UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Biologie Moléculaire et Cellulaire

(UPR 9002)

THÈSE présentée par :

Elodie MAILLER

soutenue le **22 septembre 2017**

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

Sciences de la Vie et de la Santé

Biologie moléculaire et cellulaire

Structural rearrangements of the HIV-1 genomic RNA during maturation of the viral particle

THÈSE dirigée par :

Dr. Roland Marquet	IBMC, Strasbourg
Dr. Valérie Vivet-Boudou	IBMC, Strasbourg

RAPPORTEURS :

Dr. Marylène Mougel	IRIM, Montpellier
Dr. Gilles Mirambeau	UPMC, Paris

AUTRES MEMBRES DU JURY :

Dr. Carine Tisné	IBPC, Paris
Dr. David Gilmer	IBMP, Strasbourg

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury d'avoir accepté d'évaluer mon travail de doctorat, comportant de nombreuses structures et données de cartographie chimique. Je tiens également à remercier le ministère de l'enseignement supérieur et de la recherche pour le contrat doctoral ayant financé mes trois premières années de doctorat et la Fondation pour la Recherche Médicale pour m'avoir permis de réaliser une quatrième année.

J'aimerais remercier mes co-directeurs de thèse, Valérie et Roland pour leur soutien et leur implication dans mon encadrement. Un grand merci à Roland pour tous tes conseils scientifiques et pour avoir été là dans les moments difficiles. Je tiens également à te remercier pour ton aide précieuse pour la correction de ce manuscrit. Valérie, merci de m'avoir initiée au monde de la cartographie de l'ARN. Tu m'as donné ma chance en stage d'été et permis d'intégrer cette géniale équipe. Je tiens à remercier Redmond pour avoir créé ce projet. Merci pour ton encadrement au cours de ces quatre années à la paillasse et ton aide pour le traitement des données brutes de séquençage à haut débit. Ton impressionnante connaissance des techniques et protocoles m'a été précieuse pour le développement du hSHAPE-Seq.

J'aimerais remercier tout particulièrement le seul, l'unique, je veux bien sur parler de JC (alias M. Maybe). Merci d'avoir toujours été là, que cela soit pour débattre de mes résultats, pour ta rigueur scientifique ;-), pour des problèmes techniques (et il en a eu beaucoup), pour faire office de psy et/ou pour boire un coup. Sans toi ma thèse n'aurait pas été pareil. J'aimerais remercier notre belge plus blanc que blanc, Noé, pour son soutien au cours de ces cinq dernières années. Tu es présent dans ma vie depuis le stage de master 2 et ce fameux concours. J'ai adoré nos pauses déjeuner, café, bière ;-) Merci également d'avoir tenté de me tuer plusieurs fois (désolée d'être aussi coriace). Que la force de p6 soit avec toi pour la fin de ta thèse. Merci à vous deux pour tous nos délires et fous rire !!!

Remerciements collectifs mais pas moins importants pour les autres membres de l'équipe : Serena, mon italienne fashion victime préférée, si tu te fais cambrioler et qu'il ne manque que des vêtements, accessoires, chaussures, ce sera moi mais je serai déjà loin ;-). Maud ton arrivée dans l'équipe a été un vrai bonheur et tu me surprends de jour en jour. Surtout ne change rien. Tanja, ce fut un plaisir de t'avoir comme voisine de bureau/paillasse et d'échanger sur nos malheurs de qPCR, SHAPE et streptavidine. Que la force soit avec vous deux pour vos projets. Bien sûr je n'oublie pas les anciens membres de l'équipe, Camille avec qui j'ai débuté cette aventure, Marie, Jujuuuu et Jérémy notre marseillais qui m'a fait découvrir le pakito.

J'aimerais remercier la team animation et fou rire à gogo, j'ai nommé Phiphi (épouse Paillart), Guillaume (juste un bouchon), Emma (grâce à toi j'adore l'accent alsacien), Antoine (mister infos insolites), Alexis, Safi (miss cosmo) et Marine (miss poney). Je me restreins en infos croustillantes sinon les remerciements feraient la taille du manuscrit.

J'aimerais remercier mes parents sans qui je n'en serais pas là où j'en suis aujourd'hui. Merci pour votre soutien inconditionnel, votre patience :-), votre amour et de m'avoir inculqué toutes ces valeurs qui font ce que je suis. Amandine, Logan, Gwen et Kevin, vous faites partie de ma famille. Un énorme merci pour nos repas de famille, nos parties de UNO endiablées (oui n'ayons pas peur des mots), nos fous rires et les nombreuses sessions skype à venir.

Last but not least, j'aimerais remercier Damien, un lyonnais barbu, qui est arrivé dans ma vie comme un boulet de canon et est tout simplement devenu essentiel à ma vie et mon bonheur. Merci à toi.

ABREVIATIONS

TABLE OF FIGURES

ABBREVIATIONS

*	radioactively labelled
ΔDIS	DIS defective virus
A	
A	Ala/alanine
a.a.	amino acid
Ab	antibody
AMV	Avian myeloblastosis virus
asp, ASP	anti-sense protein
Asp	asparagine
AT-2	aldrithiol-2
ATVs	atazanavir sulfate
C	
°C	Celsius degree
C	Cys/cysteine
CA	capsid
CCR5	C-C chemokine receptor type 5
CD4	cluster of differentiation 4
cryoEM	cryo-electron microscopy
cryoET	cryo-electron tomography
cpm	counts per minute
CTD	C-terminal domain
CXCR4	C-X-C chemokine receptor type 4
D	
db	double-stranded
DLS	dimer linkage structure
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	di-nucleoside tri-phosphate
E	
EM	Electron Microscopy
env, Env	envelop
ESCRT	endosomal sorting complexes required for transport
F	
Fw	forward
G	
gag, Gag	group-specific antigen
G	Gly/glycine
gRNA	genomic ribonucleic acid
H	
H	His/histidine
HBR	highly basic region
HEK293T	human embryonic kidney 293 stably expressing the SV40 large T antigen
HIV-1	human immunodeficiency virus type 1
HRP	horseradish peroxidase
I	
IC_{50}	50 % inhibitory concentration

Ig	immunoglobulin
A1/2	illumina incomplete adapter 1/2
IN	integrase
INSTI	IN strand transfer inhibitor

K

kb	kilobase
kDa	kiloDalton

L

LEDGF	lens epithelium-derived growth factor
L	Leu/leucine
LINE	long interspersed element
LPV	lopinavir
LTR	long terminal repeat
Lys	lysine
LysRS	lysyl-tRNA synthetase

M

MA	matrix
miRNA	micro RNA
MHR	major homology region
Mo-MuLV	Moloney murine leukemia virus
mRNA	messenger RNA

N

NAI	2-methylnicotinic acid imidazolide
NCp7	nucleoCapsid
ncRNA	Non-coding RNA
nef, Nef	negative factor
NGS	next-generation sequencing
NMD	non-sense-mediated mRNA decay
NMIA	N-methyl isatoic anhydride
NMR	nuclear magnetic resonance
NCINI	non-catalytic site IN inhibitor
NPC	nuclear pore complex
NTD	N terminal domain

O

ORF	open reading frame

P

P	Pro/proline
PAGE	polyacrylamide gel electrophoresis
PAS	primer activation signal
PBS	primer binding site
PCR	polymerase chain reaction
PEI	polyethylenImine
PI	protease inhibitor
$PI(4,5)P_2$	phosphatidylinositol-4,5-bisphosphate
PIC	pre-integration complex
PM	plasma membrane
pol	polymerase
Pol II	RNA polymerase II
PR	protease
PR-	PR defective virus
$Pr55^{Gag}$	Gag precursor of 55 kDalton
$Pr55^{Gag-Pol}$	GagPol precursor of 160 kDalton

R

R	repeated
rev, Rev	regulator of virion protein expression
RFU	relative fluorescence unit
RNA	ribonucleic acid
RNP	ribonucleoprotein
RRE	Rev responsive element
RT	reverse transcriptase
RTion	reverse transcription
RT-PCR	reverse transcription-polymerase chain reaction
Rv	reverse
S	
SDS	sodium dodecyl sulfate
SHAPE	selective 2'-hydroxyl acylated analysed by primer extension
SL	stem loop
snRNA	small nuclear RNA
ssRNA	single-stranded RNA
SEM	standard error of the mean
STED	stimulated emission depletion
SU-gp120	surface glycoprotein of 120 kDalton
T	
T	Thr/threonine
TAR	trans-activation response element
tat, Tat	trans-activator of transcription
TM-gp41	TransMembrane GlycoProtein of 41 kDalton
TΨC	thymidine-pseudouridine-cytidine
tRNA	transfert RNA
U	
U5	unique in 5'
U3	unique in 3'
UTR	untranslated region
V	
vif, Vif	viral infectivity factor
vpr, Vpr	viral protein R
vpu, Vpu	viral protein U
VSV	Vesicular stomatitis virus
W	
WB	western blot
Y	
Y	Tyr/tyrosine
Z	
ZF	zinc finger

TABLE OF FIGURES

INTRODUCTION

INTRODUCTION

Human immunodeficiency type 1 virus

I. HIV-1: general information

I.1 Taxonomic classification

Human immunodeficiency virus type 1 (HIV-1) belongs to the *Retroviridae*, a large family of viruses found in all vertebrates. The International Committee on Taxonomy of Viruses has divided the *retrovirus* family in two groups: simple retroviruses composed of the *alpha retroviruses*, *beta retroviruses* and *gamma retroviruses*; and complex retroviruses composed of the delta retroviruses, epsilon retroviruses, lentiviruses and spumaviruses (239). All retroviruses encode *gag* (group-specific antigen), *pol* (polymerase) and *env* (envelope) gene products, and complex retroviruses encode additional regulatory proteins (95).

HIV-1 is part of the lentiviruses and is divided in four groups: M (main), O (outlier), N (non-M, non-O) and P. The M group includes more than 98% of the HIV-1 isolates and is divided into nine clades (reviewed in (180)).

I.2 HIV-1 components and virus morphology

Retroviruses were originally classified based on their characteristic morphology. The viral particle is spherical, surrounded by a membrane derived from the host lipid membrane bilayer and comprising the glycosylated surface glycoprotein 120 (SU-gp120) and transmembrane glycoprotein 41 (TM-gp41) envelope proteins. Approximately ten trimers of Env proteins per virion are thought to be present on each particle (442).

Virions are released from the infected cell initially as immature particles of roughly 120 nm in diameter with a characteristic electron-lucent centre (57, 58). Approximately 2500 (68) to 5000 (58) Gag precursors ($Pr55^{Gag}$) are present per particle with a GagPol precursors ($Pr160^{GagPol}$) level of about one-twentieth per virion (198).

Processing of $Pr55^{Gag}$ and $Pr160^{GagPol}$ drastically changes the virus morphology. Underlying the membrane of the mature viral particles the matrix (MA) layer, which contains a fullerene cone composed of capsid (CA) proteins (155). The HIV-1 cone encompasses a ribonucleoprotein (RNP) complex consisting of two copies of genomic ribonucleic acid (gRNA) coated by nucleoprotein (NCp7), several viral proteins (reverse transcriptase (RT), integrase (IN), protease (PR), viral protein R (Vpr), viral infectivity factor (Vif) and negative factor (Nef)) and some cellular factors such as the transfert RNA Lysine 3 isoacceptor ($tRNA^{Lys,3}$) RT primer, actin, APOBEC-3G and cyclophilin A (reviewed in (317)).

I.3 Organisation of the genome

The HIV-1 proviral genome, integrated into the host genome, is flanked from either side with identical duplicated regions called 5' and 3' viral long terminal repeats (LTRs). 5' and 3' LTRs comprise U3 (Unique in 3'), R (Repeated) and U5 (Unique in 5') domains (**Figure 1B**).

The U3 domain constitutes the HIV-1 promoter that controls transcription. This region fulfils its biological purpose once the RNA genome is reverse transcribed into DNA provirus. R and U5 domains contain number of *cis*-acting elements recognised by RNAs and proteins, regulating crucial steps of the viral life cycle such as transcriptional elongation, splicing, gRNA dimerisation, packaging of the unspliced genome and RT.

Once RTion is accomplished and the provirus integrated, a positive-sense and single-stranded RNA (ssRNA) of 9,2 kilobases (kb) is produced by the host transcriptional machinery with a trimethylguanosine cap at the 5' end and a polyadenylated tail at the 3' end (436). The gRNA is flanked by untranslated regions (UTR) of approximately 350 and 550 nucleotides at the 5' and 3' ends, respectively containing R and U5 domains and U3 and R domains (**Figure 1A**).

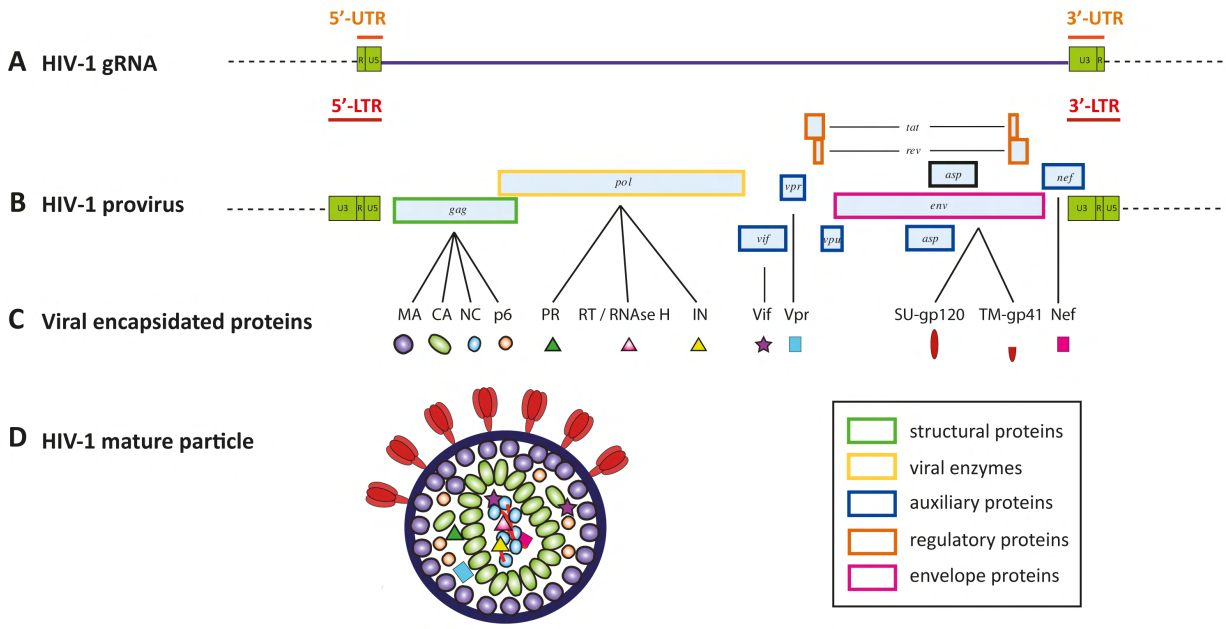


Figure 1
Overview of the organisation and expression of the HIV-1 genome.
A. Proviral genomic organisation with the coding region flanked by the two LTR regions. **B.** gRNA organisation highlighting the ten overlapping ORFs with the 5' and 3' untranslated regions. **C.** Viral proteins expressed and encapsidated inside the viral particle. **D.** Schematic representation of the mature viral particle highlighting the localisation of the different proteins.

The 5'-UTR (**Figure 2**) folds into several relatively independent structural and functional domains (28, 90, 174, 220, 319, 383, 422, 426). From 5' to 3' these are the TAR (trans-activation response element) stem-loop mediating efficient transcriptional elongation; the Poly-A stem-loop containing the repressed 5' copy of the polyadenylation signal; the PBS (primer binding site) domain crucial for annealing to the host $tRNA^{Lys,3}$ primer for initiation of RTion; the SL1 domain (stem-loop 1) containing the gRNA dimerisation initiation site (DIS) element involved in gRNA dimerisation; the SL2 domain containing the major splice donor (SD) site; the SL3 domain considered as the historical major packaging signal; SL4 containing the AUG start codon of $Pr55^{Gag}$, and forming an unstable stem-loop whose dynamic structure probably depends on the replicative step of the life cycle.

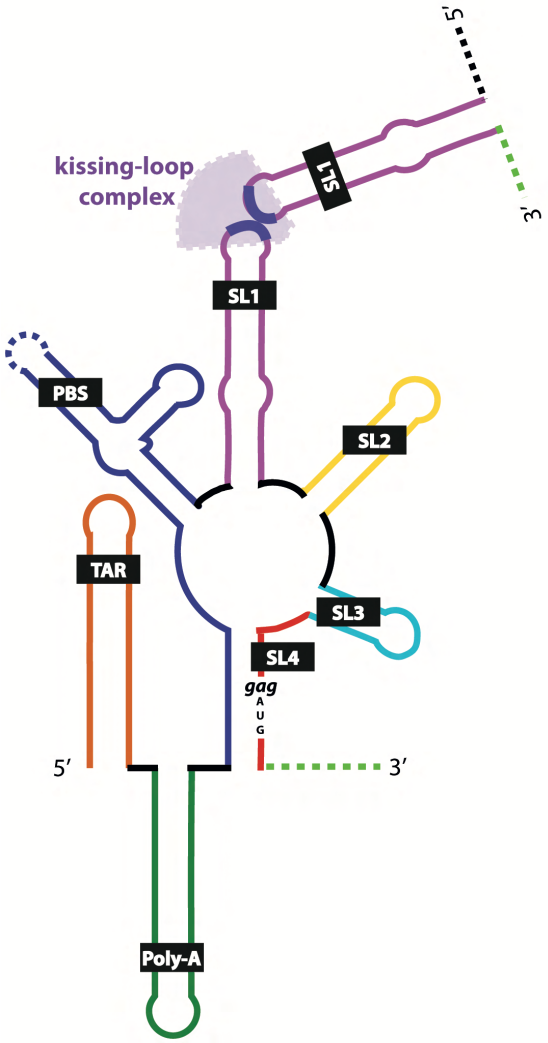


Figure 2
Model of the secondary structure of the HIV-1 gRNA 5'-UTR.
The 5'-UTR comprises the first 350 nucleotides of the gRNA. The main domains are indicated: TAR, Poly-A, PBS (nucleotides 131 to 217 are not indicated to facilitate the representation), SL1, SL2, SL3 and SL4 containing the AUG start codon of gag. The DIS is highlighted in the context of the kissing-loop complex mediating gRNA dimerisation (Model from (388)).

The 3'-UTR of gRNA comprises the U3 region followed by the R region, like the 5' end but with the fully functional poly-adenylation signal. It is important to keep in mind that the AAUAAA and GU/U-rich binding sites, localised in the Poly-A hairpin, are part of the R domain and thus located at each end of RNA. A polyadenylation enhancer localised upstream of the 3' poly-adenylation signal stabilises the binding of factors required for poly(A) addition, explaining the non-functionality of the 5' poly(A) site.

Regarding the coding region, the HIV-1 genome consists of ten overlapping open reading frames (ORFs): *gag*, *pol*, *vif*, *vpr*, *rev (*regulator of virion protein expression), *tat* (trans-activator of transcription), *vpu* (viral protein U), *env*, *asp* (antisense protein) and *nef*. The existence of *asp* has been recently demonstrated, this ORF being localised in an antisense orientation overlapping *env* (69).

The *gag* and *pol* ORFs encode $Pr55^{Gag}$ (to produce the structural proteins MA, CA, NCp7, p6) and $Pr160^{GagPol}$ (to produce the enzymatic proteins IN, PR and RT). The eight other ORFs produce SU-gp120 and TM-gp41, regulatory (Tat and Rev) and auxiliary (Vif, Vpr, Vpu, Nef and Asp) proteins, described in the following section "the HIV-1 life cycle" of the manuscript. **Figure 1C** describes viral proteins that are packaged with their location inside the virus highlighted in **Figure 1D**.

The core Psi, responsible for the correct packaging of gRNA inside viral particle, comprises the SL1 to SL4 domains. A fully functional packaging signal may also include the TAR, Poly-A and PBS domains plus the first 300 nucleotides of *gag*. This part will be discussed in detail within the section "The life-cycle of the HIV-1 Gag/RNA complex".

Packaging of a dimeric genome has two evolutionary advantages. First, a back-up template is present to correct mutations generated during the polymerisation reaction and impacting viral infectivity. This phenomenon is explained by the high replication rate and the lack of proofreading activity of the HIV-1 RT. Second, genetic diversity is conferred *via* RNA template switching generated by the RT (reviewed in (315)) (94).

II. Overview of the HIV-1 life cycle

Considered as an intracellular parasite, the HIV-1 strictly depends on the host targeted cell to reproduce itself and hijacks transcriptional and translational machinery to reproduce itself. Up to one billion virions are produced per day in an infected individual, and this viral population exists as a quasi-species due to high mutation rate of the RT. The replicative cycle of HIV-1 ends with the production of new infectious viral particles able to initiate the next round of replication. This process is usually divided into two main stages (**Figure 3**). The early phase comprises the binding of the viral particle with the host cell and its subsequent fusion, the release of the cone into the cytoplasm with RTion of gRNA, uncoating of the cone, nuclear import and DNA provirus integration into the host genome. The late phase begins with the transcription and splicing of viral RNAs, viral RNA export into the cytoplasm, the translation of viral RNAs into viral proteins, through to the assembly, release and maturation of viral particles.

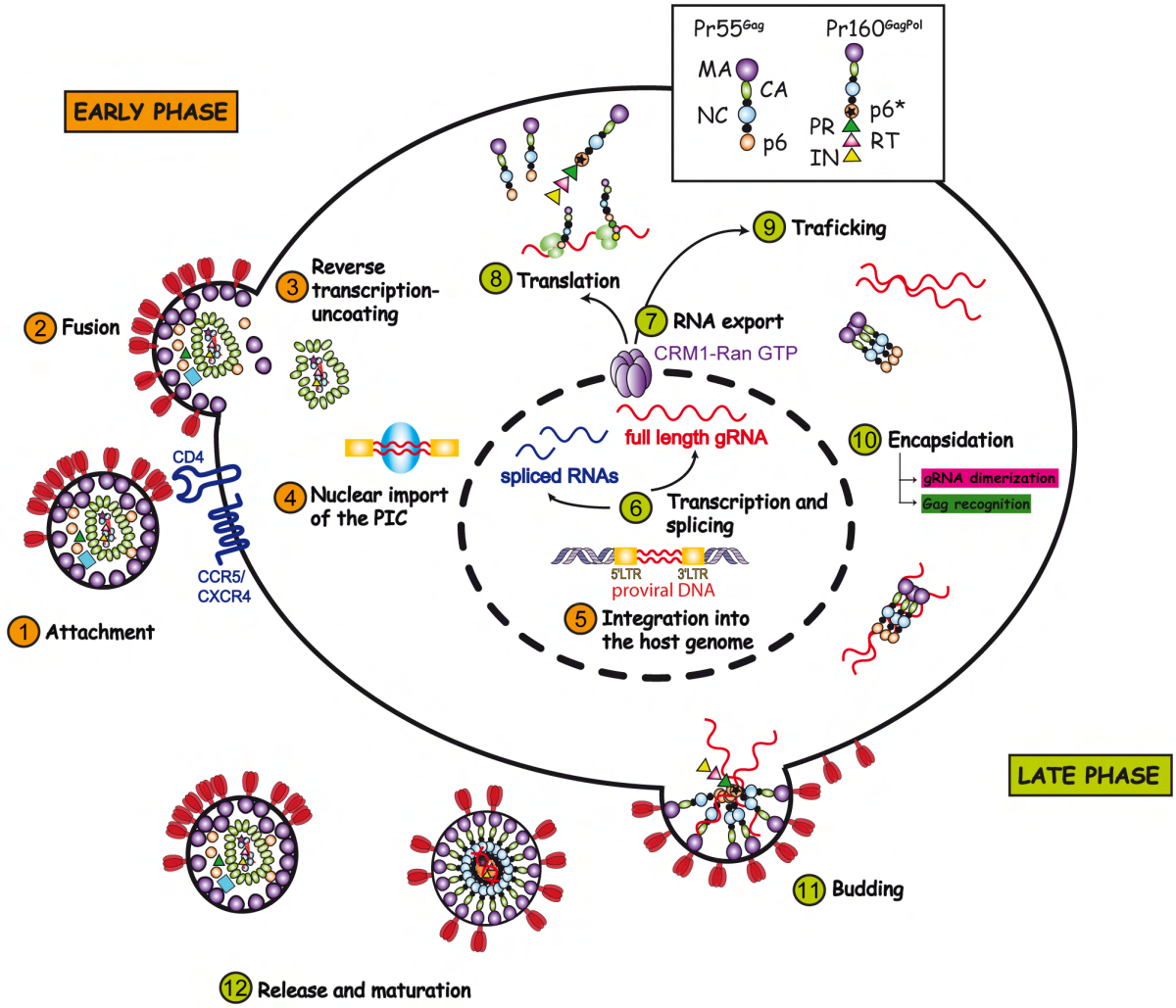


Figure 3
Schematic representation of the HIV-1 replicative cycle.
This figure highlights the different steps, from 1 to 12, crucial to produce a fully infective viral particle. Steps occurring during the early and late phases are respectively coloured in orange and green.

II.1 Early phase

II.1.1 Viral attachment and fusion

Viral entry into the host cell is mediated by interactions between the viral envelope proteins and specific cellular receptors, determining the viral cell tropism. The external SU-gp120 mediates viral attachment while the TM-gp41 promotes viral fusion. The two processed proteins form a non-covalent heterodimer (reviewed in (78)) with the gp41 TM domain anchored in the plasma membrane by a hydrophobic stop-transfer signal (47, 170). The gp120-gp41 complex is associated as a trimer in the active form (209, 274). The gp120 sub-unit is composed of five variable domains (V1-V5) interspersed with five constant domains (C1-C5) and is heavily glycosylated. Approximately half of its molecular mass is composed of approximately 20 to 35 N-linked glycans with a small additional contribution from O-linked sugars (78).

Cluster of differentiation 4 (CD4) is the primary receptor of HIV-1 entry and is associated with the two main co-receptors: CCR5 (C-C chemokine Receptor type 5) and CXCR4 (C-X-C chemokine Receptor type 4) (93, 137). Viral attachment is also performed in an unspecific manner through the binding of SU-gp120 to host cell lectins and proteoglycans. Incorporation of host cell factors into the viral envelope also contributes because they are still able to bind to their specific target on the cellular membrane (reviewed in (218)).

The infection is initiated by the SU-gp120/CD4 receptor interaction, inducing a conformational change in the Env trimer. SU-gp120 then interacts with CCR5/CXCR4 co-receptors inducing further rearrangements. TM-gp41 is finally able to insert its hydrophobic fusion peptide into the cellular membrane. This interaction allows fusion and release of the viral capsid into the cell by creating a helical bundle structure (189).

II.1.2 Reverse transcription

Partial uncoating of the capsid, which contains the viral genetic information protected by the NCp7, begins during migration of these particles from the cortical actin barrier to the microtubule network where RTion is initiated (20, 257). The timing of HIV-1 core disassembly is essential for the formation of a functional RTion complex leading to viral double-stranded DNA (dsDNA) synthesis by the packaged RT (146, 403).

Conversion of positive ssRNA into dsDNA with complete LTR harboring sites for initiation and termination of transcription requires a complex series of events (reviewed in (406)). Briefly, the $tRNA^{Lys,3}$ bound to PBS serves as a primer to initiate negative ssDNA synthesis. The RNA segment of the RNA/DNA hybrid is degraded by the RNase H activity of RT. The first DNA strand transfer occurs between the nascent negative DNA and the R region at the 3' end of the gRNA and allows the negative DNA synthesis to resume. RNA is further degraded excepted at two polypurine tracts. These sites provide the RT primers for the positive DNA strand synthesis. Removal of the $tRNA^{Lys,3}$ primer allows the positive strand DNA transfer and the central termination at the central flap located in IN ORF (73, 74).

II.1.3 Uncoating and nuclear entry

Several models of uncoating have been proposed: immediate, cytoplasmic and nuclear pore complex (NPC) uncoating (reviewed in (64)). Considering the protective function of the core regarding the RNP complex, several studies postulate that uncoating starts several hours post-entry into the cytoplasm of the infected cell and is critical for viral infectivity (188, 434). First strand transfer of RTion then induces capsid disassembly (97, 342). Nevertheless, the exact timing of uncoating is still poorly understood.

The microtubule network is necessary for the movement of the RTion complex toward the nuclear pore (20, 63), the remnant CA lattice (435) and the DNA flap (438) could be actively involved in its nuclear import. The RTion complex is then remodeled into the pre-integration complex (PIC) containing the dsDNA plus several host (high-mobility group proteins, lamina-associated polypeptide 2α, lens epithelium-derived growth factor (LEDGF)/transcriptional coactivator p75) and viral (RT, IN, MA, CA, and Vpr) proteins (reviewed in (399)). Since HIV-1 productively infects both dividing and non-dividing cells, this implies an active import of the PIC into the nucleus (reviewed in (282)). Some tRNA species (437) and CA proteins (435) modulate interaction of the PIC with the nuclear transport machinery (273).

II.2 Late phase

II.2.1 Integration

Once inside the nucleus, the dsDNA is integrated into the host chromosomal DNA, followed by gene expression and productive infection. Proviral integration into host chromatin is required for high levels of viral expression, unintegrated dsDNA supporting only limited transcription (409). The packaged IN and many cellular proteins (reviewed in (98)) facilitate insertion of the dsDNA into active transcription units (117, 267).

Integration is divided in three steps (98). First, two nucleotides at the 3' ends of dsDNA are removed in the cytoplasm by IN. This step generates CA-3' recessed hydroxyl ends, which induce an asymmetric break of the host genome and a simultaneous joining by a one-step transesterification mechanism. This strand transfer ends with the repair of each side of the viral genome, completing the stable insertion of the provirus (132).

Upon integration, IN remains associated with the viral DNA ends as part of the PIC, along with several host proteins such as LEDGF/p75. This protein complex mediates the tethering of the IN-viral DNA complex to the host chromatin (reviewed in (131)).

II.2.2 Transcription and splicing

After integration, the provirus is dependent on the cell machinery for gene expression. RNA polymerase II (Pol II), in association with cellular transcriptional trans-activating proteins such as NF-kB, NFAT and Sp1, synthesise the full-length HIV-1 RNA transcript, driven by viral promoters and enhancers localised in the U3 region of the

5' LTR. U3 is divided in three main part: the modulatory region, the enhancer and the core promoter, constituting the transcription factor binding site (reviewed in (53)).

Initial transcriptional output is low due to poor Pol II processivity. Processive transcription requires binding of the viral protein Tat to the TAR domain (251) and the association of the cellular co-factor p-TEFb allowing hyper-phosphorylation of the C-terminal domain (CTD) of Pol II.

This 9-kb transcript is polycistronic and contains multiple alternative 5' and 3' splicing sites (**Figure 4B**) (reviewed in (217)) in order to generate more than 100 different transcripts (312).

Different classes of transcripts are generated: 9 kb, 4 kb and 2 kb (128, 136, 312, 340). In addition, a new class of 1 kb transcripts has recently been proposed (312) (**Figure 4C**).

- the 9 kb class contains the unspliced viral RNA which encodes $Pr55^{Gag}$ and $Pr160^{GagPol}$ and also serves as the gRNA being packaged.

- the 4 kb class comprises singly-spliced mRNAs encoding auxiliary and Env proteins. Auxiliary proteins modify the cellular environment to render it compatible with HIV-1 replication:

- Vif counteracts the host viral restriction factor APOBEC-3G/3F
- Vpu is implicated in the CD4 downregulation and promotes the release of virions from infected cells by inhibiting the host BST2/tetherin restriction factor
- Vpr is implicated in the PIC complex and promotes viral replication in non-dividing cells
- the two envelope proteins SU-gp120 and TM-gp41 are required for binding and entry into target cells

- the 2 kb class comprises multi-spliced RNAs generated by alternative splicing and encoding auxiliary proteins that drive viral replication. These RNAs are expressed early in the viral life cycle:

- Tat enhances viral transcription
- Rev allows nuclear export of unspliced and singly-spliced RNAs
- Nef is implicated in the CD4 downregulation so the cell is less visible to the immune system and counteracts restriction factors SERINC 3/5 (355, 415)

- the 1 kb class has recently been identified and comprises 29 completely spliced RNAs, containing exons of Tat, Rev and Nef. These novel fusion products are potential candidates for acting as regulatory RNAs (312).

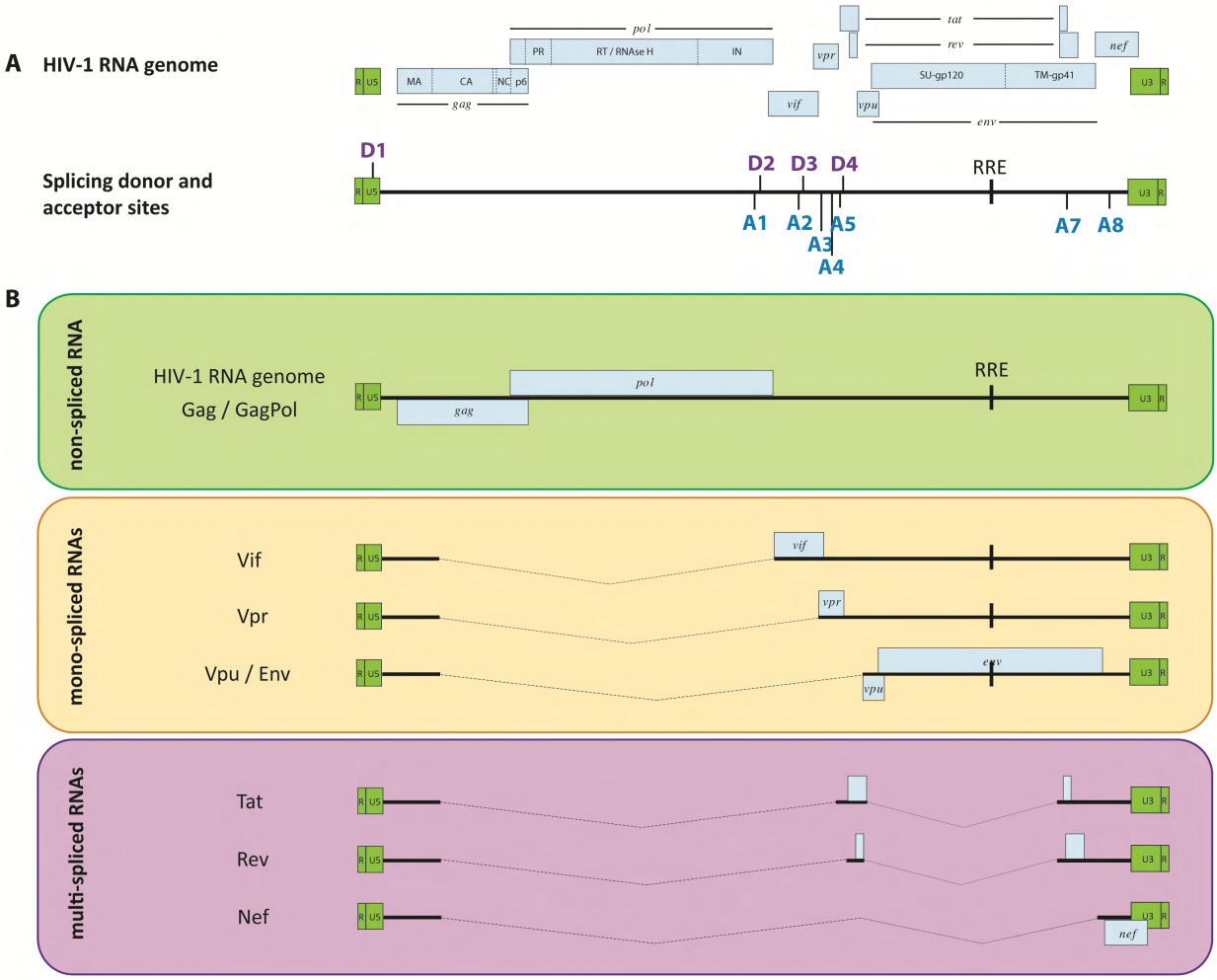


Figure 4
Overview of the splicing donor, acceptor sites and of the spliced viral RNAs derived from the gRNA.
A. Full-length gRNA highlighting splice donor (D1 to D4, purple) and acceptor (A1 to A8, blue) sites. The location of the RRE is shown. B. Representation of the different populations of viral RNAs: non-spliced of 9 kb, mono-spliced of 4 kb and multi-spliced of 2 kb coding the viral proteins.

II.2.3 Nuclear export of viral RNAs, post-transcriptional regulation translation

Rev mediates the export of unspliced and singly-spliced viral RNAs from the nucleus to the cytoplasm, as detailed in the section "The life-cycle of the HIV-1 Gag/RNA complex".

Eukaryotic cells post-transcriptionally regulate gene expression by recognising and degrading aberrant mRNAs. This quality control is ensured by multiple strategies including non-sense-mediated mRNA decay (NMD), no-go decay (NGD) and nonstop decay (NSD) (Reviewed in (382). NMD is a translation-coupled mechanism triggering degradation of polycistronic messenger RNAs (mRNAs) containing premature translation-termination codons (reviewed in (60, 279)). Given the fact that the polycistronic nature of a genome, as well as the retention of intronic sequences are key factors triggering NMD, how does HIV-1 counteract this mRNA decay machinery?

HIV-1 hijacks cellular factors involved in different mRNA decay pathways. As examples, the up-frameshift protein 1 (UPF-1) (9, 10), the Staufen protein (6, 27, 75–77, 123) and the YTH-domain containing family 2 (YTHDF2) (224) seem to be critical in determining the fate of the unspliced mRNA. The UPF-1 factor, one of the major factors of the NMD pathway, is required for HIV-1 RNA stability both in the nucleus and cytoplasm, translation (10) as well as for viral RNA nuclear export (9).

UPF1 increases the levels of viral RNAs and the expression of $Pr55^{Gag}$ proteins during the replication cycle. Indeed, UPF-1 regulates the fate of the unspliced mRNA by enhancing its association with nuclear export factors and forming a specific complex with DDX3, Rev and Chromosomal Maintenance 1 (CRM1) (9). Thus, HIV-1 ensures nuclear export of the gRNA through the recruitment of UPF-1 and by the exclusion of UPF-2. But the precise function of UPF-2 remains incompletely understood. Human Staufen-2 has also been shown to interact with HIV-1 Rev to enhance its gRNA export activity (27). DDX3 also initiates HIV-1 unspliced mRNA by binding the trimethyl cap and forming a pre-initiation complex in the absence of the initiation factor eIF4E (391). Nevertheless, the role of many host factors remains unknown as well as how exactly the recruitment of these host proteins interferes with mRNA degradation processes.

RNAs encoding Tat, Rev and Nef are expressed early in the viral life cycle to allow synthesis of the full-length genome (8, 102, 276). Structural and enzymatic proteins are expressed from this unspliced mRNA as the $Pr55^{Gag}$ and $Pr160^{GagPol}$ polyproteins, that are later cleaved into their mature subunits by the PR enzyme. In parallel, Env, Vif, Vpr and Vpu proteins are expressed from the mono-spliced mRNAs. All proteins are expressed in polysomes localised in the cytoplasm, except for the Env proteins which are synthesised in the endoplasmic reticulum (reviewed in (400)).

II.2.4 Assembly

II.2.4.1 The Life-Cycle of the HIV-1 Gag–RNA Complex

$Pr55^{Gag}$ and $Pr160^{GagPol}$ multimerise in the cytoplasm and migrate to the plasma membrane (PM) where the assembly of the newly viral particle occurs, with the concomitant recruitment of Env proteins in addition to Vif, Nef, Vpr and cellular factors. In parallel, two copies of gRNA must traffic to the PM to be packaged. The assembly process from initial $Pr55^{Gag}$ detection in the cytoplasm to the completion of $Pr55^{Gag}$ accumulation at the PM has been estimated in the range of 5 to 9 minutes (208). These aspects are further detailed in the review article “The Life-Cycle of the HIV-1 Gag–RNA Complex” published in “Viruses”.

II.2.4.2 $tRNA^{Lys}$ annealing and packaging

An important aspect of the replicative cycle which has not been developed in the review is the packaging of tRNAs into newly formed viral particles. Indeed, during HIV-1 assembly, cellular $tRNA^{Lys}$ isoacceptors $tRNA^{Lys,1}$, $tRNA^{Lys,2}$ and $tRNA^{Lys,3}$ are specifically incorporated into the viral particle (Reviewed in (237)) (205, 281). The function of $tRNA^{Lys,1}$ and $tRNA^{Lys,2}$ is elusive, these $tRNA^{Lys,1 \text{ and } 2}$ probably being encapsidated because the Lysyl-tRNA synthetase (LysRS) does not discriminates between isoacceptors, whereas $tRNA^{Lys,3}$ is the primer for initiating RTion.

Four regions of the $tRNA^{Lys,3}$ participate in its placement and stabilisation to the PBS domain (reviewed in (195)). These four interactions are: (1) the 3' end of the $tRNA^{Lys,3}$ stem acceptor anneals to an 18-base sequence located in the 5'-UTR and termed the PBS to initiate the RTion (281), (2) the variable loop of $tRNA^{Lys,3}$ interacts with the C-rich region (194, 196), (3) the thymidine-pseudouridine-cytidine (TΨC) arm of $tRNA^{Lys,3}$ interacts with the primer activation signal (PAS) (29–32, 159, 160, 387), and (4) the anti-codon loop of $tRNA^{Lys,3}$ complementary to the A-loop (160, 193, 194, 196) (**Figure 5**).

