorporated in ΔP(−)PBS, the 3HCnt is characterized by an ESIPT equilibrium time of about 80 ps (τ_1^{T*}) in good agreement with the earlier reported values for (2-furyl)-3HC (Das, Klymchenko et al. 2009) and M3HFaa (Sholokh, Zamotaiev et al. 2015). Free 3HCnt presents a single longer decay component τ_2^{T*}, while in the labeled ODNs we observe 2 or 3 additional lifetime components, indicating a conformational heterogeneity of 3HCnt in the stem and loop positions. In all labeled ODNs, the τ_2^{T*} lifetime values (or τ_3^{T*} if the shorter lifetime is present) are very close to that of the free nucleoside (1.2 ns), and are likely associated with a conformation where 3HCnt is largely exposed to solvent. Its amplitude increases in complexes with NC(11-55) in line with a binding of the protein all along ΔP(−)PBS and its ability to expose the bases to the solvent. In contrast, this amplitude decreases in duplexes, likely due to changes in the orientation of the probe in the more structured environment of double strands. The presence of the short-lived τ_2^{T*} lifetimes at positions 2, 8 and 9 could likely be related to conformations where the 3HCnt probe is strongly quenched by the neighbor nucleobases. Markedly longer τ_3^{T*}/τ_4^{T*} lifetimes of 3HCnt appear only when the probe is incorporated into the nucleic acids. These lifetimes may originate from stacking interactions of 3HCnt with other nucleobases that restrict the relative motions of the 3HC aromatic moieties, favoring the more planar and, thus, more emissive conformation of the probe. Furthermore, non-emissive "dark" species (α_0) with lifetimes shorter than the detection limit of our setup (<30 ps) could also be inferred. These species result from either static or very fast dynamic quenching and represent from 24% in position 7 to 95% in position 2. The

simplest explanation for this position-dependent quenching is that 3HCnt at position 7 has two T residues as flanking neighbors, which are poor quenchers. In contrast, 3HCnt at position 2 is located next to a G residue, which is a highly efficient quencher. Moreover, due to its position close to the 5'end, 3HCnt and its neighbors are highly dynamic, so that efficient collisional quenching is expected, explaining the low QY and high amount of dark species.

Addition of NC(11-55) induces an exposure of the 3HCnt base to the solvent for all labeling positions accompanied by a consequent decrease of the percentage of dark species. It also increases the values of the long-lived lifetimes and their amplitudes, confirming the increased rigidity of the 3HCnt molecule due to contacts with the hydrophobic platform of NC(11-55). Moreover, formation of the duplexes further increases the long-lived lifetimes as well as the amplitudes of the 3HCnt probe, likely as a result of its induced planarity when intercalated within the neighbor base pairs of the duplex that. The dark species amount varies marginally upon duplex formation for all labeling positions.

Table 3.2.1. Time-resolved fluorescence parameters of the 3HCnt nucleobase and 3HCnt-labeled ΔP(−)PBS either in its free form and in complexes with NC(11-55) or ΔP(+)PBS[a]

	QY	τ_1^{T*}	α_1^{T*}	τ_2^{T*}	α_2^{T*}	τ_3^{T*}	α_3^{T*}	τ_4^{T*}	α_4^{T*}	$\langle\tau\rangle$	α_0
3HCnt in EtOH	0.062[b]	**0.08**	-1.00	**1.2**	1	-	-	-	-	**1.2**	-
3HCnt-2ΔP(−)PBS	0.004	**0.020**	-1.00	**0.29**	0.02	**1.03**	0.02	**4.22**	0.01	**1.52**	0.95
+NC(11-55)	0.017	**0.026**	-1.00	**0.35**	0.05	**1.66**	0.05	**5.30**	0.04	**2.32**	0.86
+ΔP(+)PBS	0.005	**0.023**	-1.00	**0.17**	0.10	**0.65**	0.03	**3.63**	0.02	**0.64**	0.85
3HCnt-7ΔP(−)PBS	0.12	**0.041**	-1.00	**1.12**	0.29	**4.28**	0.47	-	-	**3.07**	0.24
+NC(11-55)	0.16	**0.044**	-1.00	**1.66**	0.33	**5.38**	0.48	-	-	**3.85**	0.19
+ΔP(+)PBS	0.31	**0.064**	-1.00	**1.34**	0.09	**7.46**	0.79	-	-	**6.85**	0.12
3HCnt-8ΔP(−)PBS	0.019	**0.029**	-1.00	**0.19**	0.05	**0.98**	0.09	**4.08**	0.08	**1.66**	0.78
+NC(11-55)	0.099	**0.031**	-1.00	**1.09**	0.31	**4.58**	0.34	-	-	**2.94**	0.35
+ΔP(+)PBS	0.03	**0.018**	-1.00	**0.17**	0.31	**0.87**	0.08	**5.21**	0.09	**1.20**	0.52
3HCnt-9ΔP(−)PBS	0.026	**0.023**	-1.00	**0.27**	0.08	**1.24**	0.08	**4.69**	0.08	**2.12**	0.76
+NC(11-55)	0.061	**0.020**	-1.00	**0.98**	0.24	**4.33**	0.22	-	-	**2.59**	0.54

+ΔP(+)PBS	0.13	**0.023**	-1.00	**0.94**	0.17	**8.09**	0.30	**-**	**-**	**5.41**	0.53
3HCnt-10ΔP(–)PBS	0.034	**0.037**	-1.00	**1.1**	0.19	**5.0**	0.09	**-**	**-**	**2.35**	0.72
+NC(11-55)	0.037	**0.028**	-1.00	**1.09**	0.16	**5.4**	0.10	**-**	**-**	**2.73**	0.74
+ΔP(+)PBS	0.027	**0.019**	-1.00	**0.62**	0.14	**3.43**	0.12	**-**	**-**	**1.91**	0.73

[a]QY is the fluorescence quantum yield, taken from Table S2, publication 2, α_0 are the amplitudes of the dark species, τ_i are the fluorescence lifetimes (ns), α_i are their amplitudes. $<\tau>$ is the mean fluorescence lifetime, calculated by using only the lifetimes associated with positive amplitudes according to $<\tau> = \sum \tau_i \times \alpha_i$. Excitation and emission wavelengths were 315 and 550 nm, respectively. The reported data are the mean of $n \geq 2$ experiments. Standard deviations are < 15% for lifetimes and amplitudes. [b]Data from Dziuba, Karpenko et al. 2014.

Thus, our time-resolved fluorescence data are fully in line with the spectroscopic data and help us to understand the observed spectroscopic changes upon binding of the labeled ODNs to NC(11-55) or ΔP(+)PBS. Taken together, our data show that the 3HCnt nucleoside can sensitively and site-specifically monitor the binding of the peptide as well as the single- to double-strand transition at nearly all positions. Due to the marginal effect of 3HCnt at position 9 on the affinity for NC(11-55) and the thermodynamic stability of the duplex, as well as due to the high sensitivity of 3HCnt at this position to peptide binding and formation of the extended duplex, this position was selected for monitoring the (–)/(+)PBS annealing reaction.

To monitor the annealing kinetics, we switched from the truncated ΔP(–)PBS to the full-length (–)PBS sequence to take into account the interactions between the overhangs. The labeled 3HCnt-9(–)PBS was mixed with the complementary non-labeled (+)PBS in pseudo first-order conditions in the absence and in the presence of NC(11-55). In both cases, addition of (+)PBS resulted in a formation of an extended duplex accompanied by an important increase of the fluorescence intensity of 3HCnt in line with the steady-state experiments (Figures 5 and 6, publication 2). In the absence of NC, the obtained kinetic traces were fitted with a bi-exponential function, yielding two observed rate constants k_{obs1} and k_{obs2} that were further plotted as a function of (+)PBS concentration (Figure 5b, c). The linear dependency of the k_{obs1} values and the hyperbolic dependency of the k_{obs2} values on the (+)PBS concentration are consistent with a 2-step model (–)PBS + (+)PBS ↔ [IC] ↔ ED with a rapidly formed intermediate complex (IC) which converts through a rate-limiting step into the extended duplex (ED). In parallel, the numerical solving software Dynafit was used to confirm the data

from the analytical calculations and additionally determine the fluorescence intensity of the intermediate products. The kinetic rate constants obtained by Dynafit perfectly matched those obtained by analytical fitting (Figure S3), giving further confidence to the obtained values. Using the numerical approach we found that the fluorescence intensity of the IC is about 70% of the fluorescence of the extended duplex, corresponding to a QY of 9.1%. This rather high QY of the IC indicates that there may be a contact between the loops in the IC, where the relative rotation of the aromatic rings of 3HCnt is restricted but less than in the final duplex. This allowed us to suggest a slightly modified mechanism of (–)/(+)PBS hybridization in the absence of NC(11-55) (Scheme 1, publication 2).

In the presence of NC, the plateau of the fluorescence intensity corresponding to the final product was reached significantly faster, confirming the ability of NC to promote the rapid annealing of the complementary strands. Due to the enhanced rate of the (–)/(+)PBS annealing reaction, the kinetic traces were monitored using both the steady-state and stopped-flow fluorescence techniques in order to obtain the complete kinetic information. As in the absence of NC, the k_{obs} values followed a linear and a hyperbolic dependence on the concentration of (+)PBS, confirming a 2-step process of the annealing reaction. All obtained kinetic parameters were found to be significantly faster than in the absence of NC, indicating that NC promotes both reaction steps. The fluorescence intensity of the IC calculated by Dynafit was found to be 87% of the value observed for the extended duplex, in agreement with the loop-loop mechanism previously reported for the NC-promoted (–)/(+)PBS annealing.

Taken together, our data show that 3HCnt introduced site-selectively into the loop of (–)PBS allows for the first time to obtain the whole set of kinetic rate constants of the annealing reaction with (+)PBS. The annealing reaction was clearly shown to proceed in two steps in both conditions. NC(11-55) was confirmed to promote a loop-loop pathway. Through this pathway, the peptide enhances both the rates of the formation of the intermediate complex and its extension into the duplex. Furthermore, in the absence of peptide, our data evidenced the presence in the IC of loop-loop contacts in the IC that were never seen before. 3HCnt at position 9 was shown to minimally affect the binding of NC to the loop and to provide kinetic constants that were consistent with the single kinetic parameter that could be recovered by using (–)PBS doubly labeled at its 5′ and 3′ ends. The 3HCnt is thus far superior to this previous approach, as it requires only one label and provides direct information on the loop,

which plays a key role in the annealing reaction. Moreover, due to its significant QY in the double stranded duplexes, the 3HCnt nucleobase is far superior to 2Ap which is nearly quenched in the duplexes and would thus require much higher PBS concentrations, leading to kinetics too fast to be properly measured. Thus, 3HCnt label appears as a reliable tool to faithfully monitor the kinetics of the (–)/(+)PBS annealing reaction and could be useful for deciphering the mechanisms of other nucleic acid chaperone proteins.

Publication 2. Full mechanism of (–)/(+)PBS annealing and its promotion by the HIV-1 nucleocapsid protein revealed by site-selective fluorescence labeling

Full mechanism of (–)/(+)PBS annealing and its promotion by the HIV-1 nucleocapsid protein revealed by site-selective fluorescence labelling

Marianna Sholokh[1,3], Natalia Grytsyk[1], Viktoriia Y. Postupalenko[1], Dmytro Dziuba[2], Nicolas P.F. Barthes[2], Benoît Y. Michel[2], Christian Boudier[1], Olga A. Zaporozhets[3], Alain Burger[2], and Yves Mély[1]

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

[2]Institut de Chimie de Nice, UMR 7272 CNRS, Université de Nice Sophia Antipolis, Parc Valrose, 06108 Nice, France

[3]Department of Chemistry, Kyiv National Taras Shevchenko University, 60 Volodymyrska street, 01033 Kyiv, Ukraine

Abstract

Fluorescent nucleoside surrogates containing environment-sensitive fluorophores are attractive tools for site-specific probing of peptide-oligonucleotide and oligonucleotide-oligonucleotide interactions. Recently synthesized 2-thienyl-3-hydroxychromone (3HCnt) nucleobase is characterized by respectable quantum yield, ratiometric response and strong environment sensitivity when incorporated in oligonucleotides. Using 3HCnt, we substituted several nucleosides of (–)PBS stem-loop, a (–)DNA copy of the HIV-1 primer binding site, in order to follow the kinetics of the (–)/(+)PBS annealing reaction, occurring during the second strand transfer of the HIV-1 reverse transcription process. This annealing reaction is chaperoned by the nucleocapsid protein NC that rearranges the PBS loops and favors a specific loop-loop mechanism. Binding of NC to the labelled (–)PBS induces substantial changes of 3HCnt fluorescence quantum yield and dual emission, in line with an increased solvent-exposure of the nucleobases. Importantly, hybridization of (–)PBS to (+)PBS further increases the emission of 3HCnt, allowing sensitive detection of the annealing kinetics at low concentrations of the fluorescent probe. At position 9 in the (-)PBS loop, 3HC was found to not alter the binding affinity of NC and provide the most sensitive response. This position was thus selected to investigate the annealing mechanism. Using steady-state and stopped-flow fluorescence measurements, we confirm the two-step mechanism of the (–)/(+)PBS annealing reaction and determine for the first time the full set of kinetic parameters. The NC-promoted

(–)/(+)PBS annealing reaction was rationalized by the increase in the kinetic rate constants governing the formation of the intermediate complex and its conversion to the extended duplex. In the absence of NC, our data further evidenced loop-loop contacts that were not seen before, and that likely stabilize the intermediate complex. The obtained kinetic parameters complete our understanding on (–)/(+)PBS hybridization mechanisms and show that 3HCnt nucleoside analog is a suitable tool to monitor annealing reactions.

Introduction

The reverse transcription step of the human immunodeficiency virus type 1 (HIV-1) life cycle is a complex process which results in the synthesis of a double-stranded linear DNA from a single-stranded genomic RNA (gRNA) (1,2). This process mediated by the HIV-1 reverse transcriptase (RT) involves two obligatory strand transfers (3). In the first strand transfer, the minus strong-stop DNA [(–)ssDNA] is transferred to the 3′ end of the gRNA template and annealed through the complementary repeat (R) sequences. Following the minus strand transfer of (–)ssDNA, the polypurine tract (PPT) and its complementary cPPT regions initiate synthesis of the plus-strand DNA [(+)ssDNA] using (–)ssDNA as a template. When the primer binding site (PBS) region of the (+)ssDNA is reconstituted, a plus-strand transfer is required to complete the synthesis of the viral DNA. Annealing of the two complementary PBS sequences of the (–) and (+) DNA strands in this second strand transfer results in a circular intermediate that further elongates into a linear double-stranded DNA. Both strand transfers are thought to be chaperoned by the nucleocapsid protein NC (4-8), a short (55 amino acids) structural HIV-1 protein endowed with nucleic acid chaperone activity. Through this activity, NC rearranges the structures of nucleic acids into their most stable conformations (7,9).

The complementary (–)PBS and (+)PBS sequences involved in the plus strand transfer are folded into short and stable stem-loops with terminal single-stranded overhangs (10) (Figure 1). The annealing mechanism was examined using either one of the PBS sequence labeled at its 5’ and 3’ ends, by the rhodamine 6G (Rh6G) fluorophore and the fluorescence quencher Dabcyl, respectively (11), or by using one of the PBS species labeled in its loop by the 2-aminopurine (2Ap) nucleoside analog (12). Using these tools, the (–)/(+)PBS hybridization was found to proceed spontaneously mainly through the flexible single-stranded overhangs in the absence of NC (11,12). In the presence of NC, the annealing reaction is accelerated and proceeds through a loop-loop kissing mechanism (11-13). This specific

pathway results from the binding of NC to the (–)PBS, which stretches the loop and flips the nucleobases to the exterior of the loop (8,12,14), allowing efficient base pairing with the complementary loop of (+)PBS. This NC-directed loop-loop pathway was found to strictly depend on the hydrophobic plateau at the top of zinc fingers of NC (12), which is thus pivotal to ensure the specificity and fidelity of the second strand transfer.

Both the loop-loop and the overhang pathways of the (–)/(+)PBS annealing reaction have been suggested to involve a fast pre-equilibrium step leading to a partially annealed intermediate complex (IC), followed by a rate-limiting conversion to the extended duplex (ED) (11). However, due to the limitations of the probes previously used to monitor the annealing reactions, only one rate constant could be extracted from the kinetic measurements (11,12), so that a complete picture of the kinetics of the two pathways could not be provided. Indeed, the PBS species doubly labeled by Rh6G and Dabcyl were found to be insensitive to the annealing of the loop regions, while the very low brightness of 2Ap (15-17) required high PBS concentrations leading to kinetics too fast to be accurately monitored. Therefore, to provide a complete mechanistic description of the (–)/(+)PBS annealing reaction and notably to describe the molecular events in the loop regions during the annealing reactions, there is a need to use fluorescent nucleobase analogs with good brightness and high environment sensitivity.

To reach this objective, we used a recently synthesized fluorescent nucleoside analog, based on 2-thienyl-3-hydroxychromone (3HCnt) (18) (Figure 1) that acts as a universal nucleoside analog able to substitute any of the natural nucleobases. 3HCnt presents a dual emission as a result of an excited-state intramolecular proton transfer (ESIPT) reaction (19) (20) between the initially excited normal form (N*) and the resulting emissive tautomeric form (T*). The ratio between the two bands was shown to be sensitive to the polarity as well as to the H-bond donor ability of the probe microenvironment. Therefore, by inserting 3HCnt at defined positions in oligonucleotides (ODNs), it is possible to site-specifically monitor the dynamics of protein/ODN interactions (18,21). In the present work, we used the environment sensitive and ratiometric 3HCnt to site-selectively label either the native (–)PBS sequence or its ΔP(–)PBS mutant lacking the 3′ overhang (Figure 1) (14) in order to characterize the (–)/(+)PBS annealing kinetics, both in the absence and the presence of NC.

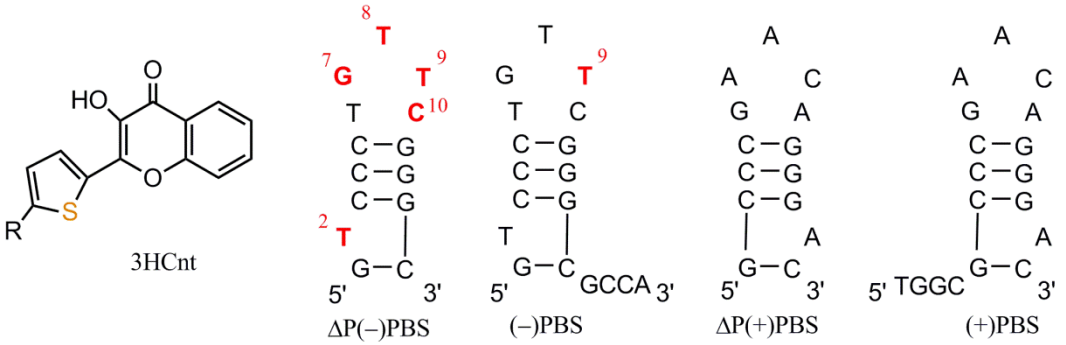


Figure 1. Structures of the 2-thienyl-3-hydroxychromone (3HCnt) nucleobase (R = 2'-deoxyribose) and sequences of the (–) or (+) primer binding site (PBS) and their truncated mutants (the residues substituted by 3HCnt are shown in red).

Materials and methods

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. A truncated form of the HIV-1 nucleocapsid protein NC(11-55) that lacks the N-terminal domain responsible of ODN aggregation was synthesized in the laboratory by the solid-phase peptide synthesis as described in (22). An extinction coefficient ε_{280} = 5700 $M^{-1}cm^{-1}$ was used to determine its concentration by absorption. NC(11-55) stock solution was prepared with 2.0 eq Zn(II) in 10 mM phosphate buffer, 30 mM NaCl (pH=6.5).

Synthesis and preparation of the labeled ODNs.

The 3HCnt nucleobase was synthesized as described in (18). Then, using solid-support ODN synthesis, 3HCnt was selectively introduced at different positions (2, 7, 8, 9, 10) in the ΔP(–)PBS DNA sequences. For the kinetic measurements, 3HCnt was introduced at position 9 of (–)PBS DNA. 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 absorption coefficient of 3HCnt (ε_{260} = 10000 $M^{-1}cm^{-1}$). 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 the water bath and then, slowly cooling down to the room temperature.

Spectroscopic measurements. Absorption spectra were recorded on a Cary 4000 UV-visible spectrophotometer (Varian). Fluorescence spectra were recorded on a FluoroMax4 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. Excitation wavelength was 360 nm for 3HCnt fluorophore. For the quantum yield (QY) measurements, free 2-thienyl-3-hydroxychromone fluorophore in phosphate buffer (QY = 0.046) (23) or quinine sulfate in 0.5 M H_2SO_4 (QY = 0.546) (24) was used as a reference. All spectroscopic measurements were performed in 10 mM phosphate buffer (pH=6.5), 30 mM NaCl, at 20 °C.

Competition experiments. To determine the relative affinity of the labeled sequences comparatively to the non-labeled sequences 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 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 or another sequence 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 the 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 thermal denaturation curves.

Kinetic measurements. Kinetic measurements were performed under pseudo-first-order conditions by mixing 50 or 100 nM labeled 3HCnt-9(–)PBS DNA with a 10- to 100-fold molar excess of non-labeled (+)PBS DNA. The experiments were performed either in the

absence or in the presence of NC(11-55) added at a 3:1 ratio to each complementary ODN prior to the start of the annealing reaction. Measurements were performed using a Fluoromax 4 (Horiba) spectrofluorimeter or a stopped-flow instrument Bio-Logic SFM-3 with 2.8 ms dead time. For the fluorimeter experiments, excitation and emission wavelengths were 360 and 540 nm, respectively. For the stopped-flow experiments, the excitation wavelength was 365 nm and the fluorescence emission was detected above 500 nm through a cutoff filter (deep yellow, Kodak).

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

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

where I_0 and I_f are the initial and final intensities of 3HCnt-9(–)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.

The kinetic rate constants of both components were then determined from the dependence of $k_{obs1,2}$ on the (+)PBS concentration according to:

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

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

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.

The Dynafit software (25,26) 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.

Results and discussion

1. **Characterization of the thermodynamic, binding and spectroscopic properties of the 3HCnt-labeled sequences**

To characterize the mechanism of the spontaneous and NC-promoted annealing of (–)/(+)PBS annealing, our first objective was to compare the thermodynamic, NC-binding and spectroscopic properties of the sequences labeled by 3HCnt at positions 2 in the stem and at

positions 7, 8, 9 or 10 in the loop, in order to identify the best labeling position for reaching our objective. To characterize the properties of the 3HCnt-labeled sequences, we used the ΔP(−)PBS stem-loop sequence (Figure 1), whose 3D structure in both free form and NC-bound form has been solved by NMR (14).

First, thermal denaturation experiments were performed on the duplexes of the labeled ΔP(−)PBS sequences with the complementary ΔP(+)PBS sequence (Figure S1, Table S1 in Supplementary Information). With the exception of position 10, the melting temperatures differed only by a few degrees from that of the non labelled ΔP(−)/(+)PBS duplex, suggesting that the 3HCnt nucleobase moderately affects the stability of the duplexes. This conclusion is consistent with the limited effect of the 3HCnt nucleobase on the stability of 15 bp model duplexes and the behavior of 3HCnt as a universal nucleobase (18).

Next, we characterized the absorption and fluorescence emission spectra of 3HCnt-labeled ΔP(−)PBS stem-loops. All 3HCnt-labeled ODNs showed absorption bands centered at 372 – 375 nm (Table S2), being 5-8 nm red-shifted as compared with the free label in buffer (23), in line with a stacking interaction of the 3HCnt nucleobase with the neighboring bases (27,28). Each labeled 3HCnt-ΔP(−)PBS sequence shows a two-band emission spectrum characteristic for 3-hydroxychromone derivatives (Figure 2). While the position of the N* band maximum (434-437 nm) was highly similar to that of the free probe (440 nm), a red-shift of 30-37 nm was observed for the position of the T* band (Table S2). Such a strong red-shift indicates that the 2-thienyl-3HC fluorophore of 3HCnt is hidden from the solvent through stacking interactions with the neighboring nucleobases. This conclusion was further substantiated by the low N*/T* ratio values (0.07-0.19, Figure 3a), confirming the low polar environment of the label in all 3HCnt-labeled sequences (18). 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 was already observed with other 3HC derivatives and was attributed to the anionic form of the probe, where the 3-OH group is deprotonated (29-31). This deprotonation may result from the interaction of the 3-OH group of 3HCnt at position 8 with a neighboring functional group of the ODN that plays the role of proton acceptor.

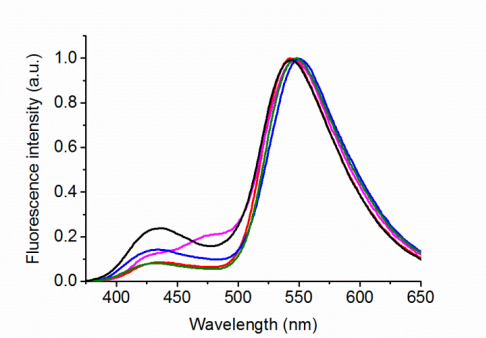


Figure 2. Fluorescence emission spectra of the labeled 3HCnt-2ΔP(–)PBS (black line), 3HCnt-7ΔP(–)PBS (red), 3HCnt-8ΔP(–)PBS (magenta), 3HCnt-9ΔP(–)PBS (blue) and 3HCnt-10ΔP(–)PBS (green) in 10 mM phosphate buffer, 30 mM NaCl, pH = 6.5. Excitation wavelength was 360 nm.

The quantum yields were about 0.4 - 4% for all positions, with the exception of position 7, for which the quantum yield was 12% (Figure 3b). From comparison with the quantum yield of 3HCnt in model sequences, the low quantum yield of 3HCnt at positions 2, 8, 9 and 10 can be rationalized by the presence of G and/or C neighbors, which both were reported to be efficient quenchers of 3HCnt fluorescence (18). In contrast, the 3HCnt residue in 3HCnt-7ΔP(–)PBS is surrounded by two T residues, which were shown to be less efficient fluorescence quenchers, explaining the higher quantum yield observed at this position.

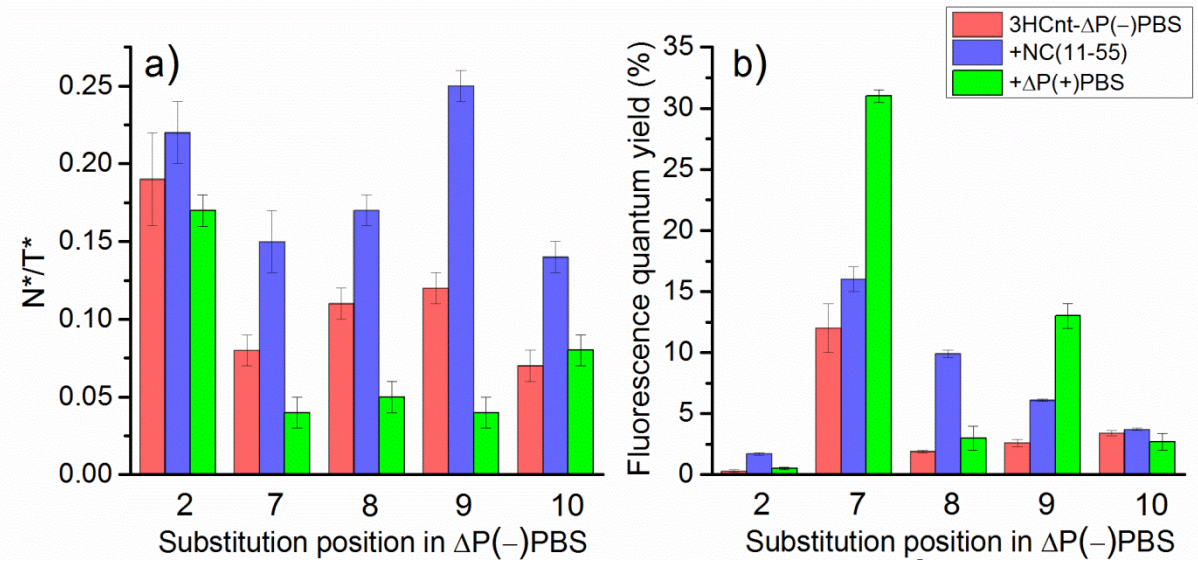


Figure 3. N*/T* ratio (a) and fluorescence quantum yield (b) for the ΔP(–)PBS sequences labeled by 3HCnt at different positions. The 3HCnt-labeled sequences were either in the form

of stem-loops in the absence (red bars) or the presence of NC(11-55) peptide (blue bars) or in the form of duplexes with the complementary ΔP(+)PBS sequence (green bars).

Then, the interaction of the labeled ODNs with NC(11-55) peptide added at saturating concentration (Figure S2) was investigated. The values of the N*/T* ratio and QY of the 3HCnt nucleobase were found to increase at all positions upon peptide binding. This confirms that NC(11-55) is able to bind all along the (–)PBS loop as well as at the 5' end of the stem (14). The increase of the N*/T* ratio values indicates that NC(11-55) binding increases the exposure of the 3HCnt nucleobases to the solvent. At positions 7 and 8, this increase is fully consistent with the NMR data (14), 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 at positions 7 and 8 to the solvent. The similar increased exposure to the solvent of the bases at positions 9 and 10 may likely be the result of the binding of NC(11-55) at its second binding site at the 3'-end of the loop (14). As observed for other 3HC-based fluorophores (32,33), the NC(11-55)-induced increase in the QY of 3HCnt is likely due to the restriction in the rotation between the thienyl and 3-HC rings, so that the two aromatic rings become co-planar. This restricted motion, which probably results from direct contacts between 3HCnt and the hydrophobic platform of NC(11-55), is fully in line with the NC(11-55)-induced restriction in the local motion of the nucleobases, as previously evidenced using 2Ap-labeled (–)PBS sequences (12).

Next, in order to determine whether 3HCnt modifies the binding of NC(11-55) to ΔP(–)PBS, competition experiments were performed that indicate the relative affinity of NC(11-55) for labeled ΔP(–)PBS versus non-labeled ΔP(–)PBS (34). To this end, NC(11-55) was mixed with equimolar concentrations of labeled and non-labeled ΔP(–)PBS. In principle, 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 is the case for 3HCnt-8ΔP(–)PBS and 3HCnt-9ΔP(–)PBS (Figure 4 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 (8,14). An opposite effect was observed when C10 was replaced by 3HCnt, suggesting an about two-fold increase in the ΔP(–)PBS affinity for the peptide. Noticeably, the rather low binding observed with 3HCnt-2ΔP(–)PBS is likely a mechanical consequence of the low affinity of the NC(11-55) peptide

for the stem, so that this binding site will be easily depopulated in the presence of the competing non-labeled sequence.

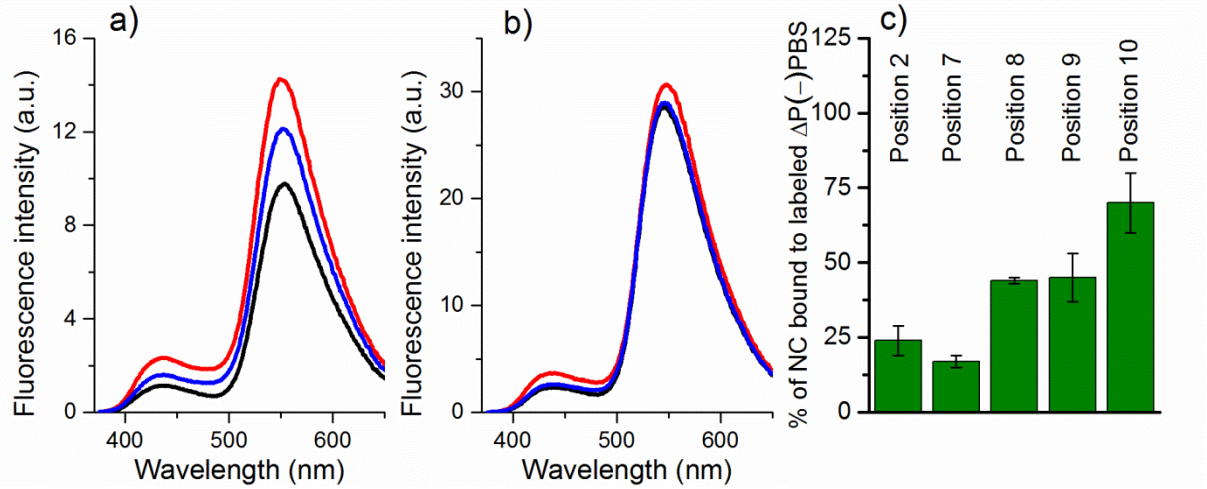


Figure 4. Binding competition experiments. To determine the relative affinity of the 3HCnt-labeled ΔP(–)PBS sequence versus the non-labeled one, NC(11-55) was added to their equimolar mixture. Fluorescence spectra of labeled 3HCnt-9ΔP(–)PBS (a) and 3HCnt-7ΔP(–)PBS (b). The spectra of the complexes of the labeled sequences with NC(11-55) were recorded in the absence (red) and in the presence (blue) of non-labeled ΔP(–)PBS. The spectra of the free 3HCnt-9ΔP(–)PBS and 3HCnt-7ΔP(–)PBS sequences are in black. Experiments were performed in 10 mM phosphate buffer, 30 mM NaCl, pH=6.5. Excitation wavelength was 360 nm. (c) Fraction of NC(11-55) protein bound to labeled ΔP(–)PBS, after addition of NC(11-55) to an equimolar mixture of labeled and non-labeled ΔP(–)PBS sequences. A fraction of 50% corresponds to an equal distribution of NC(11-55) between labeled and non-labeled ΔP(–)PBS.

Finally, we investigated the changes in the fluorescence properties of the labeled ΔP(–)PBS sequences in transitions from the stem-loop (single-strand) to the extended duplexes (Figure S4). Formation of the duplexes was accompanied by a decrease in the N*/T* ratio and a substantial increase in the QY of all labeled sequences, with the exception of the sequence labeled at position 10 that was almost insensitive to duplex formation. The observed drop of the N*/T* ratio in the duplexes is in line with the expected polarity decrease resulting from 3HCnt intercalation within the duplex. Similarly, the intercalation of the 3HCnt nucleobase in the duplex likely induces the observed QY increase (Table S2 and Figure 3b), as a result of the strong restriction of the rotation between the thienyl and 3-hydroxychromone

moieties. The only exception at position 10, where the N*/T* ratio and QY do not change significantly, is likely related to the tight stacking of the 3HCnt base with the neighbor G11 residue in ΔP(–)PBS stem-loop (14). Therefore, transition to double strand is expected to only minimally change its exposure to solvent and its rigidity.

Taken together, our data show that the 3HCnt nucleobase can sensitively and site-specifically monitor the transition between single and double strands at nearly all positions. In comparison to 2Ap that is almost non-emissive in double-stranded DNA, the QY of the 3HCnt nucleobase increases on duplex formation. This key advantage over 2Ap will be beneficially used to monitor in real time the (–)/(+)PBS annealing reaction at much lower concentrations of (–)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 extended duplex, this position was further selected for monitoring the (–)/(+)PBS annealing reaction.

2. Real-time monitoring of (–)/(+)PBS DNA annealing

Annealing of 3HCnt-9(–)PBS with (+)PBS in the absence of NC(11-55)

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 possible interactions between the overhangs. The labeled 3HCnt-9(–)PBS sequence was mixed with the complementary non-labeled (+)PBS in pseudo first-order conditions. Formation of the extended (–)/(+)PBS duplex could be monitored in real-time through the 3HCnt fluorescence increase (Figure 5a).

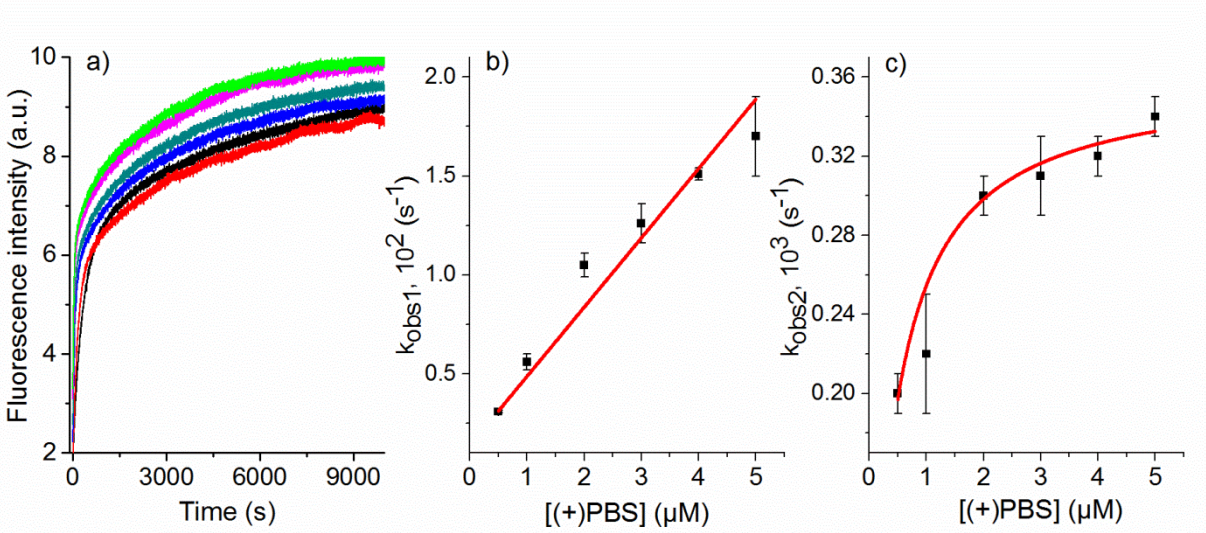


Figure 5. Kinetics of (–)/(+)PBS annealing in the absence of NC(11-55). (a) Kinetic traces of 0.05 μM 3HCnt-9(–)PBS with different concentrations of (+)PBS: 0.5 μM (black), 1 μM (red), 2 μM (blue), 3 μM (green), 4 μM (magenta) and 5 μM (light green), respectively. Excitation and emission wavelengths were 360 and 540 nm. Experiments were performed in 10 mM phosphate buffer, 30 mM NaCl, pH 6.5 at 20 °C. Dependence of the observed kinetic rate constants k_{obs1} (b) and k_{obs2} (c) of the 3HCnt-9(–)/(+)PBS annealing reaction as a function of (+)PBS concentration. The $k_{obs1,2}$ values were obtained from the fits of the curves in a) to equation 2. The $k_{obs1,2}$ values are the means from two experiments. The red lines in b) and c) correspond to the fit of the data points (black squares) to equations 3 and 4, respectively.

The kinetic traces of (–)/(+)PBS annealing (Figure 5a) could be appropriately fitted with the bi-exponential equation (1). The observed rate constants k_{obs1} and k_{obs2} obtained from the fits were then plotted as a function of (+)PBS concentration (Figure 5b, c). The fast observed rate constant k_{obs1} was found to be linearly dependent on the (+)PBS concentration, while the slow k_{obs2} was observed to follow a hyperbolic dependence on the (+)PBS concentration. These observations are consistent with a 2-step model:

$$(-)\text{PBS} + (+)\text{PBS} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} [\text{IC}] \underset{k_{-f}}{\overset{k_f}{\rightleftharpoons}} \text{ED}$$

where a rapidly formed intermediate complex (IC) is further converted through a rate-limiting step into the extended duplex (11,12). 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}

dependence on the (+)PBS concentration to eq. (2) gave k_1 and k_{-1} values of 3500±200 $M^{-1}s^{-1}$ and 1.4±0.2×10^{-3} s^{-1} (Figure 5b). By fitting the hyperbolic dependence of k_{obs2} on the (+)PBS concentration to equation 3, K_m, k_f and k_{-f} values of 2.5±0.3×10^6 M^{-1}, 3.5±0.1×10^{-4} s^{-1} and <1×10^{-4} s^{-1}, respectively, were obtained (Figure 5c). The K_m value was remarkably consistent with the 2.7×10^6 M^{-1} value obtained from the k_1/k_{-1} ratio, highlighting the internal consistency of our data and the appropriate selection of the kinetic model. Noticeably, 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 this duplex. This approach appeared much superior to the previous one based on (–)PBS labeled at its 5' and 3' ends by the rhodamine6G/Dabcyl couple (11,12), where only the fast kinetic rate constant k_1 could be recovered.

To further confirm the data from the analytical calculations, we used the numerical solving software Dynafit (25,26) that does not make any a priori assumption on the relative values of the rate constants of the kinetic model and allows determining the fluorescence intensity of the intermediate products. The kinetic rate constants obtained by Dynafit perfectly matched those obtained by analytical fitting (Table S3), giving further confidence to the obtained values. Interestingly, the numerical approach further indicated that the fluorescence intensity of the IC was about 70% of the fluorescence of the extended duplex, so that it corresponds to a QY of 9.1%. This clearly suggested that the loops also interact together in the IC, in variance with the previous model where the interaction was limited to the overhangs. This allowed us to suggest a slightly modified mechanism of (–)/(+)PBS annealing in the absence of NC(11-55) (Scheme 1).

Annealing of 3HCnt-9(–)PBS with (+)PBS in the presence of NC (11-55)

Next, we investigated the (–)/(+)PBS annealing kinetics in the presence of NC(11-55). Using steady-state fluorescence measurements, the fluorescence intensity plateau corresponding to the final product was reached significantly faster (at ~ 500s) than in the absence of NC (compare Figure 6a to Figure 5). This confirms the ability of NC to promote the rapid annealing of the complementary strands (11,12,35-38). The relative increase of the initial fluorescence intensity at the end of the annealing reaction (more than 2-fold) was identical to that obtained in the absence of NC (Table S2). Fitting of the kinetic traces with eq. (1) was only possible at the lowest concentrations of (+)PBS. Indeed, at higher concentrations, the initial part of the kinetic traces was missing due to the 10-15 s dead time

associated with this technique, so that the k_{obs1} value could not be reliably recovered. Therefore, the same annealing reactions were monitored with the stopped-flow technique, which allows monitoring fast kinetic events with a dead time of 2.8 ms. The kinetic curves (Figure 6b) were recorded during 200s. Though the kinetic traces contain complete information about the fast initial step of the annealing reaction, the formation of the final duplex was not completed. Therefore, the stopped-flow kinetic curves were fitted with eq. (1) using a fixed value of the final fluorescence intensity. Moreover, the values of the k_{obs1} constants were fixed in the bi-exponential fitting of the steady-state curves to obtain the k_{obs2} values (Table S3). The two observed reaction rate constants were plotted as a function of the (+)PBS concentration (Figure 6c, d).

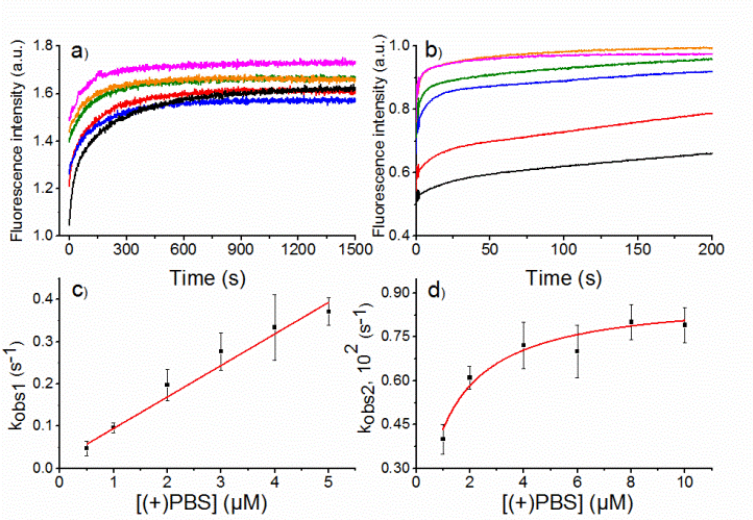


Figure 6. Kinetics of (–)/(+)PBS annealing in the presence of NC(11-55). (a) Kinetic traces monitored by steady-state spectroscopy of 0.1 μM 3HCnt-9(–)PBS with different concentrations of (+)PBS: 1 μM (black), 2 μM (red), 4 μM (blue), 6 μM (green), 8 μM (orange) and 10 μM (magenta), respectively; (b) Stopped-flow kinetic trances of 0.05 μM 3HCnt-9(–)PBS with different concentrations of (+)PBS: 0.5 μM (black), 1 μM (red), 2 μM (blue), 3 μM (green), 4 μM (orange) and 5 μM (magenta), respectively. NC(11-55) was added at a 3:1 peptide:ODN ratio. Experiments were performed in 10 mM phosphate buffer (pH = 6.5), 30 mM NaCl at 20°C. (c-d) Dependence of the observed kinetic rate constants k_{obs1} (c) and k_{obs2} (d) of the 3HCnt-9(–)/(+)PBS annealing reaction as a function of (+)PBS concentration. The $k_{obs1,2}$ values were obtained from the fits of the curves in a) and b) to equation 2. The $k_{obs1,2}$ values are the means from two experiments. The red lines in c) and d) correspond to the fit of the data points (black squares) to equations 3 and 4, respectively.

As in the absence of NC(11-55), the k_{obs1} values were linearly dependent on the (+)PBS concentration, while the k_{obs2} values showed a hyperbolic dependence on the same concentration, confirming the two-step process of the annealing reaction. The kinetic rate constants of the annealing reaction were found to be significantly faster in the presence of NC than in its absence. Indeed, the association rate constant k_1 = 75000 ± 5000 $M^{-1}s^{-1}$ and the dissociation rate constant k_{-1} = 0.020 ± 0.009 s^{-1} are respectively, 20- and 14-fold faster than in the absence of protein. Moreover, the conversion rate constant k_f = 0.90 (± 0.03)×10^{-2} of the IC to the duplex occurs 26-fold faster in the presence of NC than in its absence. Noticeably, the K_m value of 0.9 (± 0.1)×10^6 M^{-1} was consistent with the k_1/k_{-1} ratio (3.75×10^6 M^{-1}), giving confidence on the selection of the kinetic model. Interestingly, very similar values were obtained by using the Dynafit numerical software, giving further confidence in the obtained data. Moreover, the fluorescence intensity of the IC as calculated by Dynafit was found to be 87% of the value observed for the extended duplex, directly confirming the loop-loop mechanism previously inferred for the NC-promoted (–)/(+)PBS annealing reaction (11,12). Furthermore, the product $K_m \times k_f$ = 34000 $M^{-1}.s^{-1}$ was found to be in very good agreement with the single kinetic parameter k = 18400 $M^{-1}.s^{-1}$ that was obtained for the NC-promoted annealing reaction obtained by using the (-)PBS species doubly labeled by Rh6G and Dabcyl, indicating that the 3HCnt does not alter the annealing kinetics.

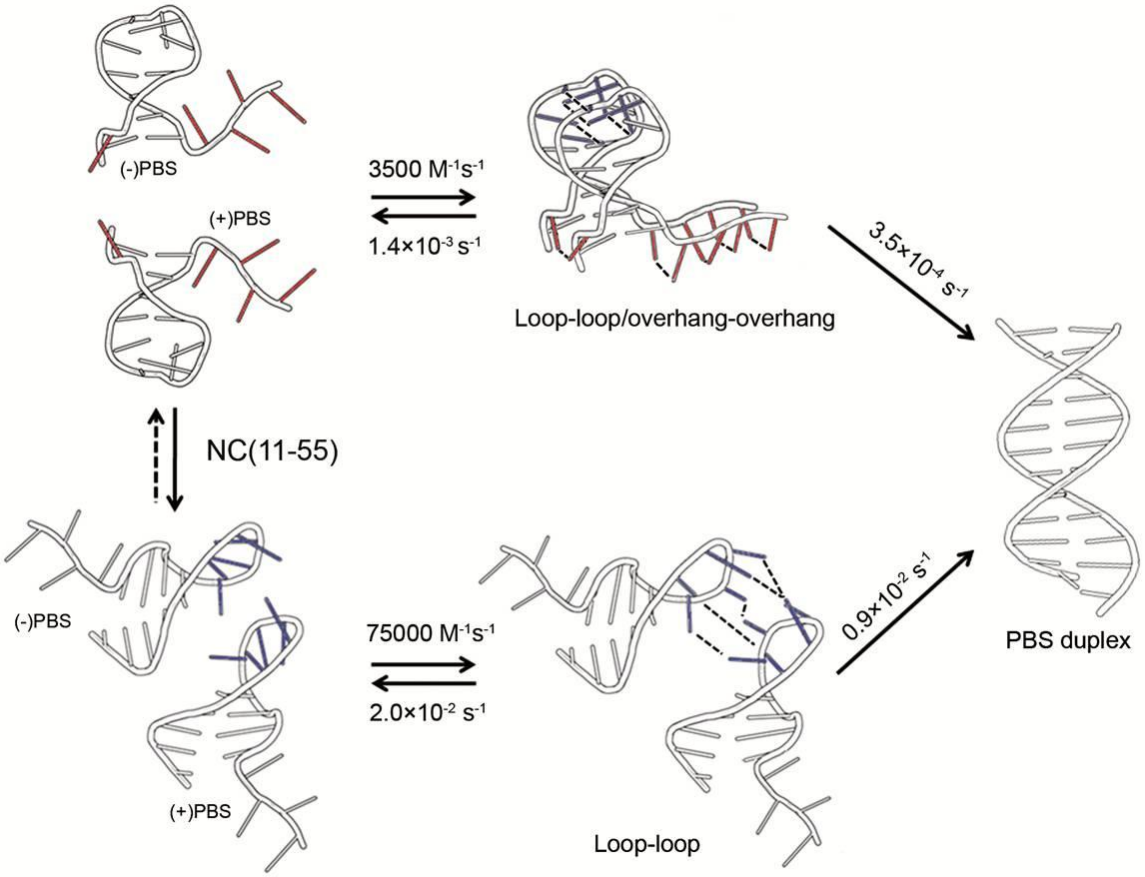
The increase in the k_1 value induced by NC may be due both to the exposure of the loop nucleobases to solvent making them competent for annealing and to the neutralization of the ODN negative charges by the positive charges of NC(11-55) that reduces their electrostatic repulsion and stabilizes binding of NC. Similarly, a 10^3-fold increase of the k_1 value was observed for TAR RNA/cTAR DNA complementary sequences, in the presence of NC (39,40). In TAR/cTAR systems, NC was reported to interact with the accessible G residues of the lower half of cTAR stem-loop, through staking interactions with the Trp37 residue (41). This interaction is thought to destabilize the secondary structure (42,43) and restrict the dynamics of the lower half of cTAR stem-loop (41), preventing fast transitions back to the stably stacked conformations. Thus, these locally destabilized sequences with probably exposed bases competent for annealing with TAR are likely responsible for the large enhancement of the k_1 rate constant in the cTAR/TAR annealing reaction.

An increase in the dissociation rate of the IC k_{-1} is likely due to the ability of NC(11-55) to destabilize imperfect duplexes (40,44-46). Interestingly, as NC stimulates the k_1 and k_{-1}

values nearly to the same extent, the ICs show similar stability in the absence and the presence of NC. The one order of magnitude increase in the conversion rate k_f in the presence of NC(11-55) may tentatively be related to the previously evidenced destabilization of the upper base pair of the stem (14) that may facilitate the annealing of the complementary sequences of (+)PBS and (–)PBS stems. Though both the overhangs and loops of (+)PBS and (–)PBS sequences anneal in the IC in the absence of NC, the stem may not be destabilized in this complex, suggesting that the conversion may be limited by the large energy barrier associated to the melting of the stems.

Conclusions

Taken together, our data show that due to its high environment sensitivity and respectable quantum yield, 3HCnt introduced at position 9 into the loop of (–)PBS allows for the first time to get the whole set of kinetic rate constants governing the (–)/(+)PBS annealing reaction both in the absence and the presence of NC(11-55) (Scheme 1). The annealing reaction was clearly shown to proceed in two steps in both conditions. NC(11-55) was clearly confirmed to promote a loop-loop pathway. Through this pathway, the peptide enhances both the rates of the formation of the intermediate complex and its extension into the duplex. Furthermore, in the absence of peptide, our data evidenced the presence in the IC of loop-loop contacts that were never seen before. 3HCnt at position 9 was shown to minimally affect the binding of NC to the loop and to provide kinetic constants that were consistent with the single kinetic parameter that could be recovered by using (–)PBS doubly labeled at its 5' and 3' ends. The 3HCnt is thus far superior to this previous approach, as it requires only one label and provides direct information on the loop, which plays a key role in the annealing reaction. Moreover, due to its significant QY in the double stranded duplexes, the 3HCnt nucleobase is far superior to 2Ap which is nearly quenched in the duplexes and would thus require much higher PBS concentrations, leading to kinetics too fast to be properly measured. Thus, 3HCnt label appears as a reliable tool to faithfully monitor the kinetics of the (–)/(+)PBS annealing reaction. Moreover, as the changes in the ratio of its two bands and the QY are from different origin, it provides two different channels to monitor the annealing reaction. Obviously, 3HCnt is a highly promising probe that could be useful for deciphering the mechanisms of other nucleic acid chaperone and annealer proteins.



Scheme 1. Proposed mechanism and kinetic rate constants for (–)/(+)PBS annealing. In the absence of NC(11-55) (upper pathway), an intermediate complex is formed through the contacts of the (–)PBS and (+)PBS overhangs and their loops. In the presence of NC(11-55) (bottom pathway), the contacts between the loops are thought to drive the formation of the intermediate loop-loop kissing complex that converts then into the final duplex. The scheme was adapted from (12).

References

1. Arts, E.J. and Wainberg, M.A. (1996) Human immunodeficiency virus type 1 reverse transcriptase and early events in reverse transcription. *Adv. Virus Res.*, **46**, 97-163.
2. Telesnitsky, A. and Goff, S.P. (1997) Reverse Transcriptase and the Generation of Retroviral DNA. In Coffin, J. M., Hughes, S. H. and Varmus, H. E. (eds.), *Retroviruses*, Cold Spring Harbor, New York.
3. Basu, V.P., Song, M., Gao, L., Rigby, S.T., Hanson, M.N. and Bambara, R.A. (2008) Strand transfer events during HIV-1 reverse transcription. *Virus Res.*, **134**, 19-38.

4. Thomas, J.A. and Gorelick, R.J. (2008) Nucleocapsid protein function in early infection processes. *Virus Res.*, **134**, 39-63.
5. Levin, J.G., Guo, J., Rouzina, I. and 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.
6. Levin, J.G., Mitra, M., Mascarenhas, A. and Musier-Forsyth, K. (2010) Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.*, **7**, 754-774.
7. Cristofari, G. and Darlix, J.L. (2002) The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, **72**, 223-268.
8. Johnson, P.E., Turner, R.B., Wu, Z.R., Hairston, L., Guo, J.H., Levin, J.G. and Summers, M.F. (2000) A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry*, **39**, 9084-9091.
9. Rein, A., Henderson, L.E. and Levin, J.G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.*, **23**, 297-301.
10. Darlix, J.L., Godet, J., Ivanyi-Nagy, R., Fosse, P., Mauffret, O. and Mely, Y. (2011) Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.*, **410**, 565-581.
11. Ramalanjaona, N., de Rocquigny, H., Millet, A., Ficheux, D., Darlix, J.L. and Mely, 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.
12. Godet, J., Ramalanjaona, N., Sharma, K.K., Richert, L., de Rocquigny, H., Darlix, J.L., Duportail, G. and Mely, 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.
13. Egele, C., Schaub, E., Ramalanjaona, N., Piemont, E., Ficheux, D., Roques, B., Darlix, J.L. and Mely, Y. (2004) HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J. Mol. Biol.*, **342**, 453-466.
14. Bourbigot, S., Ramalanjaona, N., Boudier, C., Salgado, G.F.J., Roques, B.P., Mely, Y., Bouaziz, S. and 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.

15. Guest, C.R., Hochstrasser, R.A., Sowers, L.C. and Millar, D.P. (1991) Dynamics of mismatched base-pairs in DNA. *Biochemistry*, **30**, 3271-3279.
16. Nag, N., Ramreddy, T., Kombrabail, M., Krishna Mohan, P. M., 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. *Reviews in Fluorescence*, Geddes, C., Lakowicz, J. R. (eds.), Springer Science, New York, **2006**, 311-340.
17. 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.
18. Dziuba, D., Postupalenko, V.Y., Spadafora, M., Klymchenko, A.S., Guerineau, V., Mely, Y., Benhida, R. and 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.
19. Sengupta, P.K. and Kasha, M. (1979) Excited state proton-transfer spectroscopy of 3-hydroxyflavone and quercetin. *Chem. Phys. Lett.*, **68**, 382-385.
20. Demchenko, A.P., Tang, K.C. and Chou, P.T. (2013) Excited-state proton coupled charge transfer modulated by molecular structure and media polarization. *Chem. Soc. Rev.*, **42**, 1379-1408.
21. Kuznetsova, A.A., Kuznetsov, N.A., Vorobjev, Y.N., Barthes, N.P.F., Michel, B.Y., Burger, A. and Fedorova, O.S. (2014) New environment-sensitive multichannel DNA fluorescent label for investigation of the protein-DNA interactions. *PLoS ONE*, **9**, 1-11.
22. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M.C., Darlix, J.L. and Roques, B.P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.*, **180**, 1010-1018.
23. Spadafora, M., Postupalenko, V.Y., Shvadchak, V.V., Klymchenko, A.S., Mely, Y., Burger, A. and Benhida, R. (2009) Efficient synthesis of ratiometric fluorescent nucleosides featuring 3-hydroxychromone nucleobases. *Tetrahedron*, **65**, 7809-7816.
24. Melhuish, W.H. (1961) Quantum efficiencies of fluorescence of organic substan ces: Effect of solvent and concentration of the fluorescent solute. *J. Phys. Chem.*, **65**, 229-235.

25. Kuzmic, P. (1996) Program Dynafit for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal. Biochem.*, **237**, 260-273.
26. Kuzmic, P. (2009) Dynafit – a software package for enzymology. *Methods Enzymol.*, **467**, 247-280.
27. Asseline, U., Delarue, M., Lancelot, G., Toulme, F., Thuong, N.T., Montenaygarestier, T. and Helene, C. (1984) Nucleic-acid binding-molecules with high-affinity and base sequence specificity - intercalating agents covalently linked to oligodeoxynucleotides. *Proc. Natl. Acad. Sci. U. S. A.* , **81**, 3297-3301.
28. Amann, N., Huber, R. and Wagenknecht, H.A. (2004) Phenanthridinium as an artificial base and charge donor in DNA. *Angew. Chem. Int. Ed.*, **43**, 1845-1847.
29. Mandal, P.K. and Samanta, A. (2003) Evidence of ground-state proton-transfer reaction of 3-hydroxyflavone in neutral alcoholic solvents. *J. Phys. Chem. A*, **107**, 6334-6339.
30. Shynkar, V.V., Klymchenko, A.S., Mely, Y., Duportail, G. and Pivovarenko, V.G. (2004) Anion formation of 4 '-(dimethylamino)-3-hydroxyflavone in phosphatidylglycerol vesicles induced by HEPES buffer: A steady-state and time-resolved fluorescence investigation. *J. Phys. Chem. B*, **108**, 18750-18755.
31. Tomin, V.I. (2011) Spectral properties of anionic form of 3-hydroxyflavone. *Opt. Spectrosc.*, **110**, 550-556.
32. Sholokh, M., Zamotaiev, O.M., Das, R., Postupalenko, V.Y., Richert, L., Dujardin, D., Zaporozhets, O.A., Pivovarenko, V.G., Klymchenko, A.S. and Mely, Y. (2015) Fluorescent amino acid undergoing excited state intramolecular proton transfer for site-specific probing and imaging of peptide interactions. *J. Phys. Chem. B*, **119**, 2585-2595.
33. Kilin, V., Glushonkov, O., Herdly, L., Klymchenko, A., Richert, L. and Mely, Y. (2015) Fluorescence lifetime imaging of membrane lipid order with a ratiometric fluorescent probe. *Biophys. J.*, **108**, 2521-2531.
34. Shvadchak, V.V., Klymchenko, A.S., De Rocquigny, H. and Mely, Y. (2009) Sensing peptide - oligonucleotide interactions by a two-color fluorescence label: Application to the HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, **37**, e25.
35. Darlix, J.L., Vincent, A., Gabus, C., de Rocquigny, H. and Roques, B. (1993) Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV-1 RNA. *C. R. Acad. Sci. III*, **316**, 763-771.

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

37. Liu, H.W., Zeng, Y., Landes, C.F., Kim, Y.J., Zhu, Y., Ma, X., Vo, M.N., Musier-Forsyth, K. and Barbara, P.F. (2007) Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc. Natl. Acad. Sci. U S A*, **104**, 5261-5267.

38. You, J.C. and 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.

39. Vo, M.N., Barany, G., Rouzina, I. and 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.

40. Godet, J., De Rocquigny, H., Raja, C., Glasser, N., Ficheux, D., Darlix, J.L. and Mely, 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.

41. Godet, J., Kenfack, C., Przybilla, F., Richert, L., Duportail, G. and 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.

42. Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B.P., Darlix, J.L. and Mely, 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**, 385-399.

43. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J.P., Ficheux, D., Roques, B., Darlix, J.L. and Mely, 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.

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

45. Godet, J. and Mely, Y. (2010) Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol.*, **7**, 687-699.
46. Wu, T.Y., Heilman-Miller, S.L. and Levin, J.G. (2007) Effects of nucleic acid local structure and magnesium ions on minus-strand transfer mediated by the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, **35**, 3974-3987.

Supporting information

Full mechanism of (–)/(+)PBS annealing and its promotion by the HIV-1 nucleocapsid protein revealed by site-selective fluorescence labelling

Marianna Sholokh[1,3], Natalia Grytsyk[1], Viktoriia Y. Postupalenko[1], Dmytro Dziuba[2], Nicolas P.F. Barthes[2], Benoît Y. Michel[2], Christian Boudier[1], Olga A. Zaporozhets[3], Alain Burger[2], and Yves Mély[1]

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

[2]Institut de Chimie de Nice, UMR 7272 CNRS, Université de Nice Sophia Antipolis, Parc Valrose, 06108 Nice, France

[3]Department of Chemistry, Kyiv National Taras Shevchenko University, 60 Volodymyrska street, 01033 Kyiv, Ukraine

1. Melting temperatures of the native and 3HCnt-labeled duplexes ΔP(–)/ΔP(+)PBS

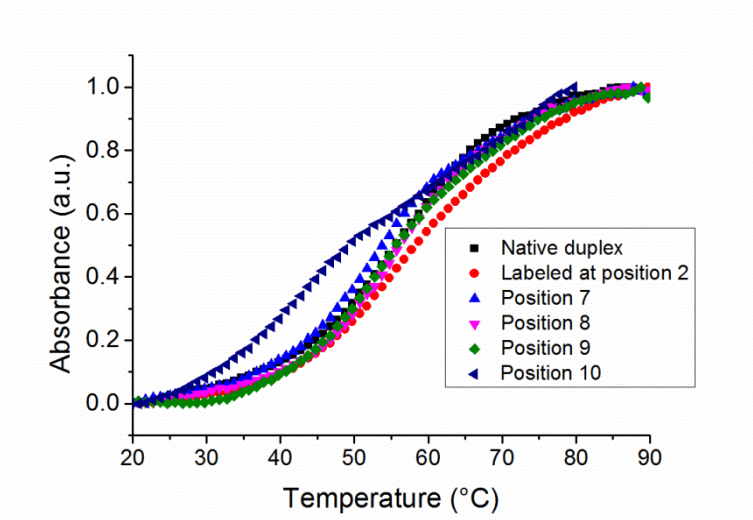


Figure S1. Normalized thermal denaturation curves of the non-labeled and 3HCnt-labeled duplexes ΔP(–)/ΔP(+)PBS. The absorbance was recorded in the forward and backward directions from 20 to 90 °C at a rate of 0.5 °C/min. Experiments were performed in 10 mM cacodylate buffer (pH=7.0), 150 mM NaCl. To avoid the evaporation of the samples, they were overlaid with mineral oil.

Table S1. Melting temperatures of the double-stranded DNA samples[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]T_m is the melting temperature. The results are expressed as mean ± standard error of the mean for at least two experiments. The concentration of ODNs was 1 μM in 10 mM cacodylate buffer, 150 mM NaCl, pH = 7.0.

Table S2. Spectroscopic properties of 3HCnt-labeled ΔP(–)PBS derivatives 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 $n \geq 2$ experiments. Standard deviations are < 15% for N*/T* and QY values. Excitation wavelength was 360 nm. Measurements were performed in 10 mM phosphate buffer, 30 mM NaCl, pH 6.5.

2. Fluorescence emission spectra of the 3HCnt - labeled ΔP(–)PBS and in complexes with NC(11-55)

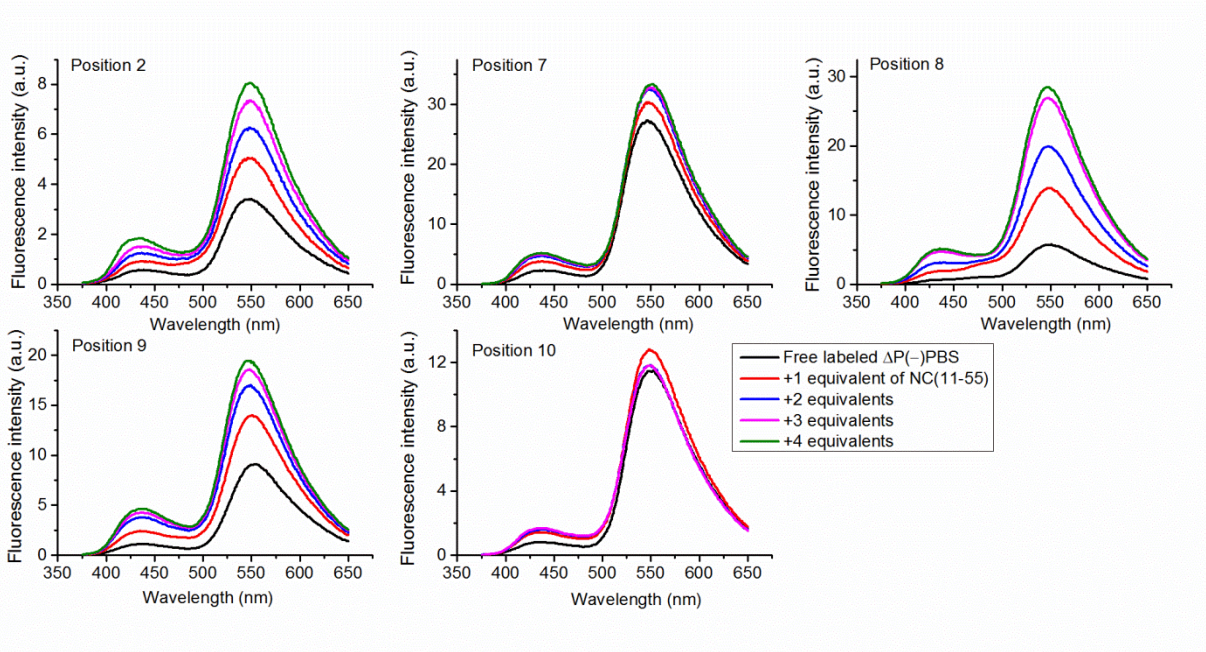


Figure S2. Fluorescence emission spectra of labeled 3HCnt-ΔP(–)PBS in the absence and in the presence of NC(11-55) at 1:1; 1:2; 1:3 and 1:4 ODN : peptide ratio. Concentrations of labeled ODNs were 2 µM. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer (pH = 6.5),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

3. Binding competition experiment for 3HCnt-labeled ΔP(–)PBS

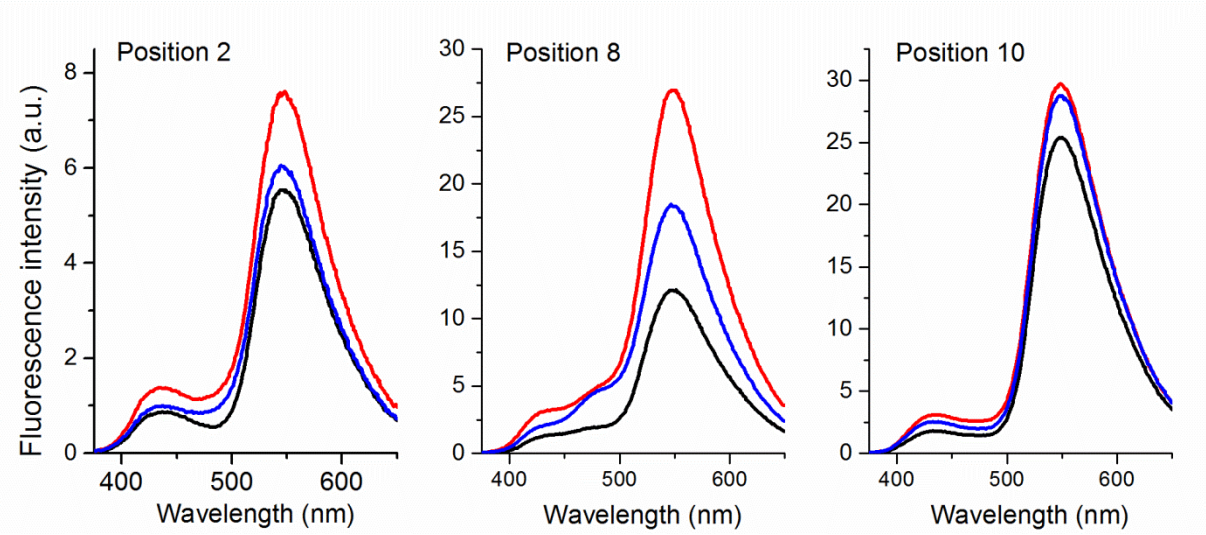


Figure S3. Competition binding experiments. Fluorescence spectra of the labeled 3HCnt-ΔP(–)PBS were performed in the absence (black line) or in the presence of NC (11-55) added

at a ratio 1:1 (red). To this mixture, one equivalent of non labeled ΔP(-)PBS (blue) was added. Concentrations of labeled ODNs were 2 µM. Excitation wavelength was 360 nm. Experiments were performed in 10 mM phosphate buffer, 30 mM NaCl (pH=6.5).

4. Fluorescence emission spectra of 3HCnt-labeled ΔP(–)/(+)PBS duplexes

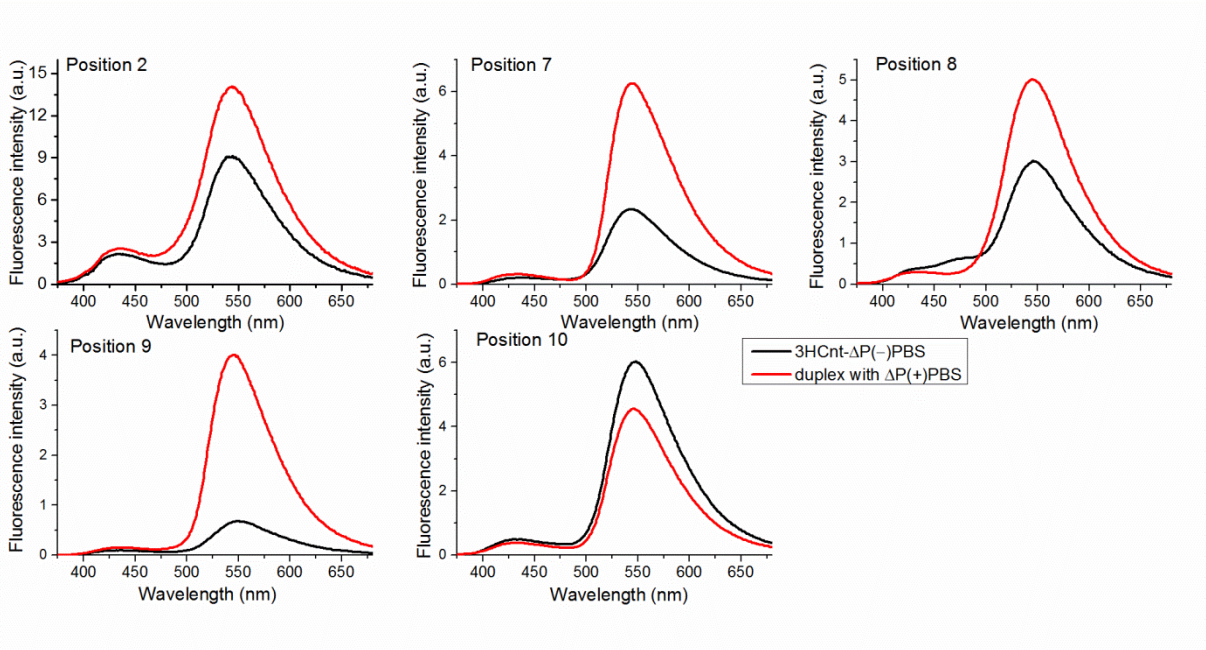


Figure S4. Fluorescence emission spectra of labeled 3HCnt-ΔP(–)PBS in the absence and in the presence of the complementary ΔP(+)PBS. ODNs were mixed at 1:3 ratio, respectively, and hybridized by heating at 85°C and cooling to room temperature. Concentrations of labeled ODNs were 2 µM. Excitation wavelength was 360 nm. Spectra were recorded in 10 mM phosphate buffer (pH = 6.5), 30 mM NaCl at 20 °C.