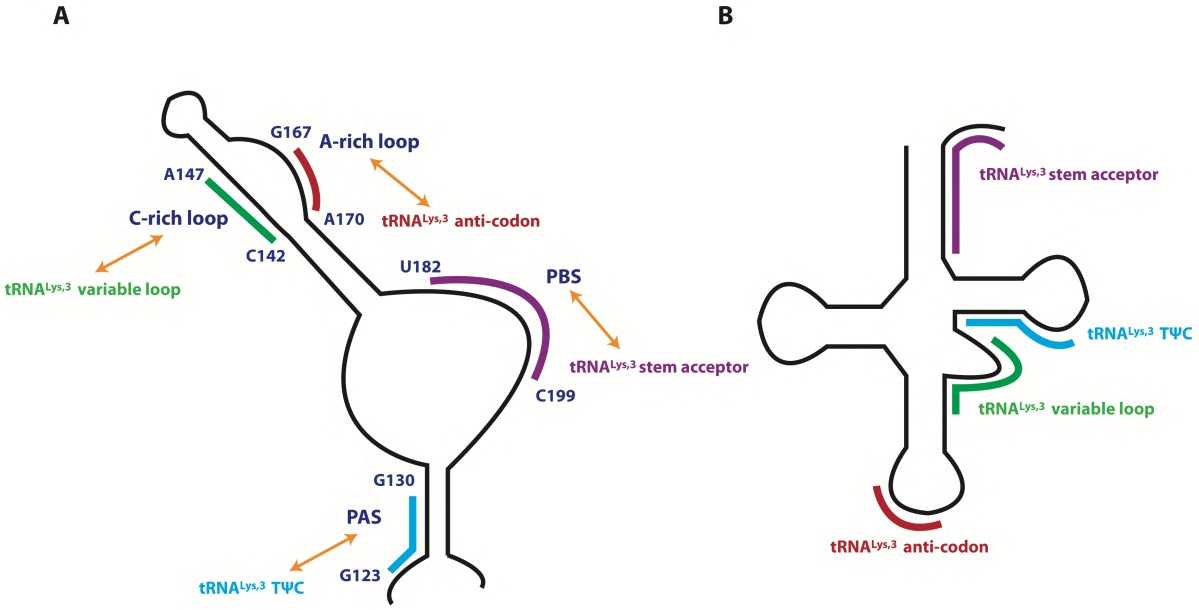


Figure 5
Description of the $tRNA^{Lys,3}$ binding to the PBS domain.
The different regions of the PBS domain (**A.**) implicated in the interaction with the four domains of the $tRNA^{Lys,3}$ (**B.**) are highlighted.

The exact mechanism underlying the specific packaging of $tRNA^{Lys}$ isoforms remains unclear, but is thought to occur independently of the gRNA despite the sequence complementarity (119, 187). The mitochondrial lysyl-tRNA synthetase (LysRS) is taken up to the virus particle in association with $tRNA^{Lys,3}$ to be the source of viral LysRS (Reviewed in (237)) (213, 240). In addition, an interaction between $tRNA^{Lys}$ and $Pr160^{GagPol}$ seems to stabilise this complex, the exact interaction domain remaining controversial with evidences for the RT domain (Khorchid et al., 2000; Saadatmand et al., 2008) as well as for the transframe and IN domains (240).

It is unknown whether $tRNA^{Lys,3}$ annealing to gRNA occurs prior to or after viral budding, nor whether $tRNA^{Lys,3}$ is still part of the LysRS-$Pr55^{Gag}$–$Pr160^{GagPol}$ complex at this stage. Nevertheless, multiple pieces of evidence suggest annealing occurs prior to $Pr55^{Gag}$ processing (138, 186, 375), followed by formation of a more stable interaction thanks to the chaperone activity of the mature NCp7 protein (reviewed in (261)) and leading to an improved ability to induce RTion (168).

The regulation of annealing seems to involve cellular and viral components. Inositol phosphates have been shown to stimulate $Pr55^{Gag}$ binding to $tRNA^{Lys,3}$ *in vitro*, suggesting that the annealing complex is formed at the PM (206). Annealing would be further aided by the viral protein Vif (182, 341) and the cellular protein RNA helicase A (433).

II.2.4.3 Packaging of cellular RNAs and their impact on assembly

Besides $tRNA^{Lys}$ isoacceptors, full-length and spliced viral RNAs, HIV-1 encapsidates a variety of host RNAs which represent approximately 30% of the total mass (reviewed in (45, 407)). Analysis of the HIV-1 RNA packageome reveals that the predominant class is non-coding RNAs (ncRNAs):

- An estimation of 50 tRNAs are encapsidated per virion with a proportion of 8 to 12 $tRNA^{Lys,3}$ (reviewed in (407)). $tRNA^{Asn}$ (as abundant in virions as $tRNA^{Lys}$ isoacceptors) / $tRNA^{Ile}$ have an unknown role in HIV-1 replication (329) but their encapsidation seems to be mediated through interaction with $Pr55^{Gag}$ and $Pr160^{GagPol}$.

- 7SL RNA constitutes the most prominent host RNA population after tRNAs with 14 to 26 copies per virion (316). 7SL RNA bind to the NC domain of $Pr55^{Gag}$ (221, 246). It has also been shown that encapsidation of 7SL is not mediated through base pairing with gRNA (reviewed in (125)) (125, 185, 229, 360). Its function in HIV-1 replication remains unknown. This has given rise to the idea that cellular RNAs play a role in the assembly process. 7SL RNA has been the subject of much speculation because it is packaged at relatively high levels. However, depletion of 7SL does not reduce viral infectivity (221).

- U1/U2 spliceosomal small nuclear RNA (snRNA) (125), fragments of endogenous retroelement RNAs called long interspersed elements (LINEs) (124), and a number of micro RNAs (miRNAs) (reviewed in (347)) (371) are also encapsidated in smaller proportions, but are probably not required for viral replication except U1 snRNA. Indeed, the SL1 of U1 snRNA is implicated in regulating the 3' polyadenylation of viral RNAs (21).

HIV-1 assembly does not specifically require gRNA packaging, but RNA appears critical for structural integrity (reviewed in (344)) (65, 309, 318, 419). Indeed, retroviruses lacking gRNA are still able to assemble through the packaging of cellular RNAs (309, 360). These RNAs presumably act as a scaffold for $Pr55^{Gag}$ and $Pr160^{GagPol}$, facilitating contacts between precursors through CA and NC domains. The NC domain of $Pr55^{Gag}$ is described as the main site of interaction with RNAs (318). A key observation is that $Pr55^{Gag}$ assembly is still efficient if the NC domain is replaced by a leucine-zipper domain (7, 99, 238, 285, 439) but viral particles contain almost no packaged RNAs. This observation suggests that interaction of RNA with the NC domain is required to mediate $Pr55^{Gag}$ multimerisation.

The spatio-temporal parameters of ncRNAs joining the RNP complex remain unresolved. Nevertheless, ncRNAs seem to be recruited early in their biogenesis (124, 156, 187).

II.2.5 Budding

The membrane scission event leading to HIV-1 particle release is mediated by the p6 domain of $Pr55^{Gag}$ (7, 162) as well as Vpu which increases the rate of release from the host cell surface (373, 408). Late domains called PTAP (Pro-Thr-Ala-Pro), YPXL (Tyr-Pro-X_n-Leu where X_n is from 1 to 3 variable residues) and PPXY (Pro-Pro-X_n-Tyr) are small peptide motifs, localised in $Pr55^{Gag}$, that recruit the host endosomal sorting complex required for transport (ESCRT) machinery to the site of budding to promote virus release (reviewed in (148)). Ubiquitylation of $Pr55^{Gag}$ proteins has been shown to help recruiting the ESCRT machinery, as several of its components contain ubiquitin-binding domains (reviewed in (417)).

The cellular machinery ESCRT is composed of four multi-subunits complexes known as ESCRT-0, I, II and III (reviewed in (190, 207)). Briefly, tumor susceptibility gene 101 (Tsg101), part of ESCRT-I, is recruited by the PTAP motif whereas ALG-2-interacting protein X (ALIX) binds to the YPX_nL domain. Binding of ESCRT-III and the recruitment of ATPase vacuolar protein sorting 4 (VPS4) lead to membrane scission, and the recycling of ESCRT factors.

II.2.6 Maturation

HIV-1 buds from the infected cell as an immature particle that must undergo maturation to become infectious. Maturation occurs concomitantly or immediately following budding and allows the internal rearrangement of the virion necessary for infectivity. This dramatic rearrangement is mediated by the sequential processing of $Pr55^{Gag}$ and $Pr160^{GagPol}$ polyproteins by the viral PR into functional units. This proteolytic processing is temporally regulated and triggers morphological rearrangements of the immature $Pr55^{Gag}$ shell in order to form a mature particle with a cone shaped core containing the gRNA in a dimeric form (reviewed in (398)).

During this time, the dimeric RNA also rearranges with a concomitant increase in dimer stability (151, 201, 313, 390). Whereas the proteolytic processing is well understood, gRNA maturation remains unclear.

Proteolytic processing and gRNA maturation are detailed in the second part of the introduction, as my PhD project is mainly focused on understanding gRNA maturation and the impact of proteolytic processing on this process.

III. Mechanism and regulation of HIV-1 virion maturation

III.1 Description of the different players

III.1.1 Pr55^{Gag} as a multitasking protein

Pr55^{Gag} is responsible for the selection of the gRNA and its subsequent assembly into viral particles (reviewed in (149, 344, 398)). It contains four major and independently folded domains: MA, CA, NC and p6 (**Figure 6**). Two flexible linkers known as Sp1 and Sp2 spacer peptides connect these domains. During the proteolytic cascade responsible for the ordered cleavage of Pr55^{Gag}, different NC-containing intermediates (NC-Sp2-p6 called NCp15 and NC-Sp2 called NCp9) are temporally present and will be further discussed in part III. 3.2.2 "From Pr55^{Gag} processing to NCp7 condensation: impact of NCp7-containing intermediates" of the manuscript.

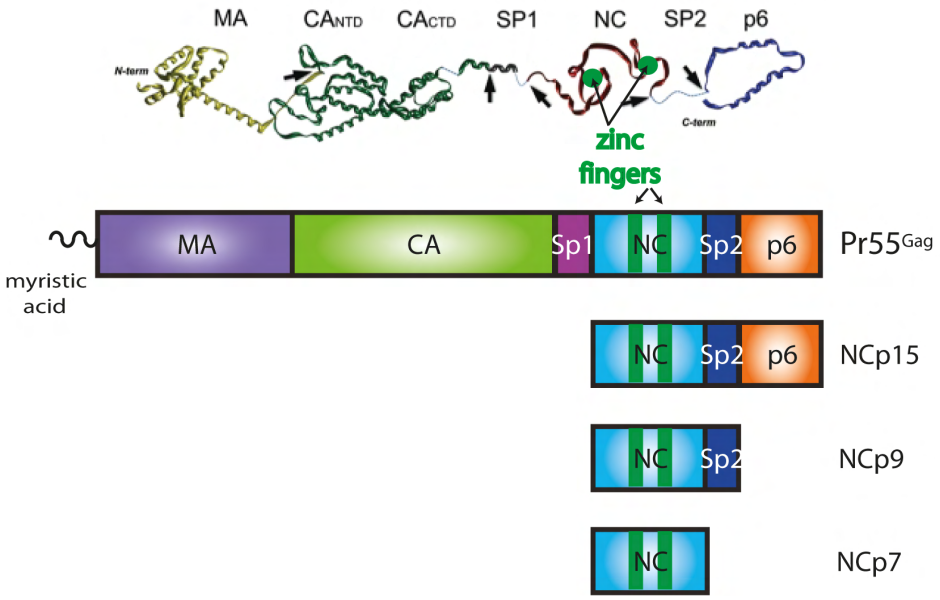


Figure 6
Description of the different domains composing Pr55^{Gag} and NC-containing intermediates.
Pr55^{Gag}, the two intermediate forms containing the NC domain, NCp15 (partial cleavage product containing NC/Sp2/p6) and NCp9 (partial cleavage product containing NC/Sp2) as well as the fully processed form, NCp7 are represented. Atomic-level structures of the individual Pr55^{Gag} domains are available. Unavailable and/or unstructured domains are represented by dashed lines. PR cleavage sites are indicated by the arrows (adapted from (294)).

III.1.1.1 Matrix (MA)

MA facilitates Pr55^{Gag} binding to the PM *via* a bipartite signal (reviewed in (86)). The highly basic region (HBR) (residues 17 to 31 comprising six lysine residues) mediates electrostatic interactions with cellular lipids like phosphatidylinositol-4,5-bisphosphate (PI(4,5)P_2) and cholesterol in the inner leaflet of the PM (61, 441). Complete depletion of PI(4,5)P_2 inhibits the formation of a pre-assembled Pr55^{Gag} lattice at the PM, revealing an important role of PI(4,5)P_2 also for HIV-1 morphogenesis.

Moreover, the post-translational modification at its N-terminus by the covalent attachment of a myristic acid moiety facilitates hydrophobic interactions of $Pr55^{Gag}$ with membranes. This 14-carbon saturated fatty acid can adopt a sequestered or an exposed conformation, constituting a reversible myristoyl switch (290, 321).

This conformational switch is proposed to prevent $Pr55^{Gag}$ interaction to intracellular membranes prior to its interaction with the PM. The HBR of MA might also regulate $Pr55^{Gag}$ binding to cellular membranes during $Pr55^{Gag}$ trafficking by binding a subset of specific tRNAs in the cytosol (246). Indeed, binding of the MA domain to RNA has previously been shown *in vitro* (reviewed in (13)) and the study of Kutluay et al. shows that deletion and substitution of lysines in the HBR domain inhibit $Pr55^{Gag}$-tRNA interaction to the same degree as deletion of the MA domain. Since the same lysine residues are involved in membrane binding and tRNA interaction, MA cannot bind simultaneously tRNAs and cell membranes. This suggests a critical role of the MA domain to regulate $Pr55^{Gag}$ localisation (85).

Recent studies highlight a second function of the MA domain which is to promote incorporation of the Env glycoproteins into the forming virions (14, 405). More specifically, MA interacts as a trimer with the cytoplasmic tail of trimeric TM-gp41 and actively guides their assembly into the budding lattice.

III.1.1.2 Capsid (CA)

CA drives $Pr55^{Gag}$ multimerisation and the assembly of the immature particle (11). CA is composed of two independently folded domains: the N-terminal domain (CA-NTD) and the C-terminal (CA-CTD) domain, connected by a flexible hinge region. CA binds host cellular factors, such as cleavage and polyadenylation factor 6 (CPSF6) (50, 254), nucleoporin NUP153 (283) and the CA-NTD interacting with NUP358 (51), the host proline isomerase cyclophilin A (147, 154, 270) (reviewed in (64). Inter-hexameric CA-NTD contacts stabilise the immature capsid (351). The CA-CTD contains the major homology region (MHR) composed of approximately 20 conserved residues (277). The function of this motif during assembly is not yet understood.

Following proteolytic processing, the mature CA protein re-assembles to form a characteristic fullerene cone (286). The crystal structure of CA monomers reveal a six fold intra-hexamer symmetry and a threefold and twofold interhexamer symmetry (165). A recent structure of CA hexamer determined by crystallography shows the presence of a size-selective pore, necessary for dNTPs entry, and bounded by a ring of six arginines with a 'molecular iris' formed by the N-terminal β- hairpin (200).

III.1.1.3 Nucleocapsid (NCp7)

From a structural point of view, NCp7 is composed of a poorly folded and highly basic N-terminal region followed by two copies of a CCHC zinc finger (ZF) motif (Cys-X2-Cys-X4-His-X4-Cys with X corresponding to amino acids other than Cys/His), ZF1 and ZF2, connected by a basic linker. The three-dimensional structure of NCp7 has been determined by nuclear magnetic resonance (NMR) methods (298, 396). Electrostatic, hydrophobic and hydrogen bonds mediate contact between NCp7 mature / as part of the $Pr55^{Gag}$ and gRNA. ZF1 and ZF2 are localised in hydrophobic pockets and have been shown to bind exposed guanosine residues, creating a simplified folding pathway (reviewed in (261)) (142, 166, 426).

Regarding the function, the NC domain in the context of $Pr55^{Gag}$, NCp15, NCp9 and NCp7 (reviewed in (294)) is involved in multiple steps of the replicative cycle, including RTion, gRNA dimerisation and maturation, selective gRNA packaging and integration.

* The selective recognition of the dimeric gRNA by $Pr55^{Gag}$ mediates its packaging inside the newly assembled virus particle.

The NC domain specifically binds gRNA presumably in the cytoplasm (140, 181, 208, 223). gRNA is recruited at the PM and assembled into viral particles (12, 46, 161). The stoichiometry is about 12 $Pr55^{Gag}$ associated with one dimer of gRNA with high affinity (48), $Pr55^{Gag}$ likely forming a trimer in solution (127). These aspects are further detailed in the article "The Life-Cycle of the HIV-1 Gag–RNA Complex".

* The unspecific binding of $Pr55^{Gag}$ through the NC domain to gRNA participates to the assembly process, gRNA acting as a scaffold. I*n vitro* assays revealed that $Pr55^{Gag}$ has a higher affinity for gRNA than NCp7, and this suggests that other domains such as MA and Sp1 would contribute to the specific interaction with the nucleic acid (reviewed in (344)) (46, 101, 105).

* Placement of primer $tRNA^{Lys,3}$ onto the PBS is initiated by the NC domain of $Pr55^{Gag}$ (70) but completion of $tRNA^{Lys,3}$ annealing requires NCp7 (reviewed in (261)) (71, 168).

* The NC domain of $Pr55^{Gag}$ has been shown to initiate gRNA dimerisation in multiple retroviruses including Harvey sarcoma virus (139), HIV-1 (151, 310) and Moloney murine leukemia virus (MuLV) (152, 166). The chaperone activity of NCp7 ensures the correct rearrangement of gRNA with an increased compaction and thermal stability of the dimeric gRNA (114, 151, 152, 201, 313). Indeed, NCp7 is an RNA chaperone that lowers the energetic barriers between RNA states to accelerate the adoption of thermodynamically stable RNA structures, and facilitates the rearrangement of misfolded states through duplex destabilisation (reviewed in (343)) (65, 87). This destabilisation is ATP-independent (23, 34, 49, 54, 178). The structural basis for destabilising activity of NCp7 is still under investigation but RNA destabilisation activity seems associated with ZFs (35, 49) and basic residues (430). The rapid kinetic of binding/annealing and dissociation/destabilisation events increases NCp7 chaperone effectiveness (100, 393).

* Following gRNA maturation, NCp7 forms a RNP complex with gRNA through unspecific contacts. NCp7 is the most prominent protein ligand of the gRNA inside the viral particle (reviewed in (345)) with a NC/RNA molar ratio of approximately one protein per 10 nucleotides.

* During the next replicative cycle, NCp7 assists the RT all along DNA synthesis by facilitating RT recruitment (258) and increasing the time of residence of active RT on gRNA (167), facilitating strand transfers (15, 203, 356) and recombination reactions (reviewed in (315)) (153, 261). NCp7 seems also to help integrating viral DNA into the host genome (62, 336), possibly by stabilising IN at the LTR ends. However, Coren et al. are not in agreement and postulate the Sp2/p6 cleavage instead of the implication of NCp7 is required for efficient viral DNA integration (96).

Several studies compared the nucleic acid chaperone properties of NC containing intermediates p15 and p9 with the mature NCp7. Impact of these different HIV-1 proteolytic intermediates on gRNA dimerisation and maturation will be further discussed in the section "genomic maturation". Retroviral mature NC proteins from different genera also exhibit different chaperone activities, with a more rapid

association/dissociation from ssRNA seen with HIV-1 and RSV NCs compared to MuLV and HTLV-1 NCs (393).

III.1.1.4 p6 domain

The p6 domain allows the correct release of the virus (7), by recruiting the ESCRT machinery. Abscission of budding virions from the PM is mediated by the p6 NTD containing the two late domains, the PTAP and YPX_nL motifs.

Additionally, the p6 CTD promotes the incorporation of the viral factor Vpr into assembling virions (242, 330) through the highly conserved LXXLF motif (241).

A recent study suggests an impact of the p6 domain on the aggregation activity of NCp15 (420). The acidic p6 domain would fold back and interact with the basic ZFs of the NC domain, explaining the weaker annealing capability of NCp15 compared to NCp9 and NCp7.

It is still unknown whether p6 has a function once it has been cleaved from $Pr55^{Gag}$, especially since p6 is present in amounts equimolar to CA in viral particles. Recently, the team of U. Schubert showed that p6 is specifically degraded by a ubiquitously expressed cytosolic metalloendopeptidase, the insulin-degrading enzyme and that removal of p6 during viral entry is important for virus replication, at least in the case of X4 tropic HIV-1 (171).

III.1.2 $Pr160^{GagPol}$

The $Pr160^{GagPol}$ precursor is translated from the same unspliced mRNA as $Pr55^{Gag}$ precursor. Since viral replication requires relatively more structural proteins compared to viral enzymes, HIV-1 uses ribosomal frameshifting to ensure a proper ratio (about 1 to 20) of $Pr160^{GagPol}$ compared to $Pr55^{Gag}$ (198). Sequences directing translational frameshifting are located at the p6 N-terminus so the p6 protein is truncated in the context of the $Pr160^{GagPol}$ and called p6*.

PR, IN and RT enzymes are generated upon processing of $Pr160^{GagPol}$ by the viral PR, once PR is auto catalytically activated.

III.1.2.1 PR

PR is part of the aspartic class of proteases and composed of two identical monomers (reviewed in (243)). Protease dimerisation is mediated by the N- and C-terminal residues (a.a. 1 to 4 and 96 to 99) and is necessary to achieve catalytic competence. The active site comprises three amino acids (Asparagine (Asp), Tyrosine (Thr) and Glycine (Gly)), also known as the catalytic triad.

III.1.2.2 RT

RT is an asymmetric heterodimer consisting of two subunits p66 and p51. The p66 subunit exhibits an RNA and DNA dependent DNA polymerase activity and an RNase H activity. The polymerase domain is composed of four subdomains: fingers, palm, thumb, and connection. The p51 subunit is obtained by further proteolytic cleavage of the p66 subunit, with the p51 subunit lacking RNase H catalytic activity (199, 369).

RT has no proofreading activity, explaining the high mutation rate in the newly synthesised proviral DNA.

III.1.2.3 IN

IN catalyses the viral DNA insertion into the genome of infected cells in the context of the intasome complex (reviewed in (164). This enzyme functions as a homo-hexadecamer, with a tetramer-of-tetramer, the structure of the intasome been determined by cryo-electron microscopy (cryoEM) (26, 328). Each protomer contains a NTD, a catalytic core domain and a CTD.

Two different classes of IN inhibitors are available and target either the enzyme catalytic site (IN strand transfer inhibitors (INSTIs)) or the IN dimer interface (non-catalytic site IN inhibitors (NCINIs)). NCINIs engage the IN dimer interface at the binding site for the LEDGF/p75 host complex.

III.2 Proteolytic processing

$Pr55^{Gag}$ and $Pr160^{GagPol}$ polyproteins generate structural and enzymatic viral proteins through a highly ordered and sequential processing of the different cleavage sites. Proteolytic processing of these two precursors is required to convert the non-infectious immature particle into a mature infectious form (reviewed in (400)) (96, 425). Proteolytic processing ensures genetic economy by using a single set of transcriptional and translational control elements.

III.2.1 Particle morphogenesis

The architecture of starting and endpoints of morphological maturation of the viral particle has been deciphered by cryoEM and cryo-electron tomography (cryoET) (38, 56, 57, 59, 372, 428, 429) (**Figure 7**). Immature particles are characterised by a semi-spherical shell of $Pr55^{Gag}$ and $Pr160^{GagPol}$ molecules (spacing of 8 nm) with the MA domain lining the lipid envelope and the NC domain extending towards the center of the virus (57). CA-CA interactions mediate $Pr55^{Gag}$ interactions. The action of the ESCRT complex to close the viral bud presumably explains the large gap in the $Pr55^{Gag}$ shell of the immature particle (68, 155, 429). Indeed, $Pr55^{Gag}$ forms a continuous but incomplete sphere in the released virion in contrast with viral assembly in cells lacking functional ESCRT with a nearly closed sphere. So, HIV-1 assembly seems to be completed in an ESCRT-dependent manner before the $Pr55^{Gag}$ sphere is complete.

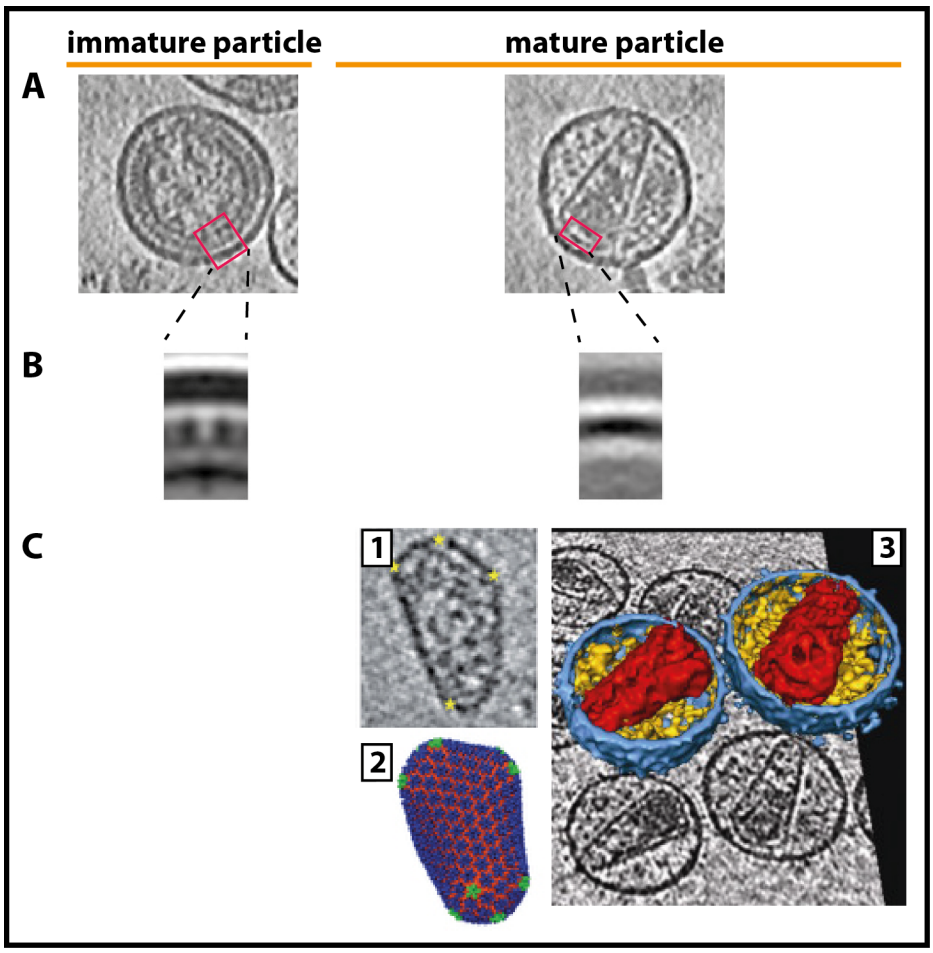


Figure 7
General morphological comparison of immature and mature particles.
A. Central slices from tomographic reconstructions of immature and mature HIV-1 virions. **B.** Radial sections through the averaged sub tomograms of immature and mature core shell accentuate the difference in CA layer thickness (adapted from (222)). **C1.** cryoET analysis of an isolated, native HIV-1 core. Yellow stars indicate locations of sharp curvature change (scale bar, 20nm). **C2.** View of a molecular dynamics equilibrated all-atom capsid model comprising 216 CA hexamers (blue, NTD; orange, CTD) and 12 CA pentamers (green) (adapted from (440)). **C3.** Three-dimensional cryoET rendering of two sample virions (scale bar, 100 nm). Blue, viral membrane; yellow, density between the membrane and the core; red, viral capsid. Half of the blue and yellow densities has been computationally removed to reveal the core. The rendered virions are shown above a central slice through the reconstruction (adapted from (56)).

In contrast, the mature particle has a MA layer lining the lipid envelope and contains a conical capsid containing the condensed nucleoprotein complex (440). The formation of a CA hexameric lattice is accomplished with 12 pentamers closing both extremities. There is a regular spacing of 9.6 nm between hexamers (59, 155, 286) so the capsid is more loosely packed than in the immature particle. Only about half of the CA proteins is used to form the mature cone (reviewed in (55)) (250). These findings suggest a complete disassembly of the immature architecture followed by a complete *de novo* reassembly of the mature lattice. This is referred to as the disassembly-assembly model (114, 222).

The HIV-1 surface is also affected during maturation with re-distribution of Env glycoproteins. Visualisation using stimulated emission depletion (STED) super resolution fluorescence microscopy has shown several patches of Env proteins on the viral surface of immature particles which convert to close clusters on the membrane of mature virions (84). These findings correlate with an active recruitment of Env trimers to budding sites due to Env CTD interactions with the MA domains of the $Pr55^{Gag}$ lattice (14, 359, 405). This maturation-induced clustering of Env trimers aids virus entry through the association of Env trimers with CD4 patches on the target cell surface.

III.2.2 Stepwise cleavage of $Pr55^{Gag}$

The different cleavage sites can be divided into three kinetic categories (135, 256, 333–335, 337) and are presented in **Figure 8**:

1. The initial processing separates MA-CA-Sp1 from NCp7-Sp2-p6, releasing the RNP complex from the thick membrane-attached MA-CA lattice. This cleavage initiates dimer gRNA rearrangement, stabilisation and condensation (114, 313, 425) with approximately 60% of dimeric RNAs (313).

2. NCp7-Sp2 is released from p6 to further continue RNP condensation without virion morphological changes (96, 114, 304, 313). MA is cleaved from CA, separating CA from the membrane-bound MA layer. The viral envelope becomes thin with the detection of some amorphous structures rather than a well-defined core (114, 255, 313, 425). Approximately 80% of dimeric RNAs are detected (313). Relative to the fastest cleavage (Sp1-NCp7), Sp2-p6 cleavage is 9 times and MA-CA cleavage 14 times slower (333, 335).

3. Sp1 and Sp2 are finally cleaved from CA and NCp7. Release of Sp1 from CA allows to obtain the mature core whereas the RNP complex becomes further condensed (96, 114, 163, 245, 425) with a similar amount and mobility profile in native conditions compared to the wild type (313). NCp7-Sp2 appears 350 times and CA-Sp1 400 times slower than the initial cleavage (333, 335). The influence of Sp1 on processing intermediates has been assessed and was shown to negatively regulate the cleavage rate of the CA/Sp1 site during sequential processing *in vitro* and *in cellula* (335).

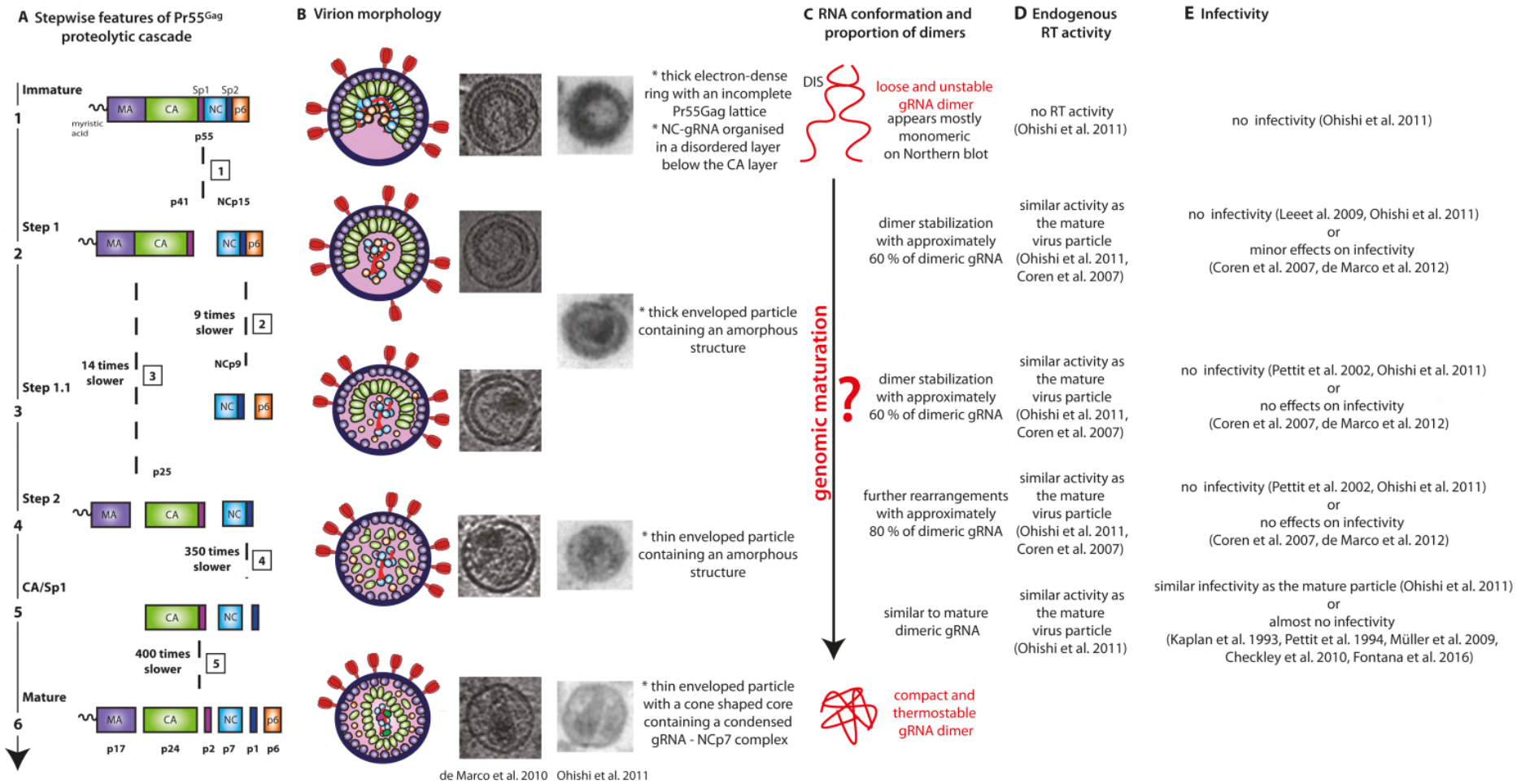


Figure 8

The Pr55^{Gag} proteolytic processing.

A. Schematic outline of the proteolytic cascade, represented from the immature (A1) to the mature viral particle (A6), with processing intermediates temporally generated at each step. Cleavage sites are highlighted with their relative cleavage rate. **B.** Schematic morphological evolution of the viral particle illustrated by transmission electronic microscopy of HIV-1 particles. (adapted from (114, 313)). **C.** Evolution of the RNA conformation depending on the Pr55^{Gag} processing with the proportion of dimeric gRNA, quantified on Northern blot (313). **D.** Endogenous RT activity depending on the Pr55^{Gag} processing (96, 313). **E.** Viral infectivity depending on the Pr55^{Gag} processing (79, 96, 113, 145, 215, 304, 313, 333, 335)

Sp1 participates in the disassembly-assembly model of capsid formation (109, 114, 255, 313, 425). Cleavage upstream and downstream of CA-Sp1 is required for immature lattice disassembly whereas CA-Sp1 cleavage initiates mature lattice assembly (222).

On the other hand, the role of Sp2 remains unclear. Several studies report the cleavage of NCp7-Sp2 as dispensable for efficient replication (96, 113, 333) whereas the study of Ohishi et al. shows a major effect on infectivity when this site is mutated (313). NCp7-Sp2-p6 as well as Sp2-p6 cleavage are reported to produce non-infectious particles (96, 113). Since NCp7 and NCp7-containing intermediates are implicated in gRNA compaction, Sp2 processing could dynamically regulate this process (294) to coordinate it to the capsid assembly.

III.2.3 Exact timing and duration of the maturation process

III.2.3.1 Initiation of the maturation process

Initiation of proteolytic processing has to be closely coordinated with virus budding (214). PR activation begins with dimerisation of PR domains within $Pr160^{GagPol}$ (331, 332). This process presumably requires high concentration of $Pr160^{GagPol}$ precursors suggesting that virus assembly could be the trigger for proteolytic processing (243).

III.2.3.2 Regulation of processing

Processing is then tightly regulated through sequential cleavage at individual sites of precursors. Cleavage sites have no amino acid sequence similarity and present different cleavage rates (256, 335, 410).

In several studies, complete cleavage appears essential for successful maturation and viral infectivity acquisition, a partially processed intermediate, in particular the late CA-Sp1 intermediate, acting as a trans-dominant negative and inhibiting infectivity (79, 145, 304, 335). Nevertheless, Ohishi et al. obtain an infectivity similar to the mature particle with the CA/Sp1 mutant (313).

The effect of the context surrounding the cleavage site on proteolytic processing has been highlighted by replacing the MA/CA cleavage site with each $Pr55^{Gag}$ cleavage site (256). The rate of the CA/Sp1 cleavage site, shown to be negatively regulated (335), significantly increases when placed in the MA/CA context. Moreover, the cleavage rate of Sp1/NC and Sp2/p6 sites is significantly decreased, indicating a possible local positive regulation in their homologous sites (256). These cleavage sites are both at the CTD end of a spacer peptide and both have been implicating in interacting with viral RNA (219, 297, 354, 381), suggesting an implication of viral RNA in this regulation. Indeed, RNA accelerates proteolytic processing by interacting directly with the PR rather than the cleavage site substrate since the proteolysis increase is independent on the substrate's ability to interact with RNAs (337). This interaction seems mediated by electrostatic forces with no specific RNA structure or sequence. This RNA-dependent processing has also been proposed to increase the processing rate at the Sp1/NC site (118).

III.2.3.3 Duration and technical issues

The duration of the maturation step to obtain a complete rearrangement of the viral particle has been matter of debate due to technical issues. Indeed, assembly and maturation occur in a short time scale (172) and maturation is asynchronous with newly particles continuously formed and released by the infected cell. Post-assembly events are thus difficult to analyse by electron microscopy (EM). This technique gives static images of mature, intermediate or immature virus produced in the presence of PR inhibitors or mutations in the PR active/cleavage sites. Since maturation intermediates have not been detected so far with wild type viruses, this process is believed to be fast.

Construction of sequential HIV-1 $Pr55^{Gag}$ mutants containing point mutations abolishing cleavages at individual or multiple cleavage sites allows the production of maturation intermediates blocked at a precise step and correlated with the ordered processing events (114, 245, 313, 335, 425). A snapshot of the morphological maturation process is thus technically possible, depending on the processing step of $Pr55^{Gag}$ and $Pr160^{GagPol}$.

Investigation of the dynamics of HIV-1 maturation using wild-type particles has further been possible by improving live-cell imaging techniques with fluorescently labeled viral particles (83). Super-resolution microscopy techniques significantly improved the spatial resolution so it is now possible to visualise sub-viral details under live-cell conditions. In parallel, an inhibitor washout strategy to synchronise the PR activation in wild-type viruses has been developed (284). This strategy has been improved and couples STED nanoscopy with a photo destructible PR inhibitor to obtain a synchronised induction of protease cleavage *in situ* (172).

A clear ring-like structure has been visualised in immature particles using cryoET (57, 68, 372, 429). Moreover, dual-color STED nanoscopy allows the discrimination between the outer membrane showing a closed ring-like structure and the $Pr55^{Gag}$ lattice covering only 70% of the inner membrane surface in the immature conformation (172) (**Figure 9**). The photodegradable PR inhibitor cleaved by UV irradiation and resulting in PR activation, enables dynamic analysis of the proteolytic processing with a defined analysis start point and the visualisation of the transition from immature to mature morphology in native and unfixed viral particles. The time-course of $Pr55^{Gag}$ proteolytic processing revealed a half-life of 29 +/- 8 minutes to switch from an immature to a condensed lattice (172), indicating a morphological conversion without further delay once PR is active. Nevertheless, it is important to remind that viral infectivity was not recovered upon inhibitor washout (284), thus questioning the infectivity of virus particles with utilisation of the photodegradable PR inhibitor.

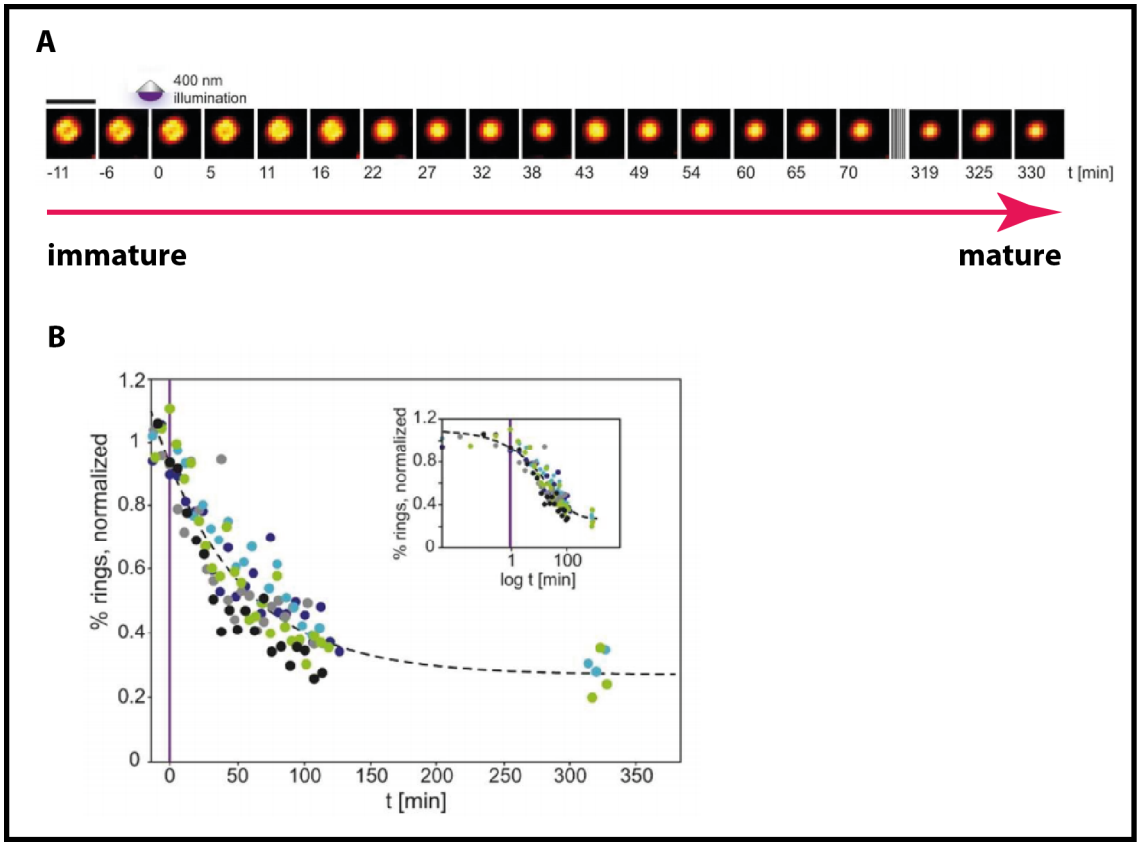


Figure 9
Dynamics of the morphological transition analysed by STED microscopy.
A. HIV particles labelled with the CLIP tag were imaged by STED nanoscopy at the indicated time points. Images shown represent overlays of particles detected at the respective time (n~500). Scale bar: 300 nm. **B.** Time-course analysis showing the proportion of rings compared to all particles detected in the respective image. The time point of illumination is indicated by the purple line. The graphs show combined data from five independent experiments, indicated by differently coloured symbols; the inset is a log transformed representation of the same data set. Dashed curves represent an exponential fit to the data set (adapted from (172)).