5. Fitting of the kinetic curves

Table S3. Kinetic parameters of the 3HCnt-labeled (–)/(+)PBS annealing in the absence of NC(11-55)[a]

Approach	k_1, $M^{-1}s^{-1}$	k_{-1}, s^{-1}	k_f, s^{-1}
Analytical	3500±200	0.0014±0.0002	0.00035±0.00001
Numerical	4500±400	0.0011±0.0001	0.00030±0.00001

[a] (-)PBS was labeled at position 9. The same set of kinetic curves were treated by either the analytical or numerical approach. k_1 and k_{-1} are the kinetic rate constants for the formation

and dissociation of the IC; k_f is the forward rate constant for the conversion of the IC into the final duplex.

Table S4. Observed kinetic rate constants of the NC(11-55)-promoted (–)/(+)PBS annealing reaction[a]

Concentration of (+)PBS, µM	k_{1obs}, s^{-1}	k_{2obs}, s^{-1}
1	0.095	0.0040
2	0.17	0.0061
4	0.32	0.0072
6	0.47	0.0070
8	0.62	0.0080
10	0.77	0.0079

[a] k_{1obs} and k_{2obs} are observed kinetic constants reported in s^{-1}. The reported values are the means from two experiments. SDs are < 20% for k_{1obs}, and < 10% for k_{2obs}.

Calculations using Dynafit

The data files reporting the fluorescence intensity changes of the 3HCnt-labeled (-)PBS as a function of time were used. These data files were used with a script that includes the mechanism of the reaction, approximate constants taken from the analytical calculations, initial concentrations of 3HCnt-9(–)PBS and (+)PBS, and the responses of the fluorescent products that are their fluorescence intensities divided by the concentration of 3HCnt-9(–)PBS.

A typical script for the 3HCnt-9(–)PBS/(+)PBS system in the absence of NC is presented below:

```
[task]
 task  = fit
 data  = progress

[mechanism]
   A + B <===> AB   :   k1  k-1
   AB <===> C  :   kf  k-f

[constants]
  k1   = 3000 ?
  k-1   = 0.0027 ?
  kf =  0.00036 ?
  k-f = 0.000001 ?
[concentrations]
```

```
 A = 50e-9

[progress]
 directory      C:\Users\msholokh\Desktop\kamaldynafit\marianna\pbs
 extension   txt

 file       500nM
 concentration   A = 50e-9 , B = 500e-9
 responses  A = 3.58e+11 , AB = 1.79e+12 ? , C = 1.79e+12 ?

 file       1000nM
 concentration   A = 50e-9 , B = 1000e-9
 responses  A = 3.49e+11 , AB = 1.74e+12 ? , C = 1.74e+12 ?

 file       2000nM
 concentration   A = 50e-9 , B = 2000e-9
 responses  A = 4.44e+11 , AB = 1.83e+12 ? , C = 1.83e+12 ?

 file       3000nM
 concentration   A = 50e-9 , B = 3000e-9
 responses  A = 5.14e+11 , AB = 1.87e+12 ? , C = 1.87e+12 ?

 file       4000nM
 concentration   A = 50e-9 , B = 4000e-9
 responses  A = 6.17e+11 , AB = 1.97e+12 ? , C = 1.97e+12 ?

 file       5000nM
 concentration   A = 50e-9 , B = 5000e-9
 responses  A = 6.59e+11 , AB = 1.98e+12 ? , C = 1.98e+12 ?
[output]
 directory   ./FP/out_rtpt
[end]
```

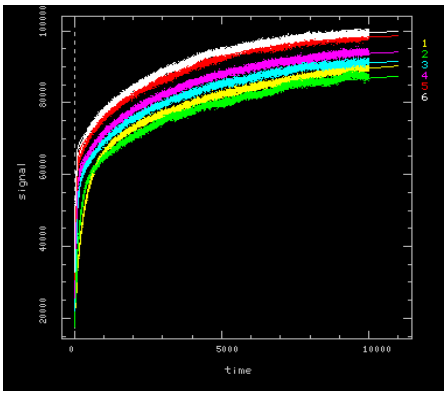


Figure S6. Fitting of the kinetic traces of (–)/(+)PBS annealing in the absence of NC (11-55). The continuous lines correspond to the best fit of the data with the Dynafit software.

3.2.2. Photophysical characterization of the isomorphic guanosine surrogate thienodeoxyguanosine as a potential G7 mimic in (–)PBS

In parallel to the investigations of the dynamics and mechanism of (–)/(+)PBS annealing reaction using 3HCnt analog, we exploited thG nucleoside analog, a thieno[3,4-d]pyrimidine-derivative, synthesized by Dongwon Shin from the group of Yitzhak Tor. In a first step, a complete photophysical analysis of the ribo-thG was performed. It revealed the existence in buffer of the two ground-state tautomeric forms (Figure 1, publication 3) with significantly shifted absorption (313 and 334 nm) and emission (400 and 470 nm) maxima (Figure 2, publication 3), but similar high QY in aqueous solution (0.49±0.03). Since the absorption spectrum of the red-shifted form was identical to the spectra of thG and dthG in dioxane, we used its molar extinction coefficient in dioxane to calculate the concentration of the red-shifted form in buffer and deduced the proportion and extinction coefficient of the blue-shifted form. Photostability experiments further showed that the red-shifted tautomer excited at 350-380 nm is highly photostable (Figure S2, publication 3) while illumination at higher energies (e.g., 325 nm), where both tautomers absorb, showed a continuously decrease in emission, suggesting that the blue-shifted tautomer is less photostable.

Spectroscopic studies of thG in various organic solvents demonstrated the presence of the two tautomers in polar protic alcohols (methanol, ethanol and n-butanol) (Figure S3, publication 3), while in polar aprotic (*N*,*N*-dimethylformamide, acetonitrile and ethyl acetate) and relatively apolar solvent (1,4-dioxane) only the red-shifted tautomer was observed. Comparison of the polarity and proticity parameters of the solvents (Table S1, publication 3) suggested that the blue-shifted tautomer may be stabilized by the H-bond donor ability of the solvent. This hypothesis was further substantiated by the deconvoluted absorption spectra of thG in polar protic solvents (Figure S6, publication 3) where the relative concentrations of the blue-shifted form was linearly increasing with solvent proticity. Therefore, the equilibrium between the two tautomers seems to mainly depend on the H-bonding stabilization of the solvent.

To further identify the two emissive tautomers, quantum chemical calculations were performed in the gas phase, dioxane and water (in collaboration with Roberto Improta). The keto-amino thG-H1 tautomer (Figure 3, publication 3) appears largely dominant over the other tautomers, except in water where it was almost isoenergetic with the keto-amino thG-H3 tautomer. thG-H1 was found to be the main contributor to the spectrum of thG in dioxane, and

thus, was assigned to the red-shifted tautomer. The mixture of the two H1 and H3 tautomers was found to be responsible of the spectroscopic properties of thG in water, and thus, the blue-shifted form was assigned to the thG-H3 tautomer. Further, the absorption and emission maxima were computed for the two forms, and were very close to the experimental ones (Figures S9, 2 and Table S1, publication 3). Altogether, our data indicate that the thG-H1 and thG-H3 tautomers, differing by the position of the protonated nitrogen atom, are responsible of the observed photophysical properties of thG.

When the dG7 residue in (–)PBS loop was replaced by dthG, both H1 and H3 tautomers were observed (Figure 4a, publication 3). In contrast, when dthG7(–)PBS was annealed to its complementary (+)PBS strand, only the emission of thG-H1 was observed, indicating that this tautomer is favored in the double-stranded form (Figure 4, publication 3). A similar switch from a two-band to a single-band emission has been previously observed in model thG- and dthG-labeled sequences upon transition from single- to double-strands, indicating that the reported tautomeric shift is not unique for (–)PBS.

Different emission spectra were observed for the fully matched and mismatched (–)/(+)PBS duplexes, where dthG was placed opposite A, T, or G (Figure 4d, publication 3). The blue-shifted tautomer was nearly absent in the fully complementary duplex, but its relative concentration increased depending on the opposite nucleobase in the following order: T<G<A. Thus, the relative emission of the two dthG tautomers is sensitive to its opposite base and can, in principle, be applied to characterize the base pairing of the substituted G residues. Next, to check the response of the tautomers to NC(11-55) binding, a titration of dthG7(–)PBS by NC(11-55) was performed, showing a strong increase of the red-shifted peak of dthG7(–)PBS, but only a marginal increase in the blue-shifted peak. This indicates that the relative emission of the two tautomers is sensitive to protein binding.

To further substantiate these biophysical observations, molecular dynamics (MD) simulations were performed using the NMR structure of ΔP(–)PBS DNA (in collaboration with Mattia Mori). In two representative π-π stacked and solvent-exposed conformations of dG7 in ΔP(–)PBS (Figure S11, publication 3), dG and the two dthG tautomers were found to show the same behavior, according to the analysis of MD trajectories and thermodynamic parameters (Figures S12, S13, publication 3). In contrast, the analysis of dG and dthG base pairing in the (–)/(+)PBS duplex clearly showed that dthG-H1 shows the same behavior as dG

(Figure S12d, publication 3), forming Watson-Crick H-bonds with dC, while dthG-H3 forms multiple non-canonical base pairs with dC (Figures S14, S15, publication 3). Taken together, the replacement of dG with dthG-H1 in (–)/(+)PBS duplex is markedly more favorable than the replacement with dthG-H3 from a thermodynamic and conformational point of view.

To summarize, either as a free nucleoside and when incorporated into nucleic acids, thG exhibits two ground-state tautomers with significantly shifted absorption and emission spectra. These two tautomers were identified as being the keto-amino thG-H1 and thG-H3 forms. MD studies further suggested that thG-H1 is highly similar to G in both the single- and double-stranded PBS sequences, whereas the thG-H3 tautomer behaves comparably to G only in (–)PBS loop. The ratio of the two tautomers and their relative emission was found to be highly sensitive to the strandedness of the ODN, the nature of the opposite base in DNA duplexes and protein binding.

Publication 3. Tautomers of a fluorescent G surrogate and their distinct photophysics provide additional information channels

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*

This work is dedicated to Professor David Shugar on the occasion of his 100th birthday.

Abstract: Thienoguanosine (thG) is an isomorphic nucleoside analog 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 (313 and 334 nm) and emission (400 and 470 nm) wavelengths, and high quantum yield in buffer (0.49±0.03). 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 environment sensitivity provide unprecedented information channels for analyzing G residues in oligonucleotides and their complexes.

[*] 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, California 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, 80136 Napoli (Italy)

Dr. M. Mori, [+] Prof. Dr. M. Botta
Dipartimento di Biotecnologie, Chimica e Farmacia
Università degli Studi di Siena
via Aldo Moro 2, I-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 d'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
1 rue Laurent Fries, 67400 Illkirch (France)

[+] These authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document.

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

In this context, emissive nucleoside analogs, which have become powerful biophysical tools,[4] provide unique prospects. A tautomerizable nucleoside analog, 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. Here we analyze the photophysics of thienoguanosine thG, a highly useful G surrogate,[5] and identify two environmentally sensitive ground state tautomeric forms (Figure 1) that display distinct absorption and emission spectra. The equilibrium between the two tautomers is mainly governed by the H-bond donor properties of the solvent. Their observed sensitivity to the microenvironment was rationalized by *ab initio* calculations. By exploring single and double stranded thG-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.

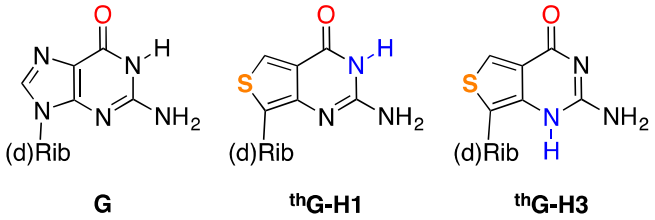


Figure 1. Structures of guanosine (G) and the two emissive tautomers of thienoguanosine (thG-H1 and thG-H3). (d)Rib = D-ribose or 2'-deoxy-D-ribose.

The emission spectra of thG or dthG in water and methanol are surprisingly complex.[5a,b] Excitation at both 360 and 380 nm gives a similar emission spectrum (Figure 2a, orange) centered

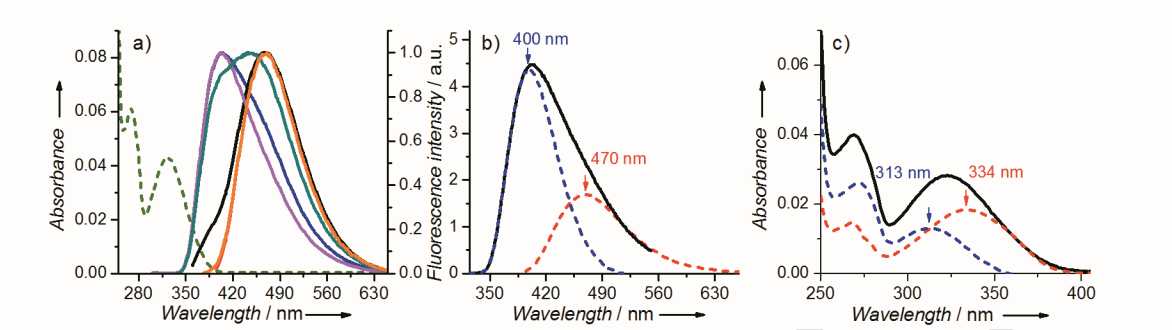


Figure 2. Absorption and emission spectra of thG in TRIS-HCl buffer 25 mM, pH=7.5, 30 mM NaCl, 0.2 mM $MgCl_2$. (a) Absorption (green dashed line) and emission spectra of thG at different excitation wavelengths: 283 nm (magenta line); 298 nm (blue); 320 nm (green); 345 nm (black); and 360 nm (orange). The emission spectra were normalized at their maxima. The normalized emission spectrum at λ_{exc} = 380 nm fully overlaps that at λ_{exc} = 360 nm and is not shown. (b) Deconvoluted emission spectrum of thG, obtained at λ_{ex}= 283 nm. (c) Deconvolution of the absorption spectrum of thG (black line) in its two ground-state forms (colors as in b).

at 468 nm. When the excitation energy is progressively increased, a blue-shifted emission with a maximum at 400 nm appears and becomes dominant for excitation below 300 nm (Figure 2a, magenta and blue). The simplest interpretation is that thG exhibits two ground-state species with shifted emission spectra. This hypothesis is supported by recording excitation spectra for various emission wavelengths (Figure S1 in 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 (290–375 nm) and close to earlier reported values,[5a,c] 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 thG and dthG in dioxane.[5a,b] Using the molar extinction coefficient of thG in dioxane (ε^{333} = 4530 $M^{-1}cm^{-1}$),[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 $M^{-1}cm^{-1}$) of the blue-shifted tautomer. Importantly, the red-shifted tautomer, excited at 350–380 nm, is highly photostable (Figure S2). In contrast, extended illumination at higher energies (e.g., 325 nm), where both tautomers absorb, shows continuously diminished emission, suggesting that the blue-shifted tautomer is less photostable and that the two tautomers are equilibrating.

The spectroscopic properties of thG were comparatively characterized in various solvents (Table S1). In methanol, ethanol and *n*-butanol (Figures S3, S4) spectra comparable to those in buffer are obtained, indicating that thG 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 (Figure S3c). Its emission maximum correlates well with the empirical polarity index $E_T(30)$ of the solvent (Figure S5). Interestingly, although *N,N*-dimethylformamide, acetonitrile and methanol are all rather polar [$E_T(30)$ > 43 kcal mol^{-1}, ε > 32], the blue-shifted tautomer is seen only in methanol, strongly suggesting H-bonding stabilization. This is further substantiated by the deconvoluted absorption spectra in polar protic solvents (Figure S6), showing that the relative concentration of the blue-shifted isomer linearly increases with solvent proticity (Figure S6d, Table S1). Thus, the equilibrium between the two thG tautomers is dependent on the H-bond donor ability of the solvent.

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

Independently of the inclusion of one solvent molecule in the computational model (Figure S7, S8 and Table S2), the thG-H1 tautomer appears as the main contributor to the observed spectroscopic properties of thG in 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 (Figure S7 and Table S2). This suggests 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, i.e. < 400 cm^{-1}) is beyond the expected accuracy of our method, explaining why the molar fraction of the thG-H3 tautomer (Table S2) does not perfectly match with the experimental value

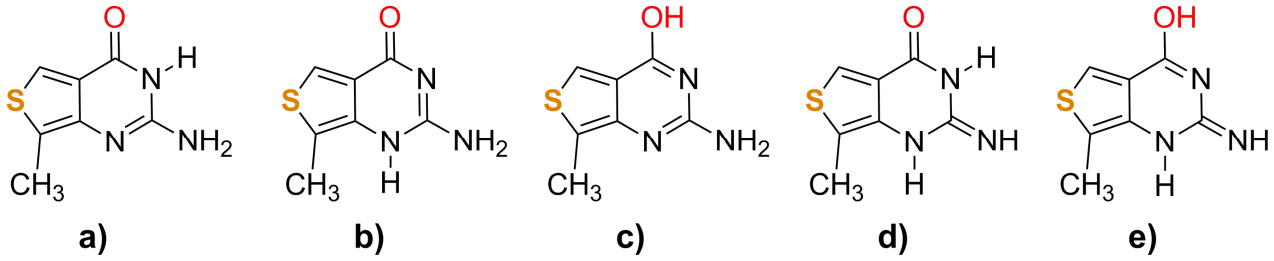


Figure 3. Schematic drawing of the thG tautomers that have been calculated: a) keto-amino thG-H1, b) keto-amino thG-H3, c) enol-amino, d) keto-imino, and e) enol-imino.

(0.44).[9] The computed vertical absorption and emission energies (Table S4)[10] indicate that, independently 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 transition.[11]

Interestingly, small differences in the shape of the frontier orbitals involved in the electronic transition (Figure S9) result in fairly large differences in the computed vertical excitation energy, so that the excitation wavelength 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 330~350 nm (depending on the solvation model) for thG-H1 and 300~310 nm for thG-H3, are very close to the experimental ones (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 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 448 and 383 nm for the thG-H1 and thG-H3 tautomers, respectively,[15] (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 (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 thG's tautomers in oligonucleotides, the DNA equivalent of the 18-mer primer binding site of HIV-1 was selected as a biologically-relevant model (Figure S11). It forms a stem-loop of known 3D structure[16] and is involved in the second strand transfer of HIV-1 reverse transcription.[17] Deoxy-thG, 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 (Figure S11), forming the dthG7(–)/(+)PBS duplex (Figure 4c, inset), the normalized emission spectra obtained at different excitation wavelengths all overlap, indicating that the dthG-H1 tautomer is predominant in the double-stranded 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] 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 375 nm and 550 nm, I_{375}/I_{550}, increases by a factor of 3 and 5 in the mismatched duplexes with opposite dG and dA, respectively (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 two-fold difference in their quantum yield (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, 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), 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.

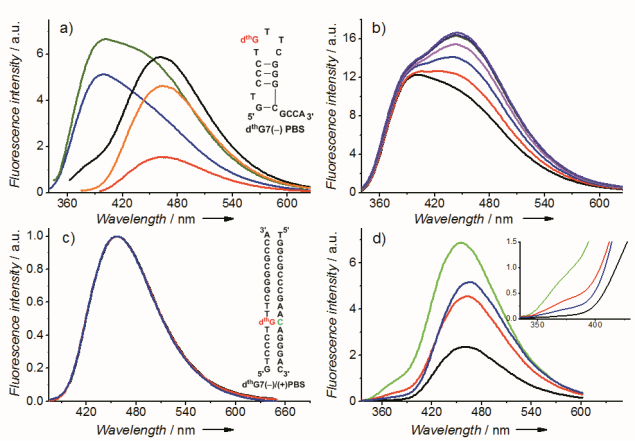


Figure 4. Emission spectra of dthG7(–)PBS (a, b) and dthG7(–)/(+)PBS duplexes (c, d). (a) Emission spectra of dthG7(–)PBS recorded at different excitation wavelengths: 298 nm (blue), 320 nm (green), 345 nm (black), 360 nm (orange), and 380 nm (red). Inset: structure of dthG7(–)PBS, the G7 residue (red) 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 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 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.

To shed light on the biophysical observations, molecular dynamics (MD) simulations using the ff12SB AMBER force field were performed using the NMR structure of ΔP(–)PBS DNA,[16b] a truncated form of (–)PBS, lacking the 3' overhang (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 dG or the two dthG tautomers in two representative ΔP(–)PBS structures (Figures S12, S13a and b, S14). In contrast, analysis of local motion within the Watson-Crick base pair established by 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 (Figure S13d). A local structural analysis of MD trajectories further confirms that dthG-H1 forms the three canonical H-bonds with dC as observed for guanine (Figure S15a, b), while dthG-H3 contacts dC in multiple non-canonical complexes (Figure S15c and Figure 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 (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 that 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.

Acknowledgements

The work was supported by a fellowship from the Ministère de la Recherche (M.S.), the European Project THINPAD "Targeting the HIV-1 Nucleocapsid Protein to fight Antiretroviral Drug Resistance" (FP7-Grant Agreement 601969), Agence Nationale de la Recherche (ANR blanc Fluometadn and FEMTOSTACK), Agence Nationale de Recherche sur le SIDA, French-Ukrainian Dnipro program, the Université de Strasbourg, the Centre National de la Recherche Scientifique (CNRS), the Institut de la Santé et de la Recherche Médicale (INSERM) and the US National Institutes of Health (GM 069773). Computing time was provided at the French national computing centers by GENCI (Grand Equipement National de Calcul Intensif) and the Meso-center for High Performance Computing at the Université de Strasbourg and supported by the project EQUIP@MESO.

Keywords: ab initio calculations • fluorescent nucleosides • molecular modelling • nucleic acids • tautomerism

[1] a) H. T. Miles, *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 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. Interact., Suppl. J. Biosci.* **1985**, *8*, 657; d) W. N. Wang, H. W. Hellinga, L. S. Beese, *Proc. Natl. Acad. Sci. U.S.A.* **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 H-bonding ability of the solvent.

[9] For the examined solvents, PBE0 and M052X provide similar indications, 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, supporting their assignment.

[13] a) M. Gedik, A. Brown, *J. Photochem. Photobiol. A: Chem.* **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 $^{th}G \bullet 6H_2O$ 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 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.

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*

[+] These authors contributed equally to the work.

Table of contents

1. Materials S2

2. Spectroscopic measurements S3

3. Quantum chemical calculations S7

4. Molecular Dynamics simulations S15

4.1 Molecular dynamics simulations using the AMBER force field S15

4.2 Molecular dynamics simulations using the CHARMM force field S21

References S24

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. thG 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 thG, by normalizing the excitation spectra at wavelengths (> 350 nm) where only the red-shifted form absorbs.

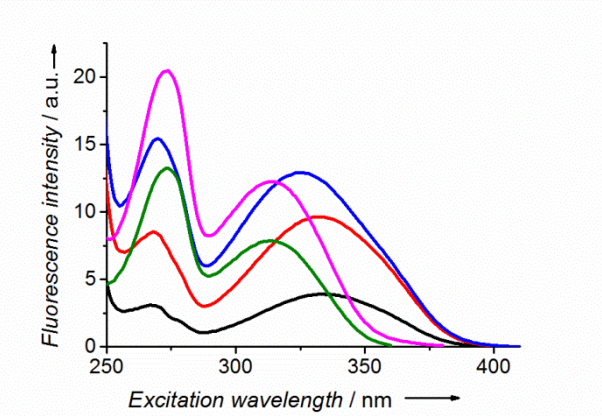


Figure S1. Excitation spectra of thG in TRIS-HCl buffer (25 mM, pH=7.5, 30 mM NaCl, 0.2 mM MgCl2) at different emission wavelengths: 550 nm (black line), 500 nm (red), 450 nm (blue), 400 nm (magenta), 375 nm (green). A maximum at 334 nm is observed when emission is recorded in the 500–550 nm range, and the excitation maximum is progressively blue-shifted down to 313 nm, for shorter wavelength emission.

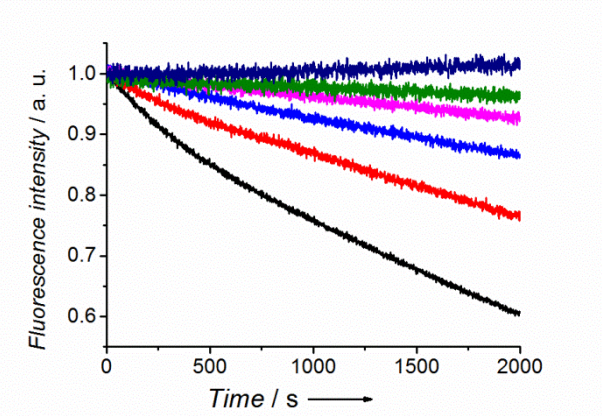


Figure S2. Photostability of the thG ground-state tautomers in buffer. Kinetics at different excitation and emission wavelengths, respectively: 325 and 400 nm (black); 325 and 525 nm (red); 350 and 525 nm (blue); 360 and 525 nm (magenta); 370 and 525 nm (green); 380 and 525 nm (dark blue). Concentration 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.

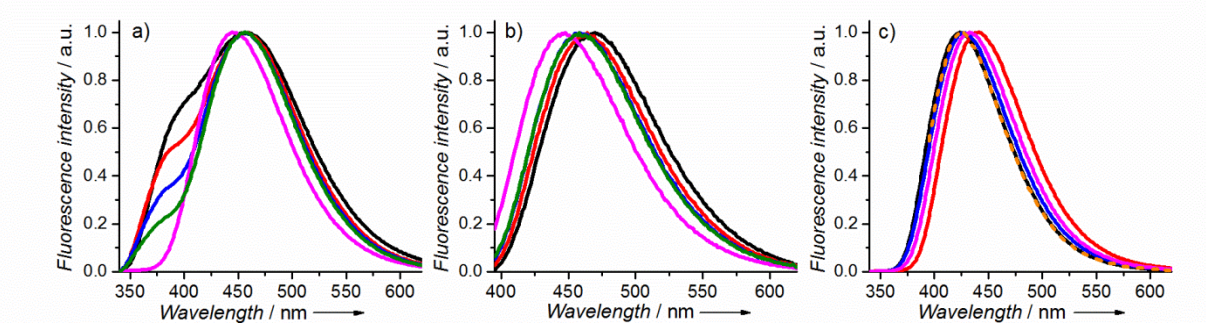


Figure S3. Normalized emission spectra of thG in various solvents. (a) Buffer (black), methanol (red), ethanol (blue), *n*-butanol (green), 2-methyl-2-butanol (magenta), λ_{ex} = 325 nm; (b) Emission spectra in the same solvents as in (a), but with λ_{ex} = 380 nm; (c) Emission spectra in 1,4-dioxane (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).

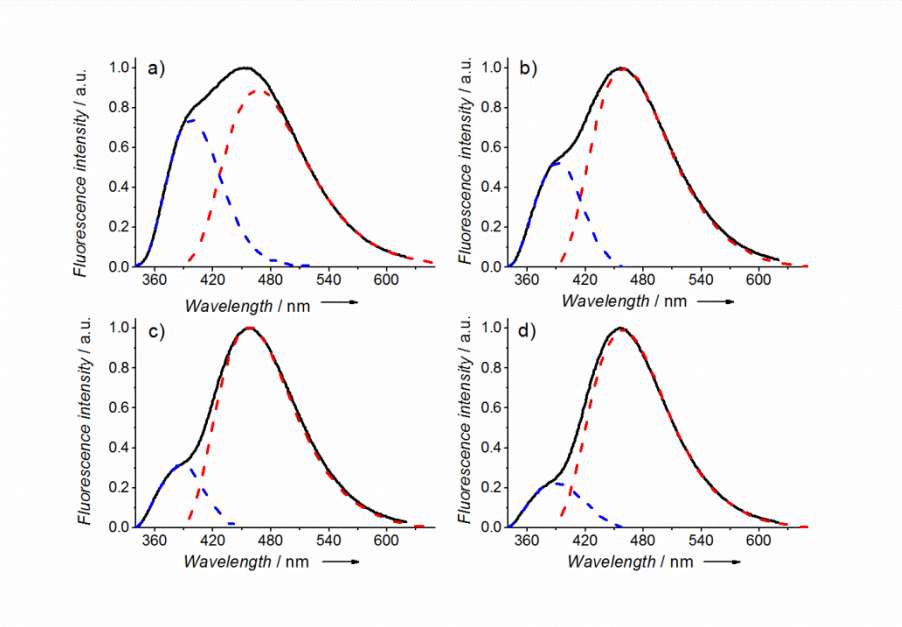


Figure S4. Deconvolution of the emission spectra of thG nucleoside in buffer (a), methanol (b), ethanol (c), and *n*-butanol (d). Excitation wavelength was 325 nm.

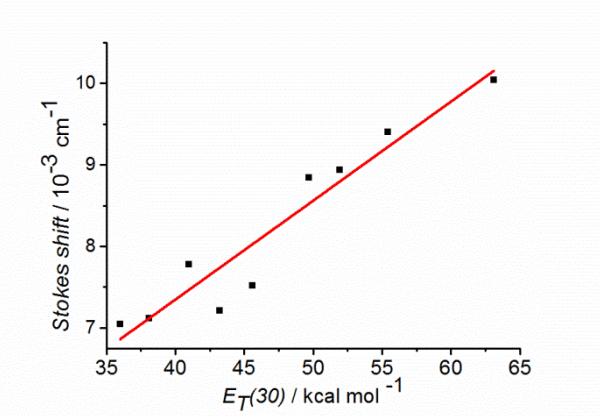


Figure S5. Dependence of the Stokes shift of the red-shifted tautomer on the empiric polarity index $E_T(30)$ (symbols). The red line represents the linear fit to the data with a slope of 0.12 ± 0.01 and an intercept of 2.4 ± 0.6, R^2 = 0.90. The $E_T(30)$ values of the solvents are given in

Table S1. The Stokes shift was calculated from the absorption and emission maxima, after correction of the emission spectra according to: Intensity ($\bar{\upsilon}$) = Intensity (λ) × λ^2.[6]

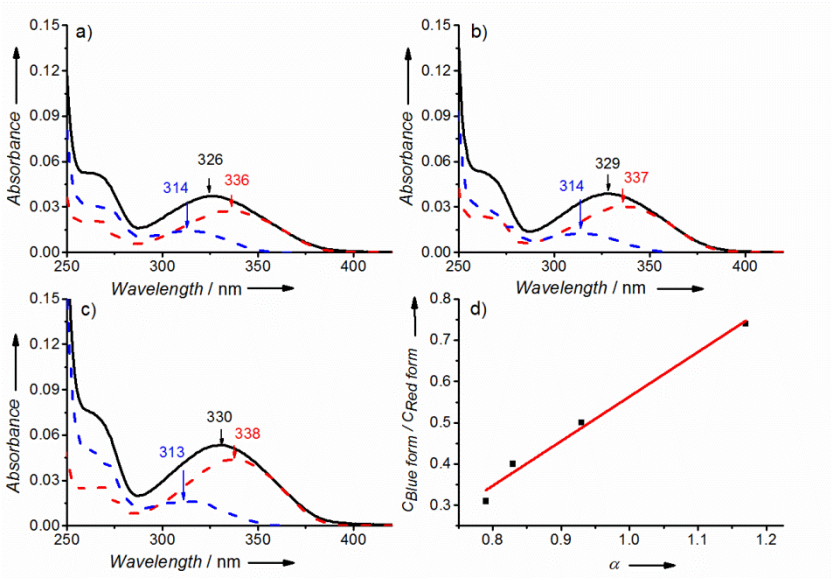


Figure S6. Deconvolution of the absorption spectra of thG in methanol (a), ethanol (b) and *n*-butanol (c) (black line) in its two ground state forms. Absorption spectra of the blue-shifted (blue dashed line) and red-shifted (red dashed line) forms. (d) Dependence of the concentration ratio of the two 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 νA and emission energies νE 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_2O (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.

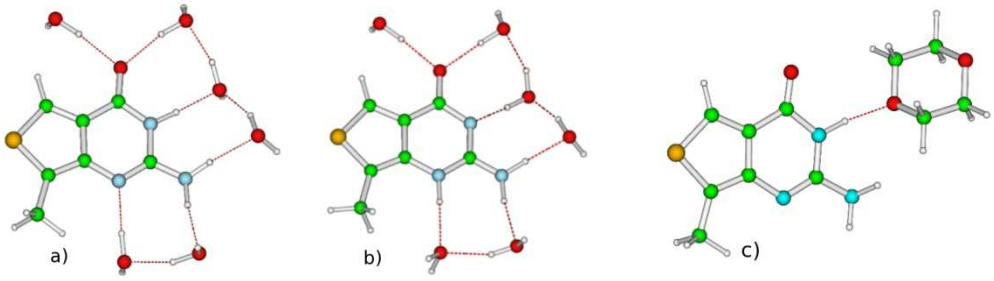


Figure S7. Computational models used to study the absorption and emission spectrum of the thG-H1 and thG-H3 tautomers in water (with six water molecules of the first solvation shell) (a and b, respectively) and thG-H1 in dioxane, with one explicit dioxane molecule (c).

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 dthG-H1 and dthG-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 dthG in a DNA environment, such as in (–)PBS.

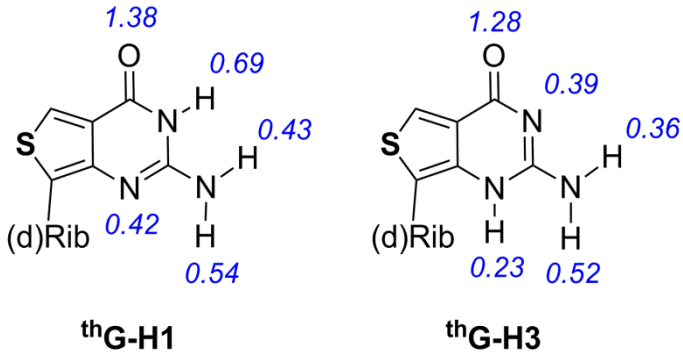


Figure S8. Average number of H-bonds calculated along MD trajectories for dthG-H1 and dthG-H3 in (–)PBS.

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 thG-H1 and thG-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]

thG-H1 tautomer				thG-H3 tautomer			
PBE0		M052X		PBE0		M052X	
Abs	Em	Abs	Em	Abs	Em	Abs	Em
Gas phase							
317 (0.08)	406 (0.06)	288 (0.11)	368 (0.08)	286 (0.11)	340 (0.08) [b]	261 (0.15)	330 (0.09) [b]
Dioxane (PCM)							
325 (0.10)	416 (0.08)	292 (0.14)	376 (0.11)	292 (0.14)	352 (0.11)	267 (0.19)	323 (0.16)
Dioxane (PCM + 1 Dioxane)							
330 (0.11)	418 (0.08)	295 (0.15)	378 (0.11)	293 (0.13)	353 (0.12)	269 (0.20)	332 (0.14)
Water (PCM)							
331 (0.09)	422 (0.08)	296 (0.16)	381 (0.11)	299 (0.12)	362 (0.10)	273 (0.16)	333 (0.13)
Water (PCM + $6H_2O$)							
350 (0.08)	448 (0.07)	308 (0.12)	402 (0.10)	309 (0.10)	383 (0.08)	279 (0.14)	349 (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.

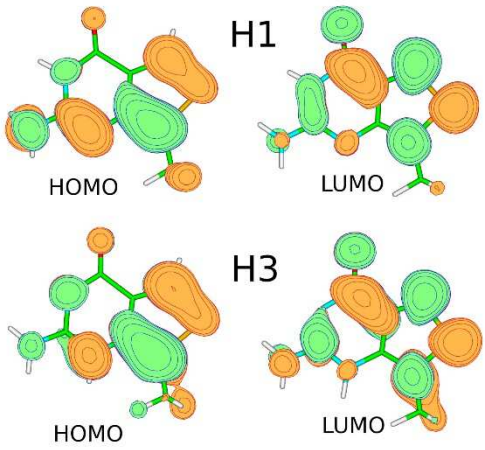


Figure S9. Frontier orbitals involved in the $S_0 \rightarrow S_1$ transition of the thG-H1 and thG-H3 tautomers.

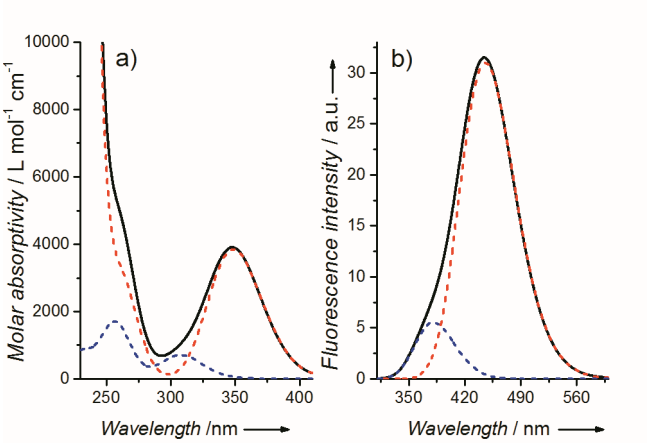


Figure S10. Absorption (a) and emission (b) spectra of the thG-H1 (red) and thG-H3 (blue) tautomers. The individual spectra and their sum weighted for their molar fraction were computed at PCM/TD-PBE0/6-311+G(2d,2p)/PCM/ PBE0/6-31G(d) level. The transitions have been convoluted by a phenomenological Gaussian with a HWHM of 0.25 eV.

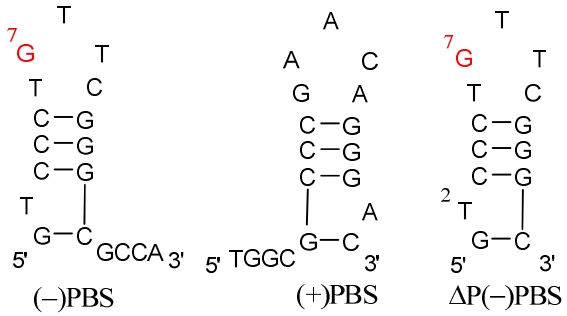


Figure S11. Structures of the primer binding site (PBS) oligonucleotides used in this study.

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 d^{th}G-H1 or d^{th}G-H3.[18] For both d^{th}G 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] 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]

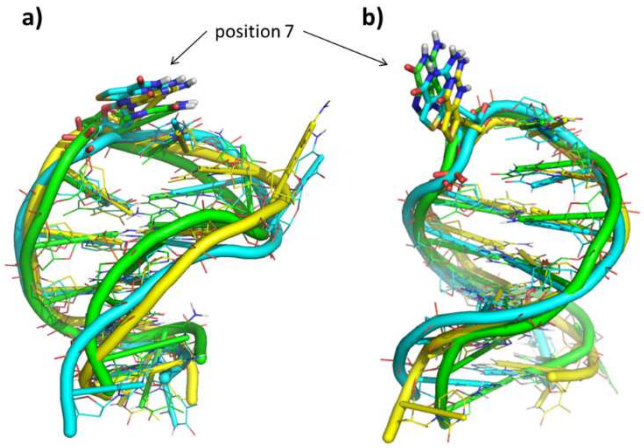


Figure S12. Superimposition of representative structures of the most populated clusters extrapolated from MD trajectories of ΔP(–)PBS single-stranded DNA. Representative structures from MD simulations of structure #1 (a) and structure #10 (b). The nucleotide at position 7 highlighted by a black arrow 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 d^{th}G-H1 or d^{th}G-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 d^{th}G-bearing systems, thus suggesting that dG replacement with d^{th}G-H1 or d^{th}G-H3 in single stranded ΔP(–)PBS does not alter the conformational preferences of ΔP(–)PBS.

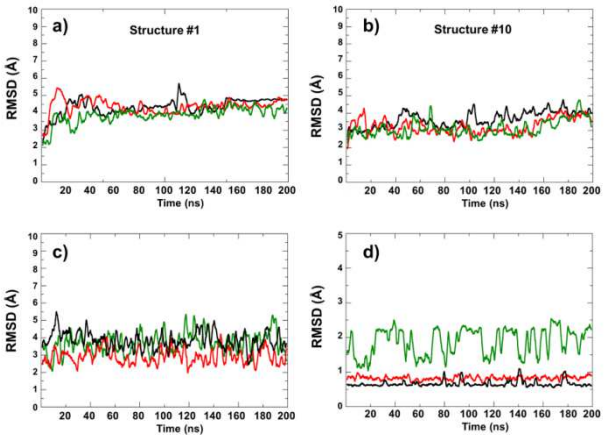


Figure S13. RMSD variation of (–)PBS and (–)/(+)PBS duplex. (a-b) RMSD variation of the single-stranded (–)PBS DNA structure #1 (a) and structure #10 (b) along 200 ns of unbiased MD trajectories. In black: unmodified (–)PBS having dG at position 7; in red: d^{th}G7(–)PBS bearing d^{th}G-H1 tautomer at position 7; in green: d^{th}G7(–)PBS bearing d^{th}G-H3 tautomer at position 7. (c-d) RMSD variation of the whole (–)/(+)PBS DNA duplex (c) and the dG (or d^{th}G)-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.

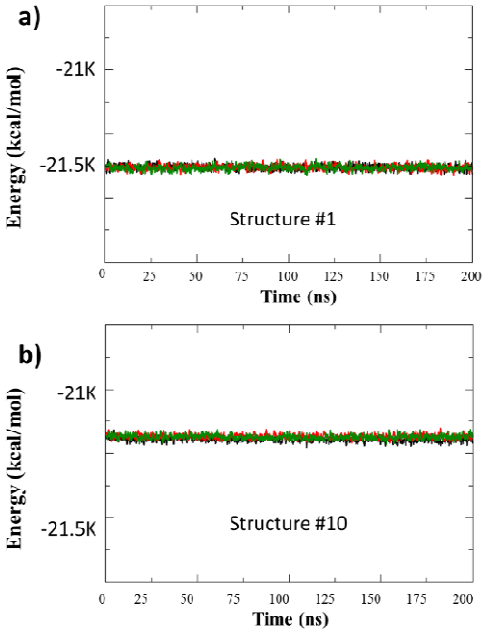


Figure S14. Total energy of the NMR-based structure #1 (a) and structure #10 (b) of ΔP(–)PBS DNA along 200 ns of unbiased MD simulation, bearing dG (black line), d^{th}G-H1 tautomer (red line) or d^{th}G-H3 tautomer (green line) at position 7.

MD simulations of (–)/(+)PBS DNA duplex. In the (–)/(+)PBS DNA duplex, the dG base at position 7 was kept unchanged or was replaced with d^{th}G-H1 or d^{th}G-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 d^{th}G-H1 or d^{th}G-H3 does not impact the overall geometry of the DNA duplex (Figure S13c). To further investigate the effects of replacing dG with d^{th}G 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 d^{th}G-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 d^{th}G-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.

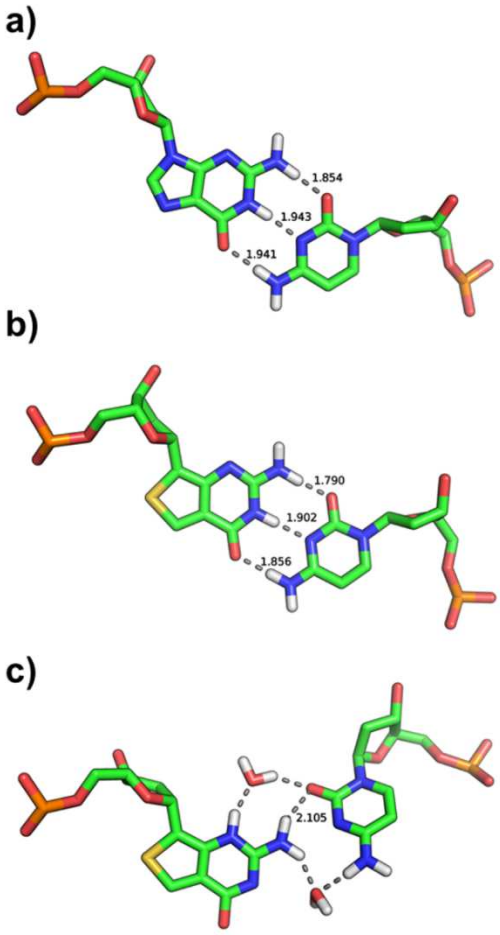


Figure S15. Structure of the base pair involving dG (a), dthG-H1 tautomer (b), and dthG-H3 tautomer (c) in the (–)/(+)PBS DNA duplex. Frames corresponding to the most representative structures of unbiased MD trajectories are shown. H-bond interactions are shown as grey dashed lines. Distances for direct and water-bridged H-bonds between bases are reported (in Å).

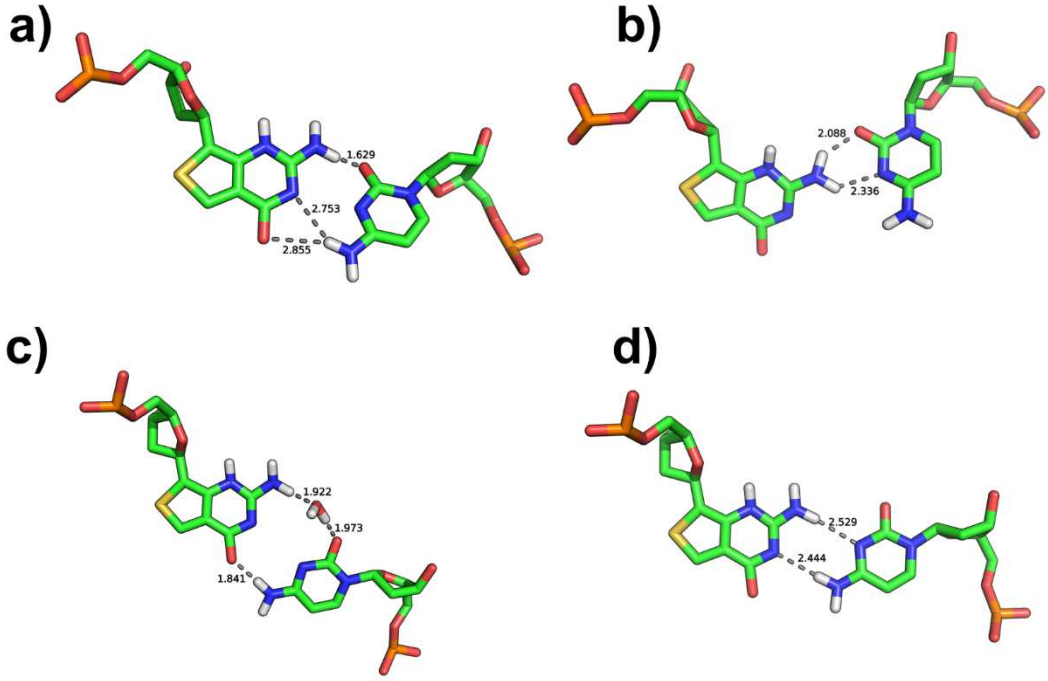


Figure S16. Four clusters of the dthG-H3 tautomer in various base-pair conformations with the opposite dC nucleobase. The most populated and stable cluster is shown in Figure S15c. H-bond interactions are showed as grey dashed lines. The distances of direct and water-bridged H-bonds between dthG and 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 d^{th}G-H3 can establish H-bond interactions with the phosphate group, which likely enhances the abundance of the *syn* conformation with respect to dG, and d^{th}G-H1.

Table S6. Percentage of the syn or anti conformation of dG, d^{th}G-H1, or d^{th}G-H3 in MD simulations of (–)/(+)PBS and (–)PBS

	(–)/(+)PBS		(–)PBS	
	anti (%)	*syn* (%)	*anti* (%)	*syn* (%)
dG	92.5	7.5	89.4	10.6
d^{th}G-H1	94.3	5.7	91.3	8.7
d^{th}G-H3	87.5	12.5	63.2	36.8

Thus, our data strongly suggest that the *anti* conformation is preferred for d^{th}G-H1 and d^{th}G-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 d^{th}G-H1 or d^{th}G-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 d^{th}G 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 d^{th}G-H1- and d^{th}G-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, d^{th}G-H1 or d^{th}G-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 d^{th}G-H3 tautomer which contributed −1.1 ± 0.6 kcal mol^{-1}. This suggests that the alternative d^{th}G-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 d^{th}G-H1 tautomer mimics very well the dG base in both single-stranded and double-stranded DNA. The d^{th}G-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 d^{th}G-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.

3.2.3. Monitoring the dynamics and conformations of the G residue in (–)PBS stem-loop and (–)/(+)PBS duplex substituted by thienodeoxyguanosine and their side-by-side comparison with 2-aminopurine

Since the red-shifted thG-H1 tautomer can be excited and detected selectively, we used its signal to perform a side-by-side comparison with the 2-aminopurine (d2Ap) nucleoside analog (Figure 1, publication 4). For this, we substituted the G7 of (–)PBS with either dthG or d2Ap (Figure 2, publication 4), and analyzed the photophysical and biophysical features of both the labeled (–)PBS stem-loop and (–)/(+)PBS duplex. As a first step, thermal denaturation experiments indicated that dthG minimally impacts the melting temperatures of the (–)PBS stem-loop and the corresponding (–)/(+)PBS duplex (Table 1, publication 4). Additionally, the dthG-dT mismatch resulting from the C12→T12 substitution in (+)PBS was found to decrease the stability of the dthG7(–)/T12(+)PBS duplex to the same extent as for the corresponding dG-dT mismatch. Interestingly, the substitution by d2Ap only slightly affected the stability of the (–)PBS stem-loop but decreased the melting temperature of the duplexes, likely due to the formation of an unstable d2Ap-dC mismatch. Remarkably, the "perfect duplex" d2Ap7(–)/T12(+)PBS was found to be 5 °C less stable than the native or the dthG7(–)/(+)PBS duplex, likely due to the lower stability of the dA-dT base pair, as compared to dG-dC. This indicates that dthG base pairs as efficiently as dG with the opposite dC and thus, appears as a much better substitute of dG than d2Ap.

Next, the photophysical properties of dthG7(–)PBS and dthG7(–)/(+)PBS were compared to the corresponding d2Ap-labeled sequences (Figure 3, Table 1, publication 4). For the d2Ap-substituted duplex, to avoid any artifact resulting from the d2Ap-dC mismatch, we used the more stable d2Ap7(–)/T12(+)PBS duplex for comparison with the dthG7(–)/(+)PBS duplex. Though the QY of dthG was significantly lower than that of the d2Ap, when inserted in the (–)PBS loop both labels exhibited similar QY (0.10 as compared to 0.08 for dthG and d2Ap, respectively). Noticeably, the QY of the dthG7(–)/(+)PBS duplex was 2-fold higher than that of the dthG7(–)PBS stem-loop and exceeded by more than one order of magnitude that of d2Ap7(–)/T12(+)PBS. This major difference indicates that in contrast to d2Ap, dthG is not efficiently quenched by its immediate neighbors in the double strand. This constitutes an obvious asset over d2Ap for monitoring the single to double strand transition and characterizing the dynamic properties of the substituted base. Our data clearly confirm that

$d^{th}G$ is not only a better structural substitute of dG residues than d2Ap, but also it is endowed with highly improved fluorescence properties, especially in DNA duplexes.

To further explore the photophysics of $d^{th}G$, time-resolved experiments were performed (Table 1, publication 4). The two emissive nucleosides exhibited a single exponential decay, while the corresponding modified ODNs showed a more complex behavior. Four decay components were observed for d2Ap(–)PBS indicating a large conformational heterogeneity of d2Ap at position 7. The three short lifetimes likely reflect conformations leading to collisional stacking of d2Ap with its neighbors, while the long-lived lifetime represents an extrahelical conformation of d2Ap, distant from the neighbor nucleobases. Moreover, comparison of the mean lifetime and QY of d2Ap7(–)PBS with the d2Ap ones indicated the presence of "dark species" (48%) where the emission of d2Ap is quenched by either static quenching or very fast dynamic quenching. The intensity decay of $d^{th}G$7(–)PBS was fitted with three components, where the long-lived lifetime reflects a minimally quenched $d^{th}G$ in (–)PBS loop, and two others represent dynamically quenched populations of $d^{th}G$ by its neighboring bases. In contrast to d2Ap, the "dark species" of $d^{th}G$ in (–)PBS was found to be negligible (<10%), which represents a distinctive advantage over d2Ap, since all $d^{th}G$ conformations in (–)PBS can be monitored by time-resolved measurements.

Even more remarkable differences between d2Ap and $d^{th}G$ were observed in the (–)/(+)PBS duplex, which may be rationalized by the differences in the ability of $d^{th}G$ and d2Ap to substitute their respective natural nucleoside. A total of 98% of the species in d2Ap-labeled duplex had lifetimes < 0.5 ns, explaining its very low QY. These commonly observed features confirm that d2Ap is a strongly limited fluorescent reporter in double strands. Indeed, due to the difference in the amino group position between dA and d2Ap, one of the Watson-Crick H-bond with the opposite T is moved from the major to the minor groove. This was shown to alter the dynamics and lower the stability of the duplex, by considerably reducing the lifetime of the d2Ap-T base pair and its immediate adjacent base pairs as compared to the corresponding duplex with the canonical dA-dT base pair. In line with a key role of conformational motions of DNA bases for the charge transfer based quenching mechanism for d2Ap, this increased dynamics of the d2Ap-dT base pair and its adjacent base pairs obviously favor efficient quenching of d2Ap, explaining the multiple and mainly short-lived fluorescence lifetimes observed for d2Ap in duplexes. In contrast to the highly complex decay of d2Ap7(–)/T12(+)PBS, the $d^{th}G$7(–)/(+)PBS decay was characterized by only two lifetimes

and marginal amounts of dark species. This indicates that $d^{th}G$ adopts a more limited amount of conformations, likely as a result of its ability to form a stable Watson-Crick base pair with dC. To further support our interpretation, we compared the decays of the matched $d^{th}G7(–)/(+)PBS$ with the mismatched $d^{th}G7(–)/T12(+)PBS$ that reveals higher conformational heterogeneity of $d^{th}G$ in the mismatched duplex, likely resulting from the reduced constraints imposed by the $d^{th}G$-dT mismatch as compared to the Watson-Crick $d^{th}G$-dT base pair.

To complete the comparison of $d^{th}G$ and d2Ap, we performed time-resolved anisotropy measurements 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, figure S7, publication 4). Two correlation times were observed for d2Ap7(–)PBS: the short one was assigned to the local motion of the d2Ap associated with the long-lived τ_4 lifetime, while the longer correlation time may correspond to a combination of the (–)PBS tumbling and segmental motion, likely associated to the loop. The local motion confirms that the unstacked d2Ap conformation likely protrudes into the solution, so that it can freely rotate. Interestingly, the values of the fast correlation time and its amplitude at position 7 were close to those observed when d2Ap was at position 6 or 8 of (–)PBS loop (Godet et al, 2011), suggesting that these three residues experience similar local mobility. These results were further substantiated by the very high bimolecular quenching constant k_q, observed for d2Ap7(–)PBS in iodide quenching experiments (Table 2, publication 4). The calculated k_q value was only 2-fold lower than that of the free d2Ap, confirming that its extrahelical conformation is highly accessible to the solvent. The anisotropy decay of $d^{th}G7(–)PBS$ was fitted to a single component that perfectly matches the theoretical correlation time of the folded ODN, indicating that the major conformation of $d^{th}G$ is rigidly held in the (–)PBS loop and only the tumbling of the whole ODN is monitored. This behavior is fully consistent with the NMR structure of (–)PBS, showing that the dG7 residue is directed toward the loop interior and well constrained by the close proximity of the T6, T8 and T9 neighbors. The internal orientation of $d^{th}G$ with poor solvent accessibility was also supported by the low k_q value observed for $d^{th}G7(–)PBS$, that was more than one order of magnitude lower than that of the free nucleoside. The correlation time of d2Ap7(–)/T12(+)PBS was found to be much shorter than the calculated theoretical time for the tumbling rotation of this duplex, and thus, likely

represents a combination of the tumbling motion of the duplex and segmental dynamics of its backbone. In fact, the emission of d2Ap7(–)/T12(+)PBS is largely dominated by the τ_2 and τ_4 lifetimes that correspond to partially stacked conformations of d2Ap. As a result of this partial stacking, the corresponding duplexes may be less stable, so that additional dynamics could be perceived, explaining thus a decrease in the value of the correlation time. The anisotropy decay of the d^{th}G7(–)/(+)PBS was also characterized by a single correlation time, that matched well with the theoretical correlation time in contrast to the d2Ap-labeled duplex. This confirms the attribution of the major long-lived lifetime component of d^{th}G to the d^{th}G-dC base pair in its optimally stacked configuration, where only the tumbling motion could be detected. In contrast, the corresponding configuration with highly stacked d2Ap-dT base pairs is thought to be non-emissive in d2Ap7(–)/T12(+)PBS, so that no information could be obtained on its rotational dynamics. Interestingly, a single correlation time was also observed for d^{th}G7(–)/T12(+)PBS describing the overall tumbling of the duplex. The absence of a short rotational correlation time indicated that the d^{th}G conformation associated with 28.2 ns lifetime component dominant in the duplex, does not exhibit any local motion and is thus, probably not extrahelical. For both d2Ap- and d^{th}G-labeled duplexes the k_q values were at least 3 orders of magnitude below those of the free nucleosides, suggesting that they predominantly adopt an intrahelical configuration.

Taken together, our results clearly illustrate that d^{th}G can faithfully substitute the critical G7 residue in (–)PBS loop and reflect its predominant conformation determined by other means, such as NMR. The isomorphic character of d^{th}G ensures a reliable base pairing with dC, while its high QYs in single- and double-stranded ODNs are a clear advantage over d2Ap-labeled ODNs in providing reliable information on dG conformations and dynamics in single- and double-stranded ODNs.

Based on the obtained information, we envision to follow the hybridization of the (–)PBS stem-loop with its complementary (+)PBS duplex in real-time in the absence and in the presence of NC(11-55) protein using the emission signal of d^{th}G in order to obtain the kinetic parameters of the annealing reaction and compare them with those obtained using the 3HCnt nucleobase. Preliminary kinetic experiments of d^{th}G7(–)PBS annealing with (+)PBS in the absence of NC show that we can follow the (–)/(+)PBS reaction in pseudo-first order conditions. The kinetic curves are different from those observed using 3HCnt-labeled (–)PBS since they show a rising and a declining component (Figure 3.2.1).

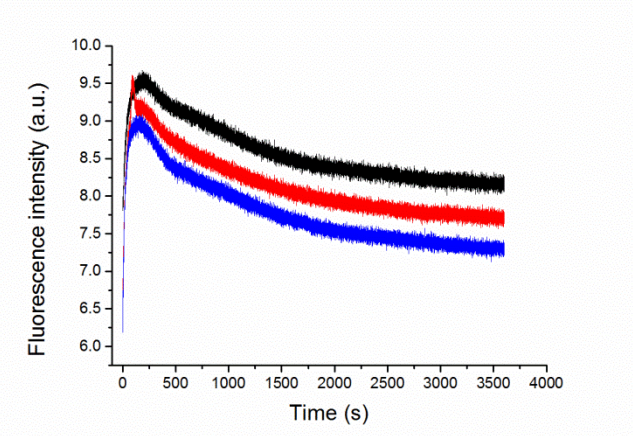


Figure 3.2.1. Kinetics of 300 nM dthG7(–)/(+)PBS with 3µM (black), 4µM (red) and 5µM (blue) (+)PBS. Excitation and emission wavelengths were 360 and 470 nm. Experiments were performed in 25 mM TRIS-HCl buffer (pH=7.5), 30 mM NaCl, 0.2 mM $MgCl_2$.

The kinetic traces were adequately fitted using a difference of two exponential functions giving two observed kinetic rate constants that are highly comparable to those obtained using the 3HCnt-labeled (–)PBS. These traces constitute the proof of principle experiments indicating that thG is well suited to monitor the annealing reactions and extract the kinetic constants. Of course, additional experiments will be needed to understand the observed changes in the fluorescence signal and to extract the full set of kinetic parameters both in the absence and the presence of NC. Comparison with the 3HCnt data will allow to confirm and/or complete the mechanism of the annealing reaction and provide information on the possible perturbation induced by the introduction of the probes.

Publication 4. Conquering 2-aminopurine's deficiencies: highly emissive isomorphic guanosine surrogate faithfully monitors guanosine conformation and dynamics in DNA

.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

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.

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.

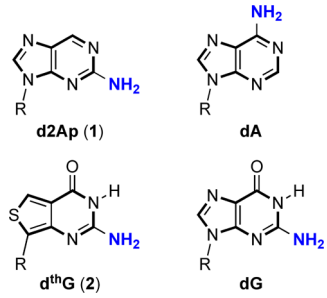


Figure 1. Structure of d2Ap (**1**), $d^{th}G$ (**2**), and the naturally occurring purines (R = 2′-deoxy-D-ribose).

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

Received: December 24, 2014
Published: February 25, 2015

DOI: 10.1021/ja513107r
J. Am. Chem. Soc. 2015, 137, 3185−3188

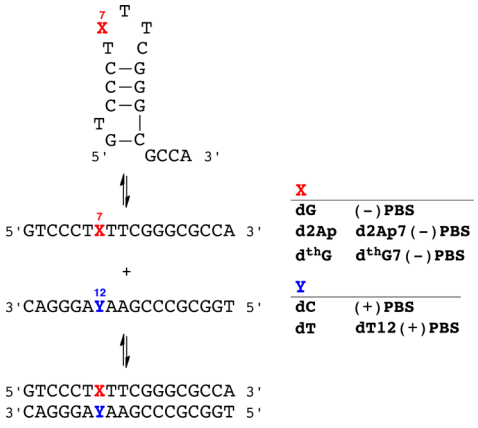


Figure 2. DNA sequence of the HIV-1 primer binding site (−)PBS shown as a single strand (middle), stem loop (top), and duplex with (+)PBS (bottom). Also shown are the site-specifically modified sequences containing d2Ap and dthG.

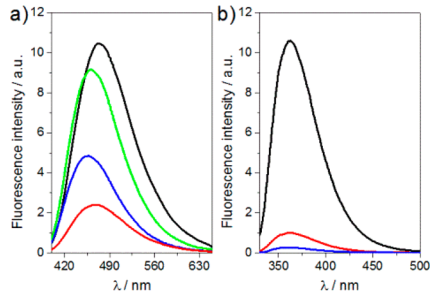


Figure 3. Emission spectra of (a) dthG- and (b) d2Ap-labeled (−)PBS sequences. Emission spectra of (a) dthG free nucleoside (black), dthG7(−)PBS (red), dthG7(−)/(+)PBS (blue), and dthG7(−)/T12(+)PBS (green); (b) d2Ap free nucleoside (black), d2Ap7(−)PBS (red), and d2Ap7(−)/T12(+)PBS (blue). Excitation 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 ± 1 and 51 ± 1 °C, respectively) (Table 1). Similarly, the identical melting temperature of the native and the dthG7(−)/(+)PBS duplexes (67 ± 1 and 67 ± 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 ΔT_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 ± 0.02 in buffer (Figure 3 and Table 1). When incorporated into position 7 in the (−)PBS loop, the QY drops to 0.10 ± 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 (τ_1–τ_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 (τ_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 ± 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 ± 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 ± 1	0.10			0.5	0.32	2.8	0.40	12.3	0.28	4.7	<0.1
dthG7(−)/(+)PBS	67 ± 2	0.20			1.1	0.17			11.3	0.83	9.6	<0.1
dthG7(−)/T12(+)PBS	61 ± 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.

DOI: 10.1021/ja513107r
J. Am. Chem. Soc. 2015, 137, 3185−3188

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(−)/

DOI: 10.1021/ja513107r
J. Am. Chem. Soc. 2015, 137, 3185−3188

T12(+)PBS. In contrast, the anisotropy decay of the dthG7(−)/(+)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 dthG-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 dthG7(−)/T12(+)PBS, indicating that the major dthG 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 dthG but not 2Ap can be used to obtain relevant information on the oligonucleotide dynamics and size. Noticeably, for both d2Ap7- and dthG7-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 dthG 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 dthG'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 dthG 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

We thank Ludovic Richert for his technical help. This work was supported by a fellowship from the Ministère de la Recherche (M.S.), the European Project THINPAD "Targeting the HIV-1 Nucleocapsid Protein to fight Antiretroviral Drug Resistance" (FP7 Grant Agreement 601969), Agence Nationale de la Recherche (ANR blanc Fluometadn), Agence Nationale de Recherche sur le SIDA, and French-Ukrainian Dnipro program. Support by the U.S. National Institutes of Health (grant GM069773 to Y.T.) is gratefully acknowledged.

■ 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.

DOI: 10.1021/ja513107r
J. Am. Chem. Soc. 2015, 137, 3185−3188

Supporting Information

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

Table of contents

1. Synthesis S2

1.1. $d^{th}G$ phosphoramidite synthesis S3

1.2. Synthesis of non-labeled and $d^{th}G$/d2Ap-labeled ODNs S7

2. UV/visible absorption and steady-state fluorescence measurements S8

3. Time-resolved fluorescence intensity decays S9

4. Time-resolved fluorescence anisotropy decays S10

5. Quenching measurements S13

References S14

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

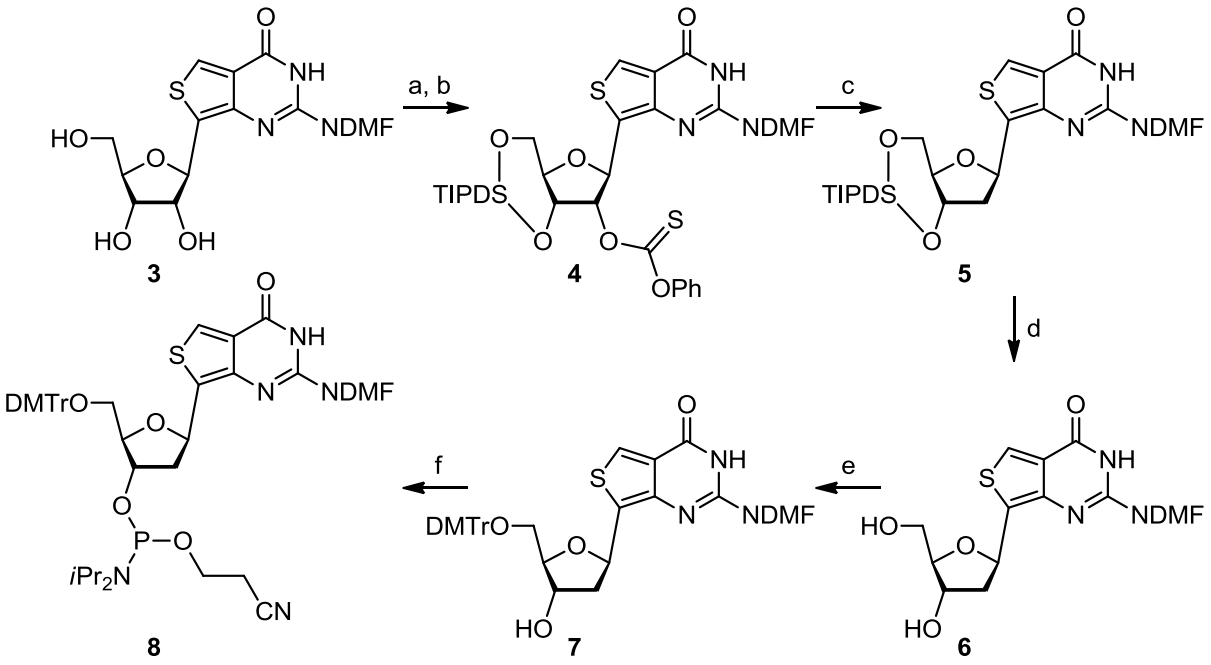


Figure S1. dthG-phosphoramidite synthesis[a]

[a] Reagents and conditions: (a) $TIPDSiCl_2$, Py, 89%; (b) PhOC(S)Cl, DMAP, Py, 78%; (c) Bu_3SnH, 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_2SO_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 $C_{26}H_{45}N_4O_6SSi_2$ $[M+H]^+$597.26, found 597.19; ESI-HRMS calculated for $C_{26}H_{45}N_4O_6SSi_2$ $[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_2SO_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 $C_{33}H_{49}N_4O_7S_2Si_2$ $[M+H]^+$733.26, found 733.22; ESI-HRMS calculated for $C_{33}H_{49}N_4O_7S_2Si_2$ $[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_3SnH (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_2SO_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 $C_{26}H_{45}N_4O_5SSi_2$ $[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 $C_{14}H_{19}N_4O_4S$ $[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 $C_{44}H_{54}N_6O_7PS$ $[M+H]^+$841.35, found 840.86; ESI-HRMS calculated for $C_{44}H_{54}N_6O_7PS$ $[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 filter right 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 C18 (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 mM NaCl, 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 $<\tau>$ 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:

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

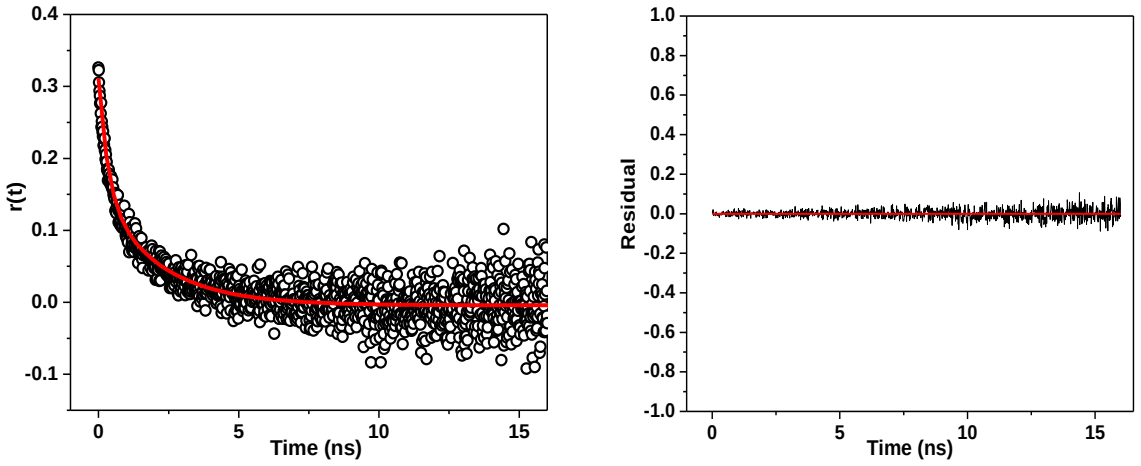


Figure S2. Anisotropy decay and corresponding residual plot for d2Ap7(–)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