III.2.4 Mechanism of action of protease inhibitors

PR inhibitors (PIs) efficiently block HIV-1 replication by inhibiting the viral PR (244). Clinical concentrations of PIs induce a severe reduction of viral infectivity with, surprisingly, minor effects on $Pr55^{Gag}$ proteolytic processing (215, 304). This inhibition of viral infectivity correlates with an accumulation of NC/Sp2 processing intermediates. In addition, these viruses, produced in the presence of PIs, have a normal *in vitro* RT activity but fail to synthesise endogenous viral DNA (304). Furthermore, PI resistance mutations are located in the NC-Sp2-p6 region of $Pr55^{Gag}$, establishing the important role of this region in determining resistance (103).

III.3 Genomic maturation

The HIV-1 genome is a single-stranded and positive-sense RNA. Two RNA copies are packaged inside the viral particle as a dimer.

III.3.1 The dimeric nature of retroviral genomes

The dimeric nature of retroviral gRNAs has been demonstrated in the early 70s, by applying biochemical techniques (analytical ultracentrifugation and sedimentation analysis) to gRNAs purified from viral particles (82, 235, 278, 350, 394). Electron microscopic studies of gRNA from several retroviruses allowed the detection of monomeric and dimeric gRNA with monomers linked in parallel orientation near their 5' ends (36, 37, 278, 311).

The mechanism of HIV-1 dimerisation has been further investigated *in vitro* using RNA fragments encompassing the 5' UTR. These truncated RNAs spontaneously dimerise *in vitro* (28, 40, 107, 280, 358, 397) allowing the identification of the dimer linkage structure (DLS). This region has been shown as highly structured (28, 174) but its structure remains unclear with several models proposed as explained in the review "The Life-Cycle of the HIV-1 Gag–RNA Complex". The contact between the two monomers is mediated by non-covalent interactions, with dimers dissociating upon heating (152, 390, 394).

The essential motif within the DLS, called the DIS, has finally been identified *in vitro* by interference of chemical modifications (386), site-directed mutagenesis (252, 322) and antisense oligonucleotides (308, 385). DIS mediates HIV-1 gRNA dimerisation through a kissing loop interaction (322). The precise six nucleotides implicated in the dimerisation initiation have been identified by interference of chemical modifications (386), *in vitro* dimerisation tests (322) and confirmed by directed mutagenesis (322). The three purines (A255, A256 and A263) from either side of the DIS are also required for the dimerisation process and stabilise the kissing-loop complex (325, 385).

III.3.2 Conformational rearrangement of the HIV-1 dimeric gRNA

III.3.2.1 Requirement of the proteolytic processing of $Pr55^{Gag}$

The team of M. Laughrea analysed the general conformation of gRNA purified from viral particles in a time course manner using non-denaturing gel electrophoresis (201, 390). After 5 minutes, viruses present about 80% of gRNA migrating as a dimer with slower electrophoretic mobility (**Figure 10 lane 2**), compared to the remainder monomer. This monomeric population disappears after 48 hours (**Figure 10 lane 1**).

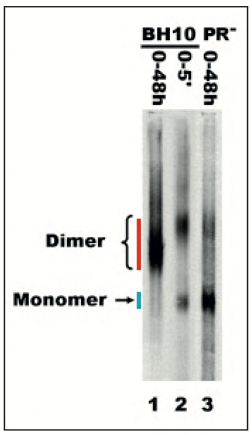


Figure 11
Dimerisation of gRNA isolated from newly released and 48 hours wild-type viruses.
gRNAs extracted from 48 hours-released wild-type (lane 1), 5 minutes-released wild-type (lane 2) and 48 hours-released PR⁻ virions were electrophoresed on a non-denaturing 1% (w/v) agarose gel and analysed by northern blotting (adapted from (390)).

The team of J. Sakuragi further studied the gRNA dimer transition, from fragile to stable, by generating Pr55^{Gag} cleavage-site mutants (**Figure 11**). These mutants allow specific gRNA extraction, thermal dissociation kinetic and densitometric analysis by native blotting in a step-wise manner (313).

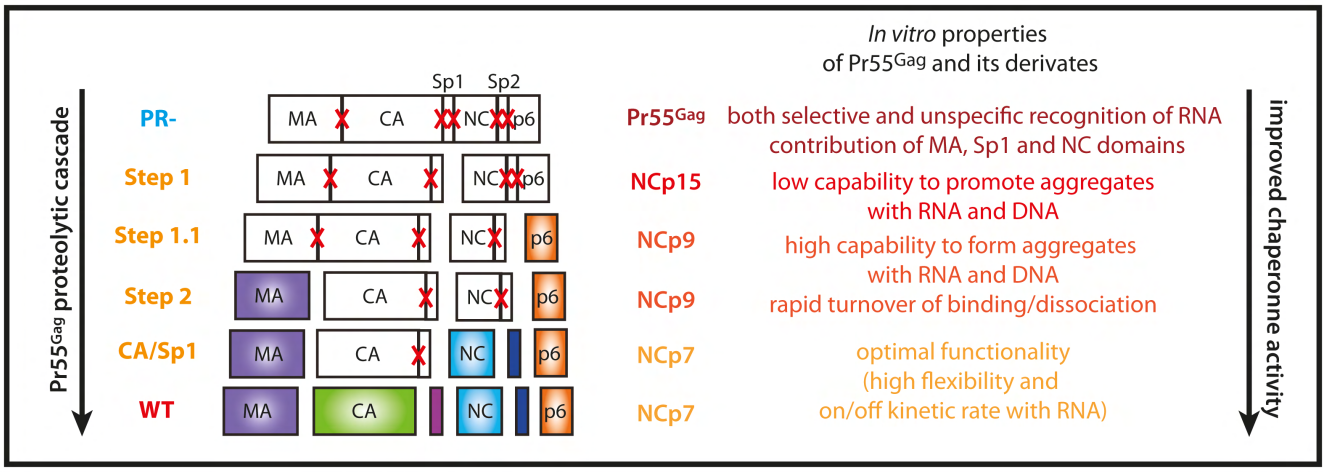


Figure 10
Description of mutants mimicking the sequential processing of HIV-1 Pr55^{Gag}.
Mutants are described from the completely immature mutant Step 0 exhibiting the 5 cleavage sites mutated (represented by red crosses) to the wild-type. White boxes represent fusion proteins resulting from mutation of the cleavage sites and colored boxes proteins in their mature form.

Thermal dissociation kinetic of gRNA dimers extracted from mature, immature and mutant particles has been compared. gRNA dimers from mature particles, as well as CA/Sp1 mutants, exhibit an increased thermal stability, with a dissociation of the dimer at approximately 44 °C (**Figure 12**) compared to immature particles where half of the gRNA population is monomeric at 40 °C. Interestingly, mutants Step 1, Step 1.1 and Step 2 have an intermediate thermal stability with a dissociation at approximately 43 °C.

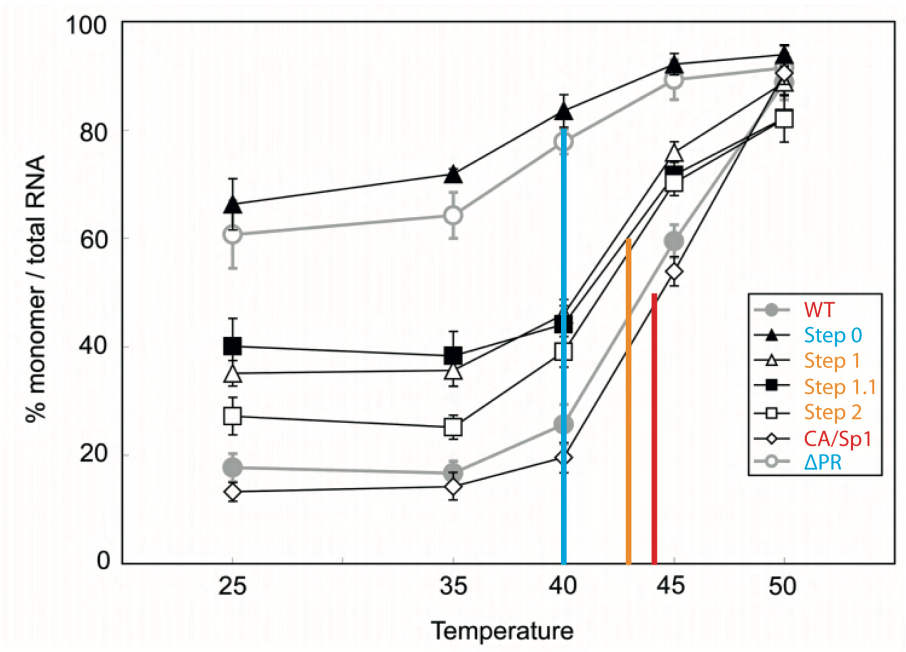


Figure 12
Thermal dissociation of gRNA isolated from wild-type, immature and intermediate mutant virions.
The level of gRNA dimerisation from extracted viral RNA were analysed by northern blot as a function of incubation temperature. Independent experiments were analysed and plotted (adapted from (313)).

Densitometric analysis of these samples' distribution highlights the lower mobility of the intermediate mutants Step 1, Step 1.1 and Step 2, with Step 1 considerably different from the wild-type (**Figure 13**). This result was not expected since the dimeric gRNA proportion is similar in Step 1 and Step 1.1 mutants (**Figure 13**).

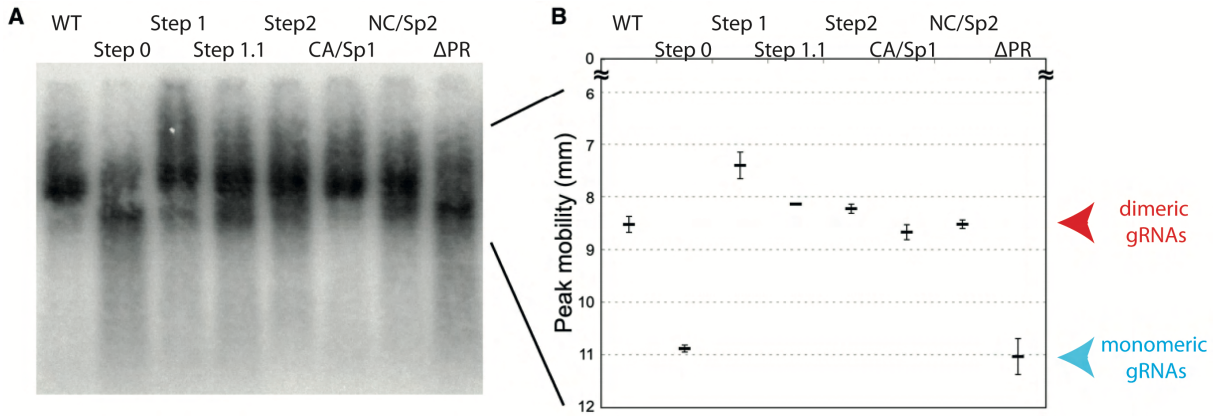


Figure 13
Densitometric analysis of gRNA samples isolated from wild-type, immature and intermediate mutant virions in native conditions.
A. RNA profile from northern blot. **B**. Mobility of the samples is compared and highlights the monomeric and dimeric gRNA populations (adapted from (313).

These evidences demonstrate that the dimeric gRNA is rearranged from a "loose" and unstable to a "tight" conformation.

Several teams looked at the HIV-1 RNA dimer isolated from WT and PR defective (PR-) particles (viruses deletion for PR) (151, 313, 377, 378, 390). A different conformation is observed and this is coherent with the previously observed conformational rearrangement of the Mo-MuLV genome upon viral release from the cell (152).

The WT HIV-1 gRNA totally migrates as a dimer (**Figure 10 lane 1**) and this proportion is reduced to 40-50% in the case of PR- (**Figure 10 lane 3**).

In addition, the dimeric gRNA isolated from PR- is thermolabile (**Figure 12**). So, the gRNA dimer of PR- viruses is unstable, much like RNA dimers isolated from immature viruses (313, 390). These experiments show that the proteolytic processing is required to observe a fully effective gRNA maturation.

Effects of mutations within the primary (Sp1/NCp7), the secondary (MA/CA and Sp2/p6) and the tertiary (NCp7/Sp2 and CA/Sp1) cleavage events have been further assessed on gRNA dimerisation. A strong influence of the primary cleavage event on thermal stability of the dimeric gRNA is shown compared to secondary and tertiary cleavage events (184, 313, 379). Moreover, the Sp1/NCp7 cleavage is more important for RNA dimerisation and stability in the context of $Pr55^{Gag}$ than in the $Pr160^{GagPol}$ and also impacts the core formation (379). So, mutation of the Sp1/NCp7 cleavage site hinders formation of stable HIV-1 gRNA dimers probably because the chaperone activity is incomplete in the context of $Pr55^{Gag}$, and also because the gRNA in association with NCp15 is still attached to MA-CA at the PM. This cleavage defect leads to the formation of defective cores.

CryoET and subtomogram averaging (114) also highlight the importance of the primary cleavage between Sp1 and NC domains of $Pr55^{Gag}$ to initiate core formation surrounding the RNP complex and dimer RNA stabilisation (Figure 6C). The cleavage of Sp2 from the NCp7 CTD is also required in order to increase the stability of the dimeric gRNA (313).

III.3.2.2 From $Pr55^{Gag}$ processing to NCp7 condensation: impact of NC-containing intermediates

Factors mediating the gRNA maturation are still unclear, the proteolytic processing of $Pr55^{Gag}$ and more precisely the release of NCp7 being considered to be the leading candidate. In the mature viral particle, NCp7 is the most prominent protein ligand of the gRNA (reviewed in (261, 345)). Indeed, the NCp7 domain in both the context of $Pr55^{Gag}$ and the NC-containing intermediates as well as the fully processed form are well-known for their nucleic acid chaperone activity (138, 150, 169, 173, 212, 345). NC has also been shown to play a role in *in vitro* RNA dimer stabilisation of Harvey sarcoma virus genome (139) by increasing its thermal stability and MoMuLV (52).

Regarding HIV-1, an average of one NCp7 for 10 nucleotides typically drives efficient co-aggregation of NCp7 with dimeric gRNAs (230) whereas selection of this dimer from a large pool of spliced viral and cellular RNAs is mediated by the NCp7 domain in the context of $Pr55^{Gag}$ (185).

The HIV-1 chaperone properties of NCp7 and NC-containing intermediates have been shown to be different. Indeed, the chaperone activity of the different processing intermediates containing the NCp7 domain is improved from $Pr55^{Gag}$ to

NCp7. This is done by increasing both the flexibility of NCp7 and the on/off kinetic rate with gRNA (100, 393).

The *in vitro* binding capacities of $Pr55^{Gag}$, NCp7 and NC-containing intermediates have been compared and showed that aggregative properties of NCp7-containing intermediates are lost with the release of the p6 domain from $Pr55^{Gag}$, NCp15 being able to promote aggregates (reviewed in (294)). The p6 domain has recently been proposed to fold-back on NCp7/Sp2 through the two ZFs of NCp7 (420). Actually, NCp9 strongly promotes gRNA aggregation with a rapid turn-over of binding/dissociation (292). The complete functionality is acquired through the release of Sp2, allowing both specific and unspecific contact of NCp7 with gRNA (292, 395).

In agreement with this, gRNA dimerisation is significantly disrupted in viral particles blocked at NCp15 processing state (Step 1 in **Figure 11**) and RTion/integration steps are highly impaired, compared to NCp9 (201, 211, 293). Surprisingly, NCp9 is the most aggregative intermediate compared to NCp15 and the mature NCp7 (292). This property is explained by its less efficient ability to rapidly bind and dissociate from nucleic acids (100). The exact purpose of this characteristic is still unknown.

Thus, the state of NC processing spatiotemporally regulates intravirion morphogenesis with RNP condensation, conferring to the viral particle an optimal architecture within the core crucial for its ensuing replication in the target cell. NCp15 and NCp9 are short-lived species undetected during viral production, the reason of their maintenance remaining unclear.

III.3.3 Other factors implicated in the maturation of the gRNA

Nevertheless, proteolytic processing of $Pr55^{Gag}$ is not sufficient for mediating gRNA maturation as demonstrated with mutants in the CA domain (211) or defective for the PTAP motif in the p6 domain (150) where structural proteins were mostly processed but with an immature particle morphology and an impaired dimeric gRNA thermal stability, similar to PR- viruses.

The $Pr55^{Gag}/Pr160^{GagPol}$ ratio is also implicated in gRNA maturation with a reduced dimeric gRNA stability in the absence of $Pr160^{GagPol}$ (378) and when the ratio is changed from 20/1 to 20/21 (377). In addition, the addition of PR *in trans* did not completely restore the stability of gRNA suggesting a role for RT and IN in gRNA maturation (227, 275, 378). Several mutations in IN coding region have been highlighted as not only affecting integration (class I mutations) but also the viral particle maturation (class II mutations) (129, 130), suggesting the possibility that integration interference is not the only mode of action of IN inhibitors.

An atypical morphological phenotype has been observed in presence of NCINIs (referred to as allosteric IN inhibitors) (25) (**Figure 14**). Defects in core morphology were observed by thin-section EM (25) and tomographic (144) analysis with a WT level of condensed RNP complex localised outside the capsid shell and an accumulation of IN oligomers. This "empty capsid" phenotype correlates with the phenotype of viral particles lacking IN (129). Presence of the majority of NCp7 proteins in close proximity with gRNA has been checked by irradiation-induced bubbling (144).

Identical conclusions were drawn from other studies (210) with a potent late-stage antiviral effect of allosteric IN inhibitors due to their inference with viral assembly, leading to a block of RT in newly infected cells.

A recent study postulates that IN initiates core morphogenesis (144), as a partial rescue of the conical core and HIV-1 infectivity occurred when defective IN particles were *trans* complemented with IN. In addition, RNP complex seems incorporated into the mature core by binding to IN (144, 227), with IN mutations that disrupt its binding to gRNA leading to an eccentric viral cone.

The detailed mechanism is still incompletely understood but could implicate IN multimerisation upon treatment with allosteric IN inhibitors that disrupts the IN-RNP complex interaction. Direct interaction between IN and gRNA has been demonstrated in virions using the CLIP-seq and appears specific for viral RNAs through structural elements such as TAR and RRE (227). Nevertheless, deletion of TAR or RRE does not noticeably impact viral morphogenesis (246), probably due to the redundancy of structural elements along the viral genome.

Given the crucial role of the core in the RTion process (5, 25, 143, 146, 204, 259), the mislocalised RNP complex is unstable and could explain the DNA synthesis defect observed with these inhibitors. Moreover, a recent study postulates the viral gRNA as well as IN from ALLINI-treated virus particles to be prematurely degraded in target cells, whereas RT remains stably associated with the core (275), further explaining the DNA synthesis defect.

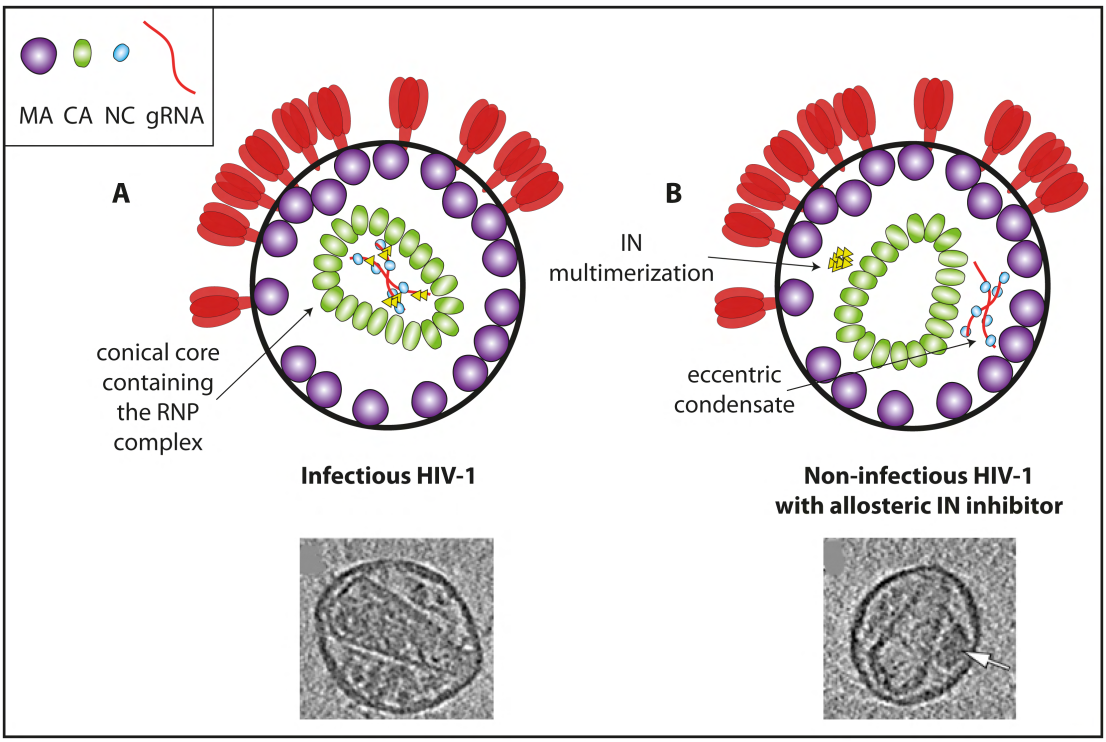


Figure 14
Model illustrating the mechanism of allosteric IN inhibitors.
Tomographic central slices of representative HIV-1 virions (bar, 50 nm) **A.** in absence or **B.** presence of allosteric IN inhibitors illustrated with schematic representations. **A.** In the WT particle, the conical cone contains the dimeric Grna associated with NCp7 proteins. **B.** In presence of allosteric IN inhibitors, the viral core is empty with an eccentric condensate relocated outside the capsid shell (highlighted with the white arrow). The IN accumulates in a multimeric conformation (cryoET slices from (144)).

III.3.4 Proteolytic processing and gRNA maturation: the other way around

The classical way to link proteolytic processing of the $Pr55^{Gag}$ and $Pr160^{GagPol}$ precursors with maturation of the gRNA is through the release of the mature NCp7 chaperone activity. Nevertheless, the opposite could also be true with an impact of gRNA maturation onto proteolytic processing. Indeed, gRNA binding to $Pr55^{Gag}$ and NCp7-containing intermediates, besides forming a scaffold mediating assembly, also stimulates PR activity and ensure complete processing of $Pr55^{Gag}$ and $Pr160^{GagPol}$ (293, 337, 381).

This hypothesis emerged from the study of a 16 nucleotides deletion mutant within SL1, impacting packaging of gRNA. Long-term replicative culture of this mutant yielded revertant viruses with a replication kinetic similar to the WT virus. This phenotype is explained by the introduction of two additional point mutations, respectively located in the Sp1 and NCp7 domains of $Pr55^{Gag}$ (265). Deletions in the DIS region slow down cleavage between CA and Sp1 and one of the compensatory mutations is in Sp1 domain. As this mutation compensates defects of SL1 mutants, it suggests an impact of SL1 RNA domain or more generally dimerisation onto proteolytic processing of the CA/Sp1 intermediate (264).

In addition, in HIV-2, genome dimerisation is mediated by the packaging signal palindromic sequence (pal) (247, 249). A mutation in pal led to an abnormal proportion of particles containing more than one core, suggesting that gRNA encapsidation, dimerisation and viral particle morphology are closely linked. The analysis of the proteolytic processing pattern of HIV-2 dimerisation mutants indicates an accumulation of the p41 (MA-CA-Sp1) intermediate (248). Thus, mutations in the non-coding region of the gRNA are linked with perturbations in $Pr55^{Gag}$ processing, which ultimately prevent correct viral core assembly and thus impact viral infectivity and replication. A compensatory mutation within the MA domain rescued viral replication of these mutants (248). In simian immunodeficiency virus infected macaques, mutants within SL1 led to an accumulation of the first and second cleavage intermediates, an atypical core architecture and diminished viral replication (424).

Altogether, gRNA dimerisation and, in general gRNA maturation, seems to have an impact on particle assembly, possibly by acting as a scaffold to allow the processing of $Pr55^{Gag}$ (313). The molecular mechanism underlying dimerisation and the role of the DIS *in vitro* are well understood. Nevertheless, the spatio-temporal parameters of the dimerisation initiation remain controversial, mainly because of technical issues due to the low gRNA amount and stability in the cell. These aspects are detailed in the review "The Life-Cycle of the HIV-1 Gag–RNA Complex".

III.3.5 Implication of other inter-molecular interactions

Whilst the DIS is the only gRNA intermolecular interaction demonstrated *in cellula*, inter-molecular interactions other than the DIS must exist since gRNA dimerisation takes place even when the DIS is non-functional (43, 183, 366, 367, 390).

gRNAs isolated from viruses containing a deletion of the DIS (ΔDIS) exhibit a 100% dimeric form (390), potentially indicating the implication of other inter-molecular interactions. Nevertheless, dimerisation is slower compared to WT viruses, with approximately 3 hours for ΔDIS viruses to reach the gRNA dimer level of newly

released WT viruses. In addition, dimeric gRNAs migrate heterogeneously with a diffuse profile. These data indicate the existence of secondary dimerisation sites, yet uncharacterised, within the genome in addition to the main DIS.

Evolution of RNAs probing methods over the last 40 years: technical aspects

---

This review is in preparation and will be submitted as a focus article in WIREs.

Evolution of RNAs probing methods over the last 40 years: technical aspects

Elodie Mailler, Jean-Christophe Paillart, Roland Marquet, Redmond Smyth* and Valérie Vivet-Boudou*

Architecture et Réactivité de l'ARN, CNRS, IBMC, Université de Strasbourg, Strasbourg, France

« Focus article » WIREs

RNA is a dynamic molecule that can fold into a vast variety of structures depending on alternative intramolecular base-pairings whose are highly dependent to the RNA environment (presence of biological macromolecules, chemical modifications or to the physicochemical environment such as pH, osmolarity or temperature). These RNA structures play a powerful role in encoding and regulating genetic information, explaining the increasing interest in studying RNA secondary structure and its higher-order tertiary conformation.

RNA structure exhibits a hierarchical organization with multiple secondary structure motifs such as hairpins, mismatches, bulges or three-way junctions which together build the tertiary structure (11). The RNA secondary structure can be determined *in silico* by phylogenetic studies where the alignment of a large number of orthologous RNA sequences allows to predict base pairs and the structure of large RNAs (28). When such an alignment is missing, an alternative de novo approach based on free-energy minimization (35) can be used to predict the thermodynamically most stable conformation which broadly corresponds to the structure that is most likely to be accurate. Beside these above *in silico* approaches, biophysical techniques such as Nuclear Magnetic Resonance (NMR) are efficient methods to characterize the structure of small RNA motifs however this technique is highly restricted by buffer composition and cannot be performed under physiological conditions. Crystallography, and more recently cryo-electron microscopy (4, 27, 29, 55), were also able to determine at the atomic level the structure of RNAs in complexes but were unable to resolve the structure of RNAs in non-compact states nor the RNA unstructured regions.

In order to gain in accuracy and in variety of RNA studied, RNA footprinting experiments were designed. The first generation of probing methods was developed 35-40 years ago (18) but remained restricted to the *in vitro* and *in vivo* analysis of short regions of RNA. A first improvement was the development of "high-throughput" methods based on capillary electrophoresis analysis (97) allowing the quantification of reactivity data and the analysis of longer RNA regions. A few years ago, a new step was reached with the coupling of probing methods to Next Generation Sequencing, NGS. These novel approaches are powerful enough to study genome-wide RNA and to characterize thousands of RNA structures in a single experiment. Since altering RNA secondary structure, RNA-RNA or RNA-protein interactions modulate RNA processing, localization, translation, and decay, these tools allow the biologists to address new questions, unthinkable only 10 years ago, concerning RNA structure / function relationships.

Among the emerging questions that can be addressed, the influence of the cellular environment onto the RNA structure (85), is of great interest and comparing RNA structuromic in native conditions to data obtained *in vitro* is highly informative. The distinction of protein-binding site versus base-pairing can also be assessed. Actually, a nucleotide protection from chemical probing modification does not directly reveal the nucleotide's interaction partner. This protection can in fact be due to base-pairing or RNA-binding protein. By comparing the *in vivo* and *ex vivo* states, the nature of the protection, i.e. protein-binding site or base-pairing, can in theory be determined. Moreover, the remodeling of RNA structure in cell by protein partners such as in the case of riboswitch structures (reviewed in (80)) or upon binding of chaperon proteins (23, 36, 57) can be visualized.

Beside the RNA secondary structure studies, advances performed in the development of novel transcriptome-wide sequencing technologies also allowed the identification of post-transcriptional RNA modification (reviewed in (32, 46)). Approximately 150 RNA modifications are together termed "epitranscriptome" (70). These modifications are present in non-coding RNAs (ribosomal RNA, transfer RNA and small nuclear RNA) and eukaryotic mRNAs. N^1- and N^6-methyladenosine, 5-methylcytidine, 5-hydroxylmethylcytidine, pseudouridine and inosine are the most predominantly RNA modification internally found in eukaryotic mRNA (reviewed in (68)). It is interesting to note that some of these RNA modifications are known to alter the secondary structure of mRNAs by changing the H-bond formation pattern or by modifying the electron density of the aromatic ring and the base stacking: m^1A at translation start site and first splice site in coding transcripts (17), m^6A favoring locally unstructured transcripts (67, 79), pseudouridine altering stop codon read-through (21) as well as m^1A crucial for human mitochondrial $tRNA^{Lys}$ to obtain the canonical cloverleaf structure (31).

The classical experimental methodologies for RNA structure determination are going to be reviewed first with a brief presentation of the different probe specificities. Since these methodologies have recently been adapted for *in vitro* and *in vivo* genome-wide analysis, insights gained from each approach will be discussed in detail. Several biases introduced during generation of the sequencing library as well as the challenging computational analysis of NGS data that can strongly infer misinterpretation of sequencing results will be highlighted and briefly discussed at the end of the review.

I. RNA PROBING REAGENTS

Enzymatic and chemical probes can be used to assess the RNAs secondary structure under single-hit kinetics conditions. Such conditions ensure on average one cleavage or modification event per RNA molecule so that there is no 5' signal decay (5) and RNA is not refolded due to multiple modifications localized in close proximity. Nucleotide accessibility or reactivity indicates their propensity to be in a paired or unpaired state, and can be determined using a variety of readout methods. The target positions of enzymatic and chemical probes are illustrated in figure 1.

I.1. Enzymatic probes

Several nucleases are commonly used to probe RNA in a wide range of physiological buffers and have different structural specificity. RNase V1 cleaves double-stranded or structured regions within RNAs without base specificity and generates fragments with a 5'-phosphate. Nucleases S1 and P1 are zinc-dependent endonucleases, unspecific of the base moiety but which cut RNAs in single stranded regions leading to 5'-phosphate RNA fragments. RNase T1 and T2 cleave specifically the bond adjacent to the 3'-phosphate of unpaired guanosine (T1) or preferentially the bond adjacent to unpaired adenosine (T2) with formation of a 3'-phosphate RNA fragment. RNase A cleavage also results in the formation of a 3' phosphate fragment but this enzyme is specific of pyrimidine residues (Figure 1). Other more exotic nucleases such as RNases U2 and Cl3 or from Neurospora crassa were also reported in the literature to address the secondary structure of RNAs but are less used (18). Nevertheless, due to their large size, enzymatic probes are sensitive to steric hindrance and do not allow the detection of small RNA structural motifs such as bulges or mismatches, and are not suited for studies where RNA is bound to proteins. Another limitation of the enzymatic mapping approach is that enzymes are restricted to *in vitro* applications because of their membrane–impermeable nature and their requirement for non-physiological conditions (18).

I.2. Chemical probes

A strength of the chemical probes is their diversity. Many probes have been developed, and they have been designed to interrogate the top of the base (Hoogsteen face), the Watson-Crick face or the sugar-phosphate backbone depending on their specificity (**Figure 1**). For a review, see (20). Due to their smaller size compared to enzymes, chemical probes allow to obtain more detailed information, at the

nucleotide level. The choice of a specific probe is also driven by the reaction time scale and its ability to enter living cells and quite often several chemical probes or a mix of chemical and enzymatic probes based experiments are required to obtain an accurate RNA structure.

- The base-specific chemical probes interrogate the Hoogsteen face and/or the Watson-Crick face of the base but have the drawback of having nucleotide specificities. For example, dimethylsulfate (DMS) which methylates N1-A, N3-C and N7-G at neutral pH is largely used to identify unpaired adenosine and cytosine nucleotides *in vitro* and in a wide variety of organisms (94). However, as the introduction of the methyl group is located on the Hoogsteen face of G an additional step is needed to detect this modification (61) and nowadays DMS is less used for guanine probing. Base specific reagents also include 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metha-p-toluene sulfonate (CMCT) which reacts primarily with N3-U and N1-G of unpaired nucleotides under basic conditions, 2-keto-3-ethoxy-butyraldehyde (kethoxal) forms an additional ring between the primary amine located on C2 and the N-1 of unpaired guanine. The modification reaction is reversible but under slightly acidic conditions or in the presence of borate ions the adduct is stable. At neutral pH, diethylpyrocarbonate (DEPC) reacts specifically with N7-A and after treatment with aniline it allows the detection of adenosine implicated in tertiary interactions.

- Other chemicals interrogate the sugar-phosphate backbone with the advantage that they are not specific to an individual nucleotide. For example, ethylnitrosourea (ENU) is an alkylating reagent specific of the oxygen atom of phosphate groups non-engaged in tertiary interactions nor in cation coordination. The alkylation results in the formation of an unstable phosphate tri-ester, which leads to RNA cleavage under mild alkaline treatment. Hydroxyl radicals also cleave the RNA backbone but after abstraction of an hydrogen atom from C4' and/or C5' ribose position, providing nucleotide-level information on solvent accessibility of the nucleic acid backbone and tertiary fold of the RNA molecule (84). Until recently, hydroxyl radical probing was limited to *in vitro* applications, since radicals were generated by Fenton reagents: hydrogen peroxide, Fe(II)-EDTA and sodium ascorbate. But this probe has been adapted to effective footprinting of RNA-protein complexes *in vivo*. In that case, a synchrotron X-ray beam generates hydroxyl radicals inside the cell with a 100 milliseconds exposure (1, 2).

In-line probing technique is not strictly speaking a chemical probing technique since it is based on the natural ability of each nucleotide inside RNA to form a 2'-3'-cyclic phosphate product leading to the cleavage of the ribose-phosphate chain. This event is mediated by the attack of the sugar 2'-OH on the adjacent phosphate di-ester (66) and occurs more frequently when the two groups are "in-line", that is when the nucleotides are flexible. Whilst this technique has been used to investigate the modification profile of RNAs implicated in a riboswitch (47), it is not adapted to study fast RNA structure transitions since the in-line probing technique takes about 40 hours to perform cleavages in mild conditions. The same mechanism, but at the minute scale, occurs with metal-induced cleavages (22). The best studied example is the probing with lead(II) ion which activates the 2'-OH of nucleotides located in single-stranded regions such as bulges and loops but also in paired regions destabilized by non-canonical interactions or distortions (8, 10).

SHAPE (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension) probes are not susceptible to the base nature and under slightly basic conditions they react with the ribose 2'-hydroxyl group of flexible nucleotides often meaning unpaired. These reagents are powerful because they are able to interrogate the local dynamic of the 4 different nucleotides at the same time (14). Moreover, the increasing number of reagents available allows the study of various biological processes where RNAs are involved. Indeed, 1-methyl-7-nitroisatoic anhydride (1M7) (26) and benzoyl cyanide (BzCN) (54) quickly react with RNA (14 s and 0.25 s, respectively) and are thus well suited for the study of dynamic RNAs. On the other hand, N-methylisatoic anhydride (NMIA) (26), 2-methylnicotinic acid imidazolide

(NAI) and 2-methyl-3-furoic acid imidazolide (FAI) (78) are less sensitive to 2'-OH attack and react within 10, 33 and 73 min, respectively, compatible with *in vivo* study of RNA structures (78, 85).

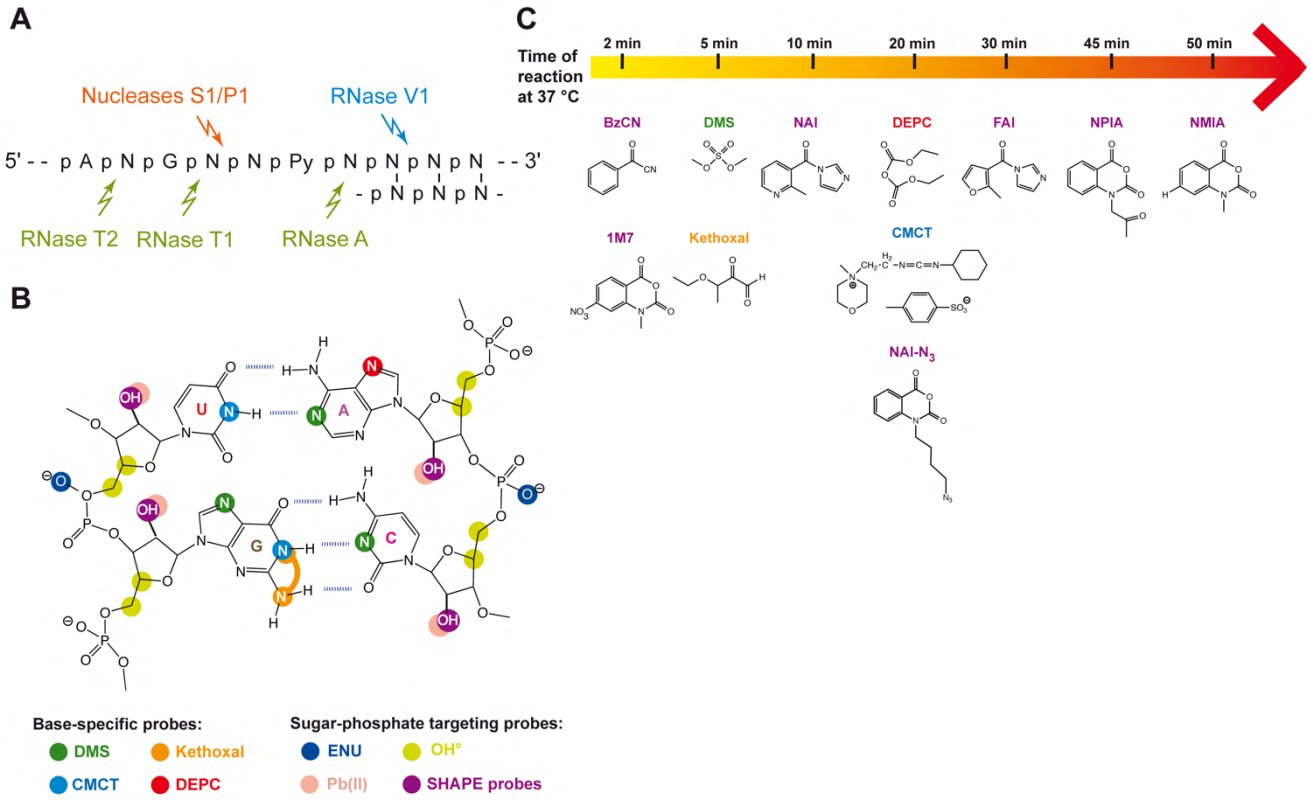


Figure 1

Mechanism of action of chemical and enzymatic probes.

A. Bases preferences of various nucleases. **B**. Canonical Watson-Crick interactions and target positions of multiple chemical probes. Hydroxyl radicals are generally generated by Fenton reagents (H_2O_2, Fe(EDTA) and sodium ascorbate) and Pb(II) is from lead acetate. **C**. Chemical structure of probes with the typical time of reaction reported in the literature (14, 20, 26, 54, 59, 78, 85) but the time reaction is highly dependent on probes' concentration and temperature.

II. CONVENTIONAL REACTIVITY READOUT METHODS

Sites of chemical adduct formation or strand scission can be detected in several ways. When the probing methods involve enzymatic or chemical strand scission, end-labeled RNAs at their 5' or 3' extremity can be used. In that case, the population of cleaved RNA fragments is sized by denaturing polyacrylamide gel electrophoresis and detection of labeled fragments allows the identification of the position of cleavages on the RNA molecule. However, this method can only be used to analyze short RNAs (<300 nucleotides). For probing the structure of large RNAs, primer extension by reverse transcriptase is a preferred alternative (63). Usually the modified RNA is hybridized with a 5'-terminal [^{32}P]-labeled oligodeoxyribonucleotide complementary to a chosen sequence in the RNA. Since the reverse transcription reaction is usually blocked when the enzyme encounters a RNA-chemical adduct or a strand scission, a population of truncated radiolabeled cDNA whose 3'-end corresponds to the nucleotide before the site of modification or to the nucleotide of the cleavage site is generated. This population is traditionally resolved by denaturing polyacrylamide gel electrophoresis and detection is done with either X-ray autoradiography or a phosphorimager apparatus. Reactivity are mapped back

to the primary RNA nucleotide sequence thanks to dideoxyribonucleotide sequencing reactions performed in parallel. Band intensity of nucleotide fragments resolved on gel can be quantified by the semi-automated footprinting analysis (SAFA) program. However, cDNA synthesis is not only blocked by chemical adducts or cleavage sites but also occurs at sites with a stable RNA structure, natural modification (such as m^2G or m^6A) or due to RNA degradation. Thus, a negative control sample must be conducted in parallel to the experiment. Data from the probed samples are then normalized by control data to obtain chemical reactivity. Nevertheless, experimental readout using radioactively labeled RT primers is labor-intensive and limited, leading to semi-quantitative data for reading windows of 150 nucleotides only due to the resolution of polyacrylamide gels.

A great improvement over this traditional approach has been the use of capillary electrophoresis (CE) for sizing the pool of cDNAs obtained in probing experiments with 2'-*O*-acylating reagents (97). This modification to the readout method (termed hSHAPE) uses fluorescently labeled RT primer and capillary electrophoresis to extend the analysis window to approximately 500 nucleotides. CE electropherograms, representing fluorescent intensity for each nucleotide, are generated and subjected to multiple bioinformatics treatments in order to be converted into single-nucleotide reactivity. Several analytical pipelines have been developed to semi-automatically convert CE electropherograms into reactivity. These include capillary automated footprinting analysis (CAFA) (51), ShapeFinder (88), high-throughput robust analysis for capillary electrophoresis (HiTRACE) (98), Fast Analysis of SHAPE traces (FAST) (60), SHAPE-CE (6) and QuShape (38). All these programs need electrophoretic traces for the chemically modified RNA, the unmodified reference and one or two sequencing reactions as input. ShapeFinder need 4 traces to be analyzed in the same capillary requiring of a DNA primer labeled with 4 different fluorescent probes, whereas QuShape needs only 2 fluorescently-labeled primers with an analysis split in two capillaries. Finally, SHAPE/FAST needs only one labeled primer and the presence of an internal standard allows the comparison of multiple capillaries. The use of decreasing number of fluorescent probes for one experiment was developed because it reduces the complexity of the data treatment and allows increased automation and thus saved time. ShapeFinder requires the user to manually select tools making data processing time consuming and requires significant user training, whereas QuShape software is easier to set with the automation of multiple treatments (data entry, pre-processing, signal alignment, sequence alignment, reactivity estimation). In all cases, numeric reactivity data is output for use as input in secondary structure prediction programs such as RNAstructure (14). Even though the pioneering hSHAPE method has been applied to the detection of other chemical or enzymatic modifications (51) and capillary electrophoresis has significantly increased the resolution and the coverage length compared to sequencing gel, this readout method is not suitable for genome-wide analysis because large amounts of RNA are still needed, as well as the design of gene-specific primers.

Recently, gel- and capillary electrophoresis techniques have given way to next-generation sequencing methods, allowing for the multiplexing of several samples within single experiments or even transcriptome-wide analysis.

III. HIGH-THROUGHPUT METHODS

The democratization of the fast and deep sequencing of genome linked to the Next Generation Sequencing (NGS) technique advent allows the outbreaks of numerous protocols to probe RNA structure special by their length (genome wide), their complexity (complex RNA mixture) or their low abundance. Despite some differences due to the RNA origin, these techniques share several core steps. Similar to the above strategies, RNA structure is interrogated using enzymatic or chemical probes and then RNA is reverse transcribed into cDNA using one of two strategies: the truncation strategy or the mutational profiling approach. Library is then prepared for sequencing by addition of two adapters flanking the cDNA and barcodes for sample multiplexing. Finally, bio-informatic processing of sequencing reads is performed for reads alignment and reactivity determination.

III.1. NGS coupled to RNA segmentation or ligation

The first methods developed for the study of RNA structures in a high-throughput manner have been derived from enzymatic probing techniques and were carried out *in vitro* on RNA mixtures extracted from cells. Parallel analysis of RNA structure (PARS) combines RNase V1 (specific of paired or stacked nucleotides) and nuclease S1 (specific of single stranded nucleic acid) treatments and reactivity is scored by the ratio of V1 to S1 number of reads for a given position (39, 89). Almost at the same time, Fragmentation sequencing (Frag-seq) was reported by Underwood *et al*. In this approach, the RNA is cleaved in single stranded regions by nuclease P1 and any natural degradation, potentially being quantified as a P1 cut is taken into account by two negative controls, one without P1 enzyme to quantify the amount of unspecific cleavages leaving a 5' phosphate and one with T4 Ligase treatment to detect cleavage products without a 5' phosphate end. (86). For a comparative study of the two methods see (96). A chemical probe leading to the cleavage of RNA was also coupled with Illumina sequencing. In the HRF-Seq technique, hydroxyl-radical footprinting was used to interrogate the solvent accessibility of the E. coli 16S ribosomal RNA followed by a random priming reverse transcription step (40). More recently, an innovative strategy called RNA proximity ligation (RPL) was developed (65) to gain information about nucleotide-nucleotide interactions or at least to determine the nucleotides close in three-dimensional space. After the treatment of *ex vivo* RNAs with RNases and an end-repair step with T4PNK under non-denaturing conditions, the nucleotides in close vicinity are linked together by T4 RNA Ligase I. Ligation junctions are mapped and counted thanks to Illumina sequencing and allow the detection of stem loop structures as well as long-range interactions.

These strategies constitute a great step in the study of RNA structures since they have been able for the first time to interrogate a complex mixture of long RNAs and in particular the human transcriptome (90). However, a major drawback is that they are restricted to *in vitro* studies.

III.2. NGS coupled to chemical probing

The **Figure 2** briefly recapitulates the high-throughput structure probing methods described in this part. In 2011, J.B. Lucks *et al.* have reported for the first time the combination of probing with chemical adduct formation and next-generation sequencing strategies (49). Compared with the previously described "cleaving" approaches, the so-called **SHAPE-Seq** technique takes the advantage of the SHAPE chemistry that interrogates all the nucleotides of RNAs in variable conditions and at high resolution. As a proof of concept, SHAPE-Seq was used to study the structures of a complex mixture of RNAs transcribed *in vitro* and that differ by single nucleotide point mutations (49). The protocol has been upgraded to **SHAPE-seq v2.0** (48) and **v2.1** (92) with the development of a strategy using a universal RT primer and a selective PCR step to amplify ligated products containing transcribed cDNAs.

Associating chemical probing to high-throughput sequencing opened the gate to *in vivo* RNA structure probing methods and allowed to study the influence of the viral / cellular environment on RNA structures.

The first *in vivo* genome-wide analysis at nucleotide resolution was developed in 2014 and applied to the study of the Arabidopsis thaliana RNAs (16). After DMS methylation of unprotected adenine and cytosine, RNAs were subjected to reverse transcription using random hexamer primers, allowing the coverage of the entire RNA. A pool of cDNAs of different sizes was generated and subjected to intermolecular ligation of Illumina adapter. This technique called **Structure-seq** provided structural information among more than 10,000 transcripts (15, 16). Several other techniques using the DMS probe slightly differ from the **Structure-seq** approach by performing a RNA fragmentation followed by the ligation of a universal adapter at the 3'-end of the RNA prior to reverse transcription. This step used in **DMS-seq** (69) and **Mod-seq** (82) was designed to ensure the random initiation of reverse transcription and is thus suitable for studying RNA of any length. Nevertheless, it also involves more technical steps on the RNA template and requires a careful analysis to discriminate modification sites from the end of unmodified RNA fragments.

At the same time, Chemical Inference of RNA Structures (**CIRS-seq**) was developed to study the impact of the cellular environment on RNA structure. Using DMS and CMCT probes, they analyzed the secondary structure of *ex vivo* mouse embryonic stem cell transcripts after a native deproteinization treatment (34). cDNAs covering the full RNA landscape were obtained by random-priming RT, ligation to Illumina adapters and sequencing. Nevertheless, since the chemical probing is performed on cell lysate after proteinase K treatment, CIRS-seq is not exactly an *in vivo* technique.

The *in vivo* probing methods at the genome-wide scale have been improved by using SHAPE reagents which interrogate the four different nucleotides rather than two with the DMS probe. The in-cell probing with 1M7 (85) and NAI (78) reagents has been set up first by capillary and denaturing gel electrophoresis, respectively and next coupled with the Illumina sequencing technology (75, 79).

Another layer of improvement was achieved with the SHAPES (62) and icSHAPE (Spitale et al. 2015) protocols which take advantage of bifunctional chemical probes such as N-propanone isatoic anhydride (NPIA) and NAI-N_3, respectively. After RNA probing, these two reagents can be coupled to a biotin molecule allowing the enrichment of the modified RNAs and/or cDNAs that terminated at a probed position, depending on the treatment protocol. This selective pool-down eliminates a large amount of unwanted side products and sequencing libraries are thus enriched for structure informative molecules. Even if the two techniques share the same idea of selection, they have their own advantage and drawbacks. The NAI-N_3 reagent used in the icSHAPE technique is interesting because, like NAI, it easily penetrates the cell membranes (78). On the other hand, the SHAPES technique involves an additional enzymatic cleavage during the selection step that removes the cDNA caused by premature termination even from reverse transcription of modified RNA. This wash only leaves on the beads the cDNA that contain the structural information and makes unnecessary to drive a negative control.

The icSHAPE technique revealed an increased reactivity of adenosine residues *in vivo* compared to *in vitro*, whereas cytidine and guanosine residues were less impacted by the environment. The authors conclude that reactivity difference between *in vivo* and *in vitro* **icSHAPE** measurements provides insights into *in vivo* RNA structural dynamics and reveals structural profiles controlling gene expression such as translation and post-transcriptional modifications (79).

Existing methods give confident answers when applied to broad questions about structure of a cellular RNA population. Yet the analysis of a specific RNA target, presumably in low abundance, or performing single cell sequencing is still technically challenging. The team of Lucks developed **in-cell SHAPE-Seq** (91), adapted from the **SHAPE-Seq v2.1** method (92), to study targeted RNAs inside the cell. Low abundance of RNAs of interest required an optimization of PCR steps necessary to add the Illumina adaptors. Indeed, a careful protocol has to be designed to obtain the right balance between sufficient number of PCR cycles required to amplify the low amount of samples and the amplification of unwanted side products (91). In addition to the determination of RNA structure, the team of Lucks addressed their function by measuring gene expression through fluorescent reporter assay.

The above described methods, based on the identification of reverse transcriptase truncation products, allowed the study of large number of RNAs *in cellula* in various environments. However, the multiple ligation steps used in these techniques to perform the cDNA sequencing represent a major bottleneck for one wants to study very low amount of biological RNA or discriminate between distinct conformations of the same RNA. Several approaches have recently been developed to overcome this problem by directly identifying the modification positions. These new strategies take the advantage of two RT enzymes that, under specific conditions, after pausing, induce mutations in the nascent cDNA when they encounter DMS or SHAPE modified nucleotides. As the mutational profiling approach does not involve stop detection, cDNAs are directly analyzed by massively parallel sequencing avoiding the complex RNA ligation steps and the biases introduced by the preparation of the library in classical probing. Another benefit to detect RNA modifications in this way is that the read-through method is

not sensitive to the background RNA degradation, nor to the RT drop-off or the 5' signal decay. Moreover, on the contrary to the truncation strategy, RT read-through of the modification site allows the analysis of multiple modified nucleotide positions in a single RNA molecule and should decrease the amount of starting material to analyze.

The superscript II (SSII) enzyme has been the first enzyme used for the read through of SHAPE modifications in the **SHAPE-MaP** technique developed by the team of K. Weeks (74, 76). Indeed, using Mn^{2+} containing buffer, this RT enzyme can incorporate a mutation or induce a deletion at a site corresponding to a modified nucleotide during cDNA synthesis. Generally, SHAPE adducts are misread as A or T but additional information are obtained from other mis-incorporations and more than 50% of modified nucleotides are detected. Massively parallel sequencing analyzed by ShapeMapper software allows the determination of a mutation rate for each nucleotide. Subtracting the number of mutations obtained for an untreated control from the experiment where RNA is treated with SHAPE reagent and normalizing with the denaturing control gave the modifications position and frequency. These reactivity plots can then be used as input in several softwares such as RNA structure, ShapeMap_Folding_pipeline (Siegfried 2014) or SuperFold (Smola 2015b) to model the RNA secondary structures. **SHAPE-MaP** has been validated with the already established structure of the E. coli thiamine pyrophosphatase riboswitch aptamer bound with its ligand and the E. coli 16S RNA. The *ex viro* structure of the entire HIV-1 genomic RNA has also been deciphered with this protocol and constitutes an update of the *in vitro* model proposed in 2009 (93). Nevertheless, it is important to keep in mind that an important quantity of extracted viral genomic RNA (1 ug) was necessary to perform this protocol (74). Later, the **SHAPE-Map** approach has been adapted to *in vivo* studies by performing in-cell probing with the 1M7 reagent and is, as **icSHAPE**, based on the subtraction of *in cellula* from *ex vivo* reactivity (75).

More recently, the thermostable group II reverse transcriptase (TGIRT) enzyme, engineered by the team of A. Lambowitz (19, 53), has been adapted in the **DMS-MaPseq** approach (101). The team of S. Rouskin compared the suitability of the SSII and the TGIRT for the **DMS-MaPseq** approach. DMS modifications detection has first been checked by performing two replicates using the yeast 25S rRNA which has endogenous m^1A modifications, one of the two sites methylated by DMS. Surprisingly, it seems that SSII under-estimates DMS modification depending on the nucleotide context (101). The authors suggest an underlying failure of SSII to robustly detect adenines compared to other set of data obtained with other protocols. Analyzing the distribution of mutation type generated reveals that TGIRT produces around 6% of insertions/deletions compared to nearly 30% by SSII, the presence of indels creating positional ambiguity when sequencing data are analyzed. Moreover certain G-U residues are highly detected suggesting a propensity for non-random errors in cDNA synthesis (101). Nevertheless, this direct comparison of the TGIRT and SSII ability in mutational profiling approaches and the robustness of this method was based on only one study and needs to be further assessed. An implemented version of **DMS-MaPseq** (101) can specifically focus on low-abundance RNA by using target-specific primers. RT primers can also be tagged with unique molecular indexes so that each RNA molecule is followed at a single-molecule resolution. Nevertheless, the targeted nature of **DMS-MaPseq** limits the size of RNA region analyzed so that this protocol cannot be applied to genome-wide purpose.

Since **icSHAPE** and *in vivo* **SHAPE-MaP** claim to measure *in vivo* structure with next-generation sequencing, the team of K. Weeks compared data of the well-characterized SRP RNPs obtained with the two methods, and showed a poor agreement (101). They hypothesize these differences to be due to the NAI-N_3 half-life, longer than the one of 1M7 so that NAI-N_3 reacts slower and does not reflect the RNA steady-state structure.

It is important to note that several substantial differences exist when comparing the two approaches. A rigorous comparison of the two techniques would have required a direct experimental testing instead of a comparative analysis of primary data. Besides the fact that used SHAPE reagents differ in

solubility and half-life, the SRP RNA has been produced in different cell types. Regarding the experimental procedure, the sequencing library is enriched with acylated RNA molecules in **icSHAPE** whereas **SHAPE-MaP** includes all generated cDNAs. Finally, the read-out of SHAPE reactivity relies on cDNA truncation or cDNA mutation. To solve this issue, the team of H. Chang directly compared the *in vitro* and *in vivo* modification ability of NAI-N_3 and 1M7 reagents using gel electrophoresis (45). Whereas NAI-N_3 and 1M7 are clearly able to *in vitro* probe RNAs with a slightly higher signal-to-background ratio for NAI-N_3, this was not shown *in vivo* at the published 1M7 probe concentration (75) and even with a ten-fold excess. These results were confirmed in other cell types and targeted RNAs (50). By permeabilizing mouse embryonic stem cells, they showed a higher modification signal generated by 1M7 compared to intact cells. The authors suggested that the poor 1M7 *in vivo* activity was due to its weak permeability to living cell barriers and that the raw data obtained *in vivo* were due to the probing of residual RNAs from cells/viruses with compromised membranes which signal was amplified during the library preparation. However even if the lipophilicity of the probes and the way they go through membranes is of great importance for *in vivo* studies, the difference of reactivity observed between NAI-N_3 and 1M7 may be linked to another important point such as the shorter half-life of 1M7 which reacts with OH groups within 14 s, when NAI-N_3 reacts in 5 min.

Method	Organism	RNA pool	Chemical probe	Specificity	Application	Readout	NGS platform	Analysis pipeline	Reference
Structure-seq	yeast	poly-A selection	DMS	A and C residues	*in vivo*	truncation strategy	Illumina HiSeq	Structure fold	Ding et al. Nature 2013
DMS-seq	yeast / human	poly-A selection	DMS	A and C residues	*in vitro / in vivo*	truncation strategy	Illumina HiSeq	SOAP	Rouskin et al. Nature 2014
Mod-seq	yeast	total RNAs	DMS	A and C residues	*in vivo*	truncation strategy	Illumina HiSeq	Mod-seeker	Talkish et al. RNA 2014
CIRS-seq	mouse	total RNAs	DMS / CMCT	A / C (DMS) and G / U (CMCT) residues	*ex vivo*	truncation strategy	Illumina HiSeq	custom scripts	Incarnato et al. Genome biol. 2014
SHAPES	bacterium	total RNAs	NMIA / NPIA	2' hydroxyl	*in vitro / ex vivo*	truncation strategy	Illumina HiSeq	custom scripts	Poulsen et al. RNA 2015
SHAPE-MaP	bacteria / viruses / human	specific targets	NMIA / 1M6 / 1M7	2' hydroxyl	*in vitro / ex vivo*	mutational profiling	Illumina Mi / HiSeq	ShapeMapper / SM folding pipeline	Siegfried et al. Methods 2014
icSHAPE	mouse	total RNAs	NAI-N3	2' hydroxyl	*in vitro / in vivo*	truncation strategy	Illumina HiSeq	custom scripts	Spitale et al. Nature 2015
in-cell SHAPE-seq	bacteria	specific targets	1M7 / DMS	A / C (DMS) and 2' hydroxyl	*in vitro / in vivo*	truncation strategy	Illumina Mi / HiSeq	Spats	Watters et al. NAR 2016
SHAPE-MaP	mouse	specific targets	1M7 / NAI-N3	2' hydroxyl	*ex vivo / in vivo*	mutational profiling	Illumina Mi / HiSeq	SHAPEMapper	Smola et al. Biochemistry 2015
DMS-MaPseq	yeast	total RNAs	DMS	A and C residues	*in vitro / in vivo*	mutational profiling	Illumina HiSeq	custom scripts	Zubradt et al. Methods 2017

Figure 2

Summary of the current chemical probing methods.

Several information are described for each method: the studied organism, RNA studied, probe, specificity, application, readout, NGS platform, analysis pipeline and reference of the protocol.

IV. BIASES GENERATED DURING LIBRARY PREPARATION

NGS requires the preparation of libraries with fusion of sequencing adapters to nucleic acid samples, followed by PCR amplification before sequencing. Regarding the wide variety of NGS library preparation protocols, library generation is differently performed and potentially contains biases that compromise the quality of NGS datasets. It is important to understand these biases to improve library quality. Several steps of the library generation are impacted:

- RNA fragmentation

RNase III and zinc-induced hydrolysis are commonly used to mediate RNA cleavage. In contrast to zinc treatment, RNase III cleaves dsRNA in a sequence and structure dependent manner. Comparison of these two techniques indicates that zinc-mediated fragmentation provides more robust and accurate transcript identification when transcriptome reassembly is performed (95).

- RT random priming

Priming of random hexamers in cDNA synthesis is known to induce sequence bias, then read counts from RNA-Seq data may not be randomly distributed (30). Mismatching is also observed (87) and these two biases can be explained by a preference for specific local nucleotide composition. This leads to a significant bias in the nucleotide composition of the first sequencing reads (7, 71, 99).

- Adapter ligation

RNA ligases are mostly influenced by RNA sequence and structure at the ligation site as well as RNA-adapter structure (24, 37, 81, 100). Sorefan et al. have reported that the addition of three degenerated bases at the 5'-end of the adapter to be ligated significantly decreased the sequence-dependent bias (77).

- Size selection

Enrichment for DNA molecules of a selected size range can be performed with solid-phase reversible immobilization (AMPure beads from Beckman Coulter) or gel extraction. A more precise size selection is obtained by gel extraction but on the other hand, the melting step of agarose gel slices in chaotropic salt buffer reduces the recovery of AT-rich sequences. This is mostly due to denaturation of these samples which have a reduced affinity for dsDNA purification columns (64).

-PCR amplification

A major bias is introduced in the sequencing library due to the fact that all samples are not amplified with the same efficiency and in particular, GC-rich and AT-rich fragments shown to be underrepresented during the library preparation (3, 58). Several optimizations of the standard Phusion PCR protocol have been described with utilization of the Kapa HiFi (58) and phi29 (13) DNA polymerase.

It is important to keep in mind that a higher amount of starting material limits PCR amplification and increases the library complexity. An alternative is to avoid PCR amplification by ligating complete adapters necessary for bridge amplification on the sequencing flow-cell (42). Nevertheless, the relatively large quantity of input material constitutes a drawback. Single cell RNA sequencing protocols are facing this issue with low amount of starting material. Digital RNA-Seq technique tries to address it by ligating an overwhelming number of unique barcode sequences to cDNA samples followed by PCR amplification of these barcoded cDNA molecules. This indexing method enables the distinction between biological and PCR-derived duplicates (73).

Moreover, there is still a clear debate in the scientific community to determine in which extent increasing the number of PCR cycles amplify the background noise by rising the chance to introduce mutations. Interestingly, the team of Lucks tested increased cycling from 15 to 20 PCR cycles on the same sample to assess this question and obtained little reactivity differences (92).

Considering these biases is crucial to avoid erroneous interpretation of sequencing data. In addition to an improvement of the library quality, development of bioinformatics tools could also help to compensate these biases.

V. READS ALIGNMENT AND COMPUTATIONAL DATA ANALYSIS

Coupling chemical probing to NGS has been a powerful tool to probe the RNA structure on a genome-wide level as well as on thousand molecules at the same time. However, these experiments generated huge amount of complex dataset and the concomitant development of informatics tools to perform the reads alignment and the data analysis was necessary. Depending on the experimental procedure used to generate the library, different issues have to be addressed. But in general the most important points are (1) Correcting the biases introduced during library preparation to avoid mis-interpretation. (2) Taking into account the inter-replicate variability in both experiments and controls. (3) Obtaining a number of reads terminating at a given nucleotide sufficiently high to get reliable reactivity values.
Several pipelines were developed, i.e. the ShapeMapper pipeline for SHAPE-seq technique (75), the SPATS pipeline for SHAPE-Seq technique (91). In their work combining DMS probing and NGS (DMS-seq), Rouskin et al. aligned the raw sequences obtained from the HiSeq2000 against the reference sequence and they kept only the unique alignments. The raw data were then normalized proportionally to the most reactive residue (69). In that respect, the treatment of mutational profiling approaches such as DMS-MaPseq (101) and SHAPE-MaP (76) are easier and reads are aligned using Toplat v2.1.0 with Bowtie2 and ShapeMap Folding_pipeline.

Mod-seq (82) and Structure-seq (16) are quite similar with a RT drop off rate calculated as the total number of reads that stop at a nucleotide divided by the total number of reads covering that nucleotide. A two channel Poisson expectation maximization (EM) algorithm is then performed to determine if the drop off observed with modified RNA is statistically above the noise. Mod-seq is more powerful since it uses a probabilistic model to map sites of chemical modification with high accuracy and shows the additional benefit to take into account the replicate information. More recently, the efficiency of Beta-uniform mixture hidden Markov Model (BUM-HMM) was reported by Selega et al. (72) and seems to give better results than Mod-seq and Structure-seq.

Using these pipelines was often challenging for scientists and user friendly interfaces were designed. Kielpinski et al. have developed an R package (RNAprobR) to standardize and simplify the treatment of experiments with massively parallel sequencing.

VI. CONCLUSION

In the cell environment context, the structure of a given RNA may vary depending on its involvement in physiological processes such as translation, RNA transport, interaction with RNA-binding proteins or maturation (or in response to environment changes due to stress in eukaryotic cells). Such local RNA structural rearrangements create a structural heterogeneity, difficult to analyze in probing experiments because of the long incubation time of chemical reagents. Thus, our understanding of *in vivo* RNA secondary structure is still incomplete despite tremendous advances made in the field with techniques analyzing structure-function relationship on a genome-wide scale.
The design of new chemical probes able to be used *in vivo* and specifically targeting different cellular compartments will improve current techniques (43). Moreover, *in vivo* hydroxyl radical probing could be a powerful alternative method but still need to be coupled to high throughput sequencing techniques (2).

Direct sequencing of RNA, without intermediate cDNA synthesis, will greatly improve the study of RNA secondary structure. A first miles stone was reached last year in that area (25) with the detection of m^6A RNA modification by Oxford Nanopore Technologies (ONT). This new technique, initially developed for DNA sequencing, allows the detection of modifications (on the base moiety) during RNA

transit through nanopores by measuring the ionic current passing through the pore. Since it was possible to detect m^6A modification at a single nucleotide resolution and from a single RNA molecule, one can imagine in a next future using ONT to detect other modifications (56) and among them the ones introduced by base-specific chemical probes improving the RNA secondary structure determination.

The development of new technical approaches to study the three-dimensional RNA structure is also required. Multidimensional chemical mapping (MCM) methods have been developed to determine how the conformation of one nucleotide is affected by chemical modification and mutational profiling (Mutate-and-map (41), RING-MaP (33) and MaP-2D (83)) and hydroxyl radical cleavage (Multiplexed •OH cleavage analysis, shortly MOHCA (12) and MOHCA-seq (9)) at every other nucleotide. Nevertheless, MCM methods with gel readout have a resolution of up to 200 nucleotides in length, improved to approximately 1000 nucleotides when the method is coupled with NGS techniques. The majority of sequencing reads corresponds to modifications at unstructured nucleotides that are not informative about RNA–RNA contacts. Recent protocols tried to improve this by filtering for proximal segment pairs prior to sequencing. The CLASH approach (Crosslinking, Ligation And Sequencing of Hybrids) targets RNA-RNA interactions, cross-linked at 254 nm. Unstructured nucleotides are removed through limited nuclease digestion followed by ligation of the remaining segments into chimeric sequences. Chimeric reads are identified bio-informatically after sequencing (44). Similar approaches focused on micro-RNAs have also been developed (reviewed in (52)).

Technical advances from these two last years allowed the investigation of numerous biological questions. Nevertheless, much work remains to be done to develop a technique powerful enough to identify long-range interactions and/or to optimize existing protocols. In addition, many *in vivo* existing protocols have been designed for a broad investigation of cellular RNAs structure but are not suited to focus on a RNA specific target, which requires the design of specific primers. The association of transcriptome-wide and targeted approaches can thus be used as complementary investigation methods.

1. Adilakshmi T, Lease RA, Woodson SA. 2006. Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res.* 34(8):e64

2. Adilakshmi T, Soper SFC, Woodson SA. 2009. Structural analysis of RNA in living cells by in vivo synchrotron X-ray footprinting. *Methods Enzymol.* 468(09):239–58

3. Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, et al. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12(2):R18

4. Amunts A, Brown A, Bai X, Llácer JL, Hussain T, et al. 2014. Structure of the yeast mitochondrial large ribosomal subunit. *Science*. 343(6178):1485–89

5. Aviran S, Pachter L. 2014. Rational experiment design for sequencing-based RNA structure mapping. *RNA*

6. Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, et al. 2011. Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. U. S. A.* 108(27):11069–74

7. Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40(10):e72

8. Brunel C, Romby P, Westhof E, Ehresmann C, Ehresmann B. 1991. Three-dimensional model of Escherichia coli ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.* 221(1):293–308

9. Cheng CY, Chou F-C, Kladwang W, Tian S, Cordero P, Das R. Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife*. 4:

10. Ciesiołka J, Michałowski D, Wrzesinski J, Krajewski J, Krzyzosiak WJ. 1998. Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.* 275(2):211–20

11. Cruz JA, Westhof E. 2009. The dynamic landscapes of RNA architecture. *Cell*. 136(4):604–9

12. Das R, Kudaravalli M, Jonikas M, Laederach A, Fong R, et al. 2008. Structural inference of native and partially folded RNA by high-throughput contact mapping. *Proc. Natl. Acad. Sci. U. S. A.* 105(11):4144–49

13. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, et al. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U. S. A.* 99(8):5261–66

14. Deigan KE, Li TW, Mathews DH, Weeks KM. 2009. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U. S. A.* 106(1):97–102

15. Ding Y, Kwok CK, Tang Y, Bevilacqua PC, Assmann SM. 2015. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* 10(7):1050–66

16. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 505(7485):696–700

17. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, et al. 2016. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. *Nature*. 530(7591):441–46

18. Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res.* 15(22):9109–28

19. Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. 2014. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. *Mob DNA*. 5:2

20. Fechter P, Parmentier D, Wu Z, Fuchsbauer O, Romby P, Marzi S. 2016. Traditional Chemical Mapping of RNA Structure In Vitro and In Vivo. *Methods Mol. Biol.* 1490:83–103

21. Fernández IS, Ng CL, Kelley AC, Wu G, Yu Y-T, Ramakrishnan V. 2013. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature*. 500(7460):107–10

22. Forconi M, Herschlag D. 2009. Metal ion-based RNA cleavage as a structural probe. *Methods Enzymol.* 468:91–106

23. Fu W, Rein A. 1993. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J. Virol.* 67(9):5443–49

24. Fuchs RT, Sun Z, Zhuang F, Robb GB. 2015. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One*. 10(5):e0126049

25. Garalde DR, Snell EA, Jachimowicz D, Heron AJ, Bruce M, et al. 2016. Highly parallel direct RNA sequencing on an array of nanopores. *bioRxiv*, p. 068809

26. Gherghe CM, Shajani Z, Wilkinson KA, Varani G, Weeks KM. 2008. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J. Am. Chem. Soc.* 130(37):12244–45

27. Greber BJ, Boehringer D, Leibundgut M, Bieri P, Leitner A, et al. 2014. The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature*. 515(7526):283–86

28. Gutell RR. 1993. Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. *Curr. Opin. Struct. Biol.* 3(3):313–22

29. Hang J, Wan R, Yan C, Shi Y. 2015. Structural basis of pre-mRNA splicing. *Science*. 349(6253):1191–98

30. Hansen KD, Brenner SE, Dudoit S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* 38(12):e131

31. Helm M, Giegé R, Florentz C. 1999. A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. *Biochemistry*. 38(40):13338–46

32. Helm M, Motorin Y. 2017. Detecting RNA modifications in the epitranscriptome: predict and validate. *Nat. Rev. Genet.* 18(5):275–91

33. Homan PJ, Favorov O V, Lavender C a, Kursun O, Ge X, et al. 2014. Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* 111(38):13858–63

34. Incarnato D, Neri F, Anselmi F, Oliviero S. 2014. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15(10):491

35. Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA.

Proc. Natl. Acad. Sci. U. S. A. 86(20):7706–10

36. Jalalirad M, Laughrea M. 2010. 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*. 407(2):225–36

37. Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R. 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* 39(21):e141

38. Karabiber F, McGinnis JL, Favorov O V, Weeks KM. 2013. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*. 19(1):63–73

39. Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, et al. 2010. Genome-wide Measurement of RNA Secondary Structure in Yeast. *Nature*. 467(7311):103–7

40. Kielpinski LJ, Vinther J. 2014. Massive parallel-sequencing-based hydroxyl radical probing of RNA accessibility. *Nucleic Acids Res.* 42(8):e70

41. Kladwang W, VanLang CC, Cordero P, Das R. 2011. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* 3(12):954–62

42. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods*. 6(4):291–95

43. Kubota M, Tran C, Spitale RC. 2015. Progress and challenges for chemical probing of RNA structure inside living cells. *Nat. Chem. Biol.* 11(12):933–41

44. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. 2011. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 108(24):10010–15

45. Lee B, Flynn RA, Kadina A, Guo JK, Kool ET, Chang HY. 2017. Comparison of SHAPE reagents for mapping RNA structures inside living cells. *RNA*. 23(2):169–74

46. Li X, Xiong X, Yi C. 2016. Epitranscriptome sequencing technologies: decoding RNA modifications. *Nat. Methods*. 14(1):23–31

47. Liberman JA, Wedekind JE. 2012. Riboswitch structure in the ligand-free state. *Wiley Interdiscip. Rev. RNA*. 3(3):369–84

48. Loughrey D, Watters KE, Settle AH, Lucks JB. 2014. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* 42(21):gku909

49. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, et al. 2011. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* 108(27):11063–68

50. McGinnis JL, Liu Q, Lavender CA, Devaraj A, McClory SP, et al. 2015. In-cell SHAPE reveals that free 30S ribosome subunits are in the inactive state. *Proc. Natl. Acad. Sci. U. S. A.* 112(8):2425–30

51. Mitra S, Shcherbakova I V., Altman RB, Brenowitz M, Laederach A. 2008. High-throughput

single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Res.* 36(11):e63

52. Mittal N, Zavolan M. 2014. Seq and CLIP through the miRNA world. *Genome Biol.* 15(1):202

53. Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, et al. 2013. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA*, pp. 958–70

54. Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129(14):4144–45

55. Nguyen THD, Galej WP, Bai X, Savva CG, Newman AJ, et al. 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*. 523(7558):47–52

56. Novoa EM, Mason CE, Mattick JS. 2017. Charting the unknown epitranscriptome. *Nat. Rev. Mol. Cell Biol.* 18(6):339–40

57. Ohishi M, Nakano T, Sakuragi S, Shioda T, Sano K, Sakuragi J. 2011. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res.* 39(8):3404–17

58. Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, et al. 2012. Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC Genomics*. 13(1):1

59. Paillart JC, Westhof E, Ehresmann C, Ehresmann B, Marquet R. 1997. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.* 270(1):36–49

60. Pang PS, Elazar M, Pham EA, Glenn JS. 2011. Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Res.* 39(22):e151

61. Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc. Natl. Acad. Sci. U. S. A.* 77(8):4679–82

62. Poulsen LD, Kielpinski LJ, Salama SR, Krogh A, Vinther J. 2015. SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA*. 21(5):1042–52

63. Qu HL, Michot B, Bachellerie JP. 1983. Improved methods for structure probing in large RNAs: a rapid “heterologous” sequencing approach is coupled to the direct mapping of nuclease accessible sites. Application to the 5’ terminal domain of eukaryotic 28S rRNA. *Nucleic Acids Res.* 11(17):5903–20

64. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, et al. 2008. A large genome center’s improvements to the Illumina sequencing system. *Nat. Methods*. 5(12):1005–10

65. Ramani V, Qiu R, Shendure J. 2015. High-throughput determination of RNA structure by proximity ligation. *Nat. Biotechnol.* 33(9):980–84

66. Regulski EE, Breaker RR. 2008. In-line probing analysis of riboswitches. *Methods Mol. Biol.* 419:53–67

67. Roost C, Lynch SR, Batista PJ, Qu K, Chang HY, Kool ET. 2015. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* 137(5):2107–15

68. Roundtree IA, Evans ME, Pan T, He C. 2017. Dynamic RNA Modifications in Gene Expression Regulation. *Cell*. 169(7):1187–1200

69. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 505(7485):701–5

70. Saletore Y, Meyer K, Korlach J, Vilfan ID, Jaffrey S, Mason CE. 2012. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biol.* 13(10):175

71. Schwartz S, Oren R, Ast G. 2011. Detection and removal of biases in the analysis of next-generation sequencing reads. *PLoS One*. 6(1):e16685

72. Selega A, Sirocchi C, Iosub I, Granneman S, Sanguinetti G. 2016. Robust statistical modeling improves sensitivity of high-throughput RNA structure probing experiments. *Nat. Methods*, pp. 1–10

73. Shiroguchi K, Jia TZ, Sims PA, Xie XS. 2012. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl. Acad. Sci. U. S. A.* 109(4):1347–52

74. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*. 11(9):959–65

75. Smola MJ, Calabrese JM, Weeks KM. 2015. Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry*. 54(46):6867–75

76. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM. 2015. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* 10(11):1643–69

77. Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, et al. 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*. 3(1):4

78. Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* 9(1):18–20

79. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, et al. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*. advance on:

80. Strobel EJ, Watters KE, Loughrey D, Lucks JB. 2016. RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Curr. Opin. Biotechnol.* 39:182–91

81. Sun G, Wu X, Wang J, Li H, Li X, et al. 2011. A bias-reducing strategy in profiling small RNAs using Solexa. *RNA*. 17(12):2256–62

82. Talkish J, May G, Lin Y, Jr JLW, Mcmanus CJ. 2014. Mod-seq : high-throughput sequencing for chemical probing of RNA structure Mod-seq : high-throughput sequencing for chemical probing of RNA structure. *RNA*. 20:0–8

83. Tian S, Cordero P, Kladwang W, Das R. 2014. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA*. 20(11):1815–26

84. Tullius TD, Greenbaum JA. 2005. Mapping nucleic acid structure by hydroxyl radical cleavage. *Curr. Opin. Chem. Biol.* 9(2):127–34

85. Tyrrell J, McGinnis JL, Weeks KM, Pielak GJ. 2013. The cellular environment stabilizes adenine riboswitch RNA structure. *Biochemistry*. 52(48):8777–85

86. Underwood JG, Uzilov A V., Katzman S, Onodera CS, Mainzer JE, et al. 2010. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*. 7(12):995–1001

87. van Gurp TP, McIntyre LM, Verhoeven KJF. 2013. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS One*. 8(12):e85583

88. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*. 14(10):1979–90

89. Wan Y, Qu K, Ouyang Z, Chang HY. 2013. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nat. Protoc.* 8(5):849–69

90. Wan Y, Qu K, Zhang QC, Flynn RA, Manor O, et al. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*. 505(7485):706–9

91. Watters KE, Abbott TR, Lucks JB. 2016. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.* 44(2):e12

92. Watters KE, Yu AM, Strobel EJ, Settle AH, Lucks JB. 2016. Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods*. 103:34–48

93. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, et al. 2009. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature*. 460(7256):711–16

94. Wells SE, Hughes JM, Igel AH, Ares M. 2000. Use of dimethyl sulfate to probe RNA structure in vivo. *Methods Enzymol.* 318:479–93

95. Wery M, Descrimes M, Thermes C, Gautheret D, Morillon A. 2013. Zinc-mediated RNA fragmentation allows robust transcript reassembly upon whole transcriptome RNA-Seq. *Methods*. 63(1):25–31

96. Westhof E, Romby P. 2010. The RNA structurome: high-throughput probing. *Nat. Methods*. 7(12):965–67

97. Wilkinson K a, Gorelick RJ, Vasa SM, Guex N, Rein A, et al. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 6(4):e96

98. Yoon S, Kim J, Hum J, Kim H, Park S, et al. 2011. HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics*. 27(13):1798–1805

99. Zheng W, Chung LM, Zhao H. 2011. Bias detection and correction in RNA-Sequencing data. *BMC Bioinformatics*. 12(1):290

100. Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB. 2012. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 40(7):e54

101. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*. 14(1):75–82

AIMS OF THE PROJECT

AIMS OF THE PROJECT

Despite the extensive researches accomplished to understand the HIV-1 maturation processes, several questions remain unresolved. The exact trigger responsible for viral PR activation is still unknown as well as the exact timing of initiation, even if viral maturation is believed to initiate immediately after particle release.

In addition, transition of viral morphology from the immature to the mature stage remains unclear as well as the relationship between proteolytic processing and gRNA maturation occurring concomitantly. The gRNA conformation is evolving during viral maturation but the different steps during gRNA rearrangement, probably involving inter- and intra-molecular interactions, are unknown.

Several studies analysed the gRNA structure by covering the whole-genome but were performed after extraction of the gRNA from HIV-1 viruses (383, 422). A similar study has been performed *in viro* but targeted only the first 900 nucleotides (427). Nevertheless, it is important to emphasise the fact that only the conformation of the mature gRNA has been investigated whereas understanding gRNA maturation requires to study the whole sequential processes.

For many years, the Paillart-Marquet's research group "Viral ribonucleoproteins, genome packaging and assembly" has carried out several very relevant achievements in the HIV-1 field. Regarding the HIV-1 genome packaging, the team investigated the gRNA dimerisation mechanism and clearly identified the 6 nucleotides involved in RNA dimerisation, the main and only demonstrated inter-molecular interaction between the two copies of gRNA. The importance of the three purines from either side of the DIS for the dimerisation process has also been demonstrated. $Pr55^{Gag}$ binding to many gRNA mutants has been studied, allowing determining *in vitro* and *in cellula* that SL1 is the primary $Pr55^{Gag}$ binding site. In addition, the team has a strong expertise in interrogating RNA structure by using in solution chemical probing, both *in vitro* and *in vivo*.

In this context, my PhD project was focused on the maturation step. I initiated this project in which we are interested in determining the different steps leading to the formation of a stable and mature dimeric gRNA. We also would like to better understand the link between proteolytic processing and gRNA maturation.

To this end, I analysed the secondary structure of the 5' region of the HIV-1 genome (comprising the first 550 nucleotides) by combining *in vitro* and *in viro* approaches. Chemical probing has assessed the structure of this region:

1. *In vitro* in the presence or absence of $Pr55^{Gag}$, GagΔp6, NC-containing intermediates (NCp15 and NCp9) and NCp7. This comparison allows us to determine these protein-binding sites and if these protection sites are conserved during the proteolytic cascade. Two conditions, RNA only and RNA-protein complexes treated with proteinase K before chemical modification, were also compared to observe the different protein-induced permanent structural rearrangements.

2. *In viro* within PR-, mutants mimicking the $Pr55^{Gag}$ proteolytic cascade and mature viral particles. These particles were also treated with a zinc ejector

(AT-2) that destabilises the NC zinc fingers. The structure of gRNA, chemically modified *in viro* has been determined by the hSHAPE-Seq approach. This technique has been developed to investigate the gRNA structural maturation with a high-throughput resolution.

The structuration of this region has also been investigated within viral particles treated with two different PIs in order to better understand the mechanism of action of these antivirals and to assess their effect on genomic RNA maturation. Indeed, we hypothesise the NC/Sp2 maturation impairment caused by PIs to block gRNA structural rearrangements and further inhibit viral replication, given the crucial role of NCp7 RNA chaperone.

This project is aimed to answer the following questions:

- How does the global conformation of the first 550 nucleotides evolve during proteolytic processing?

- Are there any conformational rearrangements during gRNA maturation? If it is the case, how do these rearrangements correlate with proteolytic processing? Where are they exactly localised and what is their nature (intra- and/or inter-molecular interaction)?

- What is the impact of $Pr55^{Gag}$ and its processing intermediates on gRNA maturation? Where are localised the different binding sites? Do $Pr55^{Gag}$ and NCp7 have a common consensus binding motif?

- What is the mechanism of action of PIs on viral particle maturation? Could it be related to gRNA maturation?

EXPERIMENTAL PROCEDURES

EXPERIMENTAL PROCEDURES

My PhD project required an important technical effort, with the development of a new technique to analyse the RNA structure *in viro*, since no suitable protocol was available at the beginning of my PhD thesis. Indeed, linking genomic RNA maturation and proteolytic processing requires working with gRNA extracted from mutant viral particles. It was thus impossible to produce virus particles using several rounds of replicative cycle, as these mutants do not replicate.

Thus, the very limited gRNA amount dictated the choice of the technique I used to study its structure and required a high-throughput sequencing analysis after chemical probing to increase the sensibility of detection.

To this end, I developed an *in viro* chemical probing technique, which we term hSHAPE-Seq, and applied it to the study of the 5' first 550 nucleotides of the HIV-1 genome. This region is indeed crucial for the regulation of many different steps of the viral life cycle. This approach allowed me to analyse several mutants mimicking the sequential processing of $Pr55^{Gag}$, and the effect of protease inhibitors (lopinavir and atazanavir sulfate) on the gRNA structure. In addition, viral particles were treated with aldritiol-2 (AT-2) in order to expel the Zn^{2+} ions from the NC zinc fingers and to identify their binding sites using a reverse footprinting approach.

In addition, I studied the *in vitro* footprinting of $Pr55^{Gag}$, GagΔp6, NCp15, NCp9 and NCp7, in order to further validate hSHAPE-Seq results and compare the chaperone property of the NC domain depending on its processed status.

Development of the hSHAPE-Seq approach

I. Biological samples analysed by hSHAPE-Seq

I.1 The $Pr55^{Gag}$ mutants mimicking the proteolytic cascade

To investigate the impact of the different $Pr55^{Gag}$ cleavage products on genomic RNA maturation, five sequential cleavage-site mutants were constructed using PCR stitch mutagenesis, by inserting substitutions of several nucleotides at $Pr55^{Gag}$ cleavage sites or PR active site (**Figure 15**). These mutants provide a snapshot of each step of the proteolytic processing.