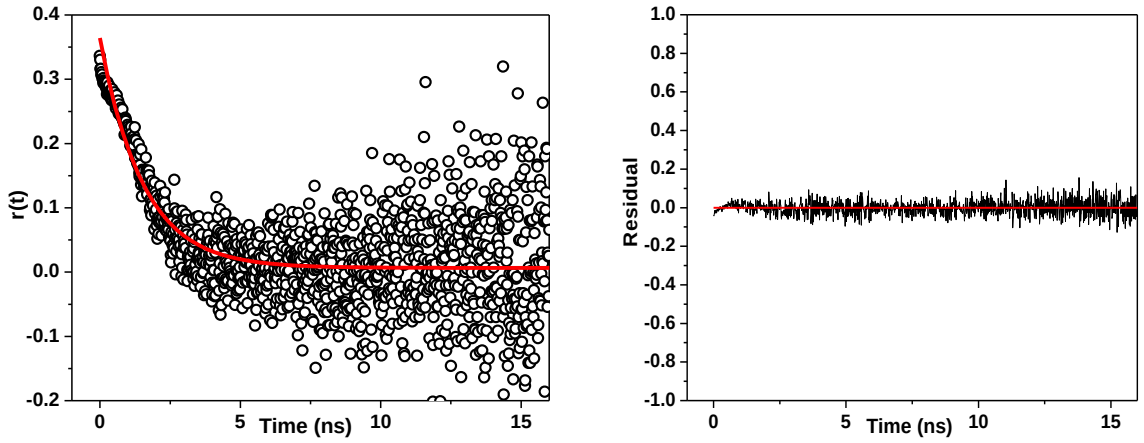


Figure S3. Anisotropy decay and corresponding residual plot for d2Ap7(–)/T12(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

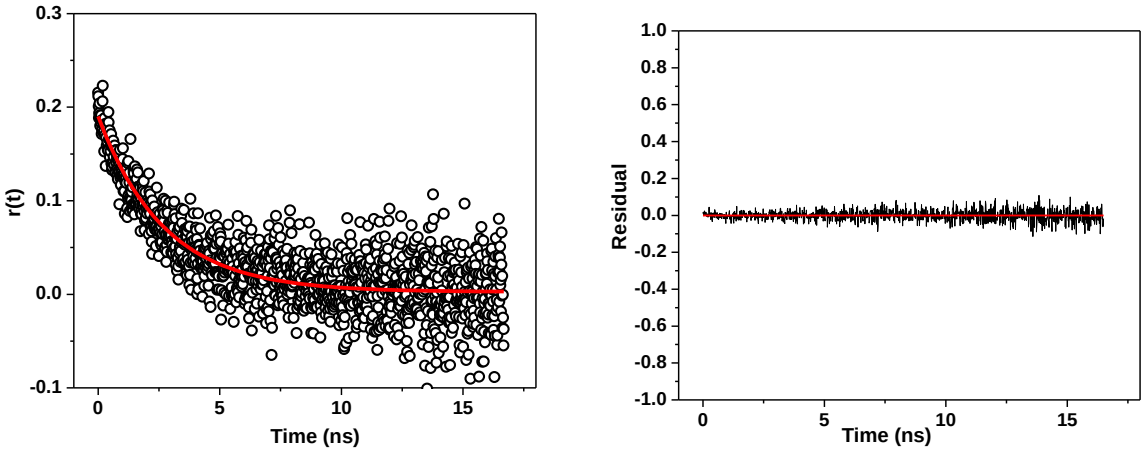


Figure S4. Anisotropy decay and corresponding residual plot for d^{th}G7(–)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

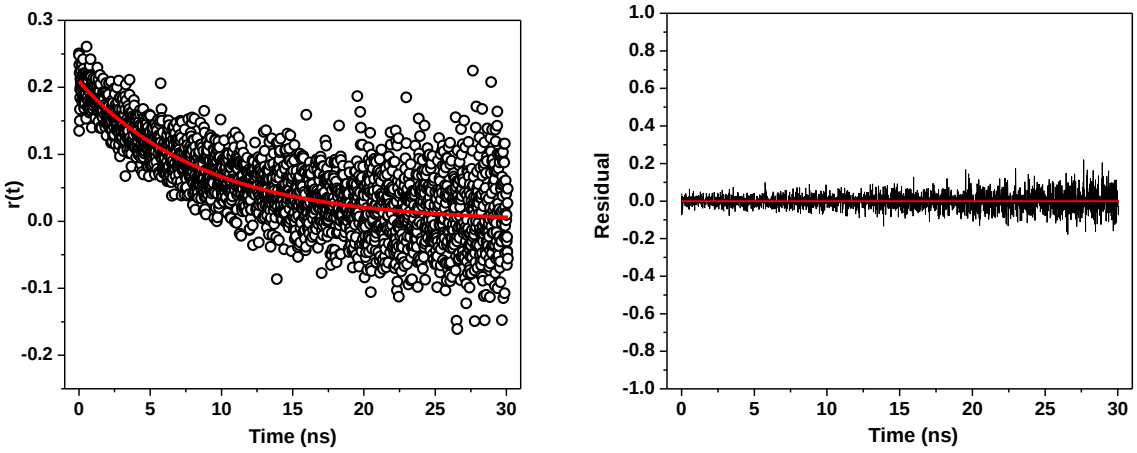


Figure S5. Anisotropy decay and corresponding residual plot for d^{th}G7(–)/(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

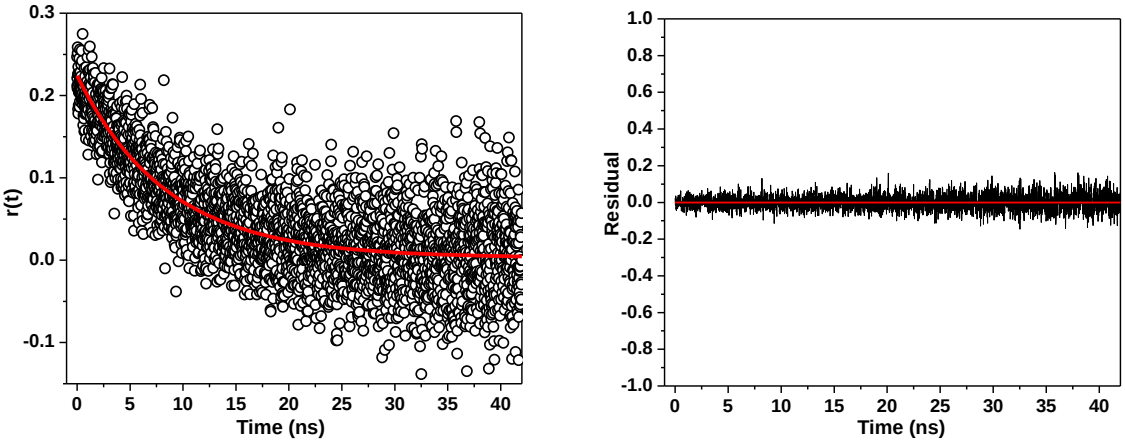


Figure S6. Anisotropy decay and corresponding residual plot for d^{th}G7(–)/T12(+)PBS.The continuous line in the left panel corresponds to the fit of the data with the parameters in Table 2. Excitation wavelength was at 315 nm.

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:

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

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.

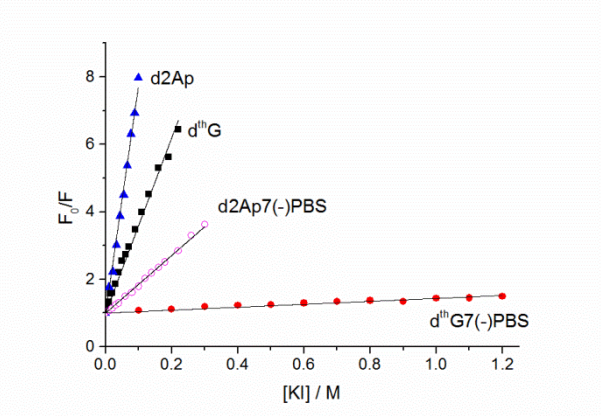


Figure S7. Stern-Volmer plots for KI quenching of dthG (black squares), dthG7(–)PBS (red disks), d2Ap (blue triangles), and d2Ap7(–)PBS (magenta circles). Black lines represent their linear fits.

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.

Chapter 4. Conclusions and Perspectives

The main objective of my research work was to characterize on the molecular level the dynamics of the interaction between the HIV-1 nucleocapsid protein and its target nucleic acids using innovative fluorescent probes. To reach this objective, either the NC protein or the nucleic acids were labeled to follow the interaction from the two sides. The first and very important task was to choose the appropriate fluorescent probes to substitute the pivotal residues of these biomolecules. Next, a detailed characterization of the probes in their free form and incorporated in NC or ODNs was performed in order to determine their photophysical properties and their sensitivity to peptide-ODN and ODN-ODN binding. Finally, we exploited these probes in order to obtain a full characterization of the NC-promoted (–)/(+)PBS annealing reaction.

To label the NC protein, a ratiometric probe of the family of 3-hydroxychromones in the form of an amino acid analog was chosen. This new amino acid showed a much stronger sensitivity of its dual emission to hydration than a previous analog and was found to constitute a nearly perfect Trp mimic. Thus, we were able to monitor the structural rearrangements of Trp37, a key residue of NC in interaction with ODNs. Both the decrease in water exposure and the decrease of the amino acid mobility are characteristic changes, when peptides interact with nucleic acids and lipid membranes. The two-photon fluorescence lifetime imaging of the labeled peptides microinjected into cells confirmed that NC(11-55) binds mainly to RNAs in the nucleoli and cytoplasm. Further kinetic studies of the labeled NC(11-55) – ODN interaction showed that the destabilization of nucleic acids by NC is a quasi-instantaneous process, accompanied by non-specific binding and sliding of the protein to its preferential binding sites. Taken together, our data validate this new amino acid based on an ESIPT dye as an efficient tool for detection and characterization of peptide/ligand interactions in solution and in living cells. The M3HF-amino acid analog may be also highly useful in establishing sensitive screening assays to identify compounds able to inhibit the binding of the labeled peptides to their ligands.

The second approach to monitor the interactions site-selectively from the side of ODNs was based on the use of emissive nucleoside analogs. Several positions of the (–)PBS loop and notably the dG7 position were targeted in order to follow their conformational changes on interaction with NC and the complementary (+)PBS. For this, two types of fluorescent nucleosides have been applied: a 3-hydroxychromone derivative (3HCnt) and an isomorphic thienodeoxyguanosine analog ($d^{th}G$) and.

Using the 3HCnt nucleoside analog we were able to monitor the ODN/NC(11-55) interaction, as well as the formation of the double-stranded ODNs. The spectroscopic properties of 3HCnt nucleobase allowed using low concentrations to monitor the (–)/(+)PBS annealing reaction, so that fast kinetic events could be evidenced by the stopped-flow technique. Our data show that 3HCnt introduced at position 9 of the loop of (–)PBS allows for the first time to get the whole set of kinetic rate constants of the annealing reaction with (+)PBS. We evidenced for the first time the presence of loop-loop contacts in the absence of NC(11-55) and confirmed the existence of a kissing complex intermediate in the presence of the peptide. Several additional experiments are still needed, such as monitoring of the annealing reaction using truncated ΔP(–)PBS and ΔP(+)PBS sequences, lacking the single-stranded overhangs, in order to assess the rate constants of the annealing reaction when only the loop-loop contacts are possible. It is also desirable to compare the obtained rate constants of (–)/(+)PBS hybridization promoted by NC(11-55) with those in the presence of NC(1-55) or $(SSHS)_2$-NC mutant in which all cysteines residues are replaced by serines to prevent zinc binding, and thus the folding of the zinc fingers. This will provide information on the role of the N-terminal domain and the folding of the zinc fingers in the NC-promoted annealing reaction. Moreover, comparison of the kinetic rate constants of (–)/(+)PBS annealing obtained using 3HCnt nucleoside with those obtained by $d^{th}G$ will provide further confidence in the established mechanism and will show its independency on the applied fluorescent probe.

On the other hand, a detailed analysis of $d^{th}G$ as a free nucleoside or when incorporated into ODNs revealed the presence of two ground-state tautomers with significantly shifted absorption and emission spectra. Quantum mechanical calculations identified them as being the keto-amino tautomers, thG-H1 and thG-H3, differing by the position of the protonated nitrogen atom. Molecular dynamics studies further suggested that, from conformational and energetic standpoints, thG-H1 behaves similarly to its native counterpart in both the single- and double-stranded structures studied in this work, whereas the thG-H3 tautomer behaves comparably to G only in the loop of a stem-loop DNA. When incorporated into double-stranded sequences, the unfavorable conformational and energetic instability of thG-H3 likely results in its tautomerization 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.

Further we investigated the ability of $d^{th}G$ to adequately substitute and monitor the conformations and dynamics of the dG7 residue in the (–)PBS stem-loop and the (–)/(+)PBS duplex. We found that $d^{th}G$ can faithfully substitute a key dG residue in (–)PBS, showing its dominant conformations revealed by NMR studies. Reliable base pairing with dC and its high fluorescence QY maintained in single- and double-stranded ODNs make $d^{th}G$ a highly improved alternative to d2Ap for substituting G residues in ODNs. $d^{th}G$ can thus be used for substituting d2Ap in numerous applications, such as monitoring RNA/DNA conformational dynamics and heterogeneity, ODN-protein interactions, dynamics,conformations of damaged ODN, mismatched base pair recognition or quadruplex formation. Due to its much higher quantum yield in double stranded DNAs as compared to d2Ap, $d^{th}G$ will be especially interesting for applications involving duplexes, since it will allow using much lower concentrations of labeled ODNs. Another promising application is the detection of base modification in mutated DNA, since replacement of the $d^{th}G$-dC base pair by a mismatch was found to shift the ratio of the two tautomers and induce large changes in the quantum yield of the red-shifted tautomer. Further work is in progress to assess this point. In the particular case of the (–)PBS sequence, due to the key role of the dG7 residue in the recognition and binding of NC protein, it will be possible with $d^{th}G7$(–)PBS to site-selectively monitor the conformational changes accompanying the binding of NC to this residue. Moreover, due to the strong differences in the spectroscopic properties of $d^{th}G7$(–)PBS and $d^{th}G7$(–)/(+)PBS, it will also be possible to monitor specifically at the level of the substituted dG residue, the NC-promoted annealing reaction with (+)PBS.

Résumé

Les progrès dans la thérapie anti-VIH ont énormément changé le cours de l'infection par le virus de l'immunodéficience humaine et donnent désormais la possibilité de traiter cette maladie grave comme une maladie chronique. Toutefois, en raison des mutations fréquentes de ce virus qui conduisent à des problèmes de résistance, la nécessité d'explorer de nouvelles cibles thérapeutiques s'impose. Du fait de sa haute conservation et de ses fonctions clés dans le virus VIH-1, la protéine de la nucléocapside NC est une cible de choix pour développer de nouveaux anti-viraux. Les fonctions de NC reposent principalement sur ses propriétés chaperonnes vis-à-vis des acides nucléiques (NAs), propriétés qui dirigent le réarrangement de ces derniers vers leur conformation la plus stable et favorisent l'hybridation de séquences complémentaires. Bien que la compréhension mécanistique des propriétés chaperonnes ait connu d'importants progrès, leurs aspects dynamiques restent mal compris, faute notamment d'outils appropriés pour leur suivi au niveau moléculaire.

Dans ce contexte, l'objectif de ce travail de thèse a été de caractériser au niveau moléculaire la dynamique des interactions de NC avec les NAs, en utilisant des outils fluorescents développés au laboratoire ou en collaborations. Nous avons utilisé des analogues fluorescents d'acides aminés introduits dans la protéine NC, ainsi que des analogues fluorescents de nucléosides introduits dans des oligonucléotides, afin de suivre l'interaction des deux côtés.

Partie 1. Interaction de NC avec ses partenaires moléculaires en utilisant des analogues fluorescents d'acides aminés basés sur les 3-hydroxychromones.

Pour marquer la NC, une sonde ratiométrique à émission duale de la famille des 3-hydroxychromones (3HC) sous la forme d'un acide aminé (M3HFaa) a été synthétisée au laboratoire (en collaboration avec le groupe de Vasyl Pivovarenko) et incorporée par synthèse peptidique à la place des résidus Trp37 dans le second doigt de zinc (W37-M3HFaa) et Ala30 dans le *linker* entre les deux doigts de zinc (NA-M3HFaa) (Figure R1). Le peptide W37-M3HFaa a montré une activité chaperonne parfaitement conservée, pratiquement identique à celle du peptide non marqué NC(11-55), ce qui indique que M3HFaa ne modifie pas la conformation du peptide et se substitue parfaitement au résidu Trp37, qui joue un rôle primordial dans la liaison de la protéine avec les NAs. Par contre, M3HFaa en position A30 de NC(11-55) a montré une possible perturbation du *linker* pouvant influencer l'orientation relative des doigts de zinc et la formation du plateau hydrophobe.

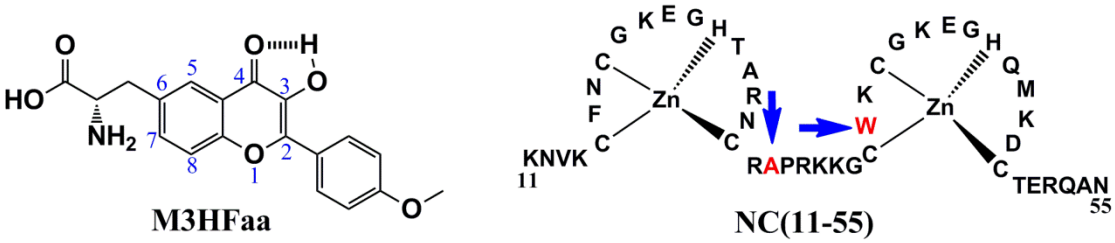


Figure R1. Analogue fluorescent d'acide aminé utilisé dans cette étude et positions de marquage de la protéine NC(11-55).

L'émission duale de M3HFaa est due à une réaction de transfert de proton à l'état excité qui génère un équilibre entre une forme normale (N*) et une forme tautomère (T*), dont les spectres d'émission sont décalés. Le rapport d'intensité N*/T* de M3HFaa est sensible à l'hydratation du microenvironnement de la sonde et indique que M3HFaa est protégé de l'eau une fois incorporé dans W37-M3HFaa alors qu'il y est davantage exposé dans le peptide A30-M3HFaa. Le rendement quantique de fluorescence de W37-M3HFaa est environ 2 fois supérieur à celui de A30-M3HFaa.

L'insertion de la sonde M3HFaa dans la NC(11-55) a permis d'obtenir des informations site-spécifiques sur la liaison NC/NAs, notamment concernant les conformations de la NC dans les complexes avec les ADNs ou ARNs. Nous avons observé que les spectres d'émission des peptides marqués présentent des différences déterminées par l'accessibilité de la sonde au solvant et le mode de liaison avec les ARNs/ADNs. La comparaison des rendements quantiques et des rapports N*/T* pour W37-M3HFaa lié à l'ARN SL2 ou à l'ADN ΔP(–)PBS suggère une interaction par empilement de M3HFaa avec les nucléobases des deux complexes, ce qui est corrélé avec les structures 3D des complexes résolues par RMN. En ce qui concerne A30-M3HFaa, les modifications du rendement quantique et du rapport N*/T* ont été interprétées par des distances différentes entre le résidu A30 et les nucléobases voisines, démontrées par RMN pour les complexes de NC(11-55) avec SL2 et ΔP(–)PBS. Nos résultats indiquent donc que M3HFaa se substitute convenablement au résidu Trp et préserve le mode d'interaction du peptide avec ses séquences cibles des NAs. Les interactions des peptides marqués avec l'ADN double-brin de thymus de veau et la séquence ARN TAR ont montré que M3HFaa en position 37 s'empile avec les nucléobases de manière identique pour tous les complexes, alors qu'en position 30 les interactions apparaissent moins bien conservées. Ceci suggère que NC se fixe différemment à l'ADN de thymus de veau et aux sites de liaison non-spécifiques de TAR en comparaison avec les sites de liaison spécifiques

de SL2 ou de ΔP(–)PBS. Enfin, la liaison de NC(11-55) marqué au M3HFaa avec des vésicules lipidiques DOPC/DOPS a également été explorée et a montré que les deux peptides marqués détectent un environnement comparable, pauvrement hydraté, du fait de leur insertion dans la bicouche lipidique significativement moins polaire et accessible à l'eau que les NAs.

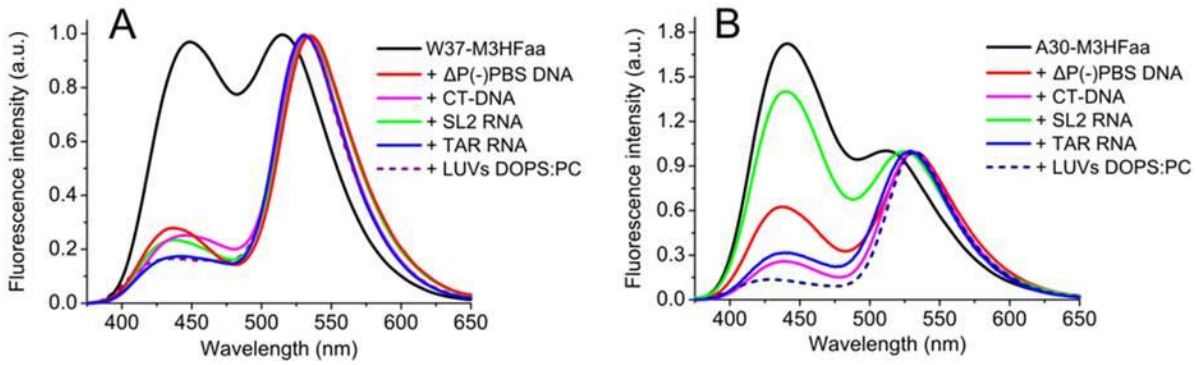


Figure R2. Effet de l'interaction d'un oligonucléotide sur les spectres de fluorescence de peptides NC(11-55) marqués au M3HFaa. Spectres de fluorescence normalisés pour les peptides W37-M3HFaa (A) et A30-M3HFaa (B).

L'interaction des peptides marqués avec des NAs induit notamment un fort changement dans leur émission duale en favorisant l'émission de la forme tautomère T*.Ce changement a été associé à l'apparition de temps de vie longs dans le tautomère, du fait de l'environnement rigide créé dans le complexe, qui limite la rotation des deux groupements aromatiques de M3HFaa. La restriction la plus forte et donc les plus longues durées de vie de fluorescence ont été observées pour la position 37, où la sonde s'empile probablement avec les bases nucléiques. Basée sur la dépendance des valeurs de durée de vie en fonction de la nature du ligand et de la position marquée, l'imagerie en temps de vie de fluorescence a été utilisée pour identifier les partenaires de liaison des peptides marqués micro-injectés dans des cellules vivantes (Figure R3). La microscopie a montré que les peptides se lient principalement aux ARNs dans le cytoplasme et les nucléoles, en accord avec d'autres études faites au laboratoire.

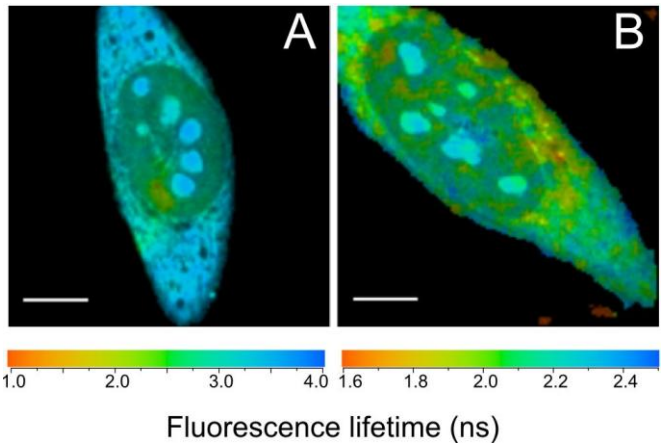


Figure R3. Imagerie en temps de vie de fluorescence (FLIM) de cellules HeLa micro-injectées avec 100 µM de peptides W37-M3HFaa (A) et A30-M3HFaa (B) après 10 min d'incubation. La couleur de chaque pixel représente la valeur du temps de vie, alors que son intensité correspond à l'intensité totale de fluorescence. L'échelle correspond à 10 µm.

Lors de l'étape suivante, nous avons suivi la cinétique du réarrangement structurel de NW-M3HFaa induit par la liaison de l'ADN ΔP(–)PBS, ceci par la technique en flux bloqué en fluorescence. La liaison entre la NW-M3HFaa et ΔP(–)PBS se forme avec une constante d'association de l'ordre 10^9 - 10^{10} $M^{-1}s^{-1}$, ce qui est supérieur à la vitesse de diffusion (10^8 $M^{-1}s^{-1}$), indiquant pour la première fois que la déstabilisation des NAs par NC est un processus quasi-instantané, accompagnant la liaison non-spécifique et le glissement de la protéine vers le site de liaison préférentielle. Aucune autre modification plus lente du signal n'a pu être mise en évidence lorsque la protéine se lie à son site de fixation spécifique. Ceci suggère soit que les changements conformationnels démontrés par RMN au niveau de ce site se produisent immédiatement lors de la fixation, soit qu'ils sont silencieux en spectroscopie.

En conclusion, l'acide aminé nouvellement synthétisé marqué au M3HF démontre une forte sensibilité au microenvironnement de son émission duale et se révèle être un excellent substitut au Trp. L'analyse de cette émission duale, du rendement quantique et des temps de vie montre que l'interaction du peptide avec des NAs ou des bicouches lipidiques induit un changement très important de l'accessibilité à l'eau ou de la flexibilité de la sonde. L'analyse en FLIM des peptides microinjectés nous a permis d'identifier les ARNs comme partenaires principaux des interactions de NC dans les nucléoles et le cytoplasme. Ce nouvel analogue d'acide aminé obtenu à partir d'une sonde ESIPT se révèle donc être un outil efficace pour la détection et la caractérisation des interactions de peptides en solution ou dans les cellules. Ce

nouvel outil peut être également très utile pour établir des criblages à haut débit afin d'identifier des composés capables d'inhiber la liaison des peptides marqués à leurs ligands.

Partie 2. Dynamique de la réaction d'hybridation de (–)/(+)PBS et de son induction par la protéine de la nucléocapside de VIH-1 en utilisant des analogues isomorphes de nucléosides et des analogues sensibles à l'environnement.

Mécanisme complet de l'hybridation de (–)/(+)PBS révélé par un marquage site-sélectif de (–)PBS par un analogue de nucléoside contenant la thienyl-3-hydroxychromone.

Une seconde approche pour suivre de manière site-sélective l'hybridation de (–)/(+)PBS, spontanée ou NC-induite, consiste à marquer l'un des NA. Jusqu'à présent, le mécanisme d'hybridation a été étudié soit en utilisant une séquence PBS marquée à ses extrémités 5' et 3' respectivement par la rhodamine 6G (Rh6G) et l'inhibiteur Dabcyl, soit en utilisant l'une des séquences PBS marquée dans sa boucle par l'analogue de nucléoside 2-aminopurine (2Ap). Cependant, les PBS doublement marqués par Rh6G et Dabcyl n'ont montré aucune sensibilité à l'hybridation des régions de la boucle, et la très faible brillance de la 2Ap nécessite de fortes concentrations en PBS résultant en des cinétiques trop rapides pour être mesurées avec précision. Une cinétique complète des voies d'hybridation ne peut donc être obtenue. Pour obtenir une bonne description du mécanisme d'hybridation de (–)/(+)PBS et notamment pour décrire les événements à l'échelle moléculaire dans les régions de la boucle pendant les réactions d'hybridation, nous avons utilisé deux nouveaux analogues : un analogue nucléosidique universel - la 2-thienyl-3-hydroxychromone (3HCnt) - et un analogue isomorphe de la guanosine - la thienodeoxyguanosine ($d^{th}G$) - (Figure R4). L'avantage des analogues universels est leur capacité à remplacer toute base naturelle, et dans le cas des dérivés de 3HC, de leur forte sensibilité au microenvironnement de la sonde. L'avantage des analogues isomorphes est leur forte ressemblance avec les nucléosides naturels à l'égard de leur encombrement et leur capacité à former des paires de bases Watson-Crick avec les bases complémentaires. Ces caractéristiques principales nous ont permis d'obtenir des informations complémentaires des deux sondes. Ces analogues ont été incorporés dans une séquence ADN (–)PBS qui est impliquée dans le deuxième transfert de brin. Nous avons caractérisé en détail

les nucléosides seuls et incorporés dans les NAs.

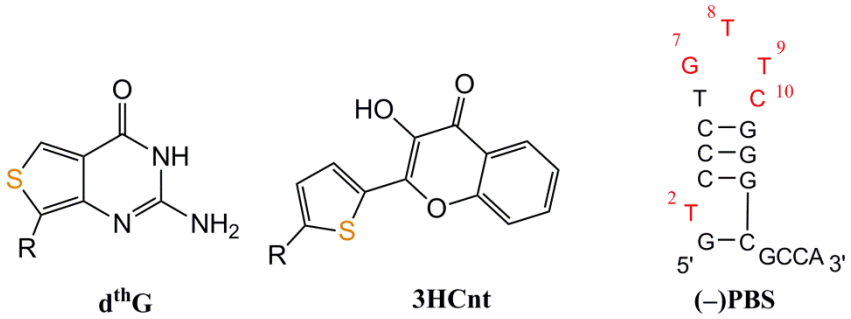


Figure R4. Analogues fluorescents de nucléosides utilisés dans cette étude et positions de marquage de l'ADN (–)PBS (R = 2'-deoxy-D-ribose).

L'analogue de nucléoside 3HCnt récemment synthétisé (en collaboration avec le groupe d'Alain Burger) possède une brillance élevée, un signal ratiométrique et une sensibilité à l'environnement qui lui permet de fournir une réponse multiparamétrique lors des interactions peptide – ADN et ADN – ADN. Afin de caractériser le mécanisme d'hybridation, tant spontanée que NC-induite, de (–)/(+)PBS, notre premier objectif fût de comparer les propriétés thermodynamiques, spectroscopiques et de liaison avec la NC des séquences de (–)PBS marquées par 3HCnt en position 2 sur la tige et en positions 7, 8, 9 ou 10 dans la boucle, et d'identifier la position optimale de marquage en vue d'atteindre cet objectif. Les expériences de dénaturation thermique ont montré que la nucléobase 3HCnt n'affecte que modérément la stabilité des duplexes, ce qui s'avère cohérent avec la nature de 3HCnt comme nucléobase universelle. Ensuite, nous avons mesuré les spectres d'absorption et d'émission du ΔP(–)PBS marqué à la 3HCnt, soit seul en solution, soit complexé avec la NC. Nous avons observé que, lors de l'interaction de NC avec (–)PBS substituée en différentes positions par 3HCnt, plusieurs bases de (–)PBS deviennent exposées au solvant. Pour toutes ces positions, nous avons également observé une augmentation substantielle du rendement quantique lors de la liaison à la protéine. Cette augmentation est probablement due à la limitation de mouvement de la sonde quand elle est en contact avec le plateau hydrophobe de NC. Ensuite, des expériences de liaison par compétition ont été réalisées afin de déterminer l'affinité relative de NC(11-55) pour la ΔP(–)PBS marquée par rapport à la non marquée afin de déterminer si l'introduction de 3HCnt modifie la liaison de NC(11-55) avec ΔP(–)PBS. Ceci a permis de montrer que 3HCnt en positions 8 et 9 ne modifie pas l'affinité de ΔP(–)PBS pour NC(11-55), ce qui ne s'avère pas le cas pour les positions 2, 7 et 10 pour lesquelles cette affinité est perturbée. Ensuite les propriétés de fluorescence des séquences marquées ont été

étudiées pendant la transition simple brin-double brin. Nous avons montré que la nucléobase 3HCnt peut mesurer de manière site-spécifique la transition simple brin – double brin dans la plupart des positions. Par comparaison avec la 2Ap qui est pratiquement non-fluorescente lorsqu'elle est insérée dans un ADN double-brin, le rendement quantique de 3HCnt augmente lors de la formation d'un duplex. Cet avantage majeur peut être utilisé avec profit pour suivre en temps réel la réaction d'hybridation de (–)/(+)PBS à des concentrations de (–)PBS beaucoup plus faibles. Afin de rationaliser davantage les propriétés spectroscopiques de ΔP(–)PBS marquée à la 3HCnt, des mesures de fluorescence résolue en temps ont été réalisées. Les paramètres alors obtenus se sont révélés être pleinement cohérents avec les données spectroscopiques et nous ont aidé à comprendre les modifications spectroscopiques observées lors des interactions entre les NAs et NC(11-55) ou ΔP(+)PBS. Comme la 3HCnt en position 9 n'affecte qu'à la marge l'affinité pour NC(11-55) et montre une forte sensibilité lors de cette interaction et de la formation d'un duplex étendu, cette position a été sélectionnée pour poursuivre l'étude de la réaction d'hybridation de (–)/(+)PBS.

3HCnt a permis pour la première fois de déterminer l'ensemble des constantes de vitesse cinétiques gouvernant la réaction d'hybridation (–)/(+)PBS tant en l'absence qu'en présence de NC(11-55) (Figure R5). Dans les deux cas, il a été clairement montré que cette réaction procède en deux étapes. En premier, le complexe intermédiaire (IC) se forme rapidement pour se transformer ensuite, par une étape limitante en temps, en un duplex (–)/(+)PBS étendu. Nos résultats ont confirmé que NC(11-55) privilégie la voie d'hybridation boucle-boucle. Par cette voie, le peptide augmente à la fois la vitesse de formation du complexe intermédiaire et son extension dans le duplex. En outre, en l'absence de peptide, nos résultats ont montré la présence de contacts boucle-boucle dans IC, ce qui n'a jamais été montré jusqu'à présent. Nous avons montré que 3HCnt en position 9 n'affecte que très peu la liaison de NC avec la boucle et permet d'obtenir des constantes cinétiques cohérentes avec l'unique paramètre cinétique obtenu en utilisant (–)PBS doublement marqué à ses extrémités 5' et 3'. Ainsi, 3HCnt présente un net avantage sur cette précédente approche, car il ne nécessite qu'un seul marqueur et procure une information directe sur la boucle, qui joue un rôle clef dans la réaction d'hybridation.

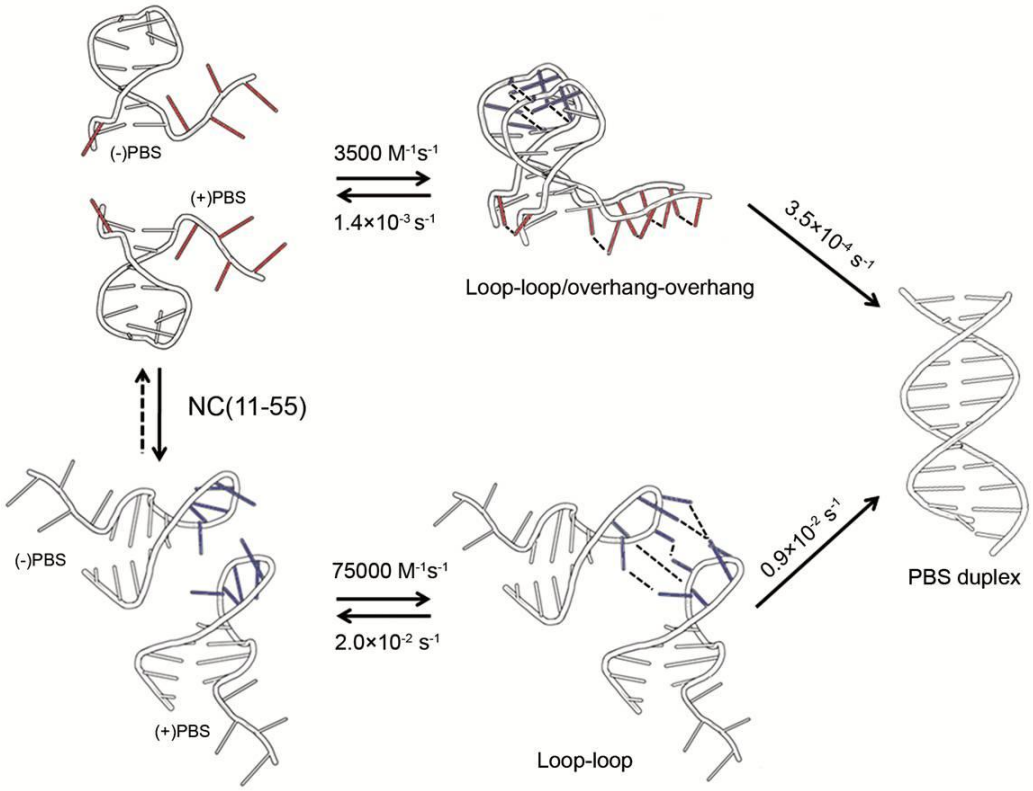


Figure R5. Mécanisme proposé et constantes de vitesse pour l'hybridation (–)(+)PBS. En l'absence de NC(11-55) (voie supérieure), un complexe intermédiaire est formé via les contacts des extensions de (–)PBS et (+)PBS et de leurs boucles. En présence de NC(11-55) (voie inférieure), les contacts entre les boucles sont supposées conduire à la formation du *kissing complex* intermédiaire entre les boucles apicales qui conduit au duplex final (adapté de Godet et al. 2011).