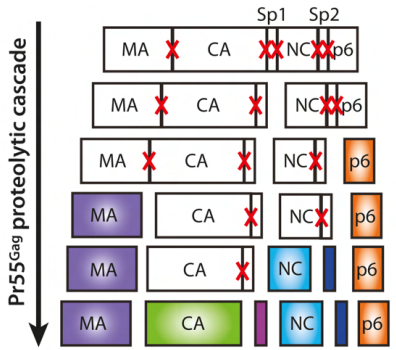
The different mutants regarding the expression of $Pr55^{Gag}$ are succinctly described:
* PR- mutant expresses unprocessed $Pr55^{Gag}$
* Step 1 mutant produces MA-CA-Sp1 (p41) and NC-Sp2-p6 (p15) fusion proteins
* Step 1.1 mutant produces MA-CA-Sp1 (p41) and NC-Sp2 (p9) fusion proteins in addition to the mature p6 protein
* Step 2 mutant produces CA-Sp1 (p25) and NC-Sp2 (p9) fusion proteins in addition to the mature MA and p6 proteins
* CA/Sp1 mutant produces CA-Sp1 (p25) fusion protein in addition to the mature MA, NCp7, Sp2 and p6 proteins.

For safety consideration, each clone harbours an inactivating deletion in the *env* gene.

The protein profile of the viral particles produced by these mutants, upon transfection of HEK-293T cells, was checked by western blot (WB).

WB protocol

An equal amount of lysed viral particles from each sample (quantified by Bradford test) was resuspended in NuPAGE Reducing Agent (10x) and LDS Sample Buffer (4x) (Life technologies). Proteins were heated for 10 min at 70°C then resolved by SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) at 200 V for 30 min using the XCell *SureLock* Mini-Cell electrophoresis system (Thermo scientific) with Criterion TGX pre-cast gels 4-15 % (Biorad). A protein ladder with a molecular weight ranging from 10 to 250 kDa was used in parallel (precision plus protein, unstained standard Biorad). Proteins were then transferred to PVDF membrane with the Trans-Blot Turbo system. The high molecular weight program was chosen, since I revealed several proteins corresponding to the $Pr160^{GagPol}$ and $Pr55^{Gag}$ cleavage products with a wide range of molecular weight (from 160 to 6 kDa). PVDF membrane were blocked with shaking during one hour at 4°C, using 5% milk resuspended in TNT solution (0.1 % triton, 150 mM NaCl, 50 mM Tris-HCl pH 7.5) followed by incubation with the primary antibody (Ab). This primary Ab is a patient serum, provided by the team of J. Mak, diluted 1: 10,000X and incubated overnight at 4°C. An anti-human immunoglobulin (Ig) G (heavy + light chains) – horseradish peroxidase (HRP) conjugate diluted 1:10,000X was used for signal detection and incubated during 1 h at 4°C. Two washing steps were performed with TNT solution for 10 min followed by a final wash with TN solution (150 mM NaCl, 50 mM Tris-HCl pH 7.5). Chemiluminescence was visualised after incubation with ECL (GE Healthcare) by mixing equal volumes of solution A and B diluted at the 1:5X. Detection was performed using the ChemiDoc Touch (Biorad) with a 1-min exposure time.



clone \ mutation's localisation	gag					pol	env
	MA/CA	CA/Sp1	Sp1/NC	NC/Sp2	Sp2/p6	PR	Env
PR-						substitution C->A G->A C->G 2326 2328 2332	deletion 6345 -> 7612
Step 0	substitution TAC->ATT 1183	substitution AA->TC T->A G->C G->C A->G A->C 1869 1876 1878 1890 1896 1902	substitution C->G AA->CG 1914 1917	substitution A->T T->A 2084 2106	substitution T->A CA->TC 1131 1142		6345 -> 7612
Step 1	TAC->ATT	AA->TC T->A G->C G->C A->G A->C		A->T T->A	T->A CA->TC		6345 -> 7612
Step 1.1	TAC->ATT	AA->TC T->A G->C G->C A->G A->C		A->T T->A			6345 -> 7612
Step 2		AA->TC T->A G->C G->C A->G A->C		A->T T->A			6345 -> 7612
CA/Sp1		AA->TC T->A G->C G->C A->G A->C					6345 -> 7612
WT							6345 -> 7612

Figure 15

Description of the different Pr55^{Gag} cleavage site mutants.

The table highlights the localisation of the mutations which are substitutions in *gag* / *pol* genes and a large deletion in *env* gene. On the left, the Pr55^{Gag} cleavage site mutants are schematically represented and follow the sequential HIV-1 cleavage cascade. Coloured boxes highlight mature proteins and white boxes, fusion proteins resulting from mutated (represented by red crosses) cleavage sites.

The protein profiles produced by the WT and the mutant viruses are in agreement with the introduced cleavage site mutations (**Figure 16**).

Immature PR- and Step 0 particles expressed Pr160^{GagPol} and Pr55^{Gag}. Pr160^{GagPol} was fully processed from the Step 2 mutant, RT and IN enzymes starting to be produced with the release of NC NTD (Step 1). The p41 fusion product was well detected in Step 1 and Step 1.1 lysed viral particles as expected. MA was released and detected from Step 2. CA was the last protein released during the proteolytic cascade. The signal in the CA/Sp1 mutant around 17 kDa corresponds to the CA-Sp1 product (p25). A faint signal around 10 kDa was visualised from Step 1.1 and mature viral lysates and could correspond to NCp9, NCp7 or p6. Faint or absent signal of these proteins with a low molecular weight would suggest that the antibody recognition epitopes were masked.

Importantly, the pattern of viral expression correlates with the same mutants generated in the team of J.-I. Sakuragi (313).

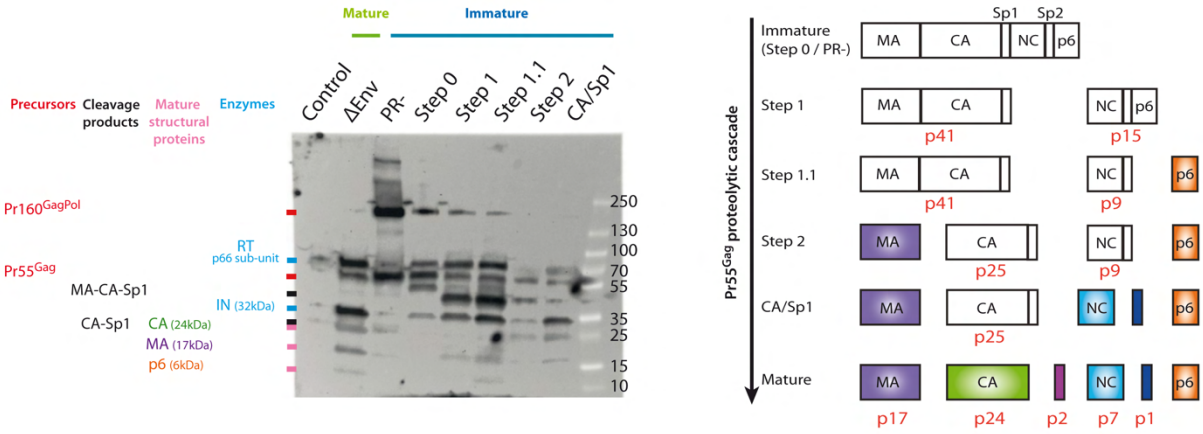


Figure 16
Validation of the different Pr55Gag mutants by WB.
Detection of the proteins from viral particles with a patient serum. A negative control with cellular extracts was run in parallel. Pr55Gag and Pr160GagPol, as well as their intermediate cleavage products and mature structural and enzymatic proteins are positioned on the gel depending on their molecular weight (in kDa).

I.2 Aldrithiol-2 treatment

One of the main limitations of *in vivo* structure mapping approaches is the presence of RNA-binding proteins, which can block chemical modification. So an unreactive position can either be located either in a double-stranded region or in a protein-binding site. In the case of HIV-1, NCp7 and NC-containing intermediates are the major proteins binding to the RNA genome through their two zinc finger structures in addition to the in-between highly basic region. By comparing *in viro* chemical reactivity in presence or absence of bound proteins, we can determine the nature of the protection, i.e. a protein-binding site or base pairing.

In order to dissociate NCp7 and NC-containing intermediates, viral particles were treated with a zinc ejector called aldrithiol 2 (AT-2) prior to *in viro* NMIA modification. AT-2 covalently bounds cysteine residues in the zinc finger domains and disrupts interactions between the zinc ion and its cysteine ligands. So, treatment of viral

particles with AT-2 promoted formation of AT-2 - NC interactions and disrupted NC – RNA interactions (348). AT-2 treated viruses are not infectious and blocked at the first steps of RT (81, 356). Moreover, AT-2 is able to enter inside the viral particle without impacting the functional integrity of the virus surface, so the chemical modification can be performed *in viro* even after AT-2 treatment of viral particles.

Two protocols have been described in the literature to inactivate HIV-1 particles with 1 mM AT-2 (stock in DMSO) either overnight at 4°C (426) or for 1 h at 37°C (356). To further investigate these differences, I compared treatment of virus particles at 4 and 37°C and their impact on viral infectivity (**Figure 18**).

Infectivity test protocol

This infectivity test is based on Tat-induced luciferase reporter gene expression after a single round of virus infection, using TZM-Bl cells. The assay was performed in 96-well culture plates. A dilution range of the virus stock (supernatant containing viruses was centrifuged at 1620 g for 30 min at 4°C, then filtered through a 0.22 μm filter) from 1:5 to 1:78,125 was tested in triplicate. Supplemented DMEM was added to each well before addition of virus dilutions (25 μl of supernatant containing virus for the first line of wells to initiate the serial dilution). In parallel, TZM-Bl cells were counted, diluted to obtain 100,000 cells/ml, and 10,000 cells/well were plated. TZM-bl cells background was measured in absence of virus infection. Forty-eight hours post-transduction, the sensitive Bright-Glo luciferase assay system (Promega) was performed following the manufacturer protocol and the luminescence intensity (Relative Luciferase Unit, RFU) was measured with a GloMax multi-detection system (Promega).

The linearity of the assay was systematically verified (**Figure 17**).

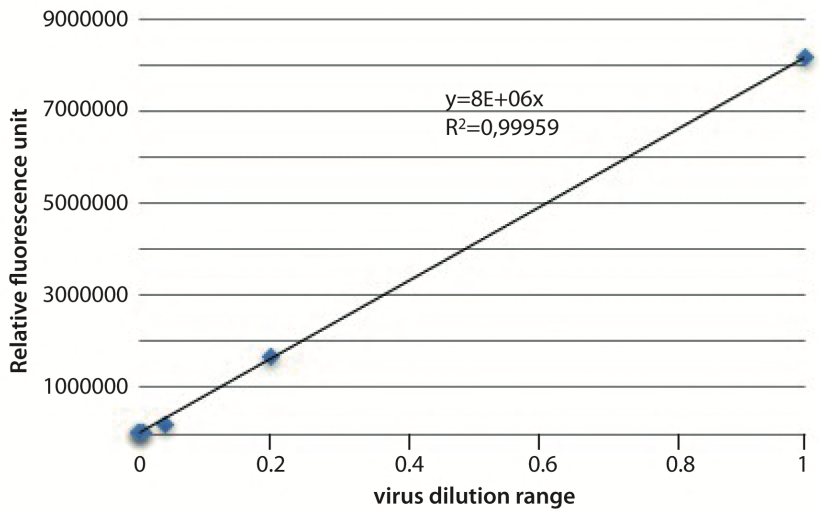


Figure 17
Linearity of the infectivity assay.
Serial dilutions of virus-containing supernatants were used to transduce TZM-Bl cells. Fluorescence intensity was measured 48 h later and plotted against virus dilutions. Linear regression of this plot yields a straight line with an intercept close to zero and a R^2 of 0,999.

The infectivity of the viruses obtained after AT-2 treatment was measured using a 1:25 virus dilution (based on the linearity of the infectivity assay). Virus particles incubated with AT-2 lost their infectivity upon treatment during 1, 2 or 15 h as expected.

Surprisingly, virus particles treated with the DMSO control for 15 h at 4°C were weakly infectious compared to the initial titre in absence of treatment. Therefore, incubation at 37°C for shorter periods was preferred.

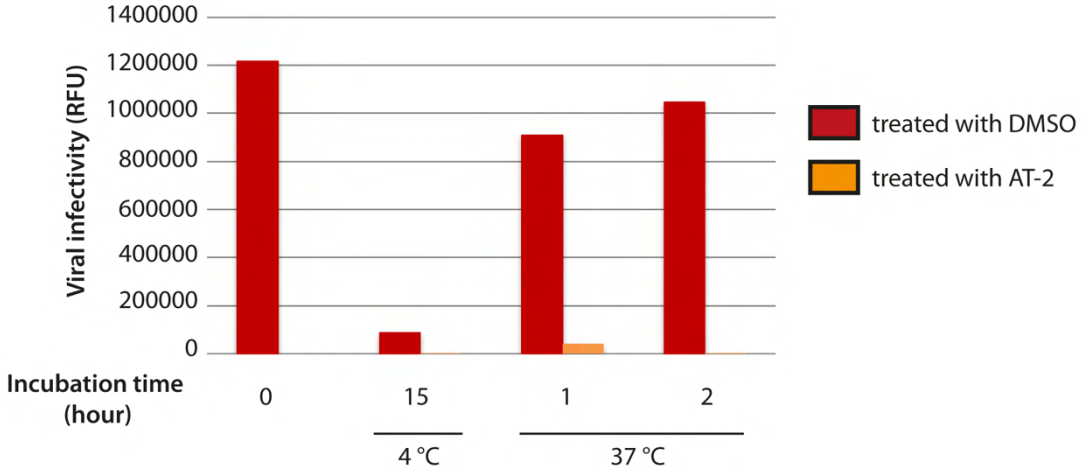


Figure 18
Impact of virus particle incubation at 4 and 37°C on infectivity.
Supernatants containing virus particles were supplemented with DMSO or AT-2 at a final concentration of 1 mM and incubated 15 h at 4°C or 1 or 2 h at 37°C. Their infectivity was measured with the TZM-Bl reporter assay based on Tat-induced luciferase expression upon integration of the HIV-1 genome into the host cell chromosome.

I also tested the AT-2 treatment on pelleted virus particles at 37°C with an AT-2 concentration range from 0.05 to 1 mM. Results of the infectivity test are presented in **Figure 19** and show a complete loss of viral infectivity at 0.5 mM after an incubation at 37°C for 30 min. So, these conditions were selected for further experiments.

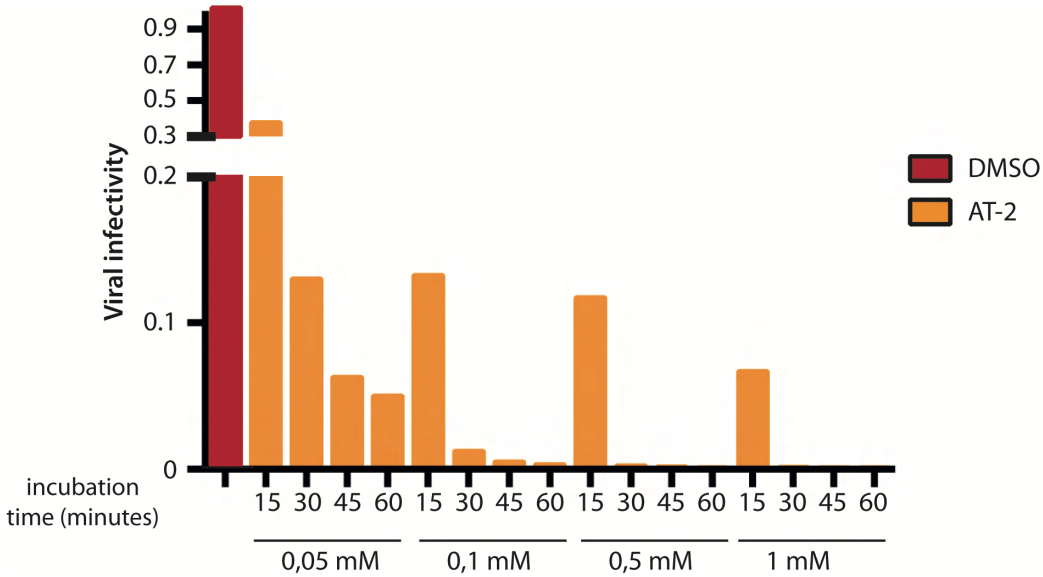


Figure 19
Determination of the optimal AT-2 treatment conditions.
Pelleted virus particles were treated at 37°C with several AT-2 concentrations (0.05, 0.1, 0.5 and 1 mM) and several incubation times (15, 30, 45 and 60 min). Their infectivity was determined with the TZM-Bl reporter assay. Infectivity (in RFU) of viruses treated with DMSO was set to 1 and samples treated with AT-2 were compared to this positive control.

The effect of the AT-2 treatment on gRNA recovery was assessed by RT-qPCR (**Figure 20**). DMSO had no influence on viral RNA recovery, while AT-2 decreased recovery by two-fold. To compensate this loss of material, I doubled the number of transfected cells to produce viral RNA in sufficient amount to complete the hSHAPE-Seq protocol. I applied this AT-2 treatment to PR- and mature virus particles.

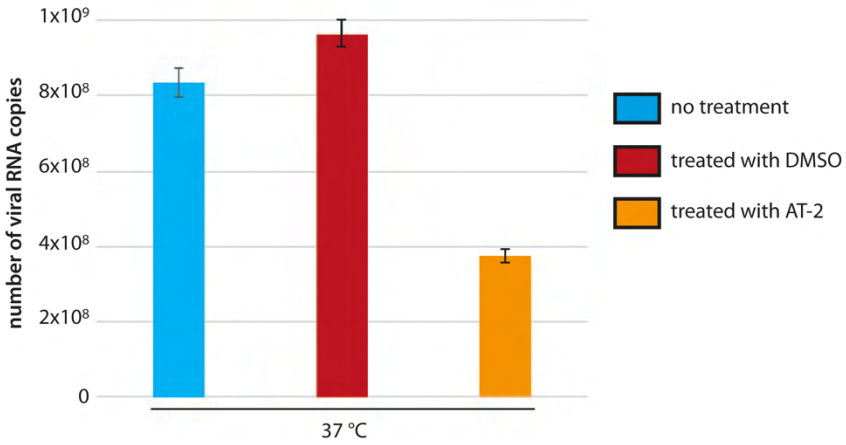


Figure 20
Influence of AT-2 on viral RNA recovery.
Viral RNA from virus particles treated for 30 min at 37°C with DMSO, 0.5 mM AT-2 or no treatment was extracted and quantified by RTqPCR. RT was performed with T-GIRT III RT and qPCR with the Taq polymerase, the Taqman probe and primers p116 Fw and p509 Rv.

AT-2 treatment of virus particles protocol

Pelleted virus particles were treated at 37°C for 1 h at a final concentration of 0.5 mM in AT-2.

I.3 Protease inhibitors

HIV-1 PIs target PR proteolytic activity through the active site of the PR enzyme. The idea behind testing PIs was to correlate their effect on $Pr55^{Gag}$ proteolytic processing (by WB), with viral infectivity (luciferase assay) and genomic RNA maturation (hSHAPE-Seq).

I chose to test two inhibitors of the PR catalytic activity designed as peptidomimetics of the PR substrates. Lopinavir (LP) (376) and atazanavir sulfate (ATVs) (352) are part of the second generation of PR inhibitors and are widely used in HIV-1 antiretroviral therapy.

LP and ATVs exhibit potent anti-HIV-1 activity with a 50% inhibitory concentration (IC_{50}) of respectively 2.83 and 0.28 nM (291). Based on these IC_{50} measured in HEK-293T cells, I tested a concentration range from 0.42 to 1600 nM in the case of LP and an extended concentration range from 0.01 to 1600 nM with AZVs.

Virus particles were produced by transfection of HEK-293T cells with PEI (as described in the following part, II.1.1 Cell culture and transfection, of the manuscript), in presence of protease inhibitors at the desired concentration (stock in DMSO). These pNL43 particles are pseudotyped with the protein G from the Stomatitis vesicular virus.

Supernatants containing virus particles were centrifuged for 30 min at 4°C and 1620 g, then filtered through a 0.22 μm filter. This stock was split in three parts to perform the following experiments:

- Viral particles were pelleted through a sucrose cushion and lysed in RIPA 1x (50 mM Tris-HCl pH 7, 150 mM NaCl, 1% sodium deoxycholate, 1% triton X-100, 0.1% SDS, 1x halt protease inhibitors cocktail (Thermo scientific)). The viral protein profile was analysed by WB with a patient serum and revealed with an Ig G – HRP conjugate (both diluted 1:10,000X and incubated at 4°C, overnight for the primary Ab and 1 h for the secondary Ab).

- Infectivity of virus particles in presence of protease inhibitors was assessed with TZM-Bl reporter assay (as previously described in the part, I.2 Aldrithiol-2 treatment, of the manuscript). Immature PR- particles were also produced and constituted a negative infectivity control, while mature particles produced with DMSO constituted the positive control.

- The hSHAPE-Seq protocol was performed to study the gRNA conformation.

II. Development of the hSHAPE-Seq protocol

The final version of the hSHAPE-Seq protocol applied to the study of the first 550 nucleotides of the HIV-1 gRNA is detailed in this part. Technical aspects are described at each step of the protocol, with the rationale for the choice of parameters, briefly emphasising the hSHAPE-Seq optimisation. Every section includes a black box highlighting the final protocol. The final version of the hSHAPE-Seq protocol is recapitulated in **Annexe 24**.

Figure 21 highlights the key steps of the hSHAPE-Seq technique. Succinctly, gRNAs were modified by a SHAPE probe inside the viral particle before extraction and DNase treatment. Following gRNA reverse transcription, RNA was degraded. Multiplexed sequencing library was then generated to be compatible with the Illumina HiSeq 2500 system.

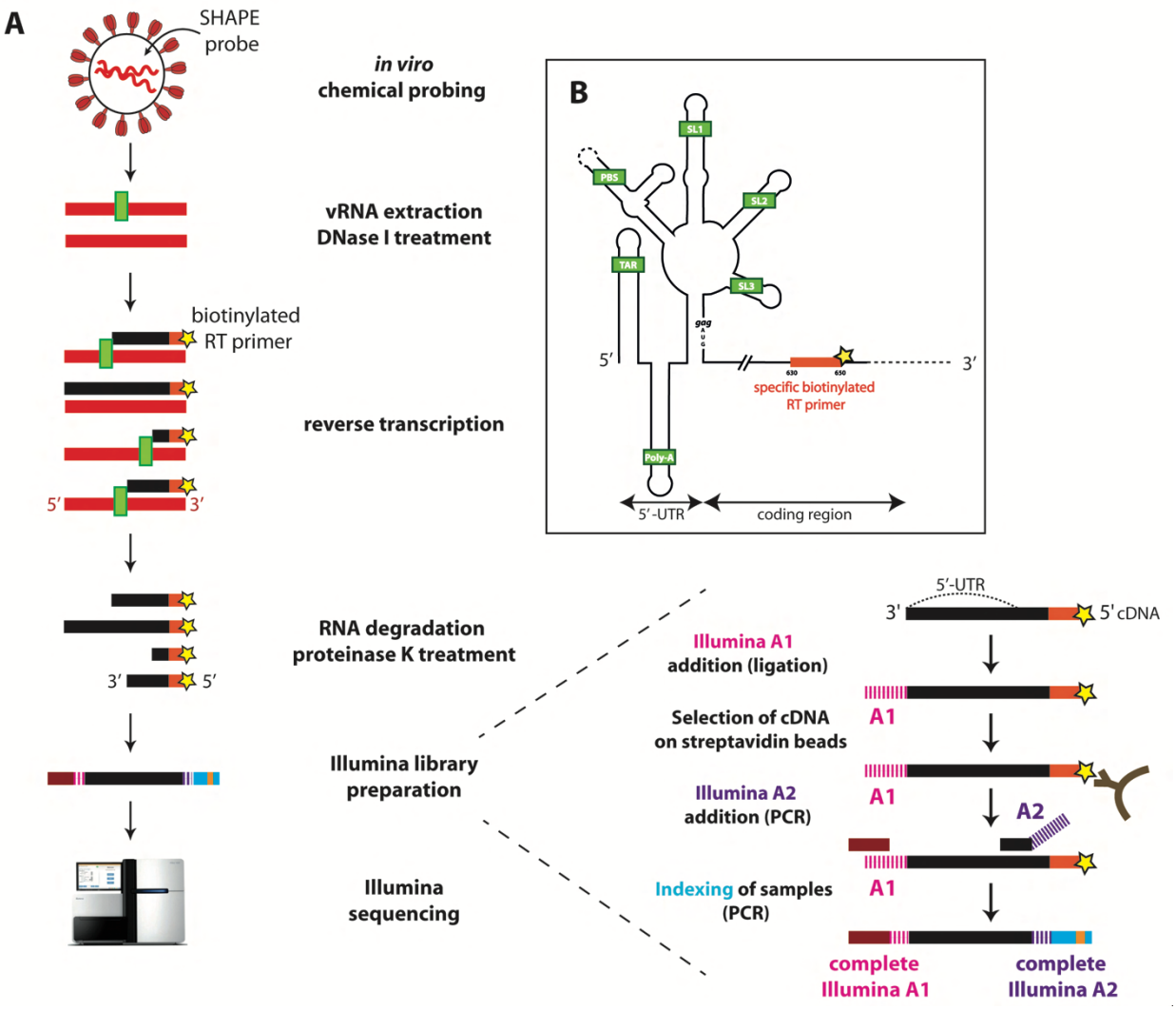


Figure 21
Overview of the different steps constituting the hSHAPE-Seq protocol.
A. The different steps from the *in viro* gRNA chemical probing to the generation of Illumina sequencing library are schematically represented. Illumina library preparation required for samples multiplexing is detailed on the right. **B.** Hybridisation position of the RT primer linked to a biotin molecule is emphasised in the whole HIV-1 genome. Primer extension covers the 5' UTR plus approximately the first 200 nucleotides of the *gag* coding region.

The different primers used in this project are described in **Table 1**.

Table 1: Summary of the primers used in the hSHAPE-Seq

RT primers	hybridisation position on HIV-1	sequence (5' to 3')
p365 Rv	Nucleotides 365-385	TTCCCATTTATCTAATTCTCC
p628 Rv	Nucleotides 628-648	TCTATCTTATCTAAGGCTTCC
p594 Rv	Nucleotides 594-614	GTGTGCATCAAAGGATAGATG

Ligated primer	sequence (5' to 3')
Short Illumina adapter A1	/Phos/NNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT/3SpC3/

Taqman probe	hybridisation position on HIV-1	sequence (5' to 3')
Taqman full length	Nucleotides 139-163	/56FAM/ATCCCTCAG/ZEN/ACCCTTTTAGTCAGTG /3IABkFQ/

qPCR and PCR primers	hybridisation position on HIV-1	sequence (5' to 3')
p116 Fw	116-136	TTGTGTGACTCTGGTAACTAG
p198 Fw	198-216	CTTGAAAGCGAAAGTAAAG
p365 Rv	365-385	TTCCCATTTATCTAATTCTCC
p509 Rv	509-535	GGATGGTTGTAGCTGTCCCAGTATTTG
A2PCR	543-565	AGACGTGTGCTCTTCCGATCTTGATCTAAGTTCTTCTGATCCTG
PCR1.0		AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG ACGCTCTTCCGATCT
Index		CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT

II.1 Production of HIV-1 viral particles

II.1.1 Cell culture and transfection

HEK-293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovin serum and antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin).

I tested several conditions of transient transfection by comparing different parameters in order to optimise viral production:
- cells counted and plated 24 h or a few min before transfection
- number of cells plated per Petri dish and diameter of Petri dishes used
- transfection reagents: polyethylenimine (PEI) (Merck) and X-tremeGENE 9 (Roche)
- DNA/transfection reagent ratio
- medium used for transfection: DMEM/Optimem
- duration of transfection.

Based on the results obtained in these different tests, I found optimal to perform transient transfection of HEK-293T in 150 mm Petri dishes with 11 millions of HEK-293T cells. Cells were plated 5 to 10 min before transfection and so were still in suspension when transfected. Regarding the choice of transfection reagent, PEI and X-tremGENE 9 had a similar transfection efficiency in my hands. PEI has been chosen because of its lower cost, with a 4.5 µg PEI per 1 µg DNA ratio. I transfected a total of 8 µg DNA per Petri dish and formed the plasmid complex with PEI in Optimem (250 µl). DNA was added to Optimem and vortexed, followed by PEI addition. The mix is briefly vortexed then incubated at room temperature for 10 min then carefully added in a drop-wise manner to cells. Seventy-two hours post-transfection, the supernatant containing viral particles was collected.

II.1.2 Virus purification

The supernatant was centrifuged for 30 min at 4°C and 1620 g, then purified with 0.22 µm filter to eliminate cellular debris. The clarified supernatant was concentrated by centrifugation in 50 ml falcon tubes through a 20% sucrose cushion in phosphate-buffered saline (PBS). Centrifugation was performed overnight at 4°C and 6360 g. Pelleted viruses from each falcon tube were resuspended in 100 µl NMIA reaction buffer (50 mM Hepes pH 8, 200 mM NaCl, 0.1 mM EDTA, and 10% fetal bovine serum) as described in the literature (426) and incubated 45 min at 4°C before recovery.

II.2 Chemical probing

We chose to use a SHAPE reagent to probe gRNA inside viral particles, as they have the advantage of interrogating the four nucleotides at the same time.

II.2.1 Evaluation of two SHAPE reagents at various concentrations

Firstly, I determined *in viro* optimal conditions for single hit kinetics by performing chemical modification with various concentrations of N-methylisatoic anhydride (NMIA) and 2-methylnicotinic acid imidazolide (NAI) reagents in PR- particles. These SHAPE probes are further detailed in the review article "Evolution of RNAs probing methods over the last 40 years: technical aspects" submitted for publication. I determined a half-life of 10 and 4 min for NMIA and NAI, respectively, by measuring their hydrolysis rate using a Uvikon spectrophotometer. NMIA (426) and NAI (392) probes are described in the literature with an optimal concentration of 10 mM and 13.5 mM, respectively. Based on this information, I tested a concentration range consisting of 3.9, 7.8, 15.6, 31.3 and 62.5 mM. In parallel, a control with DMSO only was performed to identify RT natural stops (C1) whereas the C2 control enables to verify that SHAPE reagents modify gRNA inside the viral particle rather than after extraction. This was done by adding the reagent to NMIA reaction buffer only and adding viral particles at the same time as the tri-reagent used for RNA extraction.

Following chemical modification, RTion was performed with a radioactively labelled RT primer and allowed localisation of RT stops after gel electrophoresis. For this test, I used the reverse primer p367 Rv (reverse) hybridising in *gag* (nucleotides 367-387) allowing analysis of the gRNA untranslated region.

in viro SHAPE protocol

NMIA and NAI modification of viral particles was performed in NMIA reaction buffer for 5 half-lifes of the chemical probe, i.e. respectively 50 and 20 min at 37°C. A tri-reagent RNA extraction was performed, followed by ethanol precipitation (standard protocol with 3 volumes of ethanol 100%, 0.3 M sodium acetate and 1 μg glycogen). After centrifugation at 11,000 g, pellets were washed twice with 80% ethanol, vacuum dried and resuspended in 20 μl milliQ H_2O.

Primer extension with AMV RT analysed by polyacrylamide gel electrophoresis (PAGE) protocol

Quantification and normalisation of gRNA were performed by RT-qPCR. DNA amplification was performed with a forward p198 Fw (forward) (nucleotides 198-216) and a reverse p365 Rv (nucleotides 365-385) primer, dsDNA being detected by SYBR green (Maxima Sybr green master mix Thermo Fischer). RT-primer extension was achieved with 500,000 counts per min (cpm) of the radioactively labelled stock of p365 Rv (measured with a Cherenkov detector), annealed for 2 min at 90°C followed by 2 min on ice in a final volume of 10 μl. After addition of 1x RT buffer (Life Science), 4 U RNasin (Promega), 2 U avian myeloblastosis virus (AMV) RT, 1 mM dNTPs completed with milliQ H_2O to a final volume of 15 μl, elongation was performed at 42°C for 20 min followed by 30 min at 50°C and 10 min at 60°C. Sequencing reactions were performed in parallel following Sanger's method.
All reactions were stopped by adjusting the volume to 200 μl with milliQ H_2O and proteins were extracted from the DNA pool by phenol-chloroform and chloroform treatments. Following ethanol precipitation as previously described, pellets were dried and resuspended in 8 μl of formamide loading buffer. Half of each sample was loaded onto an 8% denaturing polyacrylamide gel run at 75 W. The gel was fixed (10% acetic acid and 1 % ethanol), dried at 80°C for 1 h and revealed using a FLA-5100 (Fuji).

The modification profile was similar with NMIA and NAI (**Figures 22** and **23**) with an optimal concentration of 3.9 mM (**Figures 22** and **23, lane 3**). This concentration has been chosen because a large proportion of cDNA corresponded to full-length extension, concomitant with a visible modification profile covering the entire region of the gRNA analysed with this primer. In addition, in both NMIA and NAI experiments, the C2 control (**Figures 22** and **23, lane 1**) has the same profile as the unmodified sample (C1 control **Figures 22** and **23, lane 2**), demonstrating that gRNA modification happened inside the viral particle.

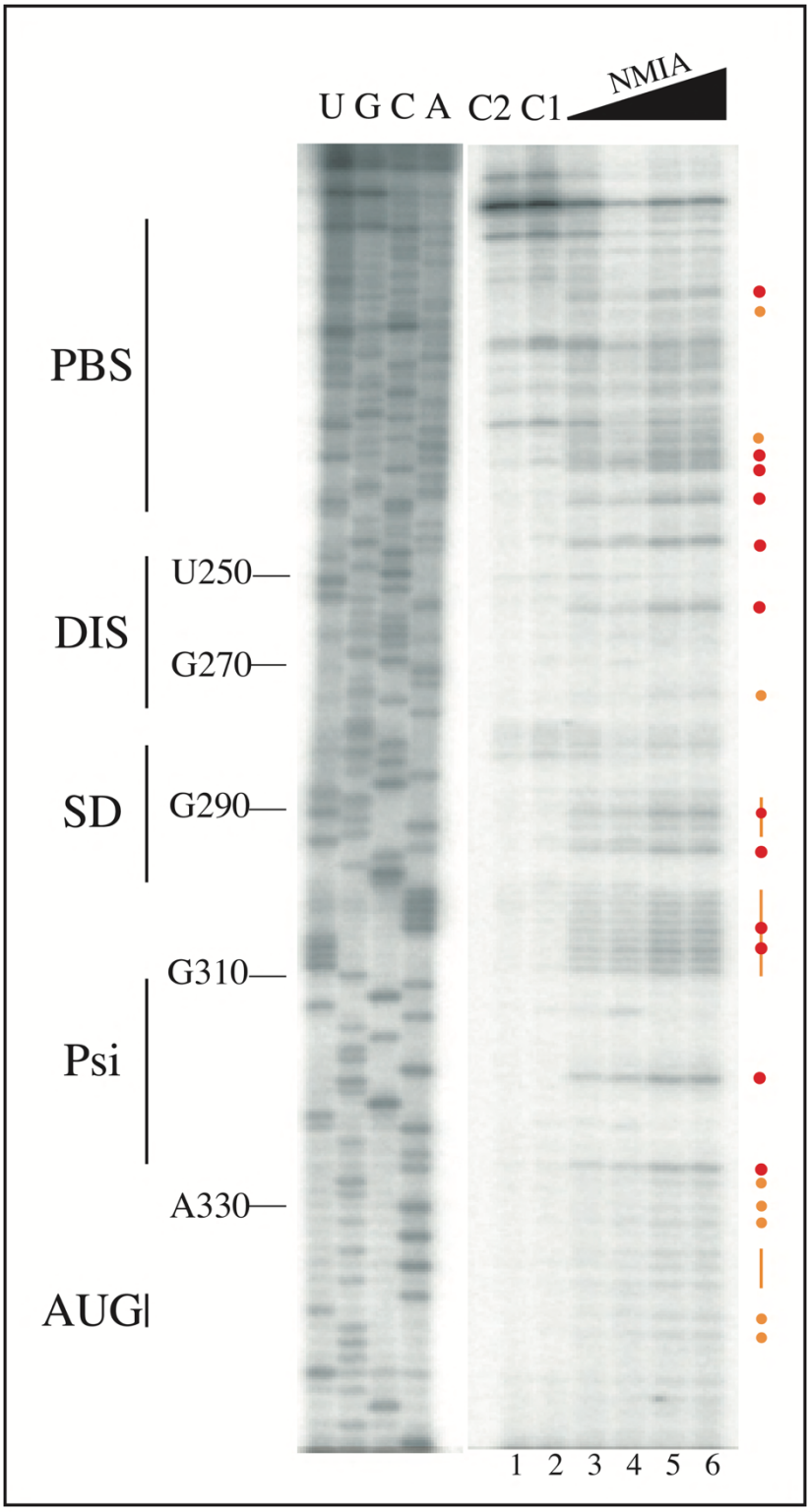


Figure 22
NMIA concentration range: determination of the *in viro* optimal NMIA concentration.
Primer extension (p365 Rv) of the untranslated gRNA region, *in viro* modified with NMIA, extracted from PR- virus particles and analysed by denaturing PAGE. Lane 1 corresponds to C2 control, ensuring that RNA modification was performed inside the virus particle whereas C1 control in lane 2 highlights natural RT stops. A concentration range comprising 3.9, 7.8, 15.6 and 31.3 mM was tested (lanes 3 to 6). Position of modified nucleotides were determined thanks to the four sequencing reactions. Strongly and weakly modified nucleotides are marked with red and orange dots, respectively.

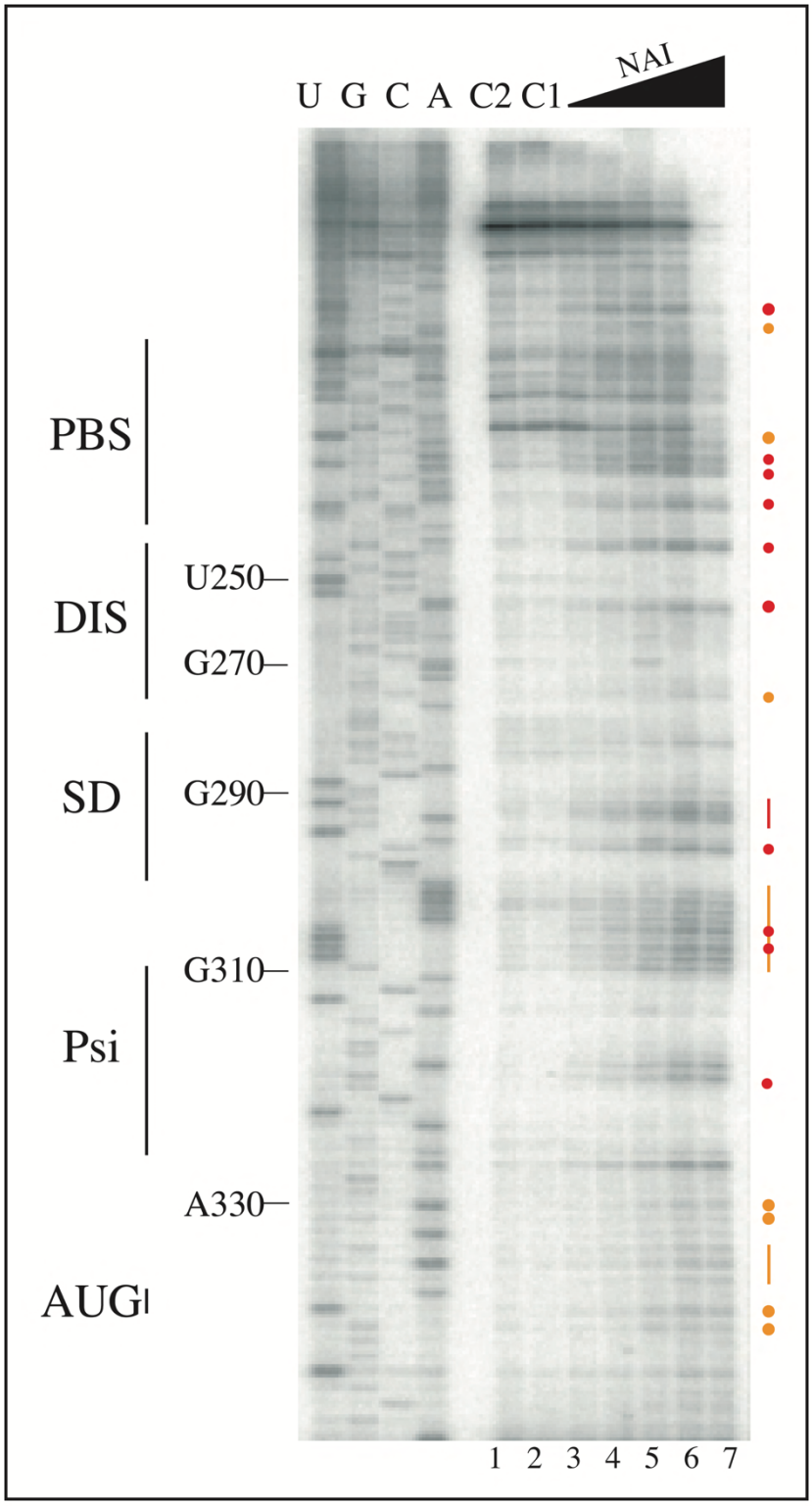


Figure 23

NAI concentration range: determination of the *in viro* optimal NAI concentration.

Primer extension (p365 Rv) of the 5' untranslated gRNA region, *in viro* modified with NAI, extracted from PR- virus particles and analysed by denaturing PAGE. Lane 1 corresponds to C2 control, ensuring that RNA modification was performed inside the virus particle whereas C1 control in lane 2 highlights natural RT stops. A concentration range comprising 3.9, 7.8, 15.6, 31.3 and 62.5 mM was tested (lanes 3 to 7). Position of modified nucleotides were determined thanks to the four sequencing reactions. Strongly and weakly modified nucleotides are marked with red and orange dots, respectively.

II.2.2 In viro RNA modification with NMIA

The NMIA reagent has been chosen to modify RNA in viral particles.

***in viro* SHAPE protocol**
Pelleted viruses resuspended in the NMIA reaction buffer were incubated with 3.9 and 7.8 mM NMIA final concentration, for 50 min at 37°C.

II.3 RNA extraction and DNase treatment

I compared several RNA extraction protocols and noticed a higher efficiency of RNA recovery with the « NucleoSpin virus » kit (Macherey-Nagel) (**Figure 24**).

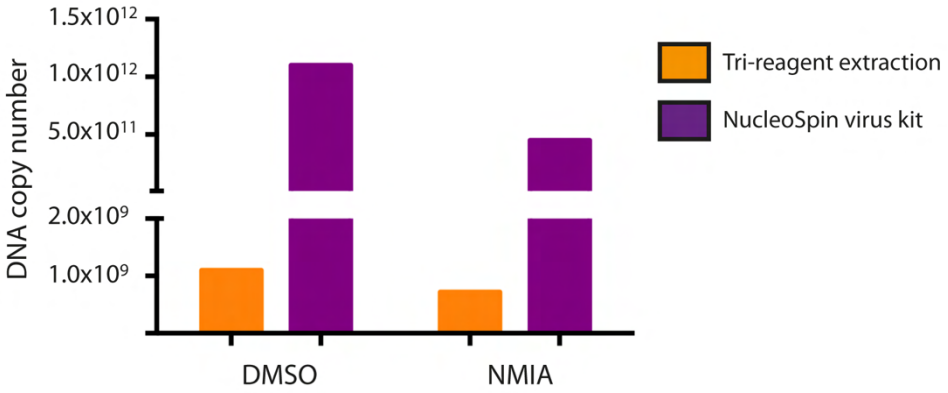


Figure 24
Comparison of two RNA extraction methods based on viral RNA recovery.
RT has been performed following AMV RT manufacturer protocol with p628 Rv RT primer. DNA amplification was performed by qPCR with p198 Fw and p365 Rv primers, dsDNA being detected using SYBR green.

Following RNA elution, a DNase I treatment was performed to eliminate transfected plasmids.

DNase I treatment protocol
RNA was incubated with 2 U of DNase I, 1x DNase buffer and milliQ H_2O in a final volume of 200 µl for 30 min at 37°C. An additional 2 U of DNase I was added at 30 min and incubated for an extra 30 min.

After phenol-chloroform and chloroform extraction, samples were ethanol precipitated and resuspended in 15.1 µl of milliQ H_2O to perform RTion.

II.4 Reverse transcription

II.4.1 Comparison of different RTs

I tested several RT, in order to compare their elongation efficiency since the 5'-UTR is a highly structured region and thus potentially difficult to reverse transcribe.

The following RT were compared by using RT-qPCR:
- AMV (company)
- Superscript III (SS III) (company)
- Group II intron RT Tel 4C (kindly provided by A.M. Lambowitz, Institute for Cellular and Molecular Biology and Department of Molecular Biosciences, Univeristy of Texas)
- Group II intron RT T-GIRT III (Ingex).

RTion has been performed following the manufacturer protocols (hybridisation of the p628 Rv RT primer at nucleotides' position 628-648). An amplification product of 421 nucleotides with primers p116 Fw (nucleotides 116-136) and p509 Rv (nucleotides 509-535) was detected by qPCR using SYBR green reagent. This first test clearly highlighted a low RT efficiency following these protocols with efficiencies lower than 10%. AMV and T-GIRT III RTs were twice more efficient than Tel 4C and SS III RTs.

Based on these results, I performed some complementary tests on AMV and T-GIRT III RTs. I checked the impact of the RNA concentration on the efficiency of these two RTs. I tested in parallel AMV and T-GIRT III RTs with a wide range of *in vitro* RNAs from 10^8 to 10^{12} copies (theoretical number of RNA copies based on RNA concentration and size) (**Figure 25**). Quite surprisingly, I noticed an effect of RNA concentration onto RT efficiency with a higher DNA production from $1.0x10^{12}$ RNA copies with AMV RT and an inversion of this tendency from $1.0x10^9$ RNA copies. A similar efficiency of AMV and T-GIRT III RTs was noticed with $1.0x10^{11}$ and $1.0x10^{10}$ RNA copies.

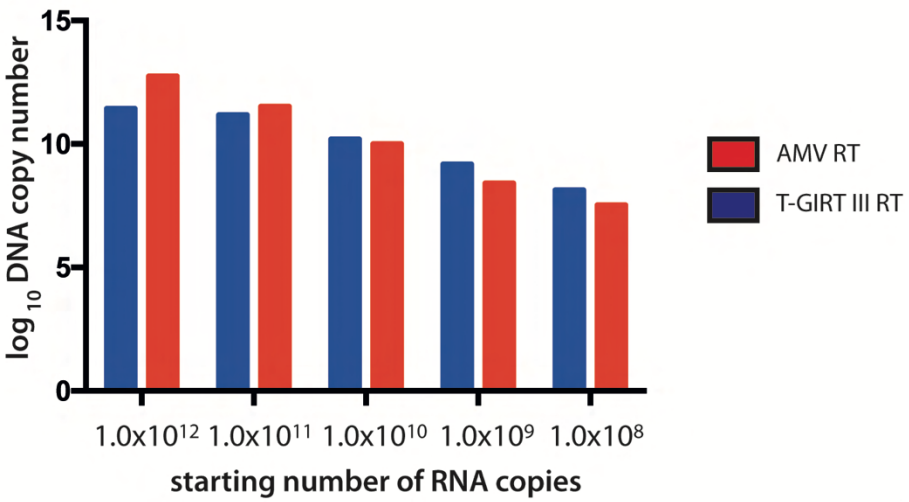


Figure 25
Effect of RNA concentration on AMV and T-GIRT III RTs.
RT was performed with p628 Rv primer on *in vitro* synthesised HIV-1 RNA (1400 nucleotides) with a concentration range from $1.0x10^8$ to $1.0x10^{12}$ RNA copies. cDNA was quantified by SYBR green based qPCR with p116 Fw and p509 Rv primers.

I also visualised by gel electrophoresis the profile of natural stops of AMV and T-GIRT III RTs. When using AMV RT, two strong RT stops were detected around nucleotide 160 and nucleotide 480 (**Figure 26 lanes 1, 2, 7** and **8**). These regions are GC-rich and are base paired in the gRNA secondary structure model of K. Weeks (426). Nevertheless, SS IV and T-GIRT III RTs significantly improved elongation by successfully passing through these highly structured regions. One explanation is the ability of SS IV and T-GIRT III RTs to withstand higher temperatures than AMV RT. Elevated temperature helps denaturing RNA with high GC content and strong secondary structure, allowing RTs to read through these sequences. The highest ratio of full length/cDNAs stopped at 480 was obtained with the T-GIRT III RT.

Based on these results, I chose to use T-GIRT III RT and optimised the protocol to increase RT efficiency, with a starting RNA amount of at least one billion RNA copies, by modifying several parameters:

- RT primer hybridisation sequence and concentration in addition to the RT primer annealing protocol
- reaction volume
- salts, dNTPs and T-GIRT RT concentration
- temperature and time of elongation.

II.4.2 T-GIRT III RT

Recently, the team of A.M. Lambowitz succeeded to engineer, express and purify highly efficient RTs encoded by mobile group II introns. These RTs have a higher thermostability, processivity, and fidelity than retroviral RTs (Lambowitz lab, Institute for Cellular and Molecular Biology and Department of Molecular Biosciences, University of Texas).

Optimised T-GIRT III RT protocol

The RT reaction was performed in a T100 thermal PCR cycler with a heated lid set at 105°C. Ethanol precipitated viral RNA was dissolved in 15.1 µl milliQ H_2O and subjected to primer annealing with 1 µl of 1 µM p628 Rv stock in addition to 100 mM Tris-HCl pH 7.5 (at 20°C) and 0.1 mM EDTA. The reaction mix was incubated at 82°C, then cooled at 25°C. The annealed RNA-primer was then pre-incubated for 30 min at room temperature with 300 U of T-GIRT III (Ingex) and 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris–HCl pH 7.5, 5 mM DTT and milliQ H_2O to a final volume of 32.9 µl. This pre-incubation step allowed the tight binding of T-GIRT III with the RNA template.
Following addition of dNTPs at a final concentration of 1.5 mM (dNTPs mix is an equimolar mixture of dATP, dCTP, dGTP, and dTTP), the RT reaction was initiated at 60°C for 1 h in a final reaction volume of 35 µl.

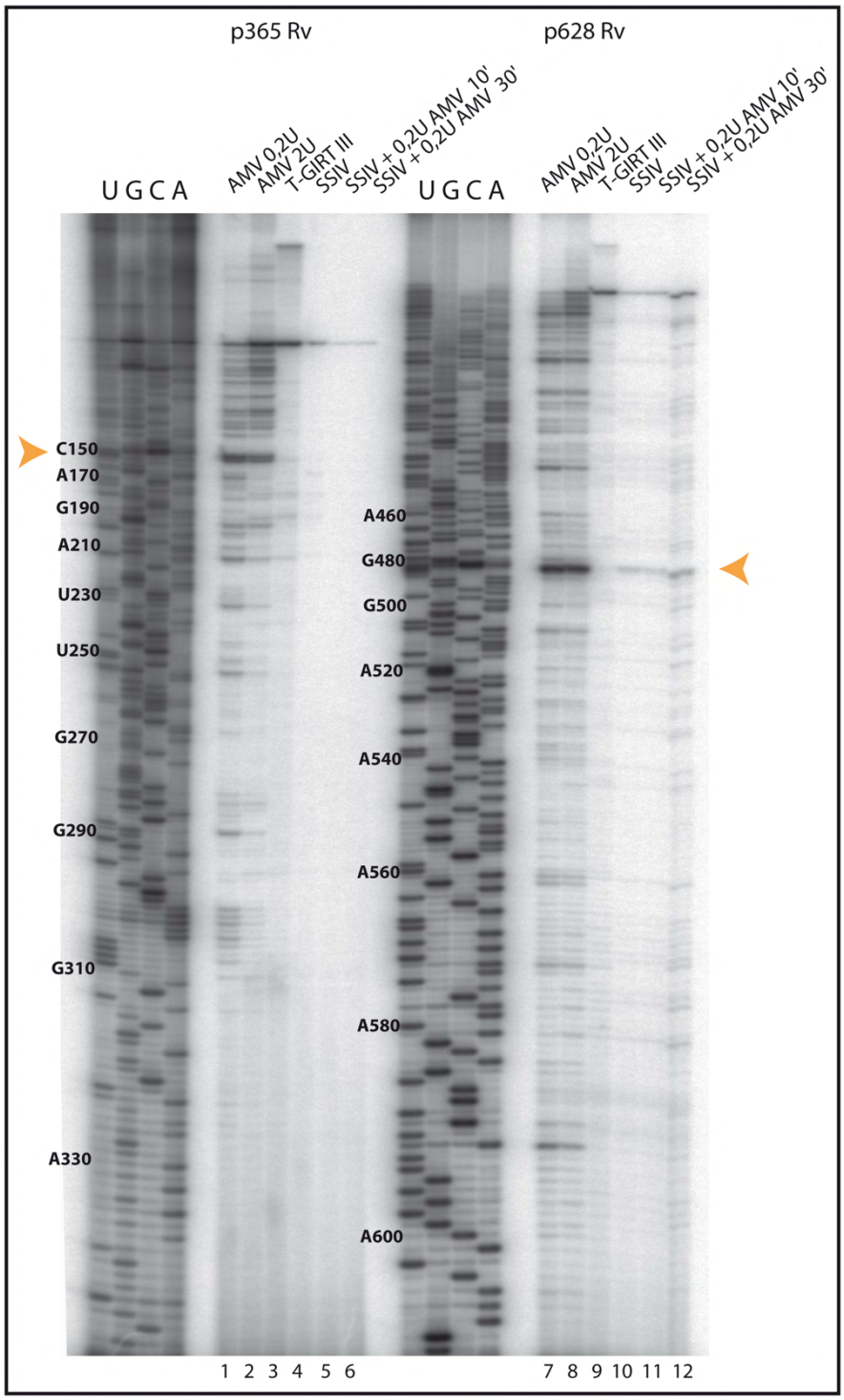


Figure 26

Comparison of RT profiles of unmodified RNAs using PAGE.

Primer extension of unmodified *in vitro* synthesised HIV-1 RNA (1400 nucleotides) with two primers (Lanes 1 to 6 with p365 Rv and lanes 7 to 12 with p628 Rv), covering the 5'-UTR gRNA plus approximately the first 200 nucleotides of the *gag* coding region. Several RTs were tested in parallel: AMV RT with 0.2 U in lanes 1 and 7 and 2 U in lanes 2 and 8, T-GIRT III RT in lanes 3 and 9, SS IV RT in lanes 4 and 10, SS IV RT + 0,2 U AMV RT for 10 mi in lanes 5 and 11 and 30 min in lane 6 and 12. Position of RT stops were determined thanks to the four sequencing reactions. Two strong RT stops are marked with orange arrowheads.

II.5 RNA degradation

Following RT reaction, degradation of the RNA template was performed to obtain ssDNAs, amenable to conversion into Illumina sequencing libraries. In addition, the T-GIRT III RT has a greater tendency to stick to RNA than retroviral RTs.

I first tested the manufacturer protocol with 0,1 M NaOH added to the 35 µl RT mix, incubated at 90°C for 3 min and neutralised with 0.1 M HCl. Since results were not convincing, I compared several other conditions: alkaline hydrolysis with sodium carbonate or NaOH and RNase A/RNase H treatments. RNase A is an endonuclease hydrolysing the phosphodiester bond between C and U nucleotides whereas RNase H specifically targets phosphodiester bonds of RNA/DNA hybrids.

In order to fully assess the impact of these treatments on RNA degradation from DNA/RNA complexes following RT, I produced radioactively labelled 1-796 RNA (RNA*) (plasmid NL4-3 digested by NsiI) by *in vitro* transcription.

Radioactive transcription protocol

Linearised DNA (2 µg) was incubated at 37°C for 1 h with 40 mM Tris-HCl pH 8, 6 mM $MgCl_2$, 10 mM NaCl, 10 mM DTT, 1 mM of each NTP except ATP which was at 0.03 mM, 2 mM spermidine, 40 U RNasin, 2.5 µl of home-made T7 RNA polymerase, 0.1 µg pyrophosphatase, 50 µCi alpha ^{32}P-ATP and milliQ H_2O in a final volume of 40 µl.
Excess free alpha ^{32}P-ATP was removed using Micro bio-spin-6 chromatography columns in Tris Buffer (Biorad). A DNase treatment was performed and following phenol-chloroform and chloroform extractions, samples were ethanol precipitated and resuspended in 20 µl of milliQ H_2O to perform RT with T-GIRT RT.
RNA* in complex with cDNA was counted and split to test the different RNA degradation protocols.

I tested two alkaline treatments and RNase A/H enzymes on RNA*/cDNA complexes (**Figure 27**). A pre-incubation at 90°C for 2 min was performed on the RNA/DNA complexes. Alkaline hydrolysis was performed at 95°C for 15 or 30 min with sodium carbonate (1x alkaline hydrolysis: 50 mM sodium carbonate pH 9.2, 1 mM EDTA) or using 0.1, 0.2, 0.3 M NaOH for 3 min. Samples were also incubated with RNase A (50 U) or RNase H (5 U), respectively at 60 and 37°C for 15, 30 and 45 min.

Following ethanol precipitation with sodium acetate, pellets were dried and resuspended in 16 µl of formamide loading buffer. Half of each sample was loaded onto 8 % denaturing PAGE. Gel was run, fixed, dried and revealed as previously described.

A clear pattern of RNA degradation was noticed with NaOH (**Figure 27 lanes 3** to **5**) and RNase A (**Figure 27 lanes 8** to **10**) with no differences regarding the NaOH concentration or the RNase A incubation time. RNA degradation with sodium carbonate was incomplete, even after 30 min (**Figure 27 lanes 6-7**) and inexistent with RNase H (**Figure 27 lanes 11** to **13**).

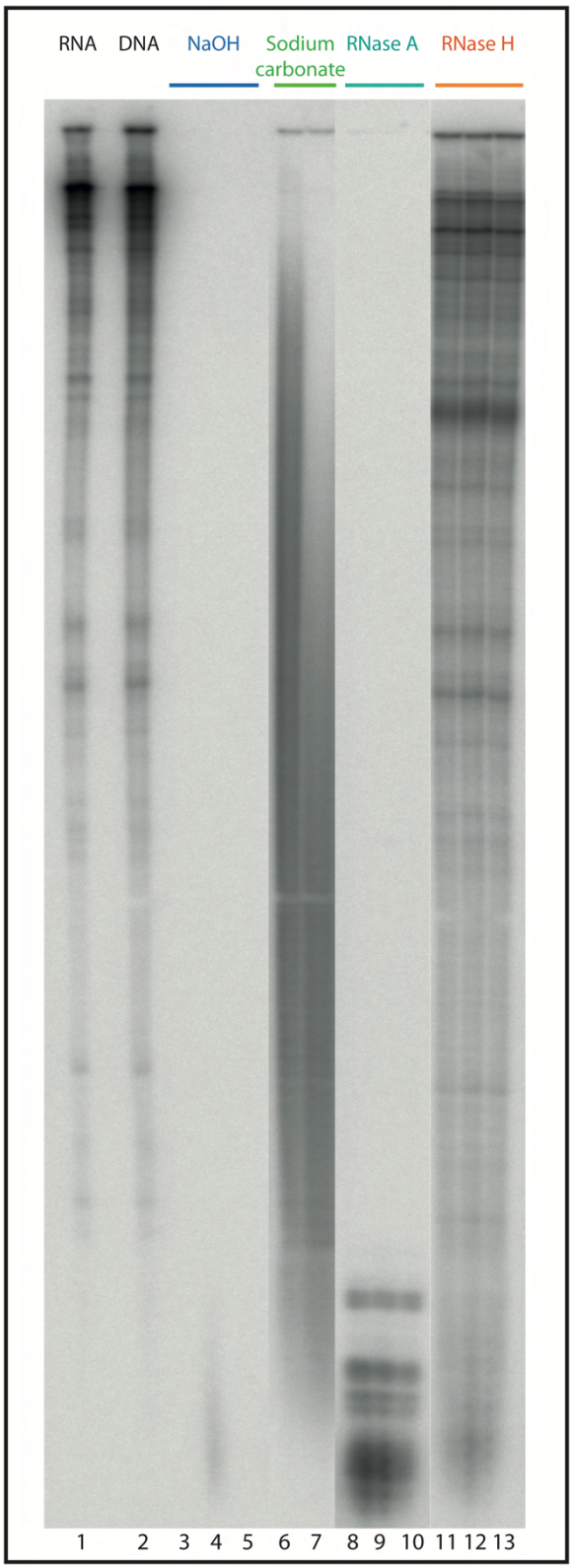


Figure 27
RNA degradation efficiency of different protocols.
Radioactive RNAs (lane 1) were reverse transcribed with T-GIRT III RT (lane 2) then submitted to different RNA degradation treatments: alkaline hydrolysis with 0.1, 0.2 and 0.3 mM NaOH for 3 min (lanes 3, 4 and 5) or 50 mM sodium carbonate for 15 and 30 min (lanes 6 and 7); 50 U RNase A for 15, 30 and 45 min (lanes 8, 9 and 10); 5 U RNase H for 15, 30 and 45 min (lanes 11, 12 and 13).

To further assess the impact of these RNA degradation treatments on the cDNA stability, I quantified cDNA recovery by qPCR. I focussed on the effects of NaOH and RNase A on DNA recovery, since these two treatments most efficiently degraded the RNA template. Quite surprisingly, NaOH treatment strongly reduced DNA recovery (**Figure 28** in blue). This was noticed even at 0.1 mM NaOH, even though the concentration instructed by T-GIRT III manufacturer is ten folds higher.

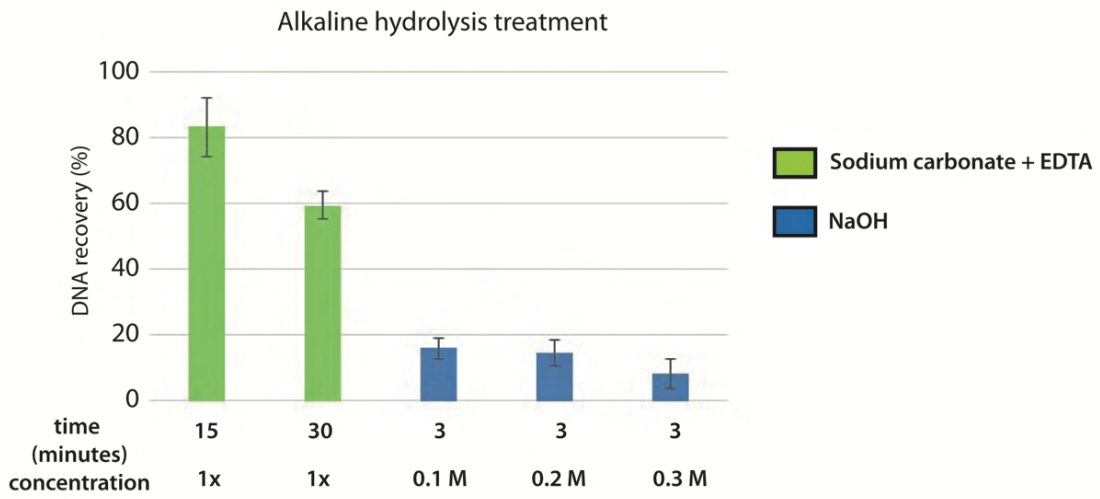


Figure 28
Impact of alkaline hydrolysis on DNA recovery.
Alkaline hydrolysis was performed with sodium carbonate or NaOH on RNA/DNA hybrids following RT with T-GIRT III RT. SYBR green based qPCR was performed with p116 Fw and p509 Rv primers. DNA recovery was calculated by dividing the DNA copy number remaining after alkaline hydrolysis by the number obtained without treatment and expressed as a percentage.

RNase A treatment had also an impact on DNA but allowed me to recover approximately 40% of DNA following RT (**Figure 29** in blue) with a 15 min incubation. Taking into account the DNA recovery yield and the absence of RNA template following treatment, I chose to perform the RNase A treatment to degrade the RNA template.

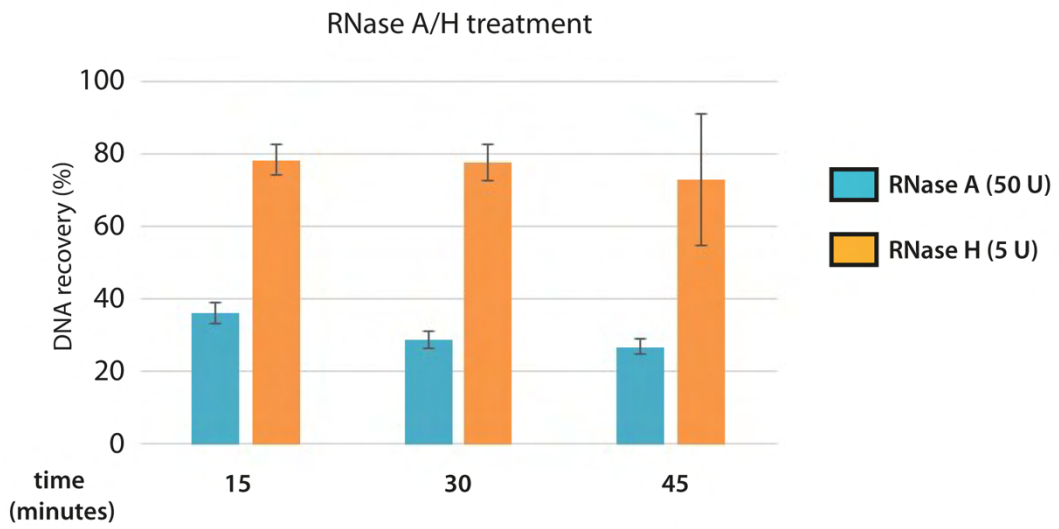


Figure 29
Impact of RNase A and RNase H treatments on DNA recovery.
RNase A and RNase H treatments were performed on RNA/DNA hybrids following RT with T-GIRT III RT. SYBR green based qPCR was performed with p116 Fw and p509 Rv primers. DNA recovery was calculated by dividing the DNA copy number remaining after RNase treatment by the number obtained without treatment and expressed as a percentage.

Treatment of DNA samples following T-GIRT III RT protocol

T-GIRT III RT reaction was ended by adding 50 U of RNase A into the 35 µl RT mix. Samples were incubated at 60°C for 15 min. RNase A was degraded by a proteinase K treatment, performed at 37°C for 30 min. A phenol-chloroform followed by a chloroform extraction were performed prior to ethanol precipitation.

II.6 Construction of the Illumina sequencing libraries

II.6.1 General overview

Illumina next-generation sequencing (NGS) technology (**Figure 30**) is able to simultaneously analyse numerous samples through multiplexing and provides a quantitative measurement of each nucleotide position with a high resolution.

This sequencing platform requires the addition of Illumina incomplete adapter 1 (A1) and 2 (A2) at the extremities of the DNA samples (step 1). The extension of A1 and A2 by PCR in order to obtain complete Illumina adapters with the incorporation of an index sequence, composed of six specific nucleotides, allows sample multiplexing and sequencing of multiple samples during a sequencing run within the same lane of the flow cell. Thanks to these adapters, sequencing templates are immobilised on the flow cell surface with complementary primers of A1 and A2 to allow solid-phase amplification. Generation of dsDNA clusters is performed by bridge amplification (step 2) with the creation of up to 1000 identical copies per single template which are denatured into ssDNA to perform sequencing by synthesis (step 3). Four fluorescently-labelled nucleotides are added during each sequencing cycle, with

the nucleotide label serving as a terminator for polymerisation. At each cycle, one nucleotide is incorporated and identified by imaging the fluorescent dye. An enzymatic cleavage of the dye initiates the start of the second cycle (step 4). Multiple chemistry cycles are performed to cover the sample (step 5).

II.6.2 Illumina library generation: several strategies

In order to add A1 and A2 adapters at either ends of DNA samples, I tested two strategies in parallel called "circularisation" (**Figure 31A**) and "ligation" (**Figure 31B**) strategies. The "ligation and PCR" strategy will be explained latter (**Figure 31C**).

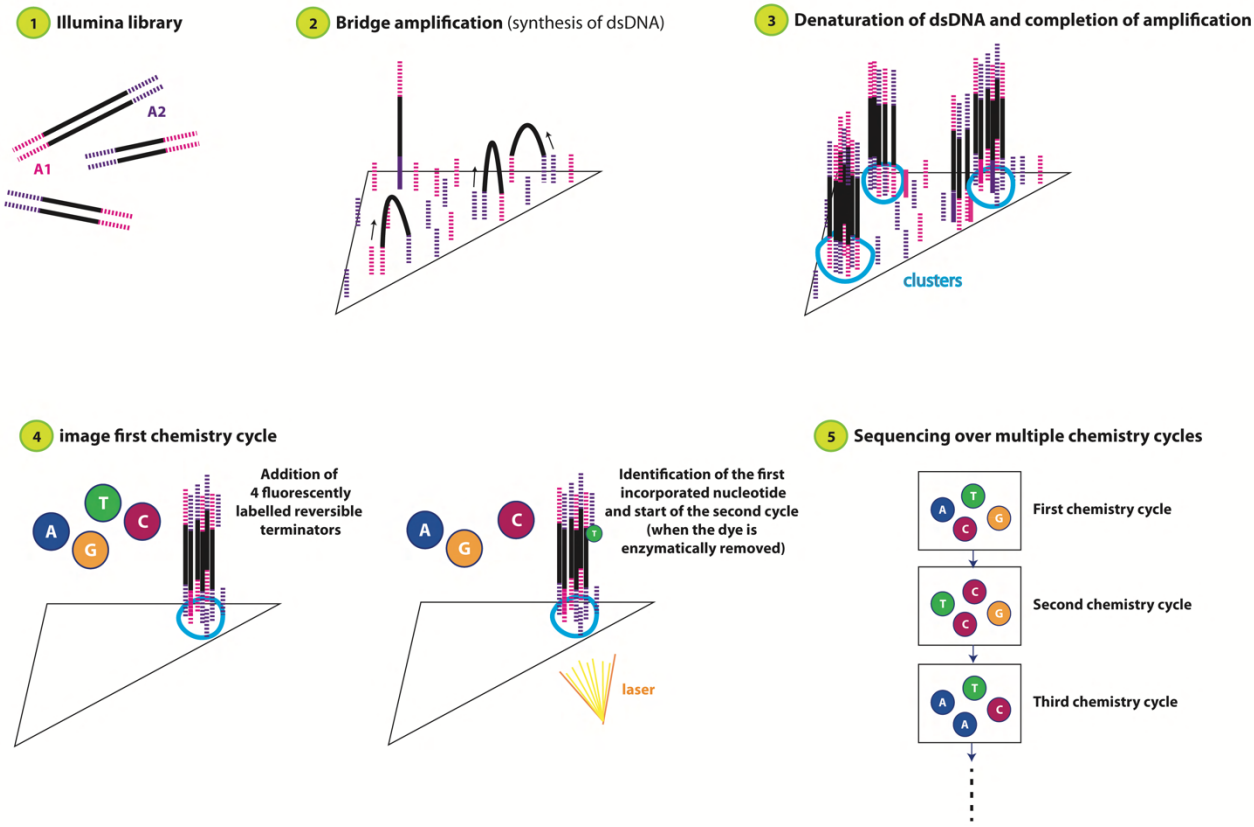


Figure 30
General overview of Illumina sequencing.
Library preparation with Illumina adapters at both ends of DNA samples is performed (step 1). ssDNA fragments randomly bind to the inside surface of the flow cell channels (step 2). Amplification of each sample is performed by bridge PCR, with generation of dsDNA clusters (step 3). Sequencing by synthesis is done with four fluorescently-labelled nucleotides, which act as reversible terminators when incorporated (step 4). After laser excitation, the incorporated nucleotide is identified thanks to its emitted fluorescence and the dye is finally cleaved. The sequencing cycles are repeated (step 5).

In the "circularisation" strategy (**Figure 31A**), a primer comprising the hybridisation sequence specific to the HIV-1 genome associated with A1 and A2 is used for the RTion. Subsequent circularisation of the DNA sample and enzymatic cleavage between A1 and A2 finally permits Illumina adapters to be at both ends of the sample. The linker between A1 and A2 is an 18-atom hexa-ethyleneglycol spacer, which can be cleaved by the APE1 enzyme. The "ligation" strategy (**Figure 31B**)

integrates a ligation step in the protocol because the RT primer is only composed of the hybridising sequence associated with A2. A1 is ligated at the 3' extremity of DNA. This primer has three random nucleotides in 5' to increase ligation efficiency and a 3' modification (C3 Spacer phosphoramidite), to prevent the formation of concatemers.

Since gRNA is chemically modified, RT generates a population of cDNAs with various lengths, from 648 nucleotides (representing the full length extension) to 21 nucleotides (p628 RT primer composed of 21 nucleotides plus one nucleotide extension). This wide range of sample sizes possibly influences circularisation and ligation efficiency.

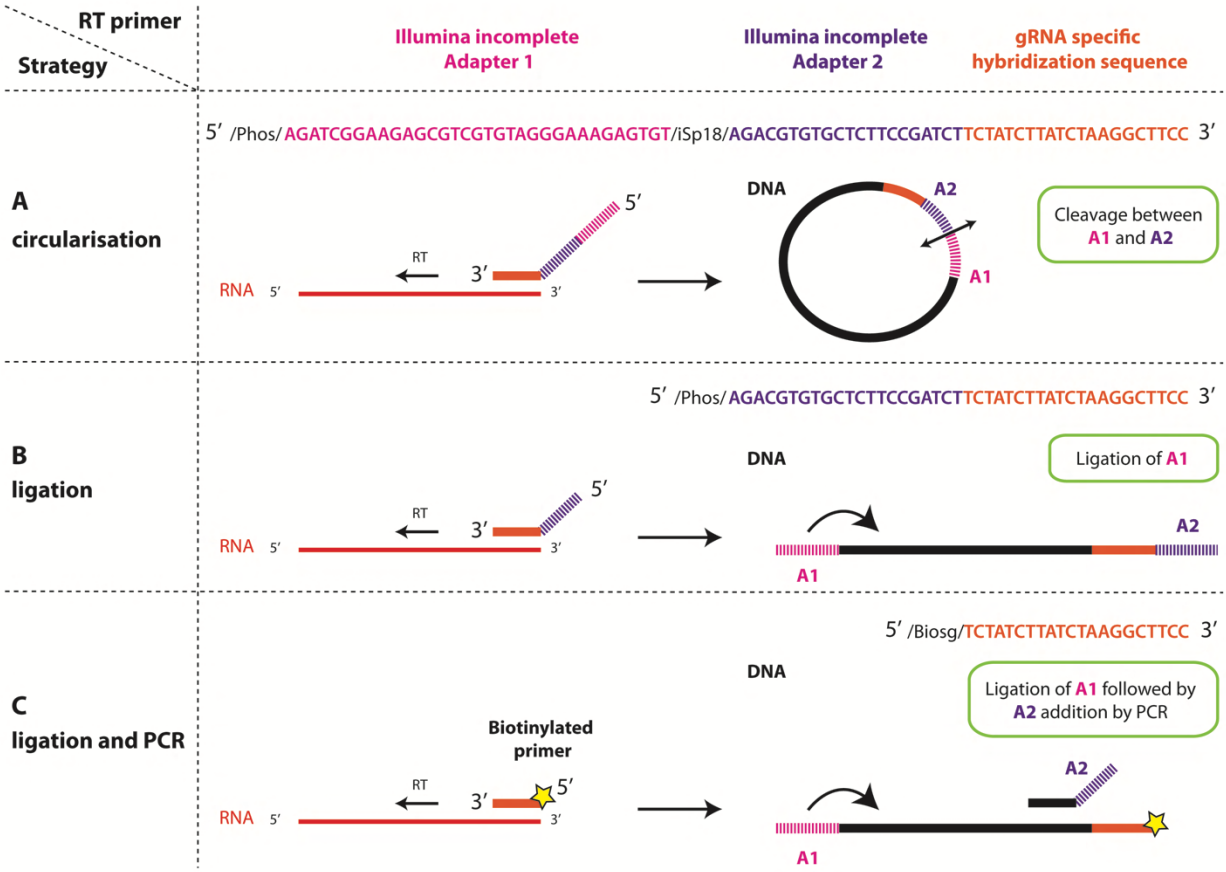


Figure 31
Description of the different strategies used to generate the Illumina library.
A. For the "circularisation strategy", the RT primer RTA2-1 is composed of a specific hybridising sequence followed by A2 and A1. The DNA sample is circularised followed by specific cleavage between A1 and A2 to linearise the sample. **B**. In the "ligation" strategy, the RT primer RTA2 is shorter with A2 associated to the specific hybridising sequence. A1 is added by 3' DNA ligation. **C**. The "ligation" strategy has been adjusted to overcome technical problems. RT is performed with a biotinylated primer (RT biotin). A1 is added as above by 3' DNA ligation followed by a selection step based on biotin-streptavidin interaction. A2 is added by PCR and associated with a specific hybridising sequence.

I have conducted both "circularisation" and "ligation" strategies in parallel, and tested different ligases:

Circularisation	Ligation
Circ ligase I (Epicentre)	Circ ligase I (Epicentre)
Circ ligase II (Epicentre)	Circ ligase II (Epicentre)
	T4 RNA ligase 1 (Thermo Fisher / NEB)
	K227Q truncated T4 RNA ligase (NEB)

Circ ligase I is a thermostable ATP-dependent ligase that catalyses intramolecular ligation (i.e. circularisation) of ssDNA substrates. With this enzyme, circularisation reaction requires a ssDNA substrate with 5'-phosphate and 3'-hydroxyl groups. I have also tested Circ ligase II which requires a ssDNA substrate with a 5'-adenylation and a 3'-hydroxyl group.

The control provided by Epicentre is a 55-oligodeoxynucleotide containing 5'-phosphate and 3'-hydroxyl ends. I first tested the Circ ligase I efficiency with the Epicentre control and RTA2-1 RT primer as templates. Under standard reaction conditions (10 pmol control oligo, 100 U Circ ligase, 2.5 mM $MnCl_2$ for 1 h at 37°C), the linear control oligo as well as my RTA2-1 primer were converted to circular ssDNA with an efficiency of approximately 50% (**Figure 32**). Circ ligase I and II enzymes are widely used in published protocols to generate NGS libraries (338, 357, 392, 401, 421, 443).

In order to improve this ligation step, I tested in parallel several protocols published in the literature to perform intermolecular ligation of a ssDNA oligo to my viral ssDNA template. The first one came from the SHAPE-Seq technique developed by the team of J.B. Lucks (268, 272). This protocol was further optimised by the team of R. Das (374). But other protocols were reported by the Assmann group and Neri (120, 192).

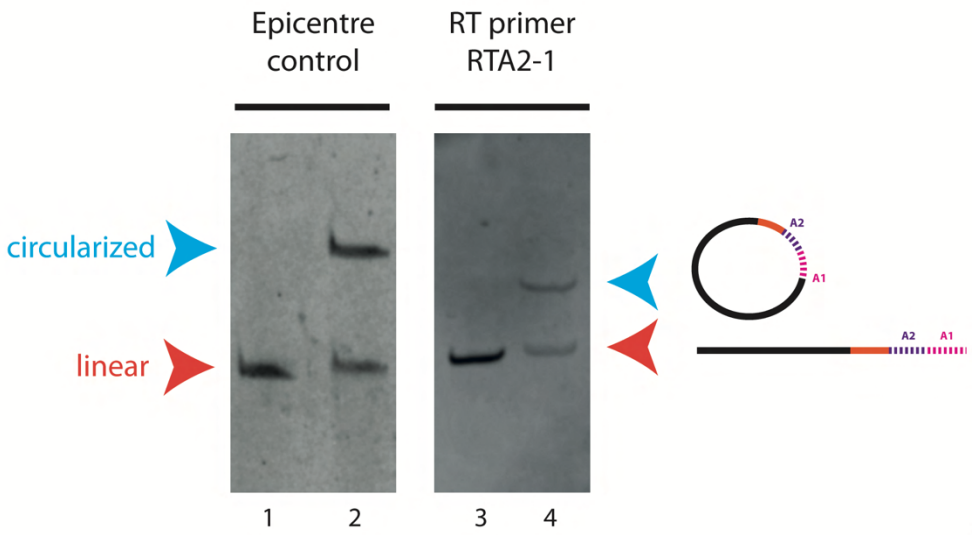


Figure 32
Evaluation of the Circ ligase I efficiency assessed by gel electrophoresis.
Circularisation was performed under standard conditions on the Epicentre control (lanes 1-2) and the RTA2-1 RT primer (lanes 3-4). A negative control without Circ ligase I enzyme (lanes 1 and 3) highlights the migration position of linear ssDNA (orange arrow head). The circularised product migrates slower (blue arrow head) compared to linear ssDNA.

Several protocols use the T4 RNA ligase 1 (401) and a truncated form of T4 RNA ligase 2 (268, 357, 392, 401, 443) to ligate ssDNA Illumina adapters to ssRNA. T4 RNA ligase 1 catalyses the ATP-dependant ligation of a 5' phosphate to a 3' hydroxyl-group through the formation of a 3' to 5' phosphodiester bond, with substrates including both ssRNA and ssDNA. T4 RNA ligase 1 has been shown to perform ssDNA/ssDNA ligation, but the reaction is slower than with RNA/RNA or RNA/DNA substrates (411). Truncated T4 RNA Ligase 2 specifically ligates the pre-adenylated 5´ end of ssDNA or ssRNA to the 3´ end of ssRNA. This enzyme has been optimised by introducing a K227Q point mutation. These enzymes reduce background ligation because only adenylated DNA oligonucleotides are used and the K227Q mutation further reduces the formation of undesired ligation products (concatemers and circles).

I tested T4 RNA ligase enzymes in parallel to Circ ligase enzymes by RT-qPCR to compare their ssDNA/ssDNA ligation efficiency. These experiments have been conducted on the same pool of reverse transcribed *in vitro* NMIA modified and unmodified RNAs. Primers used in qPCR/PCR and the size of the amplicon are indicated in **Figure 33.**

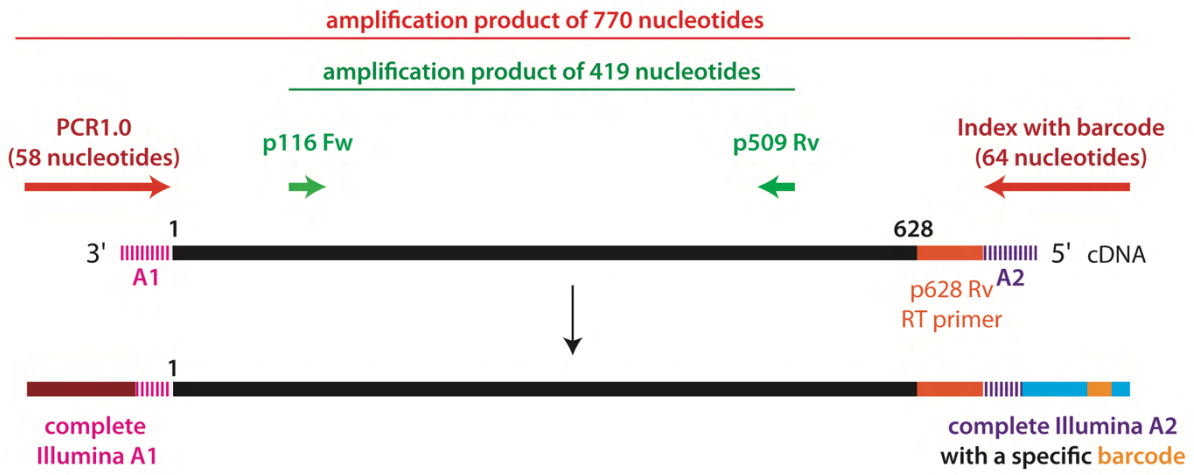


Figure 33
Description of qPCR/PCR primers: hybridisation position and amplicons.
The two qPCR primers used to quantify ssDNA template (p116 Fw – p509 Rv) and A1-ligated ssDNAs (PCR1.0 – Index) are shown with their amplicons.