Caractérisation photophysique du substitut guanosine isomorphe thienodeoxyguanosine comme mime potentiel de G7 dans (–)PBS

La dthG a été choisie pour étudier la dynamique de la transition simple brin-double brin, en remplaçant la guanosine en position 7 qui joue un rôle clé dans l'interaction de (–)PBS avec NC. dthG est la forme deoxy de thG, un dérivé thieno[3,4-d]pyrimidine, synthétisé par D. Shin du groupe d'Yitzhak Tor. Un examen approfondi des propriétés d'émission de thG/dthG a montré leur surprenante complexité (Figure R6). En solution aqueuse, deux tautomères de thG à l'état fondamental ont été observés, avec des spectres d'absorption et d'émission significativement décalés. La concentration relative de la forme hypsochrome (bleue) augmente de manière linéaire avec le caractère protique du solvant. En solution aqueuse, des concentrations pratiquement identiques des deux tautomères ont été observées. L'équilibre

entre les deux formes semble donc dépendre essentiellement d'une stabilisation par liaison hydrogène due au solvant.

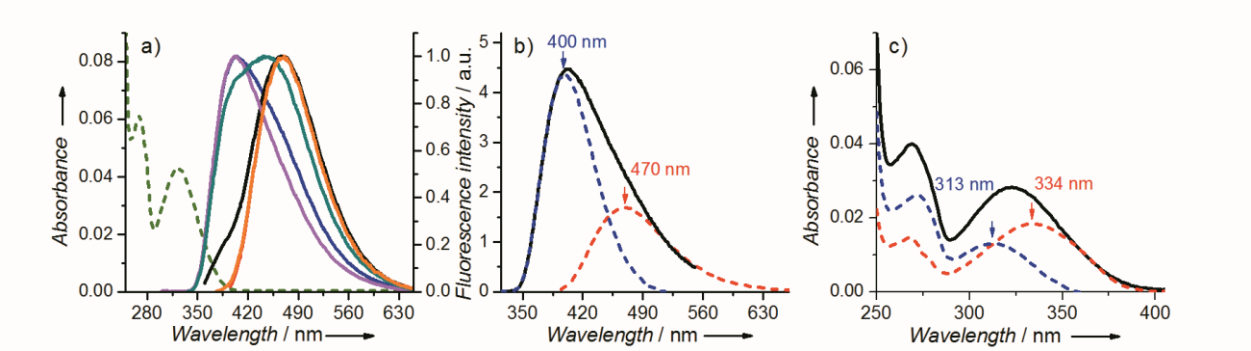
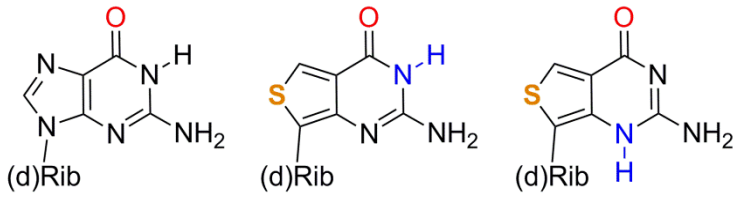


Figure R6. Spectres d'absorption et d'émission de thG en tampon TRIS-HCl, 25 mM, pH=7.5, 30 mM NaCl, 0.2 mM $MgCl_2$. (a) Spectres d'absorption (tirets verts) et d'émission de thG à différentes longueurs d'onde d'excitation: 283 nm (courbe rouge); 298 nm (bleue); 320 nm (verte); 345 nm (noire) et 360 nm (orange). Les spectres d'émission sont normalisés à leur maximum. Les spectres d'émission à λ_{exc} = 380 nm et λ_{exc} = 360 nm se superposent et ne sont pas montrés. (b) Spectre d'émission après déconvolution de thG, obtenu à λ_{ex}= 283 nm. (c) Déconvolution du spectre d'absorption de thG dans ses deux formes à l'état fondamental (les couleurs sont les mêmes qu'en b).

Afin de mieux identifier les deux formes tautomères à l'émission, des calculs de chimie quantique ont été menés en phase gazeuse, dans le dioxane et dans l'eau (collaboration avec Roberto Improta). Les deux formes ont été identifiées comme étant les keto-amines tautomères H1 et H3 (Figure R7).



G **thG-H1** **thG-H3**

Figure R7. Structures de la guanosine (G) et des deux tautomères émissifs de thiénoguanosine (thG-H1 and thG-H3). (d)Rib = D-ribose ou 2′-deoxy-D-ribose.

Lorsque le résidu dG7 dans la boucle (–)PBS est remplacé par dthG, les deux tautomères H1 and H3 sont observés (Figure R8a). Au contraire, lorsque dthG7(–)PBS est hybridé avec

son brin complémentaire (+)PBS, seule l'émission de thG-H1 est observée (Figure R8b), indiquant que ce tautomère est favorisé dans la forme double brin. Des spectres d'émission différents sont observés pour les duplexes (–)(+)PBS, totalement appariés ou non-appariés, lorsque dthG est placé en opposition avec A, T, ou G (Figure R8d). Ainsi, l'émission relative des deux tautomères dthG est sensible à la base opposée et peut, en principe, permettre de caractériser le bon appariement des résidus G-substitués. Ensuite, pour vérifier la réponse des tautomères à la liaison avec NC(11-55), une titration de dthG7(–)PBS en présence de NC(11-55) a été effectuée (Figure R8b), montrant que l'émission relative des deux tautomères est bien sensible à la liaison du peptide.

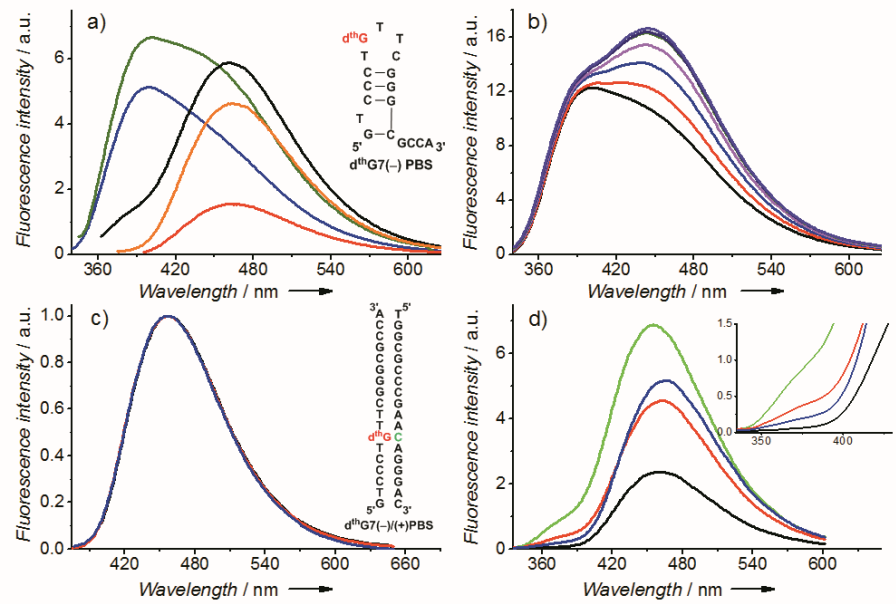


Figure R8. Spectres d'émission de dthG7(–)PBS (a, b) et de dthG7(–)/(+)PBS (c, d). (a) Spectres d'émission de dthG7(–)PBS pour différentes longueurs d'onde d'excitation: 298 nm (bleu), 320 nm (vert), 345 nm (noir), 360 nm (orange), and 380 nm (rouge). Insert: structure de dthG7(–)PBS, le résidu G7 (rouge) est remplacé par dthG. (b) Spectres d'émission de dthG7(–)PBS en absence (noir) et en présence de 1 à 6 équivalents du peptide NC(11-55) (rouge à violet) à λ_{exc} = 320 nm. (c) Spectres d'émission du duplex apparié dthG7(–)/(+)PBS aux mêmes longueurs d'onde d'excitation que (a). Insert: structure du duplex dthG7(–)/(+)PBS. Dans des duplexes non-appariés, le résidu C en vert est remplacé par A, G ou T. (d) Spectres d'émission des duplexes dthG7-labeled (–)/(+)PBS, appariés et non-appariés, à λ_{exc} = 310 nm. La base opposée à dthG est C (noir, duplex natif), T (bleu), G (rouge), ou A (vert). Insert: zoom sur le bord bleu des spectres.

Afin de préciser davantage ces observations biophysiques, des simulations en dynamique moléculaire (MD) ont été effectuées à partir de la structure RMN de ΔP(–)PBS (en collaboration avec Mattia Mori). Ces simulations ont montré que le remplacement de dG par d^{th}G-H1 dans le duplex (–)/(+)PBS est nettement plus favorable, d'un point de vue thermodynamique et conformationnel, que le remplacement par d^{th}G-H3.

Pour résumer, soit comme nucléoside libre ou incorporé dans un acide nucléique, d^{th}G présente deux états fondamentaux tautomères avec des spectres d'absorption et d'émission significativement décalés. Ces deux tautomères ont été identifiés comme étant les formes keto-amines thG-H1 et thG-H3. Les études en MD suggèrent par ailleurs que thG-H1 est très similaire à G dans les séquences PBS tant simple-brin que double-brin, alors que le tautomère thG-H3 se comporte de manière comparable à G seulement dans la boucle (–)PBS. Le rapport d'intensité de fluorescence des deux tautomères s'est révélé être très sensible à la nature double ou simple-brin de l'acide nucléique, à la nature de la base opposée dans les duplexes et à la liaison de peptides. La présence de ces tautomères et leur régulation par l'environnement apporte donc un canal supplémentaire d'information qui pourra être utilisé pour étudier les interactions des NAs avec la protéine NC ainsi que leur dynamique.

Etude dynamique et conformationnelle du résidu G de la tige boucle (–)PBS et du duplex (–)(+)PBS substitué par la thiénodéoxyguanosine et comparaison avec la 2-aminopurine.

Puisque le tautomère thG-H1 peut être excité et détecté sélectivement du fait de son déplacement bathochrome, nous avons utilisé son signal pour effectuer une comparaison point par point avec l'analogue de nucléoside 2-aminopurine (d2Ap). Dans ce but, nous avons substitué la dG7 de (–)PBS soit par d^{th}G, soit par d2Ap, et avons analysé les caractéristiques biophysiques et photophysiques de la tige-boucle (–)PBS et du duplex (–)(+)PBS ainsi marqués (Figure R9).

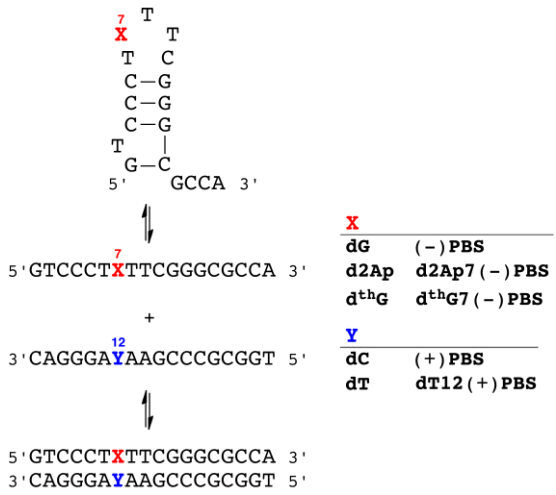


Figure R9. Séquence de l'amorce du site de liaison de (–)PBS montrée en simple brin, (milieu), tige boucle (haut) et en duplex avec (+)PBS (bas). Sont aussi montrées les séquences spécifiquement modifiées contenant d2Ap et dthG.

Dans une première étape, des expériences de dénaturation thermique ont montré que dthG influence de manière minime les températures de fusion de la tige-boucle (–)PBS et du duplex correspondant (–)/(+)PBS. Ensuite, les propriétés photophysiques de dthG7(–)PBS et de dthG7(–)/(+)PBS ont été comparées avec celles des séquences correspondantes marquées avec d2Ap (Figure R10). Le rendement quantique du duplex dthG7(–)/(+)PBS s'est révélé 2 fois supérieur à celui de dthG7(–)PBS et excède de plus d'un ordre de grandeur celui de d2Ap7(–)/T12(+)PBS. Cette différence majeure indique que contrairement à d2Ap, la fluorescence de dthG n'est pas efficacement inhibée par ses voisins immédiats dans le double-brin. Ceci constitue un atout évident comparé à d2Ap pour suivre la transition double- à simple-brin et pour caractériser les propriétés dynamiques de la base substituée.

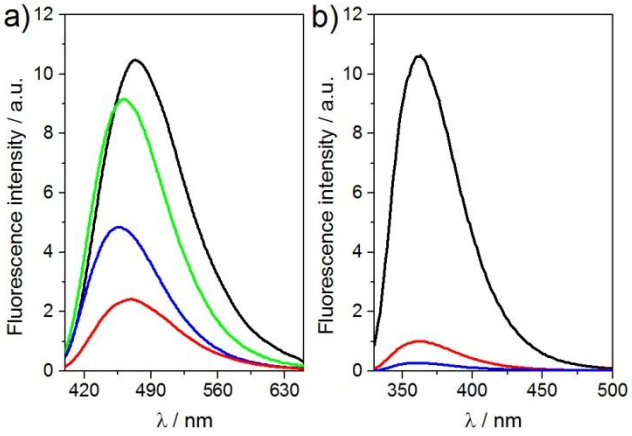


Figure R10. Spectres d'émission des séquences (–)PBS marquées avec $d^{th}G$- (a) et d2Ap- (b). (a) nucléoside $d^{th}G$ libre (noir), $d^{th}G7$(–)PBS (rouge), $d^{th}G7$(–)/(+)PBS (bleu) et $d^{th}G7$(–)/T12(+)PBS (vert); (b) nucléoside d2Ap libre (noir), d2Ap7(–)PBS (rouge), et d2Ap7(–)/T12(+)PBS (bleu).

Pour compléter la comparaison de la photophysique de $d^{th}G$ et de d2Ap, des expériences de fluorescence résolue en temps ont été réalisées. La d2Ap en position 7 de (–)PBS indique une hétérogénéité conformationnelle importante de cette sonde et l'existence "d'espèces noires" pour lesquelles l'émission de d2Ap est inhibée soit par inhibition statique, soit par une inhibition dynamique très rapide. Le déclin de fluorescence de $d^{th}G7$(–)PBS s'avère moins complexe et ne comprend qu'une quantité négligeable (<10%) de ces espèces noires, ce qui constitue un net avantage sur d2Ap, puisque toutes les conformations de $d^{th}G$ dans (–)PBS peuvent alors être suivies par ces mesures résolues en temps. Des différences encore plus remarquables entre d2Ap et $d^{th}G$ ont été observées dans le duplex (–)/(+)PBS : 98% des espèces du duplex marqué à d2Ap présentent des temps de vie très courts (<0.5 ns), ce qui explique le rendement quantique très faible et montre que la d2Ap est une sonde fluorescente très limitée pour des NAs double-brins. A l'opposé du déclin particulièrement complexe du duplex marqué à la d2Ap, le déclin de $d^{th}G7$(–)/(+)PBS est caractérisé par seulement deux temps de vie et une quantité négligeable d'espèces noires. Ceci indique que $d^{th}G$ adopte un nombre plus limité de conformations, ce qui résulte probablement de son aptitude à former une paire de bases Watson-Crick stable avec dC. Pour compléter cette comparaison entre $d^{th}G$ et d2Ap, nous avons effectué d'une part des mesures d'anisotropie de fluorescence résolue en temps afin d'obtenir des informations concernant les différents modes de mouvement de l'acide nucléique (local, segmentaire, global) et d'autre part des mesures d'inhibition de

fluorescence avec KI afin d'établir quantitativement l'exposition au solvant des nucléosides émissifs au sein de l'acide nucléique. La $d^{th}G$ s'est avérée un outil puissant pour explorer les dynamiques locales du résidu substitué par le suivi de ses paramètres de fluorescence résolues en temps et l'étude de son inhibition de fluorescence. Elle permet entre autres de suivre les conformations dominantes adoptées par G7, dirigées à l'intérieur de la boucle. Ces caractéristiques font de ce nouvel analogue fluorescent un outil parfait pour suivre fidèlement les conformations et la dynamique de résidus G dans les NAs.

En se basant sur les informations ainsi obtenues, nous envisageons de suivre l'hybridation de la tige-boucle (–)PBS avec son duplex complémentaire (+)PBS, ceci en temps réel en absence et en présence du peptide NC(11-55), en utilisant le signal d'émission de $d^{th}G$ afin d'obtenir les paramètres cinétiques de la réaction d'hybridation et de les comparer avec ceux obtenus à l'aide des nucléobases 3HCnt. Des expériences préliminaires de cinétique d'hybridation de $d^{th}G7$(–)PBS avec (+)PBS en absence de NC ont montré qu'il est possible de suivre la réaction pour (–)/(+)PBS dans des conditions de pseudo-premier ordre. Ces cinétiques sont différentes de celles obtenues avec 3HCnt- (–)PBS car elles présentent une composante ascendante et une composante décroissante (Figure R11).

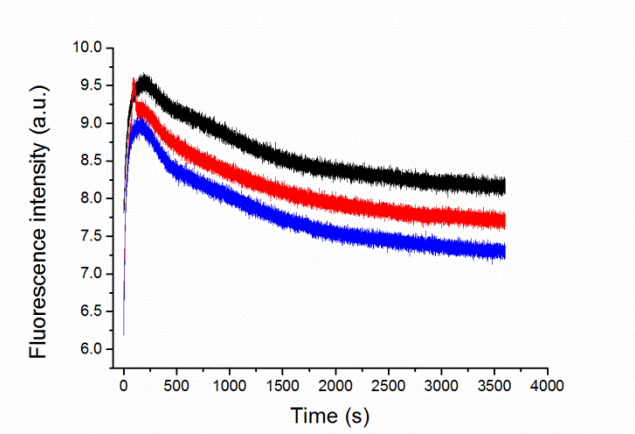


Figure R11. Cinétiques d'intensité de fluorescence pour 300 nM $d^{th}G7$(–)/(+)PBS avec 3µM (noir), 4µM (rouge) and 5µM (bleu) de (+)PBS. Longueurs d'onde d'excitation et d'émission à 360 nm et 470 nm. Tampon utilisé : 25 mM TRIS-HCl (pH=7.5), 30 mM NaCl, 0.2 mM $MgCl_2$.

Les courbes de cinétique ont été convenablement ajustées par une différence de deux fonctions exponentielles, donnant ainsi deux constantes de vitesse très comparables à celles obtenues en utilisant 3HCnt-(–)PBS. Ces expériences constituent la preuve de principe

indiquant que thG convient parfaitement pour mesurer les réactions d'hybridation et en extraire leurs constantes cinétiques. A l'évidence, des expériences complémentaires seront nécessaires pour bien comprendre les changements observés dans le signal de fluorescence et en extraire l'ensemble des paramètres cinétiques tant en l'absence qu'en présence de NC. La comparaison avec les résultats obtenus avec 3HCnt nous permettront de confirmer et/ou de compléter la connaissance du mécanisme de la réaction d'hybridation et de nous informer des possibles perturbations induites par l'introduction de ces sondes.

En conclusion, en utilisant des peptides NC et des oligonucléotides marqués en différentes positions par des sondes M3HFaa, dthG et 3HCnt, nous avons pu donner une image complète du processus dynamique qui gouverne le mécanisme chaperon de NC dans le second transfert de brin. Cette compréhension est fondamentale pour concevoir des stratégies rationnelles et cibler le rôle spécifique de NC via l'inhibition de ses interactions. En outre, la caractérisation des propriétés des sondes fluorescentes utilisées et de leurs applications nous a permis de montrer que ces sondes constituent des outils majeurs dans l'étude des interactions protéine – acide nucléique.

References

A

Abbink, T. E. and Berkhout B. A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.*, 2003, 278, 13, 11601-11611.

Abbondanzieri, E. A., Bokinsky G., Rausch J. W., Zhang J. X., Le Grice S. F. and Zhuang X. Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*, 2008, 453, 7192, 184-189.

Adamson, C. S. and Freed E. O. Human immunodeficiency virus type 1 assembly, release, and maturation. *Adv. Pharmacol.*, 2007, 55, 347-387.

Alaimo, A., Malo, C., Areso, P., Aloria, K., Millet, O., Villarroel, A. The use of dansyl-calmodulin to study interactions with channels and other proteins. *Methods Mol. Biol.*, 2013, 998, 217-231.

Aldovini, A. and Young R. A. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of non-infectious virus. *J. Virol.*, 1990, 64, 5, 1920-1926.

Alfadhli, A. and Barklis E. The roles of lipids and nucleic acids in HIV-1 assembly. *Front. Microbiol.*, 2014, 5, 253.

Altfeld, M., Fadda L., Frleta D. and Bhardwaj N. DCs and NK cells: critical effectors in the immune response to HIV-1. *Nat. Rev. Immunol.*, 2011, 11, 3, 176-186.

Amarasinghe, G. K., De Guzman R. N., Turner R. B., Chancellor K. J., Wu Z. R. and Summers M. F. 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.*, 2000, 301, 2, 491-511.

Amarasinghe, G. K., Zhou J., Miskimon M., Chancellor K. J., McDonald J. A., Matthews A. G., Miller R. R., Rouse M. D. and Summers M. F. Stem-loop SL4 of the HIV-1 psi RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition. *J. Mol. Biol.*, 2001, 314, 5, 961-970.

Ambrose, Z. and Aiken C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology*, 2014, 454-455, 371-379.

Apolonia, L., Schulz R., Curk T., Rocha P., Swanson C. M., Schaller T., Ule J. and Malim M. H. Promiscuous RNA binding ensures effective encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog.*, 2015, 11, 1.

Arts, E. J. and Wainberg M. A. Human immunodeficiency virus type 1 reverse transcriptase and early events in reverse transcription. *Adv. Virus Res.*, 1996, 46, 97-163.

Avilov, S. V., Godet J., Piemont E. and Mely Y. Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. *Biochemistry*, 2009, 48, 11, 2422-2430.

Avilov, S. V., Piemont E., Shvadchak V., de Rocquigny H. and Mely Y. Probing dynamics of HIV-1 nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. *Nucleic Acids Res.*, 2008, 36, 3, 885-896.

Azoulay, J., Clamme J. P., Darlix J. L., Roques B. P. and Mély Y. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.*, 2003, 326, 3, 691-700.

B

Balakrishnan, M., Fay P. J. and Bambara R. A. The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region. *J. Biol. Chem.*, 2001, 276, 39, 36482-36492.

Bampi, C., Bibillo A., Wendeler M., Divita G., Gorelick R. J., Le Grice S. F. J. and Darlix J. L. Nucleotide excision repair and template-independent addition by HIV-1 reverse transcriptase in the presence of nucleocapsid protein. *J. Biol. Chem.*, 2006, 281, 17, 11736-11743.

Bandwar, R. P. and Patel S. S. Peculiar 2-Aminopurine fluorescence monitors the dynamics of open complex formation by bacteriophage T7 RNA polymerase. *J. Biol. Chem.*, 2001, 276, 17, 14075-14082.

Barbara, P. F., Walsh P. K. and Brus L. E. Picosecond kinetic and vibrationally resolved spectroscopic studies of intramolecular excited-state hydrogen-atom transfer. *J. Phys. Chem.*, 1989, 93, 1, 29-34.

Barraud, P., Gaudin C., Dardel F. and Tisne C. New insights into the formation of HIV-1 reverse transcription initiation complex. *Biochimie*, 2007, 89, 10, 1204-1210.

Barré-Sinoussi, F., Chermann J. C., Rey F., Nugeyre M. T., Chamaret S., Gruest J., Dauguet C., Axlerblin C., Vezinetbrun F., Rouzioux C., Rozenbaum W. and Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune-deficiency syndrome (AIDS). *Science*, 1983, 220, 4599, 868-871.

Basu, V. P., Song M., Gao L., Rigby S. T., Hanson M. N. and Bambara R. A. Strand transfer events during HIV-1 reverse transcription. *Virus Res.*, 2008, 134, 1-2, 19-38.

Baudin, F., Marquet R., Isel C., Darlix J. L., Ehresmann B. and Ehresmann C. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.*, 1993, 229, 2, 382-397.

Bazzi, A., Zargarian L., Chaminade F., Boudier C., de Rocquigny H., René B., Mély Y., Fossé P. and Mauffret O. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: Role of sugar deoxyriboses in the binding polarity of NC. *Nucleic Acids Res.*, 2011, 39, 9, 3903-3916.

Beechem, J. M. and Brand L. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.*, 1985, 54, 43-71.

Beerens, N. and Berkhout B. The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. *J. Virol.*, 2002, 76, 5, 2329-2339.

Beerens, N., Groot F. and Berkhout B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.*, 2001, 276, 33, 31247-31256.

Beerens, N., Jepsen M. D. E., Nechyporuk-Zloy V., Kruger A. C., Darlix J. L., Kjems J. and Birkedal V. Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA*, 2013, 19, 4, 517-526.

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

Beltz, H., Clauss C., Piemont E., Ficheux D., Gorelick R. J., Roques B., Gabus C., Darlix J. L., de Rocquigny H. and Mely Y. Structural determinants of HIV-1 nucleocapsid protein for cTAR

DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J. Mol. Biol.*, 2005, 348, 5, 1113-1126.

Ben Gaied, N., Glasser N., Ramalanjaona N., Beltz H., Wolff P., Marquet R., Burger A. and Mely Y. 8-vinyl-deoxyadenosine, an alternative fluorescent nucleoside analog to 2'-deoxyribosyl-2-aminopurine with improved properties. *Nucleic Acids Res.*, 2005, 33, 3, 1031-1039.

Berg, J. M. Potential metal-binding domains in nucleic-acid binding-proteins. *Science*, 1986, 232, 4749, 485-487.

Berg, O. G., Winter R. B. and von Hippel P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, 1981, 20, 24, 6929-6948.

Berkhout, B., Klaver B. and Das A. T. A Conserved hairpin structure predicted for the poly(a) signal of human and simian immunodeficiency viruses. *Virology*, 1995, 209, 1, 284-284.

Berkhout, B., Ooms M., Beerens N., Huthoff H., Southern E. and Verhoef K. In vitro evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. *J. Biol. Chem.*, 2002, 277, 22, 19967-19975.

Berkhout, B. and Van Wamel J. L. B. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA*, 2000, 6, 2, 282-295.

Berkhout, B. and van Wamel J. L. B. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J. Virol.*, 1996, 70, 10, 6723-6732.

Berkhout, B., Vastenhouw N. L., Klasens B. I. F. and Huthoff H. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA*, 2001, 7, 8, 1097-1114.

Berkowitz, R. D. and Goff S. P. Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology*, 1994, 202, 1, 233-246.

Bernacchi, S., Piémont E., Potier N., Van Dorsselaer A. and Mély Y. Excitonic heterodimer formation in an HIV-1 oligonucleotide labeled with a donor-acceptor pair used for fluorescence resonance energy transfer. *Biophys. J.*, 2003, 84, 1, 643-654.

Bernacchi, S., Stoylov S., Piemont E., Ficheux D., Roques B. P., Darlix J. L. and Mely Y. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.*, 2002, 317, 3, 385-399.

Berry, D. A., Jung K. Y., Wise D. S., Sercel A. D., Pearson W. H., Mackie H., Randolph J. B. and Somers R. L. Pyrrolo-dC and pyrrolo-C: Fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. *Tetrahedron Lett.*, 2004, 45, 11, 2457-2461.

Blond, A., Ennifar E., Tisne C. and Micouin L. The design of RNA binders: targeting the HIV replication cycle as a case study. *ChemMedChem*, 2014, 9, 9, 1982-1996.

Bloomfield, V., Crothers, D., Tinocco, I. Jr. Nucleic acids: structures, properties and functions. *University Science Books, Sausalito, CA*, 1999.

Bochkarev, A. and Bochkareva E. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr. Opin. Struct. Biol.*, 2004, 14, 1, 36-42.

Boon, E. M. and Barton J. K. DNA electrochemistry as a probe of base pair stacking in A-, B-, and Z-form DNA. *Bioconjug. Chem.*, 2003, 14, 6, 1140-1147.

Boudier, C., Klymchenko A. S., Mely Y. and Follenius-Wund A. Local environment perturbations in alpha1-antitrypsin monitored by a ratiometric fluorescent label. *Photochem. Photobiol. Sci.*, 2009, 8, 6, 814-821.

Bourbigot, S., Ramalanjaona N., Boudier C., Salgado G. F. J., Roques B. P., Mely Y., Bouaziz S. and Morellet N. How the HIV-1 nucleocapsid protein binds and destabilises the (−)primer binding site during reverse transcription. *J. Mol. Biol.*, 2008, 383, 5, 1112-1128.

Brasey, A., Lopez-Lastra M., Ohlmann T., Beerens N., Berkhout B., Darlix J. L. and Sonenberg N. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G(2)/M phase of the cell cycle. *J. Virol.*, 2003, 77, 7, 3939-3949.

Briggs, J. A. and Krausslich H. G. The molecular architecture of HIV. *J. Mol. Biol.*, 2011, 410, 4, 491-500.

Briggs, J. A. G., Simon M. N., Gross I., Krausslich H. G., Fuller S. D., Vogt V. M. and Johnson M. C. The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.*, 2004, 11, 7, 672-675.

Brochon, J. C. Maximum entropy method of data analysis in time-resolved spectroscopy. *Methods Enzymol.*, 1994, 240, 262-311.

Broder, S., Yarchoan R., Collins J. M., Lane H. C., Markham P. D., Klecker R. W., Redfield R. R., Mitsuya H., Hoth D. F., Gelmann E. and et al. Effects of suramin on HTLV-III/LAV infection presenting as Kaposi's sarcoma or AIDS-related complex: clinical pharmacology and suppression of virus replication in vivo. *Lancet*, 1985, 2, 8456, 627-630.

Buonaguro, L., Tornesello M. L. and Buonaguro F. M. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J. Virol.*, 2007, 81, 19, 10209-10219.

Burniston, M. T., Cimarelli A., Colgan J., Curtis S. P. and Luban J. Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J. Virol.*, 1999, 73, 10, 8527-8540.

C

Cai, M., Zheng R., Caffrey M., Craigie R., Clore G. M. and Gronenborn A. M. Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.*, 1997, 4, 7, 567-577.

Campbell, E. M. and Hope T. J. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.*, 2015, 13, 8, 471-483.

Chamontin, C., Rassam P., Ferrer M., Racine P. J., Neyret A., Laine S., Milhiet P. E. and Mougel M. HIV-1 nucleocapsid and ESCRT-component Tsg101 interplay prevents HIV from turning into a DNA-containing virus. *Nucleic Acids Res.*, 2015, 43, 1, 336-347.

Charubala, R., Maurinsh J., Rösier A., Melguizo M., Jungmann O., Gottlieb M., Lehbauer J., Hawkins M. and Pfleiderer W. Pteridine nucleosides - new versatile building blocks in oligonucleotide synthesis. *Nucleosides Nucleotides*, 1997, 16, 7-9, 1369-1378.

Chatel-Chaix, L., Boulay K., Mouland A. J. and Desgroseillers L. The host protein Staufen1 interacts with the Pr55Gag zinc fingers and regulates HIV-1 assembly via its N-terminus. *Retrovirology*, 2008, 5, 41.

Chatel-Chaix, L., Clement J. F., Martel C., Beriault V., Gatignol A., Des Groseillers L. and Mouland A. J. Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles. *Mol. Cell. Biol.*, 2004, 24, 7, 2637-2648.

Chen, J., Nikolaitchik O., Singh J., Wright A., Bencsics C. E., Coffin J. M., Ni N., Lockett S., Pathak V. K. and Hu W. S. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc. Natl. Acad. Sci. U.S.A.*, 2009, 106, 32, 13535-13540.

Chen, J., Rahman S. A., Nikolaitchik O. A., Grunwald D., Sardo L., Burdick R. C., Plisov S., Liang E., Tai S., Pathak V. K. and Hu W. S. HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc. Natl. Acad. Sci. U.S.A.*, 2016, 113, 2, E201-208.

Chen, J. C., Krucinski J., Miercke L. J., Finer-Moore J. S., Tang A. H., Leavitt A. D. and Stroud R. M. Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, 15, 8233-8238.

Chen, R. F. Fluorescence quantum yields of tryptophan and tyrosine. *Anal. Lett.*, 1967, 1, 1, 35-42.

Chen, Y., Maskri O., Chaminade F., Rene B., Benkaroun J., Godet J., Mely Y., Mauffret O. and Fosse P. Structural insights into the HIV-1 minus-strand strong-stop DNA. *J. Biol. Chem.*, 2015.

Chin, M. P. S., Rhodes T. D., Chen J. B., Fu W. and Hu W. S. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc. Natl. Acad. Sci. U.S.A.*, 2005, 102, 25, 9002-9007.

Chou, P.-T., Martinez M. L. and Clements H. Reversal of excitation behavior of proton-transfer vs. charge-transfer by dielectric perturbation of electronic manifolds. *J. Phys. Chem.*, 1993, 97, 2618-2622.

Choulier, L., Shvadchak V. V., Naidoo A., Klymchenko A. S., Mély Y. and Altschuh D. A peptide-based fluorescent ratiometric sensor for quantitative detection of proteins. *Anal. Biochem.*, 2010, 401, 2, 188-195.

Cimarelli, A. and Darlix J. L. Assembling the human immunodeficiency virus type 1. *Cell. Mol. Life Sci.*, 2002, 59, 7, 1166-1184.

Cimarelli, A., Sandin S., Hoglund S. and Luban J. Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA. *J. Virol.*, 2000, 74, 7, 3046-3057.

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. and et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science*, 1986, 233, 4761, 343-346.

Clavel, F. and Orenstein J. M. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.*, 1990, 64, 10, 5230-5234.

Clever, J., Sassetti C. and Parslow T. G. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.*, 1995, 69, 4, 2101-2109.

Coffin, J. M. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science*, 1995, 267, 5197, 483-489.

Cohen, B. E., McAnaney T. B., Park E. S., Jan Y. N., Boxer S. G. and Jan L. Y. Probing protein electrostatics with a synthetic fluorescent amino acid. *Science*, 2002, 296, 5573, 1700-1703.

Cosa, G., Harbron E. J., Zeng Y., Liu H. W., O'Connor D. B., Eta-Hosokawa C., Musier-Forsyth K. and Barbara P. F. Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. *Biophys. J.*, 2004, 87, 4, 2759-2767.

Cosa, G., Zeng Y., Liu H. W., Landes C. F., Makarov D. E., Musier-Forsyth K. and Barbara P. F. Evidence for non-two-state kinetics in the nucleocapsid protein chaperoned opening of DNA hairpins. *J. Phys. Chem. B*, 2006, 110, 5, 2419-2426.

Cristofari, G. and Darlix J. L. The ubiquitous nature of RNA chaperone proteins. *Prog. Nucleic Acid Res. Mol. Biol.*, 2002, 72, 223-268.

Cruceanu, M., Gorelick R. J., Musier-Forsyth K., Rouzina I. and Williams M. C. Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J. Mol. Biol.*, 2006, 363, 5, 867-877.

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. and Williams M. C. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.*, 2006, 34, 2, 593-605.

D

D'Souza, V. and Summers M. F. How retroviruses select their genomes. *Nat. Rev. Microbiol.*, 2005, 3, 8, 643-655.

Dai, N. and Kool E. T. Fluorescent DNA-based enzyme sensors. *Chem. Soc. Rev.*, 2011, 40, 12, 5756-5770.

Dalgleish, A. G., Beverley P. C. L., Clapham P. R., Crawford D. H., Greaves M. F. and Weiss R. A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 1984, 312, 5996, 763-767.

Dallmann, A., Dehmel L., Peters T., Mügge C., Griesinger C., Tuma J. and Ernsting N. P. 2-Aminopurine incorporation perturbs the dynamics and structure of DNA. *Angew. Chem. Int. Ed. Engl.*, 2010, 49, 34, 5989-5992.

Darlix, J. L., Garrido J. L., Morellet N., Mely Y. and de Rocquigny H. Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Adv. Pharmacol.*, 2007, 55, 299-346.

Darlix, J. L., Godet J., Ivanyi-Nagy R., Fossé P., Mauffret O. and Mély Y. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.*, 2011, 410, 4, 565-581.

Darlix, J. L., Lapadat-Tapolsky M., de Rocquigny H. and Roques B. P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.*, 1995, 254, 4, 523-537.

Darlix, J. L., Vincent A., Gabus C., de Rocquigny H. and Roques B. Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C. R. Acad. Sci. III*, 1993, 316, 8, 763-771.

Das, K. and Arnold E. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. *Curr. Opin. Virol.*, 2013, 3, 2, 111-118.

Das, R., Klymchenko A. S., Duportail G. and Mély Y. Unusually slow proton transfer dynamics of a 3-hydroxychromone dye in protic solvents. *Photochem. Photobiol. Sci.*, 2009, 8, 11, 1583-1589.

Datta, K., Johnson N. P., Villani G., Marcus A. H. and von Hippel P. H. Characterization of the 6-methyl isoxanthopterin (6-MI) base analog dimer, a spectroscopic probe for monitoring guanine base conformations at specific sites in nucleic acids. *Nucleic Acids Res.*, 2012, 40, 3, 1191-1202.

Datta, S. A. K., Clark P. K., Fan L. X., Ma B. Y., Harvin D. P., Sowder R. C., Nussinov R., Wang Y. X. and Rein A. Dimerization of the SP1 region of HIV-1 Gag induces a helical conformation and association into helical bundles: implications for particle assembly. *J. Virol.*, 2016, 90, 4, 1773-1787.

Datta, S. A. K., Temeselew L. G., Crist R. M., Soheilian F., Kamata A., Mirro J., Harvin D., Nagashima K., Cachau R. E. and Rein A. On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J. Virol.*, 2011, 85, 9, 4111-4121.

Day, R. N. and Davidson M. W. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.*, 2009, 38, 10, 2887-2921.

De Clercq, E. The design of drugs for HIV and HCV. *Nat. Rev. Drug Discov.*, 2007, 6, 12, 1001-1018.

De Clercq, E. Antiretroviral drugs. *Curr. Opin. Pharmacol.*, 2010, 10, 5, 507-515.

de Guzman, R. N., Wu Z. R., Stalling C. C., Pappalardo L., Borer P. N. and Summers M. F. Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science*, 1998, 279, 5349, 384-388.

de Marco, A., Heuser A. M., Glass B., Krausslich H. G., Muller B. and Briggs J. A. G. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J. Virol.*, 2012, 86, 24, 13708-13716.

de Rocquigny, H., El Meshri S. E., Richert L., Didier P., Darlix J. L. and Mely Y. Role of the nucleocapsid region in HIV-1 Gag assembly as investigated by quantitative fluorescence-based microscopy. *Virus Res.*, 2014, 193, 78-88.

de Rocquigny, H., Ficheux D., Gabus C., Fournie-Zaluski M. C., Darlix J. L. and Roques B. P. First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.*, 1991, 180, 2, 1010-1018.

Deforges, J., Chamond N. and Sargueil B. Structural investigation of HIV-1 genomic RNA dimerization process reveals a role for the Major Splice-site Donor stem loop. *Biochimie*, 2012, 94, 7, 1481-1489.

Demchenko, A. P., Tang K. C. and Chou P. T. Excited-state proton coupled charge transfer modulated by molecular structure and media polarization. *Chem. Soc. Rev.*, 2013, 42, 3, 1379-1408.

Demene, H., Dong C. Z., Ottmann M., Rouyez M. C., Jullian N., Morellet N., Mely Y., Darlix J. L., Fourniezaluski M. C., Saragosti S. and Roques B. P. 1H NMR structure and biological studies of the His(23)-Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc-finger is critical for virus infectivity. *Biochemistry*, 1994, 33, 39, 11707-11716.

Di Nunzio, F. New insights in the role of nucleoporins: a bridge leading to concerted steps from HIV-1 nuclear entry until integration. *Virus Res.*, 2013, 178, 2, 187-196.

Donohue, J. and Trueblood K. N. Base pairing in DNA. *J. Mol. Biol.*, 1960, 2, 363-371.

Dooher, J. E., Schneider B. L., Reed J. C. and Lingappa J. R. Host ABCE1 is at plasma membrane HIV assembly sites and its dissociation from gag is linked to subsequent events of virus production. *Traffic.*, 2007, 8, 3, 195-211.

Doose, S., Neuweiler H. and Sauer M. Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *Chemphyschem*, 2009, 10, 9-10, 1389-1398.

Dorfman, T., Luban J., Goff S. P., Haseltine W. A. and Gottlinger H. G. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.*, 1993, 67, 10, 6159-6169.

Doroshenko, A. O., Sychevskaya L. B., Grygorovych A. V. and Pivovarenko V. G. Fluorescence probing of cell membranes with azacrown substituted ketocyanine dyes. *J. Fluoresc.*, 2002, 12, 3-4, 455-464.

Driscoll, S. L., Hawkins M. E., Balis F. M., Pfleiderer W. and Laws W. R. Fluorescence properties of a new guanosine analog incorporated into small oligonucleotides. *Biophys. J.*, 1997, 73, 6, 3277-3286.

Dumas, A. and Luedtke N. W. Cation-mediated energy transfer in G-quadruplexes revealed by an internal fluorescent probe. *J. Am. Chem. Soc.*, 2010, 132, 51, 18004-18007.

Dussupt, V., Javid M. P., Abou-Jaoude G., Jadwin J. A., de La Cruz J., Nagashima K. and Bouamr F. The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding. *PLoS Pathog.*, 2009, 5, 3.

Dziuba, D., Karpenko I. A., Barthes N. P. F., Michel B. Y., Klymchenko A. S., Benhida R., Demchenko A. P., Mély Y. and Burger A. Rational design of a solvatochromic fluorescent uracil analogue with a dual-band ratiometric response based on 3-hydroxychromone. *Chem. - Eur. J.*, 2014, 20, 7, 1998-2009.

Dziuba, D., Postupalenko V. Y., Spadafora M., Klymchenko A. S., Guerineau V., Mely Y., Benhida R. and Burger A. A universal nucleoside with strong two-band switchable fluorescence and sensitivity to the environment for investigating DNA interactions. *J. Am. Chem. Soc.*, 2012, 134, 24, 10209-10213.

E

Egele, C., Schaub E., Ramalanjaona N., Piemont E., Ficheux D., Roques B., Darlix J. L. and Mely Y. HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J. Mol. Biol.*, 2004, 342, 2, 453-466.

Eisinger, J. and Navon G. Fluorescence quenching and isotope effect of tryptophan. *J. Chem. Phys.*, 1969, 50, 5, 2069-2077.

El-Yazbi, A. F. and Loppnow G. R. 2-Aminopurine hairpin probes for the detection of ultraviolet-induced DNA damage. *Anal. Chim. Acta*, 2012, 726, 44-49.

El Meshri, S. E., Dujardin D., Godet J., Richert L., Boudier C., Darlix J. L., Didier P., Mely Y. and de Rocquigny H. 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.*, 2015, 427, 6 Pt B, 1480-1494.

Elson, E. L. Fluorescence correlation spectroscopy: past, present, future. *Biophys. J.*, 2011, 101, 12, 2855-2870.

Enander, K., Choulier L., Olsson A. L., Yushchenko D. A., Kanmert D., Klymchenko A. S., Demchenko A. P., Mély Y. and Altschuh D. A peptide-based, ratiometric biosensor construct for direct fluorescence detection of a protein analyte. *Bioconjug. Chem.*, 2008, 19, 9, 1864-1870.

Engelborghs, Y. Correlating protein structure and protein fluorescence. *J. Fluoresc.*, 2003, 13, 1, 9-16.

F

Fasman, G. D. *Practical handbook of biochemistry and molecular biology*. CRC Press Inc, Boca Raton, FL, 1989.

Feng, S. and Holland E. C. HIV-1 tat trans-activation requires the loop sequence within tar. *Nature*, 1988, 334, 6178, 165-167.

Feng, Y. X., Campbell S., Harvin D., Ehresmann B., Ehresmann C. and Rein A. The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.*, 1999, 73, 5, 4251-4256.

Ferreira, S. T., Stella L. and Gratton E. Conformational dynamics of bovine Cu, Zn superoxide dismutase revealed by time-resolved fluorescence spectroscopy of the single tyrosine residue. *Biophys. J.*, 1994, 66, 4, 1185-1196.

Festy, F., Ameer-Beg S. M., Ng T. and Suhling K. Imaging proteins in vivo using fluorescence lifetime microscopy. *Mol. Biosyst.*, 2007, 3, 6, 381-391.

Fiebig, T., Wan C. and Zewail A. H. Femtosecond charge transfer dynamics of a modified DNA base: 2-Aminopurine in complexes with nucleotides. *ChemPhysChem*, 2002, 3, 9, 781-788.

Fisher, R. J., Fivash M. J., Stephen A. G., Hagan N. A., Shenoy S. R., Medaglia M. V., Smith L. R., Worthy K. M., Simpson J. T., Shoemaker R., McNitt K. L., Johnson D. G., Hixson C. V., Gorelick R. J., Fabris D., Henderson L. E. and Rein A. Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. *Nucleic Acids Res.*, 2006, 34, 3, 1082-1082.

Fisher, R. J., Rein A., Fivash M., Urbaneja M. A., Casas-Finet J. R., Medaglia M. and Henderson L. E. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.*, 1998, 72, 3, 1902-1909.

Fossen, T., Wray V., Bruns K., Rachmat J., Henklein P., Tessmer U., Maczurek A., Klinger P. and Schubert U. Solution structure of the human immunodeficiency virus type 1 p6 protein. *J. Biol. Chem.*, 2005, 280, 52, 42515-42527.

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

Franklin, R. E. and Gosling R. G. Molecular configuration in sodium thymonucleate. *Nature*, 1953, 171, 4356, 740-741.

Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.*, 2015, 13, 8, 484-496.

G

Gallo, R. C., Salahuddin S. Z., Popovic M., Shearer G. M., Kaplan M., Haynes B. F., Palker T. J., Redfield R., Oleske J., Safai B., White G., Foster P. and Markham P. D. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science*, 1984, 224, 4648, 500-503.

Gamble, T. R., Vajdos F. F., Yoo S., Worthylake D. K., Houseweart M., Sundquist W. I. and Hill C. P. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*, 1996, 87, 7, 1285-1294.

Ganser-Pornillos, B. K., Yeager M. and Sundquist W. I. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.*, 2008, 18, 2, 203-217.

Gao, F., Bailes E., Robertson D. L., Chen Y., Rodenburg C. M., Michael S. F., Cummins L. B., Arthur L. O., Peeters M., Shaw G. M., Sharp P. M. and Hahn B. H. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature*, 1999, 397, 6718, 436-441.

Gao, J. M., Strassler C., Tahmassebi D. and Kool E. T. Libraries of composite polyfluors built from fluorescent deoxyribosides. *J. Am. Chem. Soc.*, 2002, 124, 39, 11590-11591.

Gao, W., Li M. and Zhang J. X. Tandem immunoprecipitation approach to identify HIV-1 Gag associated host factors. *J. Virol. Meth.*, 2014, 203, 116-119.

Geng, T., Uknalis J., Tu S. I. and Bhunia A. K. Fiber-optic biosensor employing Alexa-Fluor conjugated antibody for detection of Escherichia coli O157 : H7 from ground beef in four hours. *Sensors*, 2006, 6, 8, 796-807.

Ghisaidoobe, A. B. T. and Chung S. J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Forster resonance energy transfer techniques. *Int. J. Mol. Sci.*, 2014, 15, 12, 22518-22538.

Godet, J. *Etude du role de chaperon de la proteine NCp7 de la nucleocapside du virus VIH-1*. Université de Strasbourg, 2010.

Godet, J., de Rocquigny H., Raja C., Glasser N., Ficheux D., Darlix J. L. and Mely Y. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.*, 2006, 356, 5, 1180-1192.

Godet, J., Kenfack C., Przybilla F., Richert L., Duportail G. and Mély Y. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res.*, 2013, 41, 9, 5036-5048.

Godet, J. and Mely Y. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol.*, 2010, 7, 6, 687-699.

Godet, J., Ramalanjaona N., Sharma K. K., Richert L., de Rocquigny H., Darlix J. L., Duportail G. and Mély Y. Specific implications of the HIV-1 nucleocapsid zinc fingers in the annealing of the primer binding site complementary sequences during the obligatory plus strand transfer. *Nucleic Acids Res.*, 2011, 39, 15, 6633-6645.

Goel, T., Kumar S. and Maiti S. Thermodynamics and solvation dynamics of BIV TAR RNA-Tat peptide interaction. *Mol. Biosyst.*, 2013, 9, 1, 88-98.

Goila-Gaur, R., Khan M. A., Miyagi E., Kao S. and Strebel K. Targeting APOBEC3A to the viral nucleoprotein complex confers antiviral activity. *Retrovirology*, 2007, 4.

Goldschmidt, V. *La rétrotranscription de HIV-1: étude du complexe d'initiation et des mécanismes de résistance aux inhibiteurs nucléosidiques*. Université Louis Pasteur, 2004.

Gorelick, R. J., Chabot D. J., Rein A., Henderson L. E. and Arthur L. O. The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J. Virol.*, 1993, 67, 7, 4027-4036.

Gorelick, R. J., Nigida S. M., Jr., Bess J. W., Jr., Arthur L. O., Henderson L. E. and Rein A. Non-infectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.*, 1990, 64, 7, 3207-3211.

Goretta, S. A., Kinoshita M., Mori S., Tsuchikawa H., Matsumori N. and Murata M. Effects of chemical modification of sphingomyelin ammonium group on formation of liquid-ordered phase. *Bioorg. Med. Chem.*, 2012, 20, 13, 4012-4019.

Gotte, M., Li X. G. and Wainberg M. A. HIV-1 reverse transcription: A brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Arch. Biochem. Biophys.*, 1999, 365, 2, 199-210.

Gray, R. D., Trent J. O. and Chaires J. B. Folding and unfolding pathways of the human telomeric G-quadruplex. *J. Mol. Biol.*, 2014, 426, 8, 1629-1650.

Greco, N. J. and Tor Y. Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. *J. Am. Chem. Soc.*, 2005, 127, 31, 10784-10785.

Greco, N. J. and Tor Y. Furan decorated nucleoside analogues as fluorescent probes: synthesis, photophysical evaluation, and site-specific incorporation. *Tetrahedron*, 2007, 63, 17, 3515-3527.

Grigorov, B., Decimo D., Smagulova F., Pechoux C., Mougel M., Muriaux D. and Darlix J. L. Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers. *Retrovirology*, 2007, 4.

Guest, C. R. Dynamics of mismatched base pairs in DNA. *Biochemistry*, 1991, 30, 13, 3271-3279.

Guiot, E., Carayon K., Delelis O., Simon F., Tauc P., Zubin E., Gottikh M., Mouscadet J. F., Brochon J. C. and Deprez E. Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *J. Biol. Chem.*, 2006, 281, 32, 22707-22719.

Guo, J. H., Wu T. Y., Anderson J., Kane B. F., Johnson D. G., Gorelick R. J., Henderson L. E. and Levin J. G. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.*, 2000, 74, 19, 8980-8988.

Guo, J. H., Wu T. Y., Kane B. F., Johnson D. G., Henderson L. E., Gorelick R. J. and Levin J. G. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.*, 2002, 76, 9, 4370-4378.

Guyader, M., Emerman M., Sonigo P., Clavel F., Montagnier L. and Alizon M. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature*, 1987, 326, 6114, 662-669.

H

Ha, T. and Tinnefeld P. Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging. *Annu. Rev. Phys. Chem.*, 2012, 63, 595-617.

Hall, K. B. 2-aminopurine as a probe of RNA conformational transitions. *Methods Enzymol.*, 2009, 469, 269-285.

Harborne, J. B. *Plant flavonoids in biology and medicine II. Biochemical, cellular, and medicinal properties*. Cody, V.; Middleton, E.; Harborne, J.B.; Beretz, A. Eds.; Alan R. Liss: New York. 1988.

Hargittai, M. R. S., Gorelick R. J., Rouzina L. and Musier-Forsyth K. Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.*, 2004, 337, 4, 951-968.

Hartmann, B. and Lavery R. DNA structural forms. *Q. Rev. Biophys.*, 1996, 29, 4, 309-368.

Hawe, A., Sutter M. and Jiskoot W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharm. Res.*, 2008, 25, 7, 1487-1499.

Hawkins, M. E. Fluorescent pteridine nucleoside analogs: a window on DNA interactions. *Cell Biochem. Biophys.*, 2001, 34, 2, 257-281.

Hawkins, M. E. Fluorescent pteridine probes for nucleic acid analysis. *Fluorescence Spectroscopy*, 2008, 450, 201-231.

Hawkins, M. E., Pfleiderer W., Mazumder A., Pommier Y. G. and Falls F. M. Incorporation of a fluorescent guanosine analog into oligonucleotides and its application to a real-time assay for the HIV-1 integrase 3'-processing reaction. *Nucleic Acids Res.*, 1995, 23, 15, 2872-2880.

Hayashi, T., Ueno Y. and Okamoto T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett.*, 1993, 327, 2, 213-218.

Henderson, L. E., Bowers M. A., Sowder R. C., Serabyn S. A., Johnson D. G., Bess J. W., Arthur L. O., Bryant D. K. and Fenselau C. Gag proteins of the highly replicative MN strain of human immunodeficiency virus type-1 : posttranslational modifications, proteolytic processings, and complete amino acid sequences. *J. Virol.*, 1992, 66, 4, 1856-1865.

Heng, X., Kharytonchyk S., Garcia E. L., Lu K., Divakaruni S. S., LaCotti C., Edme K., Telesnitsky A. and Summers M. F. Identification of a minimal region of the HIV-1 5 '-leader required for RNA dimerization, NC binding, and packaging. *J. Mol. Biol.*, 2012, 417, 3, 224-239.

Hergott, C. B., Mitra M., Guo J. H., Wu T. Y., Miller J. T., Iwatani Y., Gorelick R. J. and Levin J. G. Zinc finger function of HIV-1 nucleocapsid protein is required for removal of 5 '-terminal genomic RNA fragments: a paradigm for RNA removal reactions in HIV-1 reverse transcription. *Virus Res.*, 2013, 171, 2, 346-355.

Herschel, J. F. W. On a case of superficial colour presented by a homogeneous liquid internally colourless. *Philos. Trans. R. Soc. Lond.*, 1845, 135, 143-145.

Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.*, 1995, 270, 36, 20871-20874.

Hill, C. P., Worthylake D., Bancroft D. P., Christensen A. M. and Sundquist W. I. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 1996, 93, 7, 3099-3104.

Hogue, I. B., Hoppe A. and Ono A. Quantitative fluorescence resonance energy transfer microscopy analysis of the human immunodeficiency virus type 1 Gag-Gag interaction: relative contributions of the CA and NC domains and membrane binding. *J. Virol.*, 2009, 83, 14, 7322-7336.

Holzberger, B., Strohmeier J., Siegmund V., Diederichsen U. and Marx A. Enzymatic synthesis of 8-vinyl- and 8-styryl-2'-deoxyguanosine modified DNA-novel fluorescent molecular probes. *Bioorg. Med. Chem. Lett.*, 2012, 22, 9, 3136-3139.

Hoogsteen, K. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Cryst.*, 1963, 16, 907-916.

Houzet, L., Paillart J. C., Smagulova F., Maurel S., Morichaud Z., Marquet R. and Mougel M. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res.*, 2007, 35, 8, 2695-2704.

Hulme, A. E., Perez O. and Hope T. J. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.*, 2011, 108, 24, 9975-9980.

I

Ilgu, M., Fulton D. B., Yennamalli R. M., Lamm M. H., Sen T. Z. and Nilsen-Hamilton M. An adaptable pentaloop defines a robust neomycin-B RNA aptamer with conditional ligand-bound structures. *RNA*, 2014, 20, 6, 815-824.

Inglese, J., Johnson R. L., Simeonov A., Xia M., Zheng W., Austin C. P. and Auld D. S. High-throughput screening assays for the identification of chemical probes. *Nat. Chem. Biol.*, 2007, 3, 8, 466-479.

J

Jacks, T., Power M. D., Masiarz F. R., Luciw P. A., Barr P. J. and Varmus H. E. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, 1988, 331, 6153, 280-283.

Jager, S., Kim D. Y., Hultquist J. F., Shindo K., LaRue R. S., Kwon E., Li M., Anderson B. D., Yen L., Stanley D., Mahon C., Kane J., Franks-Skiba K., Cimermancic P., Burlingame A., Sali A., Craik C. S., Harris R. S., Gross J. D. and Krogan N. J. Vif hijacks CBF-beta to degrade APOBEC3G and promote HIV-1 infection. *Nature*, 2012, 481, 7381, 371-375.

Jalalirad, M. and Laughrea M. Formation of immature and mature genomic RNA dimers in wild-type and protease-inactive HIV-1: Differential roles of the Gag polyprotein, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology*, 2010, 407, 2, 225-236.

Jameson, D. M. *Introduction to fluorescence*. CRC Press Taylor and Francis Group, 2014.

Jameson, D. M. and Ross J. A. Fluorescence polarization/anisotropy in diagnostics and imaging. *Chem. Rev.*, 2010, 110, 5, 2685-2708.

Jares-Erijman, E. A. and Jovin T. M. FRET imaging. *Nat. Biotechnol.*, 2003, 21, 11, 1387-1395.

Jean, J. M. and Hall K. B. 2-Aminopurine electronic structure and fluorescence properties in DNA. *Biochemistry*, 2002, 41, 44, 13152-13161.

Johnson, J., Okyere R., Joseph A., Musier-Forsyth K. and Kankia B. Quadruplex formation as a molecular switch to turn on intrinsically fluorescent nucleotide analogs. *Nucleic Acids Res.*, 2013, 41, 1, 220-228.

Johnson, P. E., Turner R. B., Wu Z. R., Hairston L., Guo J. H., Levin J. G. and Summers M. F. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry*, 2000, 39, 31, 9084-9091.

Jouvenet, N., Laine S., Pessel-Vivares L. and Mougel M. Cell biology of retroviral RNA packaging. *RNA Biol.*, 2011, 8, 4, 572-580.

Jouvenet, N., Simon S. M. and Bieniasz P. D. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. U.S.A.*, 2009, 106, 45, 19114-19119.

Jovin, T. M., Soumpasis D. M. and Mcintosh L. P. The Transition between B-DNA and Z-DNA. *Annu. Rev. Phys. Chem.*, 1987, 38, 521-560.

K

Kafaie, J., Song R. J., Abrahamyan L., Mouland A. J. and Laughrea M. Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging. *Virology*, 2008, 375, 2, 592-610.

Kanevsky, I., Chaminade F., Ficheux D., Moumen A., Gorelick R., Negroni M., Darlix J. L. and Fosse P. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J. Mol. Biol.*, 2005, 348, 5, 1059-1077.

Karolin, J., Johansson L. B. A., Strandberg L. and Ny T. Fluorescence and absorption spectroscopic properties of dipyrrometheneboron difluoride (Bodipy) derivatives in liquids, lipid membranes, and proteins. *J. Am. Chem. Soc.*, 1994, 116, 17, 7801-7806.

Karpel, R. L., Henderson L. E. and Oroszlan S. Interactions of retroviral structural proteins with single-stranded nucleic-acids. *J. Biol. Chem.*, 1987, 262, 11, 4961-4967.

Kasha, M. Proton-transfer spectroscopy : perturbation of the tautomerization potential. *J. Chem. Soc., Faraday Trans.*, 1986, 82, 2379-2392.

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. and Summers M. F. RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*, 2015, 348, 6237, 917-921.

Keele, B. F., Van Heuverswyn F., Li Y., Bailes E., Takehisa J., Santiago M. L., Bibollet-Ruche F., Chen Y., Wain L. V., Liegeois F., Loul S., Ngole E. M., Bienvenue Y., Delaporte E., Brookfield J. F., Sharp P. M., Shaw G. M., Peeters M. and Hahn B. H. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 2006, 313, 5786, 523-526.

Kelley, S. O. and Barton J. K. Electron transfer between bases in double helical DNA. *Science*, 1999, 283, 5400, 375-381.

Kempf, N., Postupalenko V., Bora S., Didier P., Arntz Y., de Rocquigny H. and Mely Y. The HIV-1 nucleocapsid protein recruits negatively charged lipids to ensure its optimal binding to lipid membranes. *J. Virol.*, 2015, 89, 3, 1756-1767.

Kenfack, C. A., Piémont E., Ben Gaied N., Burger A. and Mély Y. Time-resolved fluorescent properties of 8-vinyl-deoxyadenosine and 2-amino-deoxyribosylpurine exhibit different sensitivity to their opposite base in duplexes. *J. Phys. Chem. B*, 2008, 112, 32, 9736-9745.

Kimsey, I. and Al-Hashimi H. M. Increasing occurrences and functional roles for high energy purine-pyrimidine base-pairs in nucleic acids. *Curr. Opin. Struct. Biol.*, 2014, 24, 72-80.

Klasens, B. I., Thiesen M., Virtanen A. and Berkhout B. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. *Nucleic Acids Res.*, 1999, 27, 2, 446-454.

Klatzmann, D., Champagne E., Chamaret S., Gruest J., Guetard D., Hercend T., Gluckman J. C. and Montagnier L. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature*, 1984, 312, 5996, 767-768.

Klymchenko, A. S. Solvatochromic fluorescent dyes as universal tools for biological research. *Actual. Chim.*, 2012, 359, 20-26.

Klymchenko, A. S. and Demchenko A. P. Multiparametric probing of intermolecular interactions with fluorescent dye exhibiting excited state intramolecular proton transfer. *Phys. Chem. Chem. Phys.*, 2003, 5, 3, 461-468.

Klymchenko, A. S. and Demchenko A. P. 3-hydroxychromone dyes exhibiting excited-state intramolecular proton transfer in water with efficient two-band fluorescence. *New J. Chem.*, 2004, 28, 6, 687-692.

Klymchenko, A. S. and Mely Y. Fluorescent environment-sensitive dyes as reporters of biomolecular interactions. *Prog. Mol. Biol. Transl. Sci.*, 2013, 113, 35-58.

Klymchenko, A. S., Pivovarenko V. G. and Demchenko A. P. Perturbation of planarity as the possible mechanism of solvent-dependent variations of fluorescence quantum yield in 2-aryl-3-hydroxychromones. *Spectrochim. Acta A. Mol. Biomol. Spectrosc.*, 2003, 59, 4, 787-792.

Kowada, T., Maeda H. and Kikuchi K. BODIPY-based probes for the fluorescence imaging of biomolecules in living cells. *Chem. Soc. Rev.*, 2015, 44, 14, 4953-4972.

Krishnamoorthy, G., Roques B., Darlix J. L. and Mely Y. DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection. *Nucleic Acids Res.*, 2003, 31, 18, 5425-5432.

Krueger, A. T. and Imperiali B. Fluorescent amino acids: modular building blocks for the assembly of new tools for chemical biology. *ChemBioChem*, 2013, 14, 7, 788-799.

Krueger, A. T. and Kool E. T. Fluorescence of size-expanded DNA bases: reporting on DNA sequence and structure with an unnatural genetic set. *J. Am. Chem. Soc.*, 2008, 130, 12, 3989-3999.

Kucherak, O. A., Didier P., Mely Y. and Klymchenko A. S. Fluorene analogues of prodan with superior fluorescence brightness and solvatochromism. *J. Phys. Chem. Lett.*, 2010, 1, 3, 616-620.

Kutluay, S. B. and Bieniasz P. D. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog.*, 2010, 6, 11, e1001200.

Kuzembayeva, M., Dilley K., Sardo L. and Hu W. S. Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology*, 2014, 454, 362-370.

Kuzmic, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, 1996, 237, 2, 260-273.

Kuzmic, P. Dynafit – a software package for enzymology. *Methods Enzymol.*, 2009, 467, 247-280.

Kuznetsova, A. A., Kuznetsov N. A., Vorobjev Y. N., Barthes N. P. F., Michel B. Y., Burger A. and Fedorova O. S. New environment-sensitive multichannel DNA fluorescent label for investigation of the protein-DNA interactions. *PLoS ONE*, 2014, 9, 6.

L

Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Kluwer Academic, New York, 2006.

Lapadat-Tapolsky, M., de Rocquigny H., Van Gent D., Roques B., Plasterk R. and Darlix J.-L. Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life-cycle *Nucleic Acids Res.*, 1993, 21, 8, 2024-2024.

Larsen, O. F. A., Van Stokkum I. H. M., Gobets B., Van Grondelle R. and Van Amerongen H. Probing the structure and dynamics of a DNA hairpin by ultrafast quenching and fluorescence depolarization. *Biophys. J.*, 2001, 81, 2, 1115-1126.

Laughrea, M., Jette L., Mak J., Kleiman L., Liang C. and Wainberg M. A. Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol.*, 1997, 71, 5, 3397-3406.

Lavis, L. D. and Raines R. T. Bright ideas for chemical biology. *ACS Chem. Biol.*, 2008, 3, 3, 142-155.

Law, S. M., Eritja R., Goodman M. F. and Breslauer K. J. Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine. *Biochemistry*, 1996, 35, 38, 12329-12337.

Le Cam, E., Coulaud D., Delain E., Petitjean P., Roques B. P., Gerard D., Stoylova E., Vuilleumier C., Stoylov S. P. and Mely Y. Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. *Biopolymers*, 1998, 45, 3, 217-229.

Lee, B. M., De Guzman R. N., Turner B. G., Tjandra N. and Summers M. F. Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.*, 1998, 279, 3, 633-649.

Lee, H. T., Waters L., Olsen C. M., Khutsishvili I. and Marky L. A. Probing the temperature unfolding of a variety of DNA secondary structures using the fluorescence properties of 2-aminopurine. *Acta Chim. Slov.*, 2012, 59, 3, 443-453.

Leonard, N. L., Morrice A. G. and Sprecker M. A. Linear benzoadenine. A stretched-out analog of adenine. *J. Org. Chem.*, 1975, 40, 3, 356-363.

Lever, A., Gottlinger H., Haseltine W. and Sodroski J. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.*, 1989, 63, 9, 4085-4087.

Levin, J. G., Guo J., Rouzina I. and Musier-Forsyth K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.*, 2005, 80, 217-286.

Levin, J. G., Mitra M., Mascarenhas A. and Musier-Forsyth K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.*, 2010, 7, 6, 754-774.

Levy, J. A., Hoffman A. D., Kramer S. M., Landis J. A. and Shimabukuro J. M. Isolation of lymphocytopathic retroviruses from San-Francisco patients with AIDS. *Science*, 1984, 225, 4664, 840-842.

Li, J. W. J., Fang X. H., Schuster S. M. and Tan W. H. Molecular beacons: a novel approach to detect protein - DNA interactions. *Angew. Chem. Int. Ed.*, 2000, 39, 6, 1049.

Li, M., Mizuuchi M., Burke T. R. and Craigie R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.*, 2006, 25, 6, 1295-1304.

Li, Y., Xuan S., Feng Y. and Yan A. Targeting HIV-1 integrase with strand transfer inhibitors. *Drug Discov. Today*, 2015, 20, 4, 435-449.

Lingappa, J. R., Dooher J. E., Newman M. A., Kiser P. K. and Klein K. C. Basic residues in the nucleocapsid domain of Gag are required for interaction of HIV-1 gag with ABCE1 (HP68), a cellular protein important for HIV-1 capsid assembly. *J. Biol. Chem.*, 2006, 281, 7, 3773-3784.

Liu, B. D., Dai R. K., Tian C. J., Dawson L., Gorelick R. and Yu X. F. Interaction of the human immunodeficiency virus type 1 nucleocapsid with actin. *J. Virol.*, 1999, 73, 4, 2901-2908.

Liu, C. and Martin C. T. Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *J. Mol. Biol.*, 2001, 308, 3, 465-475.

Liu, H., Gao J., Maynard L., Saito Y. D. and Kool E. T. Toward a new genetic system with expanded dimensions: size-expanded analogues of deoxyadenosine and thymidine. *J. Am. Chem. Soc.*, 2004, 126, 4, 1102-1109.

Liu, H. W., Cosa G., Landes C. F., Zeng Y., Kovaleski B. J., Mullen D. G., Barany G., Musier-Forsyth K. and Barbara P. F. Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys. J.*, 2005, 89, 5, 3470-3479.

Liu, H. W., Zeng Y., Landes C. F., Kim Y. J., Zhu Y., Ma X., Vo M. N., Musier-Forsyth K. and Barbara P. F. Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc. Natl. Acad. Sci. U S A*, 2007, 104, 13, 5261-5267.

Loudet, A. and Burgess K. BODIPY dyes and their derivatives: syntheses and spectroscopic properties. *Chem. Rev.*, 2007, 107, 11, 4891-4932.

Loving, G. and Imperiali B. A versatile amino acid analogue of the solvatochromic fluorophore 4-N,N-dimethylamino-1,8-naphthalimide: a powerful tool for the study of dynamic protein interactions. *J. Am. Chem. Soc.*, 2008, 130, 41, 13630-13638.

Loving, G. S., Sainlos M. and Imperiali B. Monitoring protein interactions and dynamics with solvatochromic fluorophores. *Trends Biotechnol.*, 2010, 28, 2, 73-83.

Lu, H., Krueger A. T., Gao J., Liu H. and Kool E. T. Toward a designed genetic system with biochemical function: polymerase synthesis of single and multiple size-expanded DNA base pairs. *Org. Biomol. Chem.*, 2010, 8, 12, 2704-2710.

Lu, K., Heng X., Garyu L., Monti S., Garcia E. L., Kharytonchyk S., Dorjsuren B., Kulandaivel G., Jones S., Hiremath A., Divakaruni S. S., LaCotti C., Barton S., Tummillo D., Hosic A., Edme K., Albrecht S., Telesnitsky A. and Summers M. F. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*, 2011, 334, 6053, 242-245.

Lu, K., Heng X. and Summers M. F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.*, 2011, 410, 4, 609-633.

Luo, K., Liu B. D., Xiao Z. X., Yu Y. K., Yu X. H., Gorelick R. and Yu X. F. Amino-terminal region of the human immunodeficiency virus type 1 nucleocapsid is required for human APOBEC3G packaging. *J. Virol.*, 2004, 78, 21, 11841-11852.

Luo, K., Wang T., Liu B. D., Tian C. J., Xiao Z. X., Kappes J. and Yu X. F. Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J. Virol.*, 2007, 81, 13, 7238-7248.

M

Malim, M. H. and Emerman M. HIV-1 accessory proteins : ensuring viral survival in a hostile environment. *Cell Host Microbe*, 2008, 3, 6, 388-398.

Mark-Danieli, M., Laham N., Kenan-Eichler M., Castiel A., Melamed D., Landau M., Bouvier N. M., Evans M. J. and Bacharach E. Single point mutations in the zinc finger motifs of the human immunodeficiency virus type 1 nucleocapsid alter RNA binding specificities of the gag protein and enhance packaging and infectivity. *J. Virol.*, 2005, 79, 12, 7756-7767.

Massiah, M. A., Starich M. R., Paschall C., Summers M. F., Christensen A. M. and Sundquist W. I. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. *J. Mol. Biol.*, 1994, 244, 2, 198-223.

McBride, M. S. and Panganiban A. T. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J. Virol.*, 1997, 71, 3, 2050-2058.

McCoy, L. S., Shin D. and Tor Y. Isomorphic emissive GTP surrogate facilitates initiation and elongation of in vitro transcription reactions. *J. Am. Chem. Soc.*, 2014, 136, 43, 15176-15184.