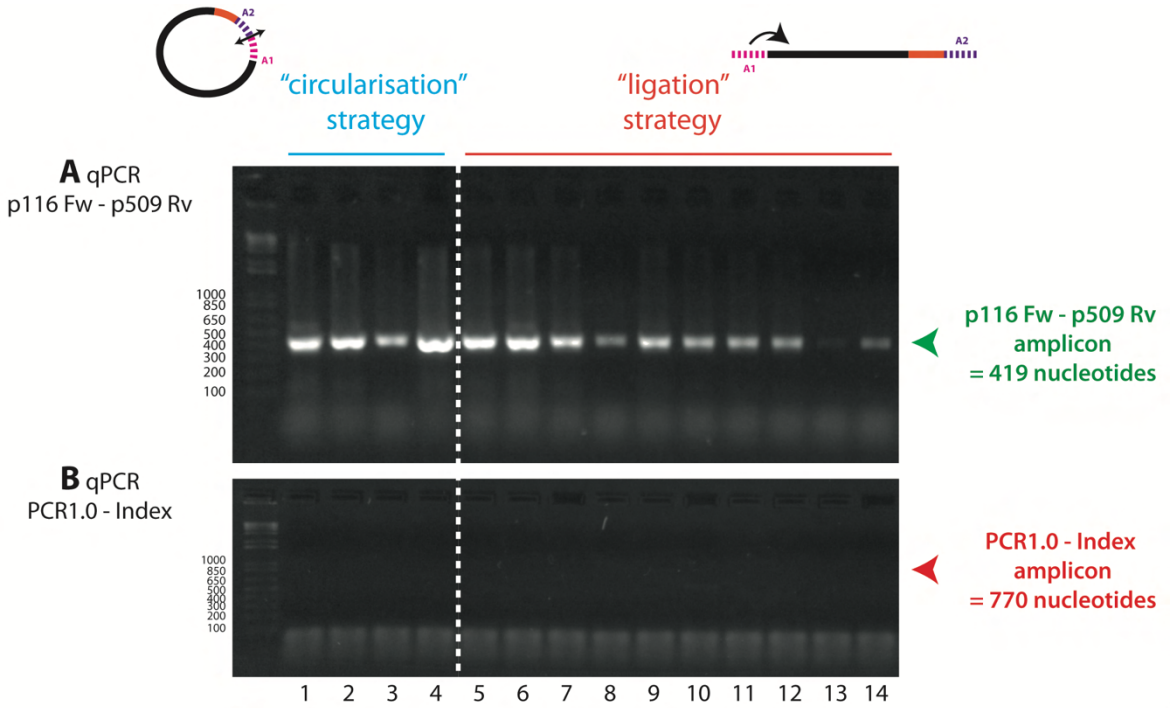


Figure 34
Comparison of "circularisation" and "ligation" strategies by RTqPCR.
Seven billion copies of NMIA modified (3.9 mM) and unmodified *in vitro* synthesised RNAs were reverse transcribed by T-GIRT III RT followed by RNase A treatment, phenol-chloroform extraction, precipitation and split into seven fractions. The "circularisation" strategy has been tested with Circ ligase I (lanes 1 and 2) and II (lanes 3 and 4). The "ligation" strategy has been performed with Circ ligase I (lanes 5 and 6) and II (lanes 7 and 8), T4 RNA ligase from Thermo Fischer (lanes 9 and 10) and NEB (lanes 11 and 12), and K227Q truncated T4 RNA ligase (lanes 13 and 14). Samples were quantified by SYBR green qPCR. The cDNA template was amplified with p116 Fw-p509 Rv primers (**A**) whereas A1 ligated cDNAs were detected with PCR1.0-Index primers (**B**). These primers and the expected amplicons are depicted in **Figure 33**. Depending on the qPCR primer pair used, location of the full-length amplicon is indicated by green (**A**) and red (**B**) arrowheads.

Results presented in **Figure 34** clearly emphasise the inefficiency of all enzymes both in "circularisation" and "ligation" strategies, with no full-length amplification as well as no smear containing the ssDNA population of different sizes. Several reasons could explain these results:

- Ligation sequence bias (in the case of Circ ligase enzymes)

One study reports a poorer ligation efficiency with the fully extended ssDNA compared to smaller products and explains it by a possible sequence bias (374). Indeed, Epicentre indicates that the sequence of the ssDNA can strongly influence the efficiency of the circularisation reaction. A transcript beginning with GG is apparently less efficiently ligated because is reverse transcribed into CC.

- Formation and preferential amplification of A1-A2 products

An amplified product of approximately 150 nucleotides in length was detected by qPCR. This product is probably obtained by circularisation of the RTA2A1 primer in "circularisation" conditions or ligation of A1 primer and RT primer containing A2 in "ligation" conditions followed by amplification with PCR1.0 and Index primers.

Confirmation that A1 primer ligated to RT primer containing A2 and that this product was amplified during PCR with PCR1.0 and Index primers has been obtained by performing PCR with A1 primer and RT primer containing A2 without DNA samples (**Figure 35**). This band has been isolated on agarose gel, cloned into Pjet and sequenced, confirming the origin of this by-product.

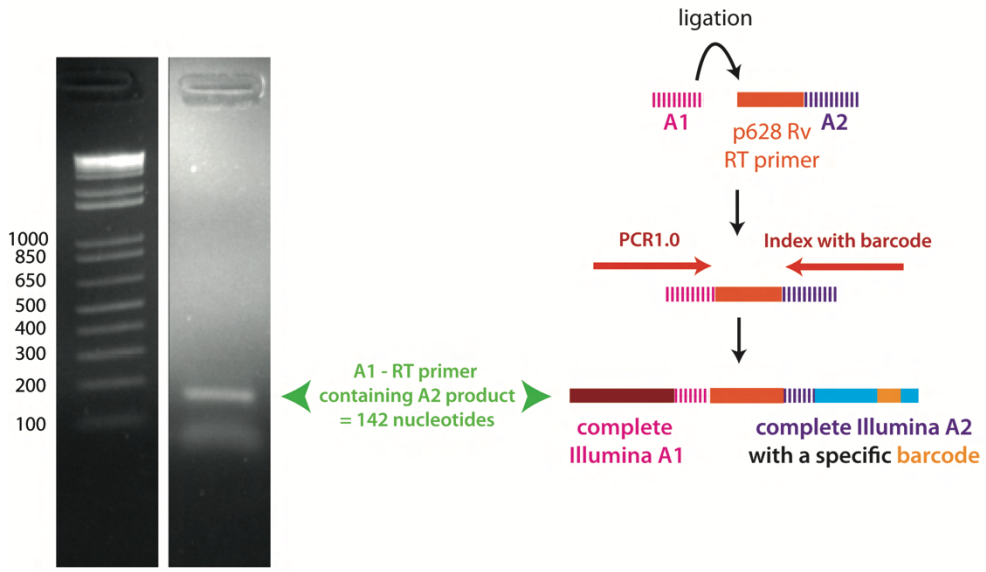


Figure 35
Identification of the PCR product of approximately 150 nucleotides.
PCR with the Phusion polymerase and PCR1.0 and Index primers was performed with a mix containing A1 primer and RT primer containing A2. A PCR product with a size of approximately 150 nucleotides was amplified. This sample has been run on an agarose gel (these two lanes are from the same agarose gel).

- Heterogeneity of ssDNA sizes

The wide range of ssDNA sizes, due to chemical modification of RNAs, influences the ligation efficiency which is higher with small DNA fragments.

- DNA secondary structure

DNA structures at the ends can inhibit ligation. Addition of 10% (v/v) DMSO can increase ligation in these cases.

- Starting amount of ssDNA

The standard CircLigase reaction uses 10 pmol of linear ssDNA whereas the different reactions were performed with 0,002 pmol (theoretical amount prior to RT).

To assess this eventuality, I performed the ligation with two different starting ssDNA amounts (40-fold increase). T4 RNA ligase is inefficient at ligating the sample with less than one million copies whereas a ligated product is detected when the ligation is performed with two billions of DNA copies. However, the ligation efficiency is poor with 0.7% (**Figure 36**).

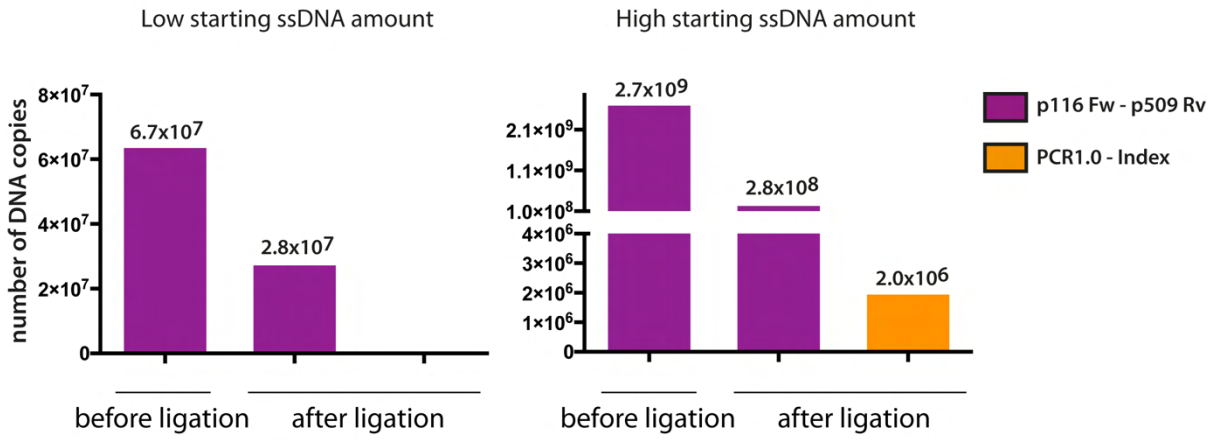


Figure 36

Influence of ssDNA concentration on ligation efficiency.

RNAs extracted from mature viral particle were reverse transcribed with T-GIRT III RT, followed by RNase A treatment, phenol-chloroform extraction, precipitation and divided into two parts with a 40-fold RNA concentration difference. Ligation of A1 primer was performed with T4 RNA ligase 1 from NEB. Template DNA was quantified by Taqman qPCR with p116 Fw - p509 Rv primers (in purple) before and after ligation whereas A1 ligated ssDNAs are detected with PCR1.0 – Index (in orange).

I have tried to adapt the manufacturer's protocols as well as published protocols to my biological samples by changing several parameters of the circularisation/ligation reaction:

- Duration of reaction
- Temperature of reaction
- Amount of ssDNA template and ratio viral ssDNA versus Illumina ssDNA adapter
- Ligases, ATP, defolding (betaine, DMSO) and crowding agents (polyethylene glycol) concentrations.

I did not succeed in increasing the yield of circular ssDNA by modifying these parameters. From the different protocols and enzymes tested as part of the "ligation" strategy, I only succeeded to obtain the full length ssDNA product of 770 nucleotides with the T4 RNA ligase from NEB (**Figure 36** and **37**).

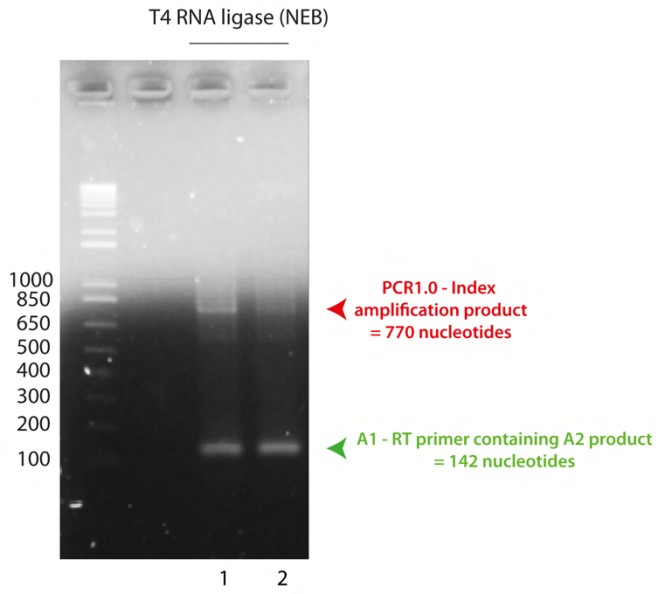


Figure 37
"Ligation strategy" performed with T4 RNA ligase 1 from NEB.
Ten billion copies of NMIA modified (3.9 mM) and unmodified *in vitro* synthesised RNAs were reverse transcribed with T-GIRT III RT followed by RNase A treatment, phenol-chloroform extraction and precipitation. The "ligation" strategy was performed with T4 RNA ligase 1 from NEB. A1 ligated cDNAs were amplified by Phusion PCR with PCR1.0-Index primers. Full-length cDNA of unmodified samples (lane 1) and in a smaller proportion for the 3.9 mM NMIA modified RNA (lane 2) is highlighted with the red arrowhead. Amplification of the A1 primer ligated to RT primer containing A2 with primers PCR1.0-Index is marked with the green arrowhead.

T4 RNA ligase 1 (NEB) protocol

ssDNA samples were incubated at 90°C for 2 min with the A1 primer in order to denature both DNA molecules. Then, 1x T4 RNA ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 1 mM DTT), 25% PEG 8000, 1 mM ATP, 1 mM DTT (in addition to the 1 mM from the commercial buffer) and 10 U T4 RNA ligase 1 were added and samples incubated at 25°C for 16 h. Following ligation, samples were ethanol precipitated and resuspended in 10 µl milliQ H_2O.

Phusion PCR protocol

A mix of 50 µl containing DNA, 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of each primer (Index and PCR1.0 to amplify A1 ligated ssDNAs), 0.01 U Phusion (Thermo Scientific) and milliQ H_2O was prepared. Samples were denatured at 98°C for 3 min followed by 29 PCR cycles composed of 15 s at 98°C (denaturation), 15 s at 56°C (hybridisation) and 2 mi at 72°C (elongation). PCR elongation reaction was completed by a final incubation for 5 mi at 72°C.

The presence of full-length cDNA in both unmodified and NMIA modified samples as well as a smear containing smaller ligated DNA molecules is observed. Nevertheless, the side-product A1-RT primer containing A2 is predominantly amplified compared to the full length cDNA.

I have tried to solve this problem by improving the PCR conditions and by adding a size selection step, prior to PCR, on gel (polyacrylamide, agarose) or using kits with size exclusion columns but the DNA recovery was very low with an elution efficiency of 10% on average.

Based on these results, I decided to adapt the protocol of Illumina library generation (**Figure 31C**). The protocol is presented in details in **Figure 38**.

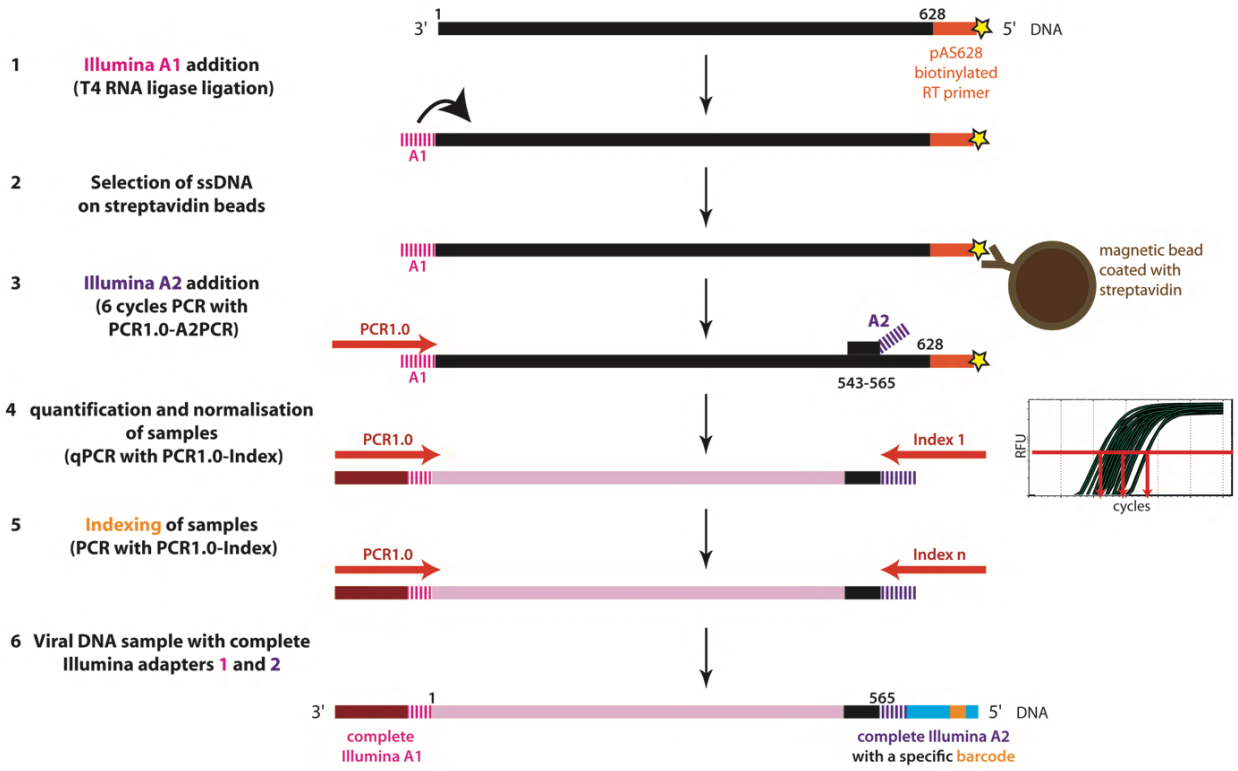


Figure 38
General overview of the Illumina library generation.
A1 was added at the 3' end of ssDNA by ligation using T4 RNA ligase 1 (step 1). ssDNAs were selected based on the biotin-streptavidin interaction, thanks to the biotinylation of the RT primer (step 2). A2 was associated with a HIV-1 specific hybridisation sequence and added by PCR (A2PCR primer) (step 3). Samples were quantified by qPCR with PCR1.0-Index1 in order to normalise the different samples (step 4). Indexing of the samples was performed by PCR with primers PCR1.0 and Index n (n corresponding to an Index primer with a unique barcode) (step 5). DNA samples with complete Illumina adapters at both ends are ready to be sequenced on an Illumina sequencing platform (step 6).

Formation and amplification of A1-A2 products was impossible with this protocol since A2 was added after excess of non-ligated A1 was removed. Washing steps have been included thanks to the selection step with magnetic beads coated with streptavidin with a final wash with milliQ H_2O to remove salts prior to PCR. Several types of magnetic beads have been tested in parallel and the most efficient capture of biotinylated molecules, balanced with non-specific binding to magnetic beads, was obtained with dynabeads MyOne magnetic C1 (Thermo Scientific).

Selection with magnetic beads protocol

The required amount of beads was calculated based on their binding capacity, transferred into a new tube and washed according to the supplier protocol. The tubes (5 µl for each sample) were placed on a magnet for 2 min followed by aspiration of the supernatant while the tubes were on the magnet. Beads were resuspended with 200 µl of 1x washing buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl). Following 3 washes, beads were resuspended with 2x washing buffer to a final concentration of 5 µg/µl (twice the original volume). Ligated samples were precipitated and resuspended in 20 µl milliQ H_2O. An equal volume of the 2x washing buffer was added and samples were incubated for 15 min at room temperature using gentle agitation. The biotinylated ssDNA coated beads were separated with a magnet for 3 min then washed 3 times with 150 µl of 1x washing buffer. An additional wash was performed with 150 µl milliQ H_2O. Beads with the immobilized ssDNA molecules were resuspended with 15 µl milliQ H_2O.

Following selection of the ssDNA samples and depletion of non-ligated A1 primers during the washing procedure, I tried to release the immobilised biotinylated molecules with the manufacturer protocol (2 min at 90°C in 10 mM EDTA pH 8.2 with 95% formamide) or by competition with free biotin molecules. Since, I never succeeded in efficiently recovering biotinylated ssDNA samples, I thus decided to set up PCR conditions with biotinylated ssDNA samples still bound to magnetic beads.

Phusion PCR protocol

A mix of 25 µl containing 1 µl of magnetic beads coated with biotinylated ssDNA samples was prepared with 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of PCR1.0 primer and 0.5 µM of A2PCR primer, 0.02 U Phusion polymerase and milliQ H_2O. Samples were denatured at 98°C for 3 min followed by five PCR cycles composed of 15 s at 98°C (denaturation), 15 s at 61.2°C (hybridisation) and 2 min at 72°C (elongation).

Beads coated with my samples were washed and resuspended in milliQ H_2O to prevent perturbation of the PCR reaction by salts. For the same reason, PCR reactions were performed with 1/15th of the sample. Samples were subjected to only six PCR cycles to avoid over-amplification and preferential amplification of smaller DNA molecules.

Prior to indexing, samples were quantified by qPCR. The PCR1.0-Index1 qPCR reaction was performed with the addition of Evagreen qPCR dye to the final PCR1.0-Index PCR mix. This allowed me to quantify PCR amplification of the different samples in parallel and in the same conditions as for the PCR1.0-Index PCR. Then, the number of PCR cycles performed on each sample for indexing was determined by setting the threshold at half of the exponential phase. So, the Threshold Cycle (Ct) reflected the number of PCR cycles required to reach half of the exponential phase (**Figure 39**).

Evagreen based qPCR protocol

A mix of 25 µl containing 1 µl of PCR PCR1.0-A2PCR was prepared with 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of PCR1.0 and Index1 primers, 0.5x Evagreen dye, 0.02 U Phusion polymerase and milliQ H_2O. Samples were denatured at 98°C for 3 min and 29 PCR cycles composed of 15 s at 98°C, 15 s at 56°C and 2 min at 72°C were performed.

Following addition of A2 at the 5' end, indexing of samples was performed by PCR with PCR1.0 and an Index primer containing a specific barcode, following the PCR protocol described for the PCR1.0-A2PCR primers with 0.25 µM of each primer and a PCR cycle composed of 15 s at 98°C, 15 s at 56°C and 2 mi at 72°C. Based on the qPCR quantification, samples were normalised by performing a specific number of PCR1.0-Index PCR cycles. Thus, a comparable amount of each sample is multiplexed in the Illumina library and amplification is stopped before reaching the plateau phase, thus preventing over-amplification.

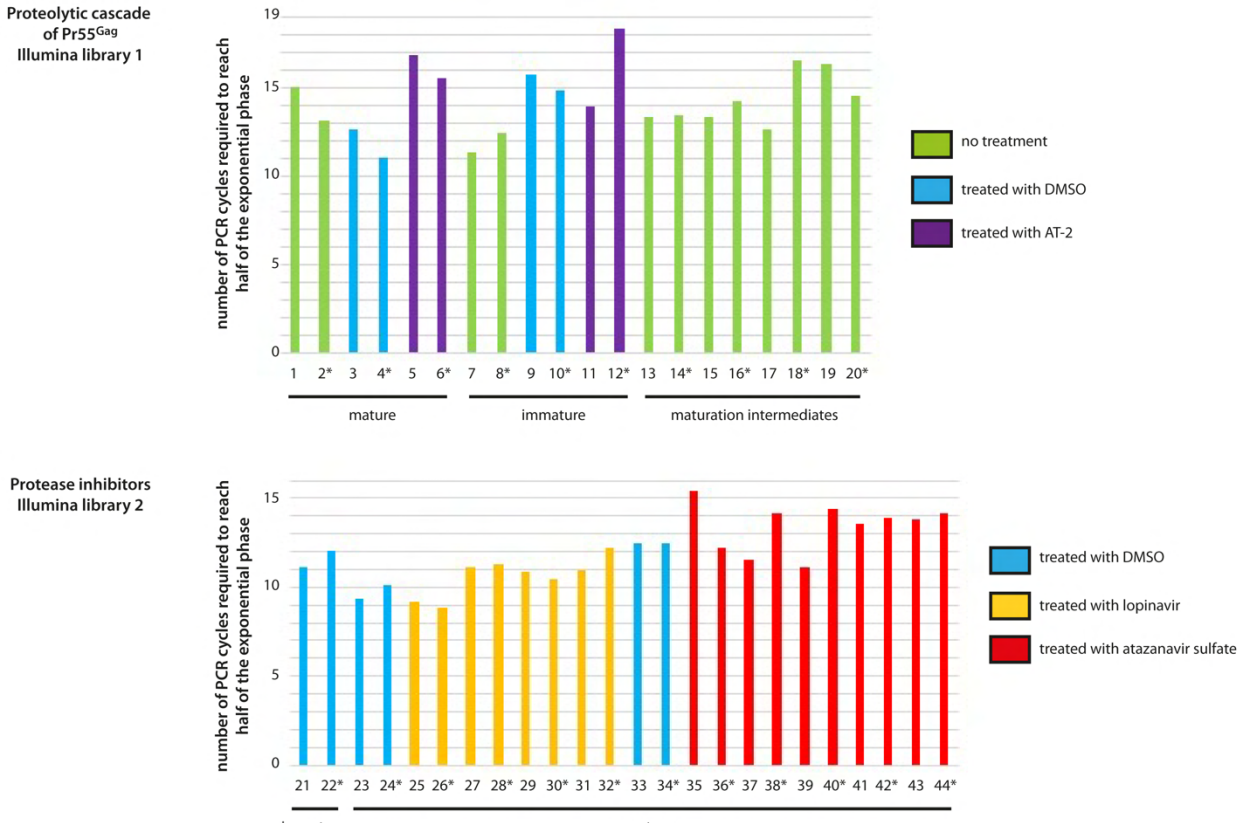


Figure 39
Determination of the number of PCR cycles to apply for normalising and indexing Illumina library samples.
All Illumina library samples were subjected to Evagreen based qPCR with primers PCR1.0-Index1, with a threshold set up at half of the exponential phase.
Illumina library 1 contains samples representing the proteolytic cascade of $Pr55^{Gag}$ with immature (lanes 7 to 12), maturation intermediates (lanes 13-14 for Step 1, 15-16 for Step 1.1, 17-18 for Step 2 and 19-20 for CA/p2) and mature samples (lanes 1 to 6). Illumina library 2 contains samples treated with protease inhibitors: increasing concentrations of LPV with 1.1 nM (lanes 25-26), 6.6 nM (lanes 27-28), 41 nM (lanes 29-30), 1600 nM (lanes 31-32) and increasing concentrations of AZVs with 0.2 nM (lanes 35-36), 1.1 nM (lanes 37-38), 2.6 nM (lanes 39-40), 16.4 nM (lanes 41-42), 1600 nM (lanes 43-44). Each condition has an unmodified sample associated with a NMIA modified (3.9 mM) sample indicated with a star *.

The profile of the different samples after indexing PCR was analysed by agarose gel electrophoresis (**Figure 40 and 41**).

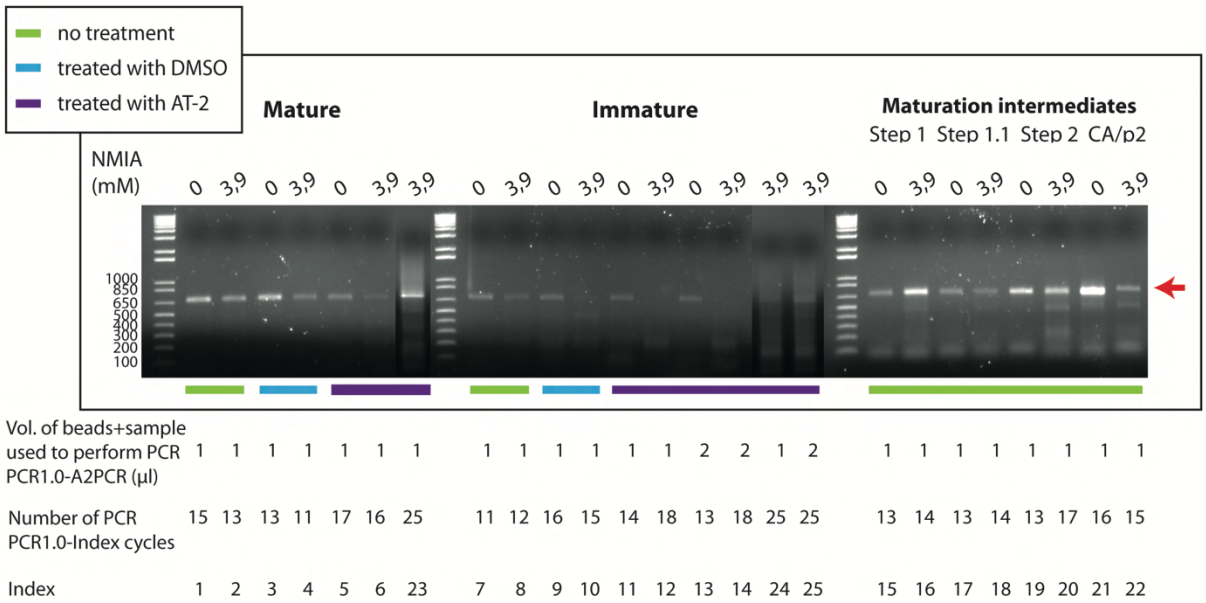


Figure 40
Profile of the samples constituting the Illumina library 1 (Proteolytic cascade of Pr55^{Gag}) after indexing PCR on agarose gel.
Samples were run on a 1% agarose gel TBE 0.5x (1/25th of indexing PCR mix). The profile is coherent with full-length products at 687 nucleotides. The volume of beads used to perform PCR1.0-A2PCR is indicated, as well as the number of cycles for the indexing PCR and the index used for each sample.

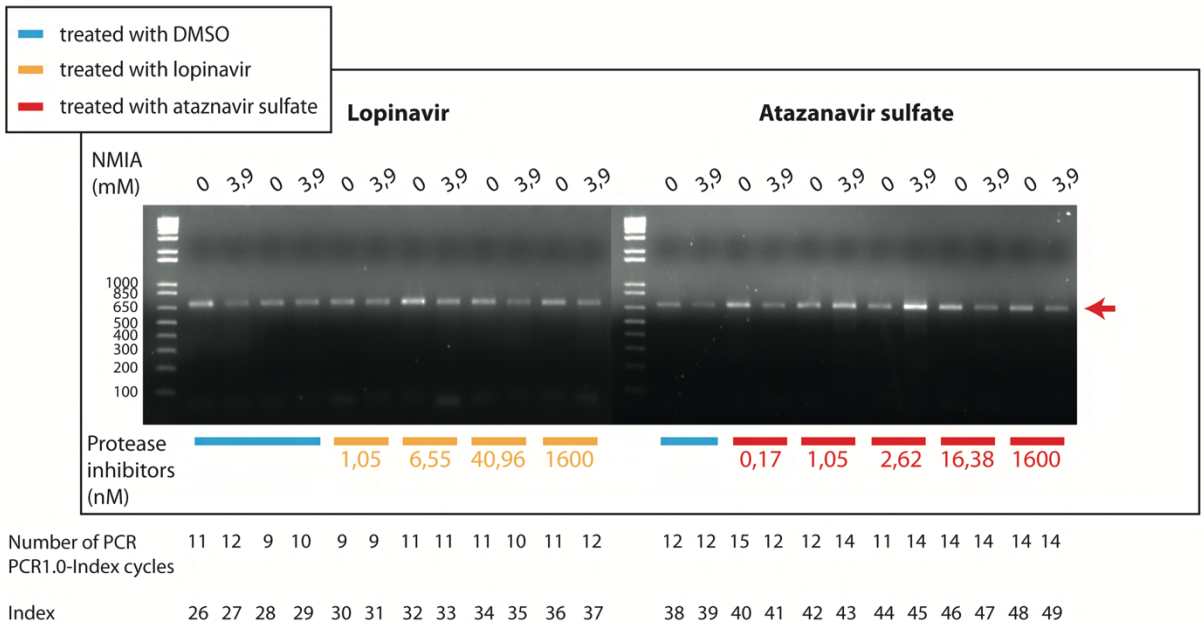


Figure 41
Profile of the samples constituting the Illumina library 2 (Protease inhibitors) after indexing PCR on agarose gel.
Samples were run on a 1% agarose gel TBE 0.5x (1/25th of indexing PCR mix). The profile is coherent with full-length products at 687 nucleotides. The number of cycles for the indexing PCR is indicated, as well as the index used for each sample.

The full-length amplification product was visible in almost all samples. In the case of sample treated with AT-2 and modified with NMIA, the full-length product was less abundant in the case of the mature sample and not visible in the immature sample. This tendency is coherent with the quantification I performed by Taqman qPCR with a specific Taqman probe and amplification performed with primers p116 Fw – p509 Rv quantifying the DNA template and PCR1.0 – p509 Rv measuring A1-ligated samples. The primers and probe are depicted in **Figure 42** and the PCR results are shown in **Figure 43**.

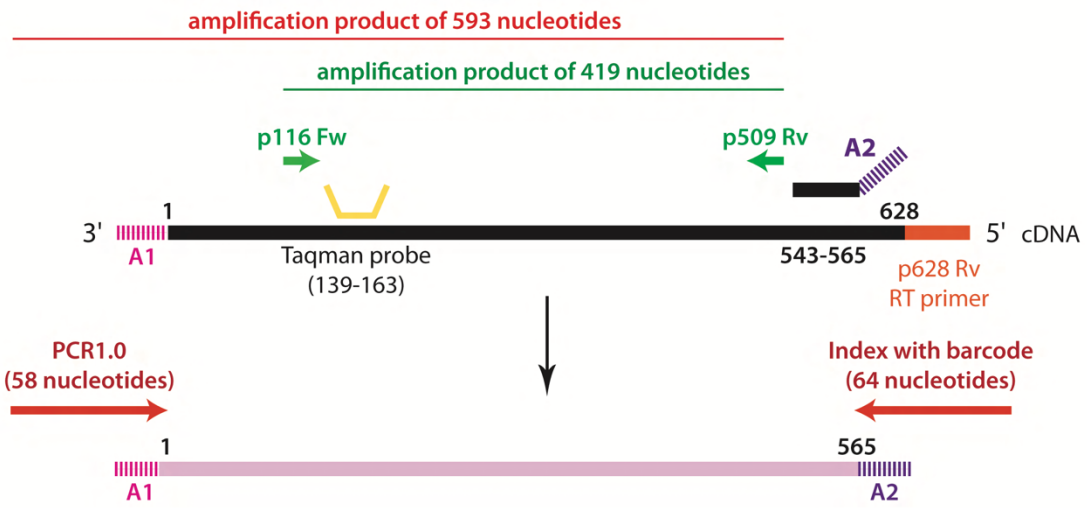


Figure 42
Description of qPCR/PCR primers, hybridisation position, taqman probe and amplification products.
The two qPCR primers used to quantify ssDNA template (p116 Fw – p509 Rv) and A1-ligated ssDNAs (PCR1.0 – p509 Rv) are shown with their amplification products, as well as the Taqman probe.

Therefore, I have tried to boost the amplification of modified mature and immature samples treated with AT-2 at two levels. I increased the number of PCR cycles during indexing PCR and I also tried in parallel to perform the PCR1.0 – A2PCR PCR with twice more DNA template followed by qPCR quantification and application of a specific cycle number required for indexing (**Figure 39**).

Following indexing, samples were pooled and send to the sequencing platform. I tested several protocols to purify PCR samples from dNTPs and other components of the PCR reaction mix.

The classical approach in this case was to use kits containing columns combined or not with agarose gel extraction (Gel and PCR clean-up from Macherey-Nagel, Wizar SV gel and PCR clean-up from Promega, Costar spin x centrifuge tube 0,22 µm filter from Sigma-Aldrich). I also tested the concentration of my samples by using centricons followed by gel purification, and classical and low-melting agarose types at various percentages. In addition, I have tried to purify the Illumina library by HPLC (high pressure liquid chromatography) with a GSK 2000 column, on Agencourt ampure xp beads (Beckman coulter) and by using the E-gel CloneWell 20% agarose gel system (Invitrogen).

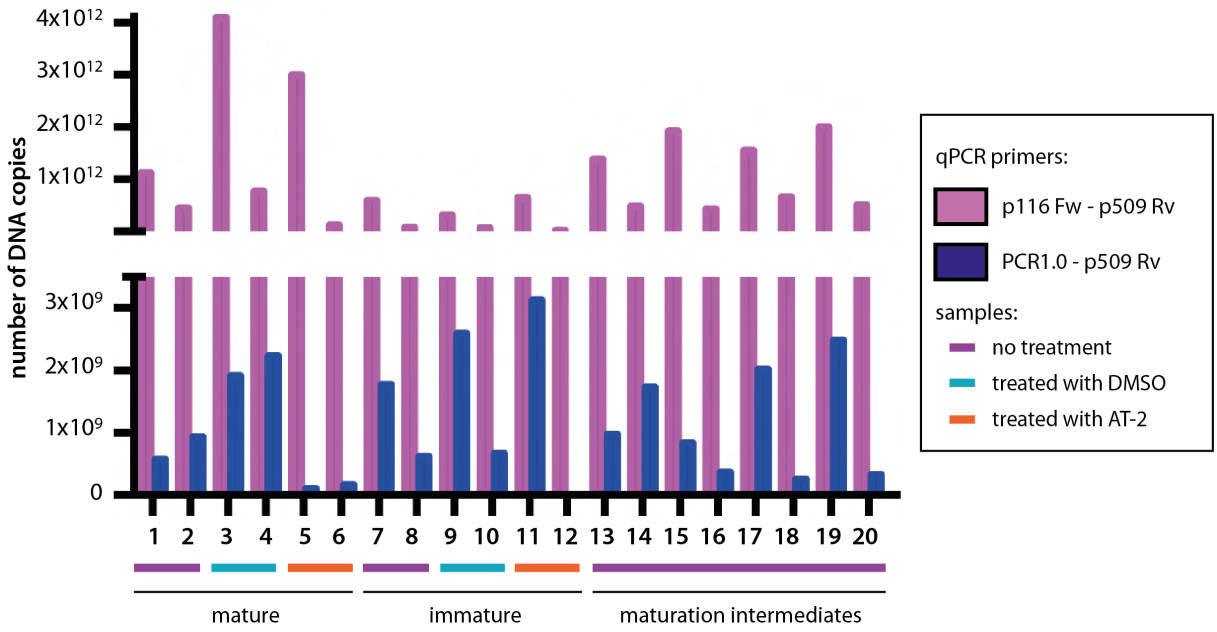


Figure 43
Quantification of the proportion of A1-ligated ssDNA samples from the Illumina library 1 by qPCR.
DNA template and A1-ligated ssDNAs were quantified by Taqman qPCR with primer pairs p116 Fw – p509 Rv and PCR1.0 – p509 Rv, respectively.

None of these techniques allowed me to recover the Illumina library after purification, with either the complete loss of the sample or an aberrant size selection. Thus, I simply pooled half of each PCR1.0-Index PCR reaction from each sample, ethanol precipitated and resuspended the Illumina library in 30 µl milliQ H_2O. The profile of each Illumina sequencing library is shown on agarose gel (**Figure 44**).

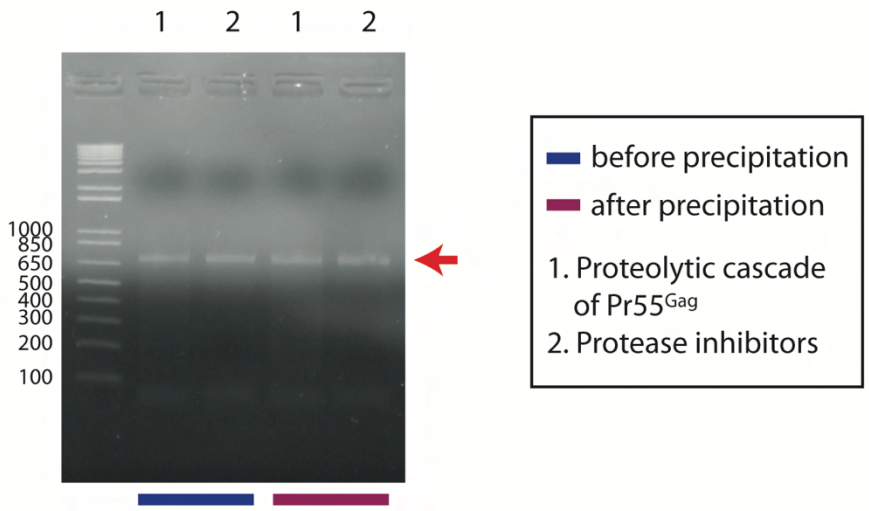


Figure 44
Profile of the two Illumina sequencing libraries on agarose gel.
The indexed samples were pooled, ethanol precipitated and run on a 1% agarose gel TBE 0.5x. The profile of the two Illumina sequencing libraries before and after precipitation is coherent with full-length products of 687 nucleotides. The smear containing the DNA population with different sizes due to RNA chemical modification was also detected.

In vitro study of the chaperone property of Pr55^{Gag} and NC-containing intermediates

---

I. Protein expression, purification and characterisation

Expression, purification and characterisation of NL4.3 Pr55^{Gag} and GagΔp6 with an appended C-terminal His$_6$-tag were performed as recently described (289) by the team of J. Mak (Deakin University, Department of Infectious Diseases, Melbourne, Australia).
NCp15, NCp9 and NCp7 proteins were expressed and purified by the team of C. Tisné (UMR8261 Expression génétique microbienne CNRS, Université Paris-Diderot Paris 7).

The purity of all proteins was confirmed using SDS-PAGE (**Figure 45**). In addition, proteins were characterised by Dynamic Light Scattering (DLS) analysis. Intensities of the scattered light and correlation times were measured using a Zetasizer Nano S apparatus (Malvern, UK).

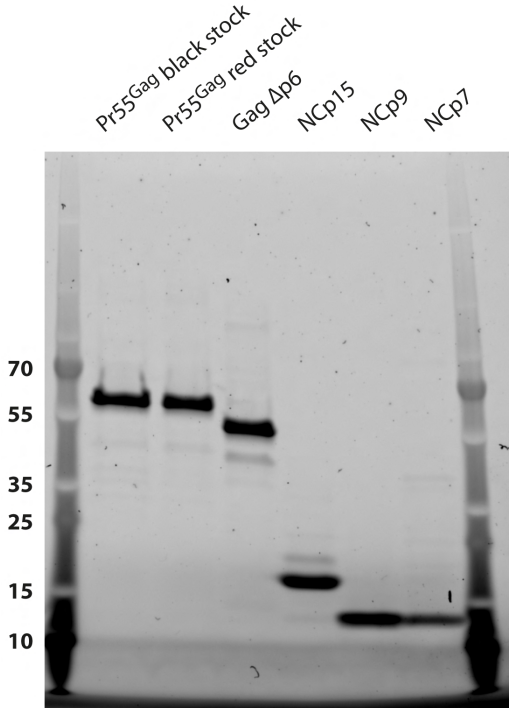


Figure 45
Characterisation of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7 proteins.
The purity of the different proteins has been assessed by using 10% SDS PAGE. Pr55^{Gag} black and red stock are originated from a unique production and purification of the protein with the black stock corresponding to the peak fractions of elution and the red stock from fractions bordering the elution peak.

II. Formation of RNA-protein complexes

The purified *in vitro* transcribed NL43 1-1400 WT RNA (plasmid pDR4607 (384)) has first been refolded prior to be mixed with proteins. Briefly, 2 pmoles of RNA in 7 µl milli-Q H_2O was denatured for 2 min at 90 °C and then chilled for 2 min on ice. RNA was refolded by the addition of 2 µl x5 binding buffer (30 mM Tris pH 7.5; 300 mM NaCl; 5 mM $MgCl_2$), 2 µg of total yeast tRNA (Sigma-Aldrich) and 5 U of RNasine (Promega) followed by incubation for 15 min at 37°C. In parallel, proteins (0.24 nmoles final concentration) were refolded by the addition of 10 µl x5 binding buffer, 1.2 µl of 10 mg/ml BSA at, 1.2 µl of 0,5 M DTT. The volume was adjusted to 50 µl with milli-Q H_2O followed by incubation for 15 min on ice. 10 µl of folded RNA and 50 µl of folded protein were then mixed to achieve a final RNA:protein stoichiometry of ≈ 1:125 and incubated at room temperature during 10 min with mixing then stabilised on ice for an extra 10 min. Depending on the condition, samples were treated with 1.6 U of proteinase K (NEB) and incubated for 1 h at 37°C. The proteinase K treatment was supplemented with 1% SDS in the case of NCp9 and GagΔp6.