Melhuish, W. H. Quantum efficiencies of fluorescence of organic substan ces: effect of solvent and concentration of the fluorescent solute. *J. Phys. Chem.*, 1961, 65, 2, 229-235.

Mely, Y., Cornille F., Fourniezaluski M. C., Darlix J. L., Roques B. P. and Gerard D. Investigation of zinc-binding affinities of moloney murine leukemia virus nucleocapsid protein and its related zinc finger and modified peptides. *Biopolymers*, 1991, 31, 7, 899-906.

Mely, Y., de Rocquigny H., Morellet N., Roques B. P. and Gerad D. Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. *Biochemistry*, 1996, 35, 16, 5175-5182.

Mely, Y., Jullian N., Morellet N., de Rocquigny H., Dong C. Z., Piemont E., Roques B. P. and Gerard D. Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer. *Biochemistry*, 1994, 33, 40, 12085-12091.

Merrifield, R. B. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.*, 1963, 85, 14, 2149-2154.

Mirambeau, G., Lyonnais S., Coulaud D., Hameau L., Lafosse S., Jeusset J., Justome A., Delain E., Gorelick R. J. and Le Cam E. Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J. Mol. Biol.*, 2006, 364, 3, 496-511.

Montalti, M., Cantelli A. and Battistelli G. Nanodiamonds and silicon quantum dots: ultrastable and biocompatible luminescent nanoprobes for long-term bioimaging. *Chem. Soc. Rev.*, 2015, 44, 14, 4853-4921.

Moore, M. D., Nikolaitchik O. A., Chen J., Hammarskjold M. L., Rekosh D. and Hu W. S. Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog.*, 2009, 5, 10, e1000627.

Morellet, N., Déméné H., Teilleux V., Huynh-Dinh T., de Rocquigny H., Fournié-Zaluski M. C. and Roques B. P. Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J. Mol. Biol.*, 1998, 283, 2, 419-434.

Morellet, N., de Rocquigny H., Mely Y., Jullian N., Demene H., Ottmann M., Gerard D., Darlix J. L., Fourniezaluski M. C. and Roques B. P. Conformational behavior of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR. *J. Mol. Biol.*, 1994, 235, 1, 287-301.

Morellet, N., Jullian N., de Rocquigny H., Maigret B., Darlix J. L. and Roques B. P. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type-1 by 1H NMR. *EMBO J.*, 1992, 11, 8, 3059-3065.

Mori, M., Kovalenko L., Lyonnais S., Antaki D., Torbett B. E., Botta M., Mirambeau G. and Mely Y. Nucleocapsid protein: a desirable target for future therapies against HIV-1. *Curr. Top. Microbiol. Immunol.*, 2015, 389, 53-92.

Mougel, M., Houzet L. and Darlix J. L. When is it time for reverse transcription to start and go? *Retrovirology*, 2009, 6.

Mouland, A. J., Mercier J., Luo M., Bernier L., Des Groseillers L. and Cohen E. A. The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *J. Virol.*, 2000, 74, 12, 5441-5451.

Muriaux, D. and Darlix J. L. Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol.*, 2010, 7, 6, 744-753.

Muriaux, D., Darlix J. L. and Cimarelli A. Targeting the assembly of the human immunodeficiency virus type I. *Curr. Pharm. Des.*, 2004, 10, 30, 3725-3739.

Muriaux, D., de Rocquigny H., Roques B. P. and Paoletti J. NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem.*, 1996, 271, 52, 33686-33692.

Muriaux, D., Fosse P. and Paoletti J. A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry*, 1996, 35, 15, 5075-5082.

N

Nadler, A., Strohmeier J. and Diederichsen U. 8-vinyl-2′-deoxyguanosine as a fluorescent 2′-deoxyguanosine mimic for investigating DNA hybridization and topology. *Angew. Chem. Int. Ed. Engl.*, 2011, 50, 23, 5392-5396.

Nienhaus, K. and Nienhaus G. U. Fluorescent proteins for live-cell imaging with super-resolution. *Chem. Soc. Rev.*, 2014, 43, 4, 1088-1106.

Nikolova, E. N., Kim E., Wise A. A., O'Brien P. J., Andricioaei I. and Al-Hashimi H. M. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature*, 2011, 470, 7335, 498-U484.

Nitz, M., Mezo A. R., Ali M. H. and Imperiali B. Enantioselective synthesis and application of the highly fluorescent and environment-sensitive amino acid 6-(2-dimethylaminonaphthoyl) alanine (DANA). *Chem. Commun.*, 2002, 17, 1912-1913.

Nordlund, T. M., Xu D. and Evans K. O. Excitation energy transfer in DNA: duplex melting and transfer from normal bases to 2-aminopurine. *Biochemistry*, 1993, 32, 45, 12090-12095.

Nyamweya, S., Hegedus A., Jaye A., Rowland-Jones S., Flanagan K. L. and Macallan D. C. Comparing HIV-1 and HIV-2 infection: lessons for viral immunopathogenesis. *Rev. Med. Virol.*, 2013, 23, 4, 221-240.

O

O'Carroll, I. P., Soheilian F., Kamata A., Nagashima K. and Rein A. Elements in HIV-1 Gag contributing to virus particle assembly. *Virus Res.*, 2013, 171, 2, 341-345.

O'Neill, M. A. and Barton J. K. 2-Aminopurine: a probe of structural dynamics and charge transfer in DNA and DNA:RNA hybrids. *J. Am. Chem. Soc.*, 2002, 124, 44, 13053-13066.

Okamoto, A., Kanatani K. and Saito I. Pyrene-labeled base-discriminating fluorescent DNA probes for homogeneous SNP typing. *J. Am. Chem. Soc.*, 2004, 126, 15, 4820-4827.

Okamoto, A., Tainaka K., Nishiza K. and Saito I. Monitoring DNA structures by dual fluorescence of pyrene derivatives. *J. Am. Chem. Soc.*, 2005, 127, 38, 13128-13129.

Okamoto, A., Tainaka K., Ochi Y., Kanatani K. and Saito I. Simple SNP typing assay using a base-discriminating fluorescent probe. *Mol. Biosyst.*, 2006, 2, 2, 122-126.

Ono, A., Ablan S. D., Lockett S. J., Nagashima K. and Freed E. O. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 gag targeting to the plasma membrane. *Proc. Natl. Acad. Sci. U.S.A.*, 2004, 101, 41, 14889-14894.

Oreilly, M. M., Mcnally M. T. and Beemon K. L. 2 Strong 5'-splice sites and competing, suboptimal 3'-splice sites involved in alternative splicing of human immunodeficiency virus type-1 RNA. *Virology*, 1995, 213, 2, 373-385.

Ormo, M., Cubitt A. B., Kallio K., Gross L. A., Tsien R. Y. and Remington S. J. Crystal structure of the Aequorea victoria green fluorescent protein. *Science*, 1996, 273, 5280, 1392-1395.

Ott, D. E., Coren L. V. and Shatzer T. 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.*, 2009, 83, 15, 7718-7727.

Ottmann, M., Gabus C. and Darlix J. L. The central globular domain of the nucleocapsid protein of human immunodeficiency virus type-1 is critical for virion structure and infectivity. *J. Virol.*, 1995, 69, 3, 1778-1784.

P

Paillart, J. C., Berthoux L., Ottmann M., Darlix J. L., Marquet R., Ehresmann B. and Ehresmann C. A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J. Virol.*, 1996, 70, 12, 8348-8354.

Paillart, J. C., Marquet R., Skripkin E., Ehresmann B. and Ehresmann C. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type-1 genomic RNA. *J. Biol. Chem.*, 1994, 269, 44, 27486-27493.

Paillart, J. C., Shehu-Xhilaga M., Marquet R. and Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.*, 2004, 2, 6, 461-472.

Paillart, J. C., Skripkin E., Ehresmann B., Ehresmann C. and Marquet R. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. U.S.A.*, 1996, 93, 11, 5572-5577.

Pan, X. Y., Baldauf H. M., Keppler O. T. and Fackler O. T. Restrictions to HIV-1 replication in resting CD4+ T lymphocytes. *Cell Res.*, 2013, 23, 7, 876-885.

Parasassi, T., Krasnowska E. K., Bagatolli L. and Gratton E. Laurdan and prodan as polarity-sensitive fluorescent membrane probes. *J. Fluoresc.*, 1998, 8, 4, 365-373.

Park, S., Otomo H., Zheng L. and Sugiyama H. Highly emissive deoxyguanosine analogue capable of direct visualization of B-Z transition. *Chem. Commun. (Camb.)*, 2014, 50, 13, 1573-1575.

Peterlin, B. M., Luciw P. A., Barr P. J. and Walker M. D. Elevated levels of messenger-RNA can account for the transactivation of human immunodeficiency virus. *Proc. Natl. Acad. Sci. U.S.A.*, 1986, 83, 24, 9734-9738.

Petryayeva, E., Algar W. R. and Medintz I. L. Quantum dots in bioanalysis: a review of applications across various platforms for fluorescence spectroscopy and imaging. *Appl. Spectrosc.*, 2013, 67, 3, 215-252.

Pettit, S. C., Lindquist J. N., Kaplan A. H. and Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology*, 2005, 2, 66.

Pierce, D. W. and Boxer S. G. Stark effect spectroscopy of tryptophan. *Biophys. J.*, 1995, 68, 4, 1583-1591.

Popov, S., Popova E., Inoue M. and Gottlinger H. G. Human immunodeficiency virus type 1 Gag engages the Bro1 domain of ALIX/AIP1 through the nucleocapsid. *J. Virol.*, 2008, 82, 3, 1389-1398.

Popovic, M., Sarin P. S., Robert-Gurroff M., Kalyanaraman V. S., Mann D., Minowada J. and Gallo R. C. Isolation and transmission of human retrovirus (human T-cell leukemia virus). *Science*, 1983, 219, 4586, 856-859.

Postupalenko, V. Y., Shvadchak V. V., Duportail G., Pivovarenko V. G., Klymchenko A. S. and Mély Y. Monitoring membrane binding and insertion of peptides by two-color fluorescent label. *Biochim. Biophys. Acta*, 2011, 1808, 1, 424-432.

Postupalenko, V. Y., Zamotaiev O. M., Shvadchak V. V., Strizhak A. V., Pivovarenko V. G., Klymchenko A. S. and Mely Y. Dual-fluorescence L-amino acid reports insertion and orientation of melittin peptide in cell membranes. *Bioconjug. Chem.*, 2013, 24, 12, 1998-2007.

Prendergast, F. G., Meyer M., Carlson G. L., Iida S. and Potter J. D. Synthesis, spectral properties, and use of 6-acryloyl-2-dimethylaminonaphthalene (Acrylodan) : a thiol-selective, polarity-sensitive fluorescent probe. *J. Biol. Chem.*, 1983, 258, 12, 7541-7544.

Purcell, D. F. J. and Martin M. A. Alternative splicing of human immunodeficiency virus type-1 messenger-RNA modulates viral protein expression, replication, and infectivity. *J. Virol.*, 1993, 67, 11, 6365-6378.

R

Rachofsky, E. L., Alexander Ross J. B. and Osman R. Conformation and dynamics of normal and damaged DNA. *Comb. Chem. High Throughput Screen.*, 2001, 4, 8, 675-706.

Rachofsky, E. L., Osman R. and Ross J. B. A. Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry*, 2001, 40, 4, 946-956.

Ramalanjaona, N., de Rocquigny H., Millet A., Ficheux D., Darlix J. L. and Mely Y. Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J. Mol. Biol.*, 2007, 374, 4, 1041-1053.

Rausch, J. W. and Le Grice S. F. HIV Rev assembly on the Rev Response Element (RRE): a structural perspective. *Viruses*, 2015, 7, 6, 3053-3075.

Rein, A. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol.*, 2010, 7, 6, 700-705.

Rein, A., Henderson L. E. and Levin J. G. Nucleic acid chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.*, 1998, 23, 8, 297-301.

Ren, X. F., Schweitzer B. A., Sheils C. J. and Kool E. T. Formation of stable DNA loops by incorporation of nonpolar, non-hydrogen-bonding nucleoside isosteres. *Angew. Chem. Int. Ed.*, 1996, 35, 7, 743-746.

Ross, J. A. and Jameson D. M. Time-resolved methods in biophysics. Frequency domain fluorometry: applications to intrinsic protein fluorescence. *Photochem. Photobiol. Sci.*, 2008, 7, 11, 1301-1312.

Rovira, A. R., Fin A. and Tor Y. Chemical mutagenesis of an emissive RNA alphabet. *J. Am. Chem. Soc.*, 2015, 137, 46, 14602-14605.

Royer, C. A. Probing protein folding and conformational transitions with fluorescence. *Chem. Rev.*, 2006, 106, 5, 1769-1784.

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

S

Saad, J. S., Miller J., Tai J., Kim A., Ghanam R. H. and Summers M. F. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 2006, 103, 30, 11364-11369.

Sakakibara, Y., Abeysirigunawardena S. C., Duc A. C. E., Dremann D. N. and Chow C. S. Ligand- and pH-induced conformational changes of RNA domain helix 69 revealed by 2-aminopurine fluorescence. *Angew. Chem. Int. Ed. Engl.*, 2012, 51, 48, 12095-12098.

Samanta, P. K. and Pati S. K. Structural, electronic and photophysical properties of analogous RNA nucleosides: a theoretical study. *New J. Chem.*, 2013, 37, 11, 3640-3646.

Sapsford, K. E., Berti L. and Medintz I. L. Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. *Angew. Chem. Int. Ed.*, 2006, 45, 28, 4562-4589.

Saroja, G., Soujanya T., Ramachandram B. and Samanta A. 4-Aminophthalimide derivatives as environment-sensitive probes. *J. Fluoresc.*, 1998, 8, 4, 405-410.

Schafer, A., Bogerd H. P. and Cullen B. R. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. *Virology*, 2004, 328, 2, 163-168.

Schroeder, R., Barta A. and Semrad K. Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell. Biol.*, 2004, 5, 11, 908-919.

Schweitzer, B. A. and Kool E. T. Aromatic nonpolar nucleosides as hydrophobic isosteres of pyrimidine and purine nucleosides. *J. Org. Chem.*, 1994, 59, 24, 7238-7242.

Scopes, D. I., Barrio J. R. and Leonard N. J. Defined dimensional changes in enzyme cofactors: fluorescent "stretched-out" analogs of adenine nucleotides. *Science*, 1977, 195, 4275, 296-298.

Seela, F., Jiang D. W. and Xu K. Y. 8-Aza-2'-deoxyguanosine: base pairing, mismatch discrimination and nucleobase anion fluorescence sensing in single-stranded and duplex DNA. *Org. Biomol. Chem.*, 2009, 7, 17, 3463-3473.

Sengupta, B., Reilly S. M., Davis D. E., Harris K., Wadkins R. M., Ward D., Gholar D. and Hampton C. Excited state proton transfer of natural flavonoids and their chromophores in duplex and tetraplex DNAs. *J. Phys. Chem. B*, 2015, 119, 6, 2546-2556.

Sengupta, P. K. and Kasha M. Excited state proton-transfer spectroscopy of 3-hydroxyflavone and quercetin. *Chem. Phys. Lett.*, 1979, 68, 2-3, 382-385.

Sette, P., Dussupt V. and Bouamr F. Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA. *J. Virol.*, 2012, 86, 21, 11608-11615.

Sharma, A., Doucette C., Biro F. N. and Hingorani M. M. Slow conformational changes in MutS and DNA direct ordered transitions between mismatch search, recognition and signaling of DNA repair. *J. Mol. Biol.*, 2013, 425, 22, 4192-4205.

Sharma, K. K., Przybilla F., Restle T., Boudier C., Godet J. and Mely Y. Reverse transcriptase in action: FRET-based assay for monitoring flipping and polymerase activity in real time. *Anal. Chem.*, 2015, 87, 15, 7690-7697.

Shehu-Xhilaga, M., Kraeusslich H. G., Pettit S., Swanstrom R., Lee J. Y., Marshall J. A., Crowe S. M. and Mak J. Proteolytic processing of the P2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J. Virol.*, 2001, 75, 19, 9156-9164.

Shen, N., Jette L., Liang C., Wainberg M. A. and Laughrea M. Impact of human immunodeficiency virus type 1 RNA dimerization on viral infectivity and of stem-loop B on RNA dimerization and reverse transcription and dissociation of dimerization from packaging. *J. Virol.*, 2000, 74, 12, 5729-5735.

Shimomura, O., Johnson F. H. and Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.*, 1962, 59, 223-239.

Shin, D., Sinkeldam R. W. and Tor Y. Emissive RNA alphabet. *J. Am. Chem. Soc.*, 2011, 133, 38, 14912-14915.

Sholokh, M., Sharma R., Shin D., Das R., Zaporozhets O. A., Tor Y. and Mely Y. Conquering 2-aminopurine's deficiencies: highly emissive isomorphic guanosine surrogate faithfully monitors guanosine conformation and dynamics in DNA. *J. Am. Chem. Soc.*, 2015, 137, 9, 3185-3188.

Sholokh, M., Zamotaiev O. M., Das R., Postupalenko V. Y., Richert L., Dujardin D., Zaporozhets O. A., Pivovarenko V. G., Klymchenko A. S. and Mely Y. Fluorescent amino acid undergoing excited state intramolecular proton transfer for site-specific probing and imaging of peptide interactions. *J. Phys. Chem. B*, 2015, 119, 6, 2585-2595.

Shynkar, V. V., Mely Y., Duportail G., Piemont E., Klymchenko A. S. and Demchenko A. P. Picosecond time-resolved fluorescence studies are consistent with reversible excited-state intramolecular proton transfer in 4'-(dialkylaminoe)-3-hydroxyflavones. *J. Phys. Chem. A*, 2003, 107, 45, 9522-9529.

Silvin, A. and Manel N. Innate immune sensing of HIV infection. *Curr. Opin. Immunol.*, 2015, 32, 54-60.

Sinkeldam, R. W., Greco N. J. and Tor Y. Fluorescent analogs of biomolecular building blocks: design, properties, and applications. *Chem. Rev.*, 2010, 110, 5, 2579-2619.

Skripkin, E., Paillart J. C., Marquet R., Ehresmann B. and Ehresmann C. Identification of the primary site of the human immunodeficiency virus type-1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, 1994, 91, 11, 4945-4949.

Slavik, J. Anilinonaphthalene sulfonate as a probe of membrane composition and function. *Biochim. Biophys. Acta*, 1982, 694, 1, 1-25.

Sleiman, D., Barraud P., Brachet F. and Tisne C. The interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. *PLoS ONE*, 2013, 8, 6.

Sleiman, D., Goldschmidt V., Barraud P., Marquet R., Paillart J. C. and Tisne C. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res.*, 2012, 169, 2, 324-339.

Soujanya, T., Fessenden R. W. and Samanta A. Role of non-fluorescent twisted intramolecular charge transfer state on the photophysical behavior of aminophthalimide dyes. *J. Phys. Chem.*, 1996, 100, 9, 3507-3512.

Sowers, L. C., Fazakerley G. V., Eritja R., Kaplan B. E. and Goodman M. F. Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl. Acad. Sci. U.S.A.*, 1986, 83, 15, 5434-5438.

Spriggs, S., Garyu L., Connor R. and Summers M. F. Potential intra- and intermolecular interactions involving the unique 5' region of the HIV-1 5'-UTR. *Biochemistry*, 2008, 47, 49, 13064-13073.

Srivatsan, S. G. and Tor Y. Fluorescent pyrimidine ribonucleotide: synthesis, enzymatic incorporation, and utilization. *J. Am. Chem. Soc.*, 2007, 129, 7, 2044-2053.

St-Pierre, P., McCluskey K., Shaw E., Penedo J. C. and Lafontaine D. A. Fluorescence tools to investigate riboswitch structural dynamics. *Biochim. Biophys. Acta*, 2014, 1839, 10, 1005-1019.

Stauffer, S., Rahman S. A., de Marco A., Carlson L. A., Glass B., Oberwinkler H., Herold N., Briggs J. A. G., Muller B., Grunewald K. and Krausslicha H. G. The nucleocapsid domain of Gag is dispensable for actin incorporation into HIV-1 and for association of viral budding sites with cortical F-actin. *J. Virol.*, 2014, 88, 14, 7893-7903.

Stewart-Maynard, K. M., Cruceanu M., Wang F., Vo M. N., Gorelick R. J., Williams M. C., Rouzina I. and Musier-Forsyth K. Retroviral nucleocapsid proteins display non-equivalent levels of nucleic acid chaperone activity. *J. Virol.*, 2008, 82, 20, 10129-10142.

Stivers, J. T. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Res.*, 1998, 26, 16, 3837-3844.

Stote, R. H., Kellenberger E., Muller H., Bombarda E., Roques B. P., Kieffer B. and Mely Y. Structure of the His44 – Ala single point mutant of the distal finger motif of HIV-1 nucleocapsid protein: a combined NMR, molecular dynamics simulation, and fluorescence study. *Biochemistry*, 2004, 43, 24, 7687-7697.

Stoylov, S. P., Vuilleumier C., Stoylova E., de Rocquigny H., Roques B. P., Gerard D. and Mely Y. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*, 1997, 41, 3, 301-312.

Strizhak, A. V., Postupalenko V. Y., Shvadchak V. V., Morellet N., Guittet E., Pivovarenko V. G., Klymchenko A. S. and Mély Y. Two-color fluorescent L-amino acid mimic of tryptophan for probing peptide-nucleic acid complexes. *Bioconjug. Chem.*, 2012, 23, 12, 2434-2443.

Summerer, D., Chen S., Wu N., Deiters A., Chin J. W. and Schultz P. G. A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U.S.A.*, 2006, 103, 26, 9785-9789.

Summerer, D. and Marx A. A molecular beacon for quantitative monitoring of the DNA polymerase reaction in real-time. *Angew. Chem. Int. Ed.*, 2002, 41, 19, 3620-3622, 3516.

Summers, M. F., Henderson L. E., Chance M. R., Bess J. W., South T. L., Blake P. R., Sagi I., Perezalvarado G., Sowder R. C., Hare D. R. and Arthur L. O. Nucleocapsid zinc fingers detected in retroviruses : exafs studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein Sci.*, 1992, 1, 5, 563-574.

Sundquist, W. I. and Krausslich H. G. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.*, 2012, 2, 7, a006924.

Swanstrom, R. and Wills J. W. *Synthesis, assembly, and processing of viral proteins*. In: Coffin JM, Hughes SH, Varmus HE, eds. Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 263–334, 1997.

Swinney, T. C. and Kelley D. F. Proton transfer dynamics in substituted 3-hydroxyflavones: solvent polarization effects. *J. Chem. Phys.*, 1993, 99, 1, 211-221.

T

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

Tanaka, M., Robinson B. A., Chutiraka K., Geary C. D., Reed J. C. and Lingappa J. R. Mutations of conserved residues in the major homology region arrest assembling HIV-1 Gag as a membrane-targeted intermediate containing genomic RNA and cellular proteins. *J. Virol.*, 2015, 90, 4, 1944-1963.

Tanchou, V., Gabus C., Rogemond V. and Darlix J. L. Formation of stable and functional HIV-1 nucleoprotein complexes in vitro. *J. Mol. Biol.*, 1995, 252, 5, 563-571.

Teale, F. W. and Weber G. Ultraviolet fluorescence of the aromatic amino acids. *Biochem. J.*, 1957, 65, 3, 476-482.

Telesnitsky, A. and Goff S. P. *Reverse transcriptase and the generation of retroviral DNA*. In: Retroviruses. J. M. Coffin, S. H. Hughes and H. E. Varmus, Cold Spring Harbor, NY, 1997.

Theilleux-Delalande, V., Girard F., Huynh-Dinh T., Lancelot G. and Paoletti J. The HIV-1Lai RNA dimerization : thermodynamic parameters associated with the transition from the kissing complex to the extended dimer. *Eur. J. Biochem.*, 2000, 267, 9, 2711-2719.

Thomas, J. A. and Gorelick R. J. Nucleocapsid protein function in early infection processes. *Virus Res.*, 2008, 134, 1-2, 39-63.

Tisne, C., Roques B. P. and Dardel F. Specific recognition of primer tRNALys 3 by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension. *Biochimie*, 2003, 85, 5, 557-561.

Tisne, C., Roques B. P. and Dardel F. The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J. Biol. Chem.*, 2004, 279, 5, 3588-3595.

Tomin, V. I., Demchenko A. P. and Chou P. T. Thermodynamic vs. kinetic control of excited-state proton transfer reactions. *J. Photochem. Photobiol. C, Photochem. Rev.*, 2015, 22, 1-18.

Tomin, V. I., Oncul S., Smolarczyk G. and Demchenko A. P. Dynamic quenching as a simple test for the mechanism of excited-state reaction. *Chem. Phys.*, 2007, 342, 1-3, 126-134.

Tor, Y. Exploring RNA-ligand interactions. *Pure Appl. Chem.*, 2009, 81, 2, 263-272.

Trono, D. Partial Reverse Transcripts in virions from human immunodeficiency and murine leukemia viruses. *J. Virol.*, 1992, 66, 8, 4893-4900.

Tyagi, S. and Kramer F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, 1996, 14, 3, 303-308.

U

Ueno, T., Tokunaga K., Sawa H., Maeda M., Chiba J., Kojima A., Hasegawa H., Shoya Y., Sata T., Kurata T. and Takahashi H. Nucleolin and the packaging signal, psi, promote the budding of human immunodeficiency virus type-1 (HIV-1). *Microbiol. Immunol.*, 2004, 48, 2, 111-118.

Ulrich, G., Ziessel R. and Harriman A. The chemistry of fluorescent bodipy dyes: versatility unsurpassed. *Angew. Chem. Int. Ed.*, 2008, 47, 7, 1184-1201.

V

Valeur, B. *Invitation à la fluorescence meléculaire*. De Boeck and Larcier, 2004.

Vazquez, M. E., Blanco J. B. and Imperiali B. Photophysics and biological applications of the environment-sensitive fluorophore 6-N,N-Dimethylamino-2,3-naphthalimide. *J. Am. Chem. Soc.*, 2005, 127, 4, 1300-1306.

Vazquez, M. E., Nitz M., Stehn J., Yaffe M. B. and Imperiali B. Fluorescent caged phosphoserine peptides as probes to investigate phosphorylation-dependent protein associations. *J. Am. Chem. Soc.*, 2003, 125, 34, 10150-10151.

Vazquez, M. E., Rothman D. M. and Imperiali B. A new environment-sensitive fluorescent amino acid for Fmoc-based solid phase peptide synthesis. *Org. Biomol. Chem.*, 2004, 2, 14, 1965-1966.

Venzke, S. and Keppler O. T. Role of macrophages in HIV infection and persistence. *Expert Rev. Clin. Immunol.*, 2006, 2, 4, 613-626.

Vo, M. N., Barany G., Rouzina I. and Musier-Forsyth K. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J. Mol. Biol.*, 2006, 363, 1, 244-261.

Vo, M. N., Barany G., Rouzina I. and Musier-Forsyth K. HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J. Mol. Biol.*, 2009, 386, 3, 789-801.

Votteler, J. and Sundquist W. I. Virus budding and the ESCRT pathway. *Cell Host Microbe.*, 2013, 14, 3, 232-241.

Vuilleumier, C., Bombarda E., Morellet N., Gerard D., Roques B. P. and Mely Y. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: A fluorescence study. *Biochemistry*, 1999, 38, 51, 16816-16825.

W

Wahl, M. C. and Sundaralingam M. Crystal structures of A-DNA duplexes. *Biopolymers*, 1997, 44, 1, 45-63.

Walker, C. L., Lukyanov K. A., Yampolsky I. V., Mishin A. S., Bommarius A. S., Duraj-Thatte A. M., Azizi B., Tolbert L. M. and Solntsev K. M. Fluorescence imaging using synthetic GFP chromophores. *Curr. Opin. Chem. Biol.*, 2015, 27, 64-74.

Wan, C., Fiebig T., Schiemann O., Barton J. K. and Zewail A. H. Femtosecond direct observation of charge transfer between bases in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 2000, 97, 26, 14052-14055.

Ward, D. C., Reich E. and Stryer L. Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J. Biol. Chem.*, 1969, 244, 5, 1228-1237.

Weber, G. and Farris F. J. Synthesis and spectral properties of a hydrophobic fluorescent probe: 6-propionyl-2-(dimethylamino)naphthalene. *Biochemistry*, 1979, 18, 14, 3075-3078.

Wegner, K. D. and Hildebrandt N. Quantum dots: bright and versatile in vitro and in vivo fluorescence imaging biosensors. *Chem. Soc. Rev.*, 2015, 44, 14, 4792-4834.

Widom, J. R., Rappoport D., Perdomo-Ortiz A., Thomsen H., Johnson N. P., Von Hippel P. H., Aspuru-Guzik A. and Marcus A. H. Electronic transition moments of 6-methyl

isoxanthopterin : a fluorescent analogue of the nucleic acid base guanine. *Nucleic Acids Res.*, 2013, 41, 2, 995-1004.

Wilhelmsson, L. M. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.*, 2010, 43, 2, 159-183.

Wilk, T., Gowen B. and Fuller S. D. Actin associates with the nucleocapsid domain of the human immunodeficiency virus Gag polyprotein. *J. Virol.*, 1999, 73, 3, 1931-1940.

Williams, M. C., Rouzina I., Wenner J. R., Gorelick R. J., Musier-Forsyth K. and Bloomfield V. A. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl. Acad. Sci. U S A*, 2001, 98, 11, 6121-6126.

Wilson, J. N. and Kool E. T. Fluorescent DNA base replacements: reporters and sensors for biological systems. *Org. Biomol. Chem.*, 2006, 4, 23, 4265-4274.

Winnacker, M. and Kool E. T. Artificial genetic sets composed of size-expanded base pairs. *Angew. Chem. Int. Ed.*, 2013, 52, 48, 12498-12508.

Winter, R. B., Berg O. G. and von Hippel P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. The Escherichia coli lac repressor – operator interaction: kinetic measurements and conclusions. *Biochemistry*, 1981, 20, 24, 6961-6977.

Wu, H., Mitra M., McCauley M. J., Thomas J. A., Rouzina I., Musier-Forsyth K., Williams M. C. and Gorelick R. J. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res.*, 2013, 171, 2, 263-277.

Wu, H., Rouzina I. and Williams M. C. Single-molecule stretching studies of RNA chaperones. *RNA Biology*, 2010, 7, 6, 712-723.

X

Xie, Y., Dix A. V. and Tor Y. FRET enabled real time detection of RNA-small molecule binding. *J. Am. Chem. Soc.*, 2009, 131, 48, 17605-17614.

Y

Yang, H., Jeffrey P. D., Miller J., Kinnucan E., Sun Y., Thoma N. H., Zheng N., Chen P. L., Lee W. H. and Pavletich N. P. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science*, 2002, 297, 5588, 1837-1848.

Yang, Z., Cao J., He Y., Yang J. H., Kim T., Peng X. and Kim J. S. Macro-/micro-environment-sensitive chemosensing and biological imaging. *Chem. Soc. Rev.*, 2014, 43, 13, 4563-4601.

Yedavalli, V. S. and Jeang K. T. Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, 2010, 107, 33, 14787-14792.

Yengo, C. M. and Berger C. L. Fluorescence anisotropy and resonance energy transfer: powerful tools for measuring real time protein dynamics in a physiological environment. *Curr. Opin. Pharmacol.*, 2010, 10, 6, 731-737.

You, J. C. and McHenry C. S. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.*, 1994, 269, 50, 31491-31495.

Z

Zamotaiev, O. M., Postupalenko V. Y., Shvadchak V. V., Pivovarenko V. G., Klymchenko A. S. and Mély Y. Improved hydration-sensitive dual-fluorescence labels for monitoring peptide-nucleic acid interactions. *Bioconjug. Chem.*, 2011, 22, 1, 101-107.

Zargarian, L., Tisne C., Barraud P., Xu X. Q., Morellet N., Rene B., Mely Y., Fosse P. and Mauffret O. Dynamics of linker residues modulate the nucleic acid binding properties of the HIV-1 nucleocapsid protein zinc fingers. *PLoS ONE*, 2014, 9, 7.

Zeng, Y., Liu H. W., Landes C. F., Kim Y. J., Ma X., Zhu Y., Musier-Forsyth K. and Barbara P. F. Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer. *Proc. Natl. Acad. Sci. U.S.A.*, 2007, 104, 31, 12651-12656.

Zhang, G., Zheng S. Q., Liu H. P. and Chen P. R. Illuminating biological processes through site-specific protein labeling. *Chem. Soc. Rev.*, 2015, 44, 11, 3405-3417.

Zhang, Y., Qian H., Love Z. and Barklis E. Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain. *J. Virol.*, 1998, 72, 3, 1782-1789.

Zhao, L. and Xia T. Probing RNA conformational dynamics and heterogeneity using femtosecond time-resolved fluorescence spectroscopy. *Methods*, 2009, 49, 2, 128-135.

Zheng, R., Jenkins T. M. and Craigie R. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc. Natl. Acad. Sci. U.S.A.*, 1996, 93, 24, 13659-13664.

Zimmerman, C., Klein K. C., Kiser P. K., Singh A. R., Firestein B. L., Riba S. C. and Lingappa J. R. Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature*, 2002, 415, 6867, 88-92.

Zimmerman, S. B. The 3-Dimensional Structure of DNA. *Annu. Rev. Biochem.*, 1982, 51, 395-427.

Zuniga, S., Sola I., Cruz J. L. and Enjuanes L. Role of RNA chaperones in virus replication. *Virus Res.*, 2009, 139, 2, 253-266.

Marianna SHOLOKH

Caractérisation site-sélective de la dynamique des propriétés chaperonnes de la protéine de la nucléocapside de VIH-1 vis-à-vis de ses cibles nucléiques, à l'aide de sondes fluorescentes innovantes

Résumé

Du fait de sa haute conservation et de ses fonctions clés dans le virus VIH-1, la protéine de la nucléocapside NC est une cible de choix pour développer de nouveaux anti-viraux. Bien que la compréhension mécanistique des propriétés chaperonnes de NC vis-à-vis des acides nucléiques ait connu d'importants progrès, les aspects dynamiques de ces propriétés restent mal comprises, faute notamment d'outils appropriés pour les suivre au niveau moléculaire.

L'objectif de ce travail de thèse a été de caractériser au niveau moléculaire la dynamique des interactions de NC avec les acides nucléiques, à l'aide d'outils fluorescents développés au laboratoire ou en collaboration. En utilisant des peptides NC et des oligonucléotides marqués en différentes positions par des analogues fluorescents d'acides aminés et de nucléosides, respectivement, nous avons pu donner une image complète du processus dynamique sous-tendant le mécanisme chaperon de NC dans le second transfert de brins de la transcription inverse du cycle du VIH-1. Cette compréhension est fondamentale pour concevoir des stratégies rationnelles afin de cibler le rôle spécifique de NC dans ses interactions avec ses cibles nucléiques.

Mots-clés : protéine de la nucléocapside, propriétés chaperonnes, transcription inverse, second transfert de brins, analogues fluorescents, 3-hydroxychromone, thiénodéoxyguanosine.

Résumé en anglais

Due to its high conservation and key functions in the HIV-1 virus, the nucleocapsid protein NC is a potential target for the development of new anti-viral drugs. Although the mechanistic understanding of the NC nucleic acid chaperone properties has achieved a significant progress, the dynamic aspects of these properties remain poorly understood, mainly due to the lack of the appropriate tools to monitor them at the molecular level.

The objective of this thesis was to characterize the dynamic interactions of NC with nucleic acids at the molecular level using new fluorescent tools developed in the laboratory or in collaboration. Using NC peptides and its target oligonucleotides labeled in different positions by fluorescent amino acid and nucleoside analogs, respectively, we were able to give a complete picture of the dynamic processes underlying the chaperone activity of NC in the second strand transfer of HIV-1 reverse transcription. This understanding is fundamental to design rational strategies in order to target the specific role of NC in interaction with its nucleic acid targets.

Keywords: nucleocapsid protein, chaperone properties, reverse transcription, second strand transfer, fluorescent analogs, 3-hydroxychromone, thienodeoxyguanosine.