III. SHAPE methodology

The different conditions (RNA only, RNA-protein complex treated with proteinase K / AT-2 and RNA-protein complex) were subjected to high-throughput SHAPE (h-SHAPE) using the NMIA probe. Briefly, each sample (60 µl total volume) was divided in two and 7.2 µl of 250 mM NMIA solution in anhydrous DMSO or an equal volume of DMSO (negative control sample) was added. After 50 min at 37°C, samples were treated with proteinase K as previously described. The reaction was stopped by adding milli-Q H_2O to a final volume of 200 µl. Samples were then subjected to phenol-chloroform followed by chloroform extraction and ethanol precipitation. Samples were precipitated using 1 µg of glycogen, 1/10 volume of 3 M sodium acetate (pH 6.5) and 3 volumes of ethanol for 30 min on dry ice. Sodium chloride instead of sodium acetate was used for samples with NCp9 and GagΔp6 at a final concentration of 0.2 M to prevent precipitation of SDS. The nucleic acid precipitates were collected by centrifugation at 20,817 g for 30 min at 4°C. The RNA pellets were washed twice with 1 mL of 70% ethanol, dried in a vacuum dryer during 5 min and resuspended in 7 µl milli-Q H_2O.

Sites of NMIA modification were identified as stop to RTion, using the Rv primer 5'-AGCTCCCTGCTTGCCCATACT-3' (nucleotides 436-457), labelled either with VIC or NED fluorophores. NMIA-modified and unmodified RNAs were annealed to 1 µl of 1 µM VIC-labelled primer for 2 min at 90°C and 2 min on ice. After addition of 2 µl of 5x RT buffer (Life Science), the samples were then incubated for 10 min at room temperature. Elongation reaction was performed at 42°C during 20 min followed by 30 min at 50°C and 10 min at 60°C in elongation buffer (1 µl 5x RT buffer, 3 µl 2.5 mM dNTPs mix, 2 U AMV RT (Life Science) in a total volume of 10.5 µl). A ddG sequencing ladder was prepared using 2 pmol of untreated RNA and 1 µl of 2 mM NED-labelled primer in 8 µl milli-Q H_2O. Annealing was performed as described above. Two µl x10 RT buffer was added followed by incubation for 15 min at room temperature. The RNA sample was aliquoted into two tubes, and the elongation reaction was performed with 1 µl 10x RT buffer, 3 µl A10 (0.25 mM dATP, 1 mM dGTP, 1 mM dCTP, 1 mM dTTP), 1 µl of 100 mM ddA and 1 U AMV RT. All the reactions were stopped by

adjusting the volume to 200 μl with milli-Q H_2O, and proteins were extracted with a phenol–chloroform followed by a chloroform treatment. For each experiment, the modified and unmodified samples were pooled with a ddG sequencing ladder. Samples were ethanol precipitated with sodium acetate as previously described. Pellets were dried and resuspended in 10 μl HiDi formamide (ABI), heat denatured at 90°C and iced for 5 min and centrifuged at 20,817 g for 15 min at 4°C before loading on the 96-well plates for sequencing on an Applied Biosystems 3130xl genetic analyser. The results were generated in the form of electropherograms, which were analysed using the Qu-SHAPE software (216).

RESULTS

RESULTS

I. Optimising Illumina sequencing library generation

The protocol for Illumina library preparation has been extensively optimised as explained in the previous section and a total of three Illumina sequencing libraries (containing samples covering the proteolytic processing of Pr55^{Gag}) have been prepared, sequenced on an Illumina sequencing HiSeq 2500 platform (50 base-pair single-end sequencing, First and second attempt at the genomic platform of IGBMC institute, Strasbourg and third attempt at the genomic platform of Imagine institute, Paris) and analysed. The protocol used to generate these three libraries is briefly summarised in **Figure 46** with differences highlighted.

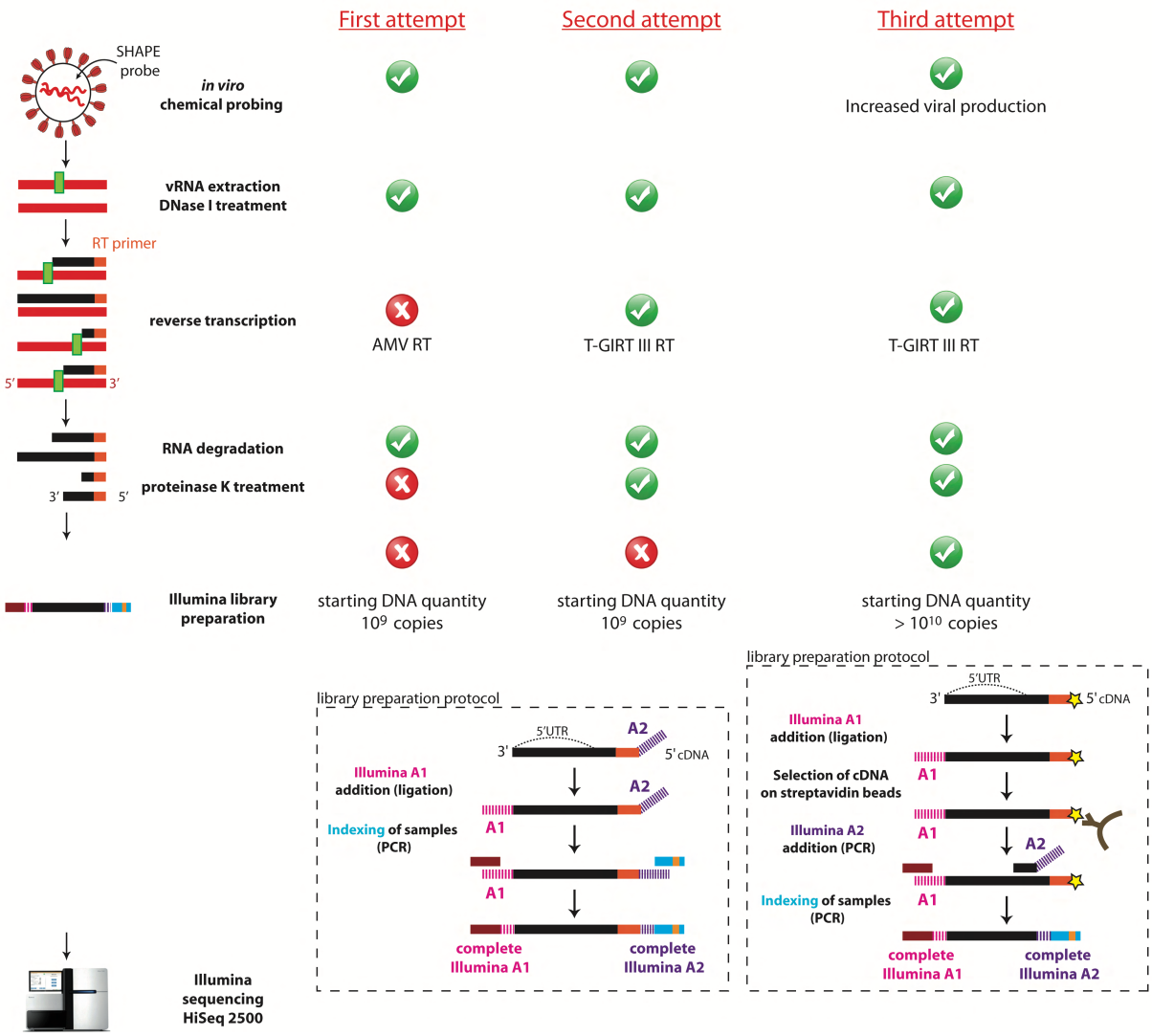


Figure 46
Evolution of the hSHAPE-Seq protocol.
The main steps constituting the hSHAPE-Seq protocol are presented on the left. Three successive versions of the hSHAPE-Seq technique with differences highlighted in the figure have yielded the production of three Illumina sequencing libraries.

The cumulative number of reads has been calculated for the unmodified mature sample in the three different Illumina sequencing libraries (**Figure 47**). The cumulative number of reads is compared and, with this representation, the number of reads decreases from 100 to 0 %.

With the first Illumina sequencing library, the number of reads covering the first 60 nucleotides is high but drastically decreases at position U485 and quickly goes to 100 %, corresponding to very few reads left beyond that point (blue trace in **Figure 47**). This profile is coherent with the AMV RT profile visualised using gel electrophoresis (previously shown in **Figure 26** lanes 1-2 and 7-8), which highlighted a strong RT pause early in the sequence.

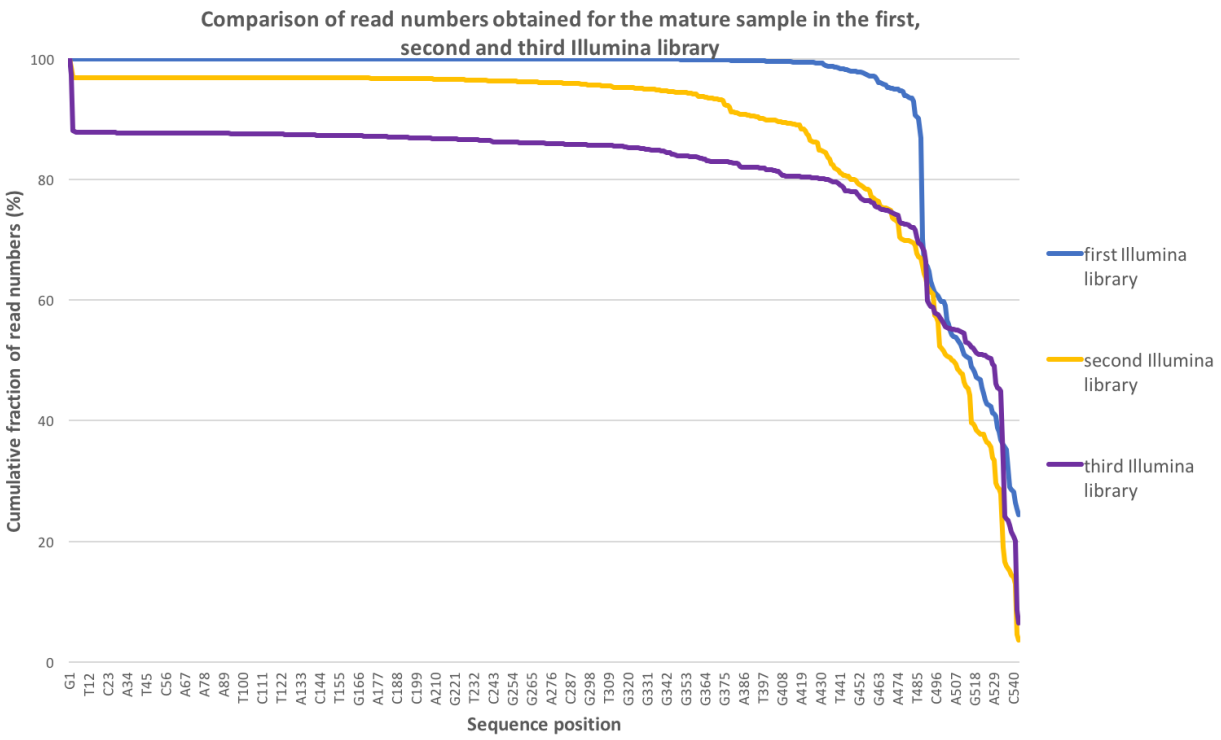


Figure 47
Comparison of read numbers obtained for the mature sample in the first, second and third attempts to generate an Illumina sequencing library.
The cumulative fraction (expressed as percentage) of read numbers in the unmodified mature sample from the three different Illumina libraries is compared. The first Illumina library was generated with the AMV RT, the second and the third libraries with the T-GIRT III RT but with different library preparation protocols. The cumulative fraction of reads was calculated by adding the number of reads at or after the position under consideration divided by the total number of reads aligned for this sample and converted into a percentage.

The second Illumina library has been prepared with a lower vDNA starting concentration and no purification step to eliminate the excess of non-ligated Illumina primers. A higher number of cycles has been applied when performing the PCR1.0-A2PCR PCR, compared to the third Illumina library with results presented below. In addition, samples were normalised at the very end of the library preparation before pooling samples together so the indexing PCR was performed on different DNA concentrations depending on the sample. Thus, the amplification was not achieved under optimal and similar conditions for all the samples of the library.

The profile of the unmodified mature sample from the second Illumina library is displayed in orange in **Figure 47**. The number of natural RT stops due to the GC-rich double-stranded region near position U485 decreased with T-GIRT III compared to AMV RT, with reads covering up to position G1 corresponding to the full-length cDNA product. The low proportion of the DNA full-length product compared to smaller DNA fragments is probably due to over-amplification of these fragments during the indexing PCR. In an attempt to increase the proportion of full-length cDNA product, a purification step and improved sample normalisation and PCR cycling conditions have been included during generation of the third Illumina library. The result of these optimisations is a three-fold increase of reads corresponding to the full-length cDNA product, with a slight enrichment of smaller DNA fragments in the first 3' 60 nucleotides (purple trace in **Figure 47**).

II. Bioinformatic analysis of sequencing data: benchmarking analytical methods

The reads of each sample were deconvoluated thanks to their barcode and aligned to the HIV-1 genome as the reference sequence. Aligned reads are converted to fastq format and analysed with an adapted pipeline to convert raw RT stops (**Figure 48**) into nucleotide reactivity.

A number of softwares have been developed to convert sequencing reads into reactivity. These analysis pipelines have been created depending on the experimental design used to generate the library. Two of them, Structure fold (120, 404) and Spats (268), have been created to analyse Illumina libraries with a profile similar to those generated by hSHAPE-Seq (population of DNA molecules with different sizes). We used Structure fold and Spats to analyse raw data and, in addition, scripts specifically written for the project by Dr Redmond Smyth.

* Spats has been designed to perform bioinformatics read alignment from paired-end fastq files. This parameter has been adapted since we generate single-end fastq files. A maximum-likelihood-based signal decay correction (22, 299) is applied to calculate SHAPE-Seq reactivity values for each nucleotide.

Signal decay is a decrease of the number of reads from the 3' to the 5' end of the RNA molecule with increasing cDNA length and is due to RT fall-off and chemical modification in NMIA modified samples.

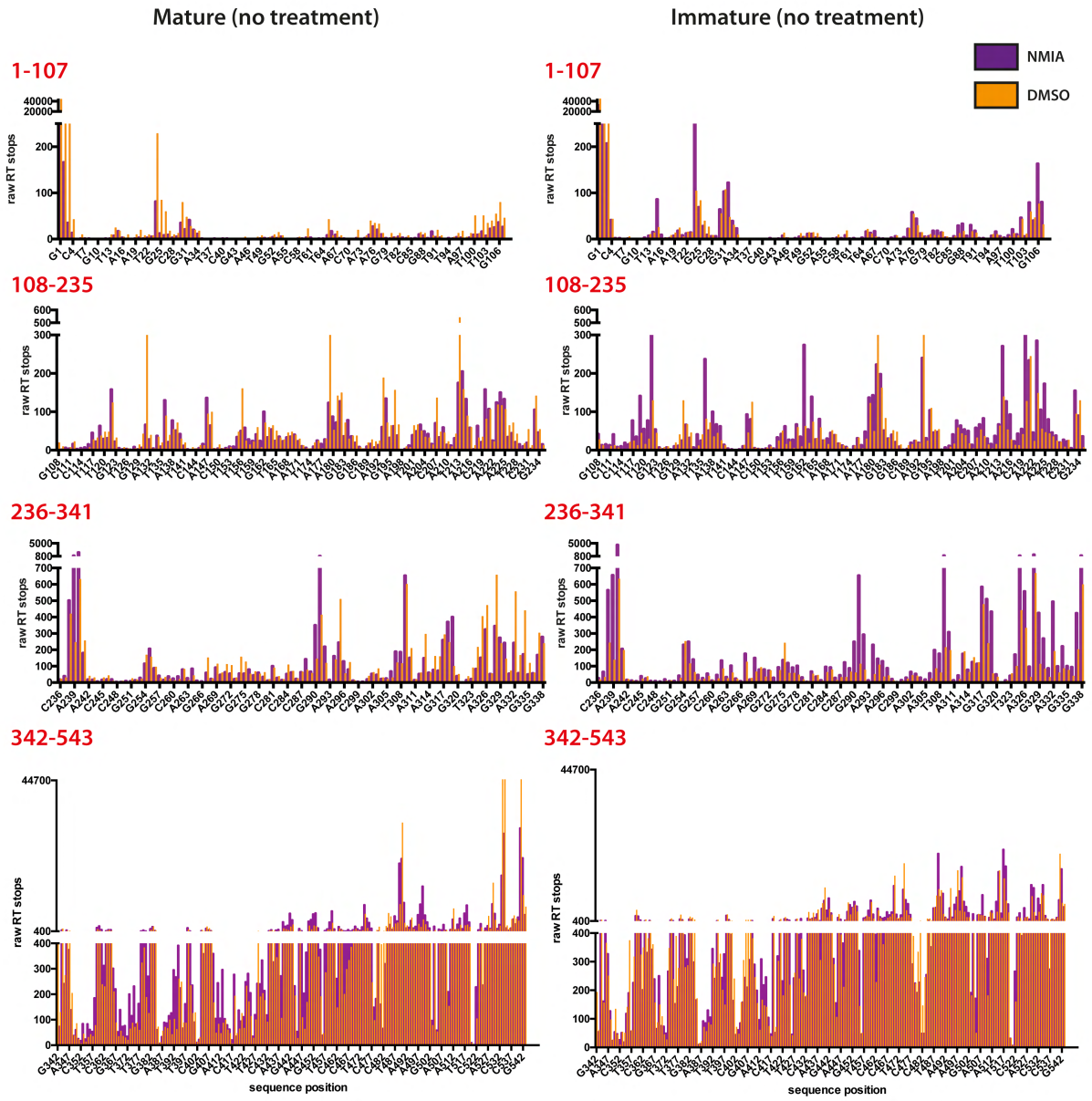


Figure 48

Raw RT stop profile of untreated mature and immature samples.

The number of raw RT stops of NMIA modified/unmodified mature and immature samples (without treatment of virus particles prior to NMIA modification) is presented. The different samples have been normalised to be compared. The sum of reads covering each position of a sample has been calculated for each sample. Then, each nucleotide has been multiplied by the sum of reads of the lowest reads' sum and divided by the actual reads' sum of this sample. RT stops due to a chemical modification are represented in purple whereas natural RT stops are in orange.

The typical profile obtained is a level of full length cDNA higher in the umodified sample, while in the region covering the first 100 nucleotides the signal is higher in the modified sample. As a consequence of signal decay the untreated signal becoming higher than the treated one from the middle of the sample until the 3' end, explaining why signal decay correction is absolutely required (**Figure 49**).

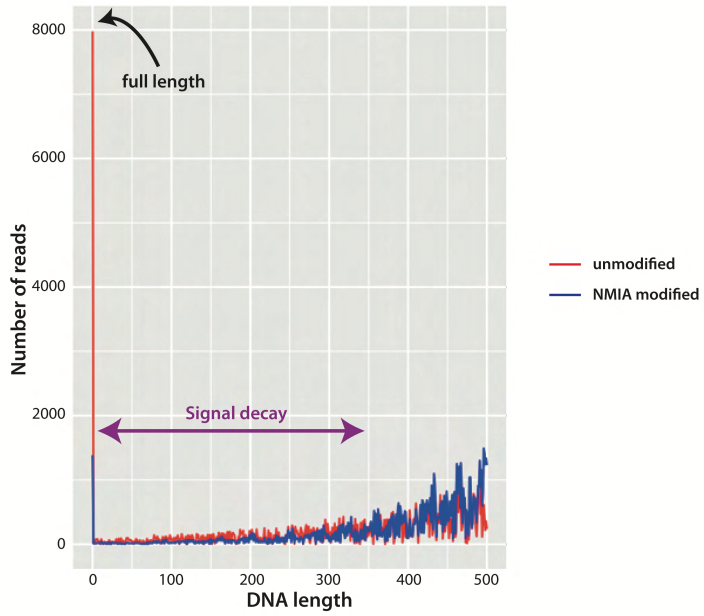


Figure 49
Illustration of the signal decay.
The number of RT stops at each nucleotide with a covering window of 500 nucleotides is represented. The signal decay phenomenon is highlighted by the lower read density of the modified sample from the middle until the 3' end compared to the unmodified sample.

* Structure-fold performs a simple treatment of data by mapping reads on a reference transcript, calculating the number of RT stops, normalising data and calculating reactivity. The signal decay is not corrected by the structure-fold approach, probably because this software was designed for the analysis of samples in which RTion was performed using random hexamers. The effect of signal decay is thus probably lower compared to approaches using a sequence-specific RT primer.

* The bioinformatics pipeline created by Dr Redmond Smyth is able to automatically convert raw sequencing data into SHAPE reactivity. Raw reads are demultiplexed to assign each index to the correct sample. These reads are then 5' trimmed to remove the first three nucleotides of the adaptor (NNN) and 3' trimmed for quality. Indeed, no mismatch is allowed in order to eliminate the possibility that the RT stop at the 3' end is not due to a RT mutation instead of a chemical modification. Reads are then aligned to the HIV-1 5' region and the 5' position of the alignment is considered to be a RT stop. Raw RT stops were normalised between 'plus' and 'minus', and then corrected to remove signal decay. Signal decay was corrected by calculating from 5' to 3' for each position of the RNA sequence the cumulated number of stops up to this position, and dividing this number by the total number of stops observed for the entire sample. SHAPE reactivity were calculated from these normalised values by subtracting the number of 'minus' stops (unmodified sample with DMSO only) from 'plus' stops (modified sample with 3.9 mM NMIA). Negative values were set to 0 and reactivity were normalised by dividing by the average reactivity of the top 2-8% of highly reactive nucleotides.

Firstly, I compared the reactivity profile obtained with the different analysis pipelines to raw data in the case of modified/unmodified mature and immature PR- samples. This comparison allowed me to choose the appropriate analysis pipeline, based on raw data obtained by the hSHAPE-Seq approach.

Raw data of these samples have been normalised by the sample with the lowest number of reads. These nucleotides reactivity have been reported on a HIV-1 RNA secondary structure model published by the team of K. Weeks, based on *in viro* chemical probing of mature gRNA by hSHAPE (426).

Three regions of the 5'-region of the HIV-1 gRNA have been taken as examples to highlight the different analysis pipelines:

- Nucleotides 146-164

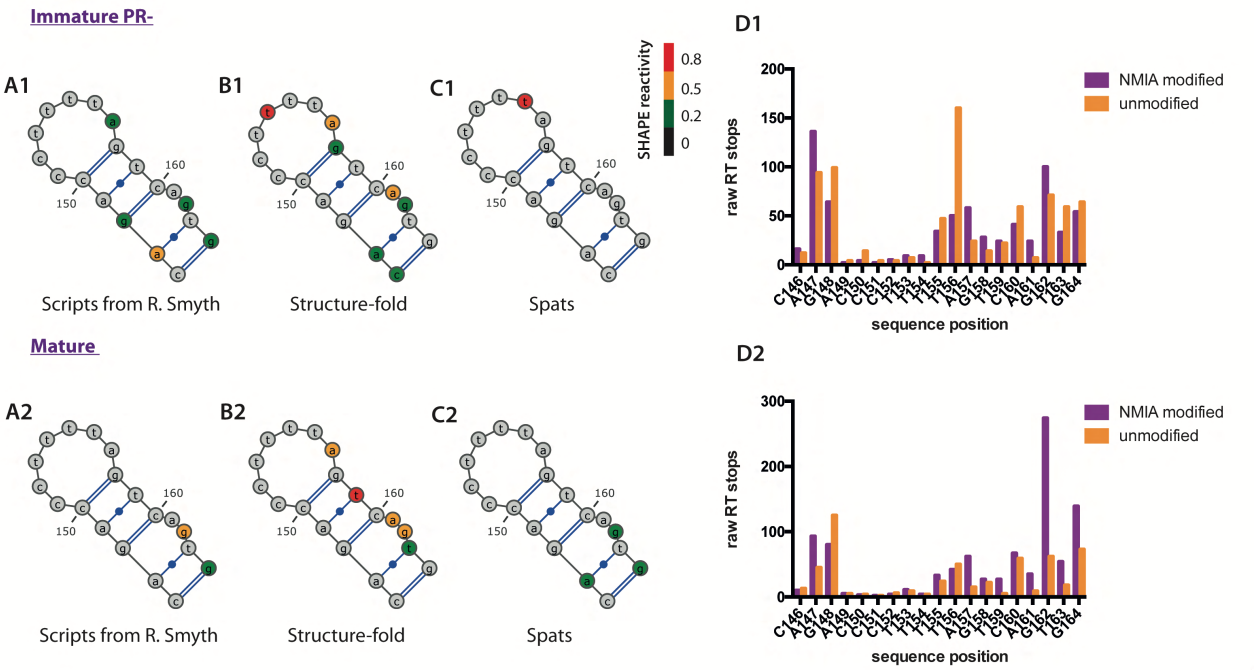


Figure 50
Comparison of the nucleotide 146-164 reactivity profile of immature PR- and mature samples, obtained with our scripts, Structure-fold and Spats.
Reactivity intensities were placed on a published model (426) with nucleotides coloured according to their SHAPE reactivity (from unreactive represented in black to highly reactive in red). The three reactivity profiles are compared with raw data representing the number of RT stops at each position in the modified and control samples.

This region is localised approximately 400 nucleotides from the 5' end and the signal decay is not corrected in raw data (**Figure 50**). This needs to be taken into account when comparing the level of RT stops between modified and unmodified samples.

In PR- virions, according to the raw data, A147 and G162 seem reactive when looking at the ratio NMIA/control samples (**Figure 50 D1**), which correlates with our scripts and Structure-fold (**Figure 50 A1** and **B1**), whereas no reactivity is displayed at these positions by Spats (**Figure 50 C1**). Conversely, U156 indicated as highly reactive with Spats (**Figure 50 C1**), is not highlighted in the two other analysis pipelines (**Figure 50 A1** and **B1**) and has a number of RT stops much higher in the control sample (**Figure 50 D1**).

The mature sample has a similar pattern of reactivity as the PR- sample. Structure-fold, in general, provides a reactivity profile higher (**Figure 50 B2**) than the two others analysis with nucleotides A161-U163 reactive whereas only G162 is detected as being reactive using our scripts and Spats (**Figure 50 A2** and **C2**). Surprisingly, only Structure-fold detects U159 as reactive (**Figure 50 B2**) and with a higher reactivity (R=0.87) than G162 (R=0.63), which has as a level of RT stops much higher than U159 (**Figure 50 D2**).

- Nucleotides 286-296

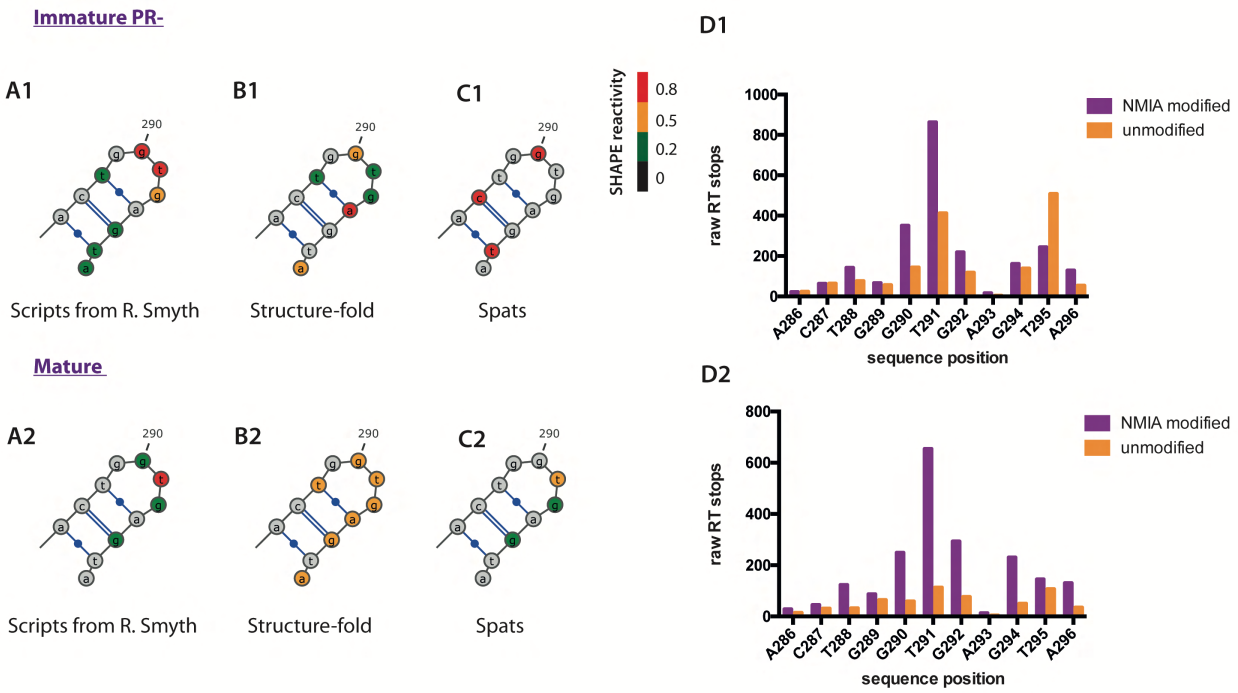


Figure 51
Comparison of the nucleotide 286-296 reactivity profile of immature PR- and mature samples, obtained with our scripts, Structure-fold and Spats.
Reactivity intensities were placed on a published model (426) with nucleotides coloured according to their SHAPE reactivity (from unreactive represented in black to highly reactive in red). The three reactivity profiles are compared with raw data representing the number of RT stops at each position in the modified and control samples.

This region is located approximately 250 nucleotides from the 5' end and the apical loop appears more reactive in the PR- compared to the mature sample (**Figure 51 1** and **2**). A clear difference of intensity is observed when comparing reactivity intensities obtained after analysis with our scripts and Structure-fold (**Figure 51 A1, B1, A2** and **B2**), with a higher level using Structure-fold (**Figure 51 C1** and **C2**). Whereas nucleotides G290-G292 have a higher number of RT stops in the PR- modified sample, indicating a reactive site (**Figure 51 D1**), Spats only detects G290 as reactive (**Figure 51 C1**).

- Nucleotides 506-521

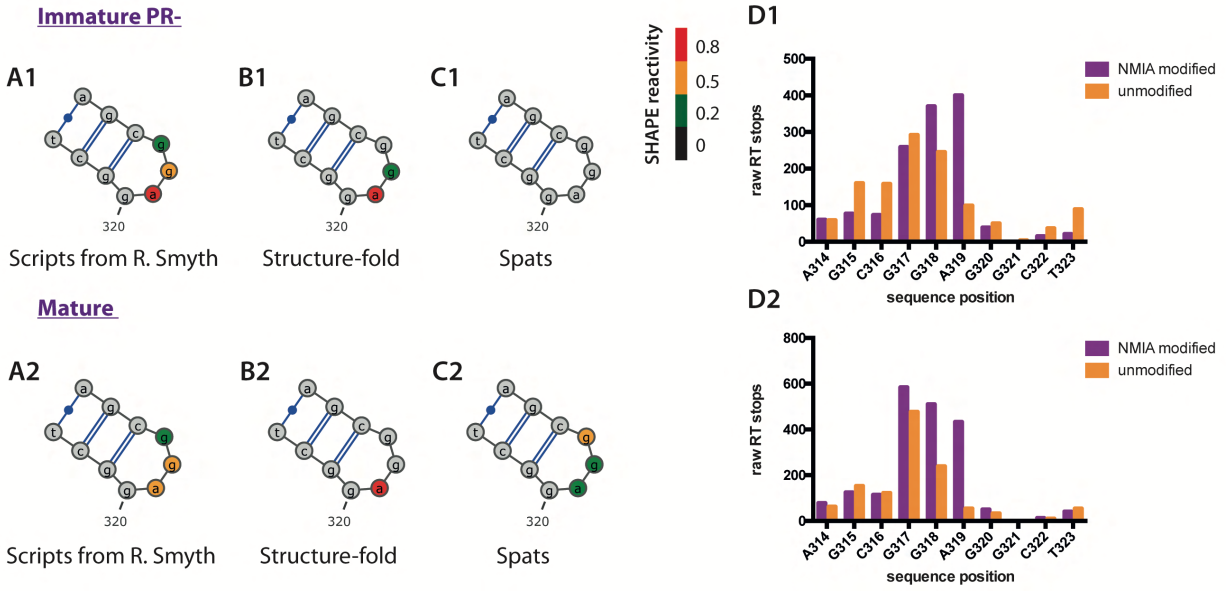


Figure 52
Comparison of the nucleotide 506-521 reactivity profile of immature PR- and mature samples, obtained with our scripts, Structure-fold and Spats.
Reactivity intensities were placed on a published model (426) with nucleotides coloured according to their SHAPE reactivity (from unreactive represented in black to highly reactive in red). The three reactivity profiles are compared with raw data representing the number of RT stops at each position in the modified and control samples.

This region is located in the *gag* coding region, approximately 20 nucleotides from the 5' end of the region of the HIV-1 RNA I analysed (**Figure 52**). No marked difference is observed between PR- and mature samples when raw data are converted into reactivity by our scripts (**Figure 52 A1** and **A2**). A slightly better correlation of the PR- reactivity profile with raw data is observed when generated by our scripts and Spats (**Figure 52 A1** and **C1**). In the mature sample, Structure-fold seems to underestimate reactivity with C516 being poorly reactive and U517 not reactive (**Figure 52 B2**), whereas raw RT stops are significantly higher in the modified sample compared to the negative control (**Figure 52 D2**). In addition, these nucleotides are close to the RT primer so they should not suffer from signal decay.

In conclusion, significant variations are observed when comparing nucleotide reactivity intensities generated from raw data by scripts written by R. Smyth, Structure-fold developed by the team of S. Assmann and Spats from the team of J. Lucks. These variations are potentially due to signal decay correction for Spats and the absence of correction of this parameter for Structure-fold. Nevertheless, this hypothesis cannot explain the difference of reactivity intensities obtained for nucleotides 506-521, since this region is close to the RT primer.

Based on these observations, raw data have been converted into reactivity using the script specifically generated by Dr Redmond Smyth in our team to treat hSHAPE-Seq raw data.

III. Analysis of Illumina library generated by hSHAPE-Seq: Proteolytic cascade of Pr55^{Gag}

The structure of the HIV-1 gRNA has been interrogated inside wild-type virus particles, as well as PR- particles and mutants mimicking the Pr55^{Gag} proteolytic cascade (**Figure 53**). This study allows us to follow the conformational rearrangements occurring within the first 550 nucleotides of the HIV-1 genome during maturation of the viral particles. In addition, RNA-protein interactions are identified thanks to the AT-2 treatment, selectively disrupting protein interactions mediated by the NC domain. This AT-2 treatment has been performed on mature and immature PR- viral particles. As this compound was dissolved in DMSO, addition of DMSO without AT-2 was used as negative control.

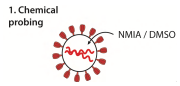
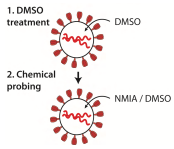
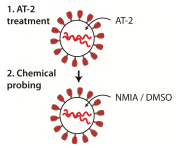
	No treatment	**DMSO**	**AT-2**
			
Immature PR-	0 / 3.9 mM NMIA 1 biological replicate	0 / 3.9 mM NMIA 2 biological replicates	0 / 3.9 mM NMIA 1 biological replicate
Step 1	0 / 3.9 mM NMIA 1 biological replicate		
Step 1.1	0 / 3.9 mM NMIA 1 biological replicate		
Step 2	0 / 3.9 mM NMIA 1 biological replicate		
CA/Sp1	0 / 3.9 mM NMIA 1 biological replicate		
Mature	0 / 3.9 mM NMIA 1 biological replicate	0 / 3.9 mM NMIA 3 biological replicates	0 / 3.9 mM NMIA 1 biological replicate

Figure 53
Recapitulating list of samples composing the Illumina library 3.
The Illumina library 3 has been generated to study the proteolytic cascade of Pr55^{Gag} and contains PR- and mature samples as well as the processing intermediates of the cascade, Step 1, Step 1.1, Step 2 and CA/Sp1. Viral gRNA was modified (NMIA with a final concentration of 3.9 mM in DMSO) inside virus particles for all samples. A negative control was systematically performed with DMSO only. In addition, PR- and mature particles have been incubated prior to chemical probing with AT-2 (final concentration of 0.5 mM in DMSO) or with DMSO alone as a negative control.

III.1 Reproducibility of hSHAPE-Seq results and impact of the DMSO treatment

Biological replicates of mature (n=3) and PR- immature (n=2) samples were generated by hSHAPE-Seq. The averaged reactivity profile is presented in **Figure 54** and **55**. Variability within the samples is highlighted by the standard error of the mean (SEM). In addition, the averaged gRNA reactivity profile of particles treated with DMSO is compared with untreated mature and PR- immature samples, obtained in absence of treatment (n=1). Indeed, the DMSO-treated condition is the negative control of the AT-2-treated condition but the mutants mimicking the $Pr55^{Gag}$ proteolytic cascade have not been treated with DMSO and were compared with untreated PR- and mature particles. Thus, the comparison of DMSO-treated and untreated conditions is required to further analyse results.

Slight differences are observed when the intensity of reactivity is compared between the untreated and DMSO-treated conditions. The same reactive sites are globally found when treating or not the virus particle with DMSO prior to performing chemical probing except four sites in mature samples (**Figure 54**) and three sites in PR- immature samples (**Figure 55**). In the case of mature samples, nucleotides C150-G158 and G406-U427 are reactive in DMSO-treated particles but not reactive in untreated particles. The opposite is observed for nucleotides G317-G320 and G325-U341. Positions C150-G158 also exhibit a higher reactivity in DMSO-treated samples in the case of PR- samples whereas nucleotides G317-G320 and C343-G348 are less reactive than noticed in untreated particles.

Therefore, these results highlight the reproducibility of the biological samples and of the hSHAPE-Seq technique. In addition, the DMSO does not seem to globally influence the nucleotide accessibility, except at some localised regions where a difference is noticed. It is important to remember that these differences are observed when comparing the average of replicates for the DMSO-treated condition and a single replicate of the untreated condition. Therefore, these changes could be reduced by producing new biological replicates for the untreated condition.

Regarding reproducibility, some sites with a characteristic reactivity profile are highlighted in **Figure 54** and **55**. Larger SEMs are noticed at positions where the reactivity is mostly superior to 1 so corresponding to highly modified positions. In addition, these replicates are biological and independent regarding the virus production, therefore could explain the observed variability.

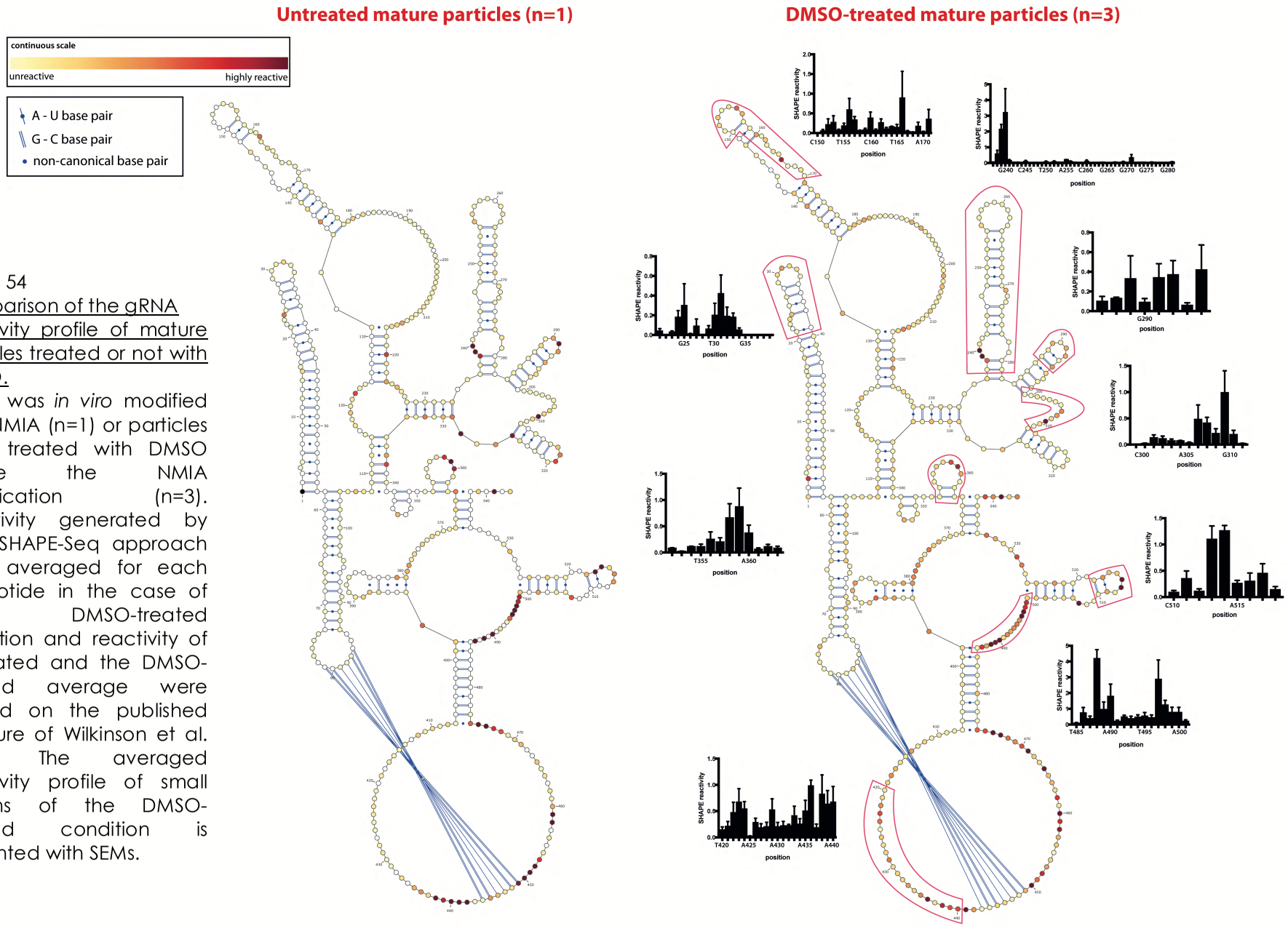


Figure 54
Comparison of the gRNA reactivity profile of mature particles treated or not with DMSO.
gRNA was *in viro* modified with NMIA (n=1) or particles were treated with DMSO before the NMIA modification (n=3). Reactivity generated by the hSHAPE-Seq approach were averaged for each nucleotide in the case of the DMSO-treated condition and reactivity of untreated and the DMSO-treated average were placed on the published structure of Wilkinson et al. (426). The averaged reactivity profile of small regions of the DMSO-treated condition is presented with SEMs.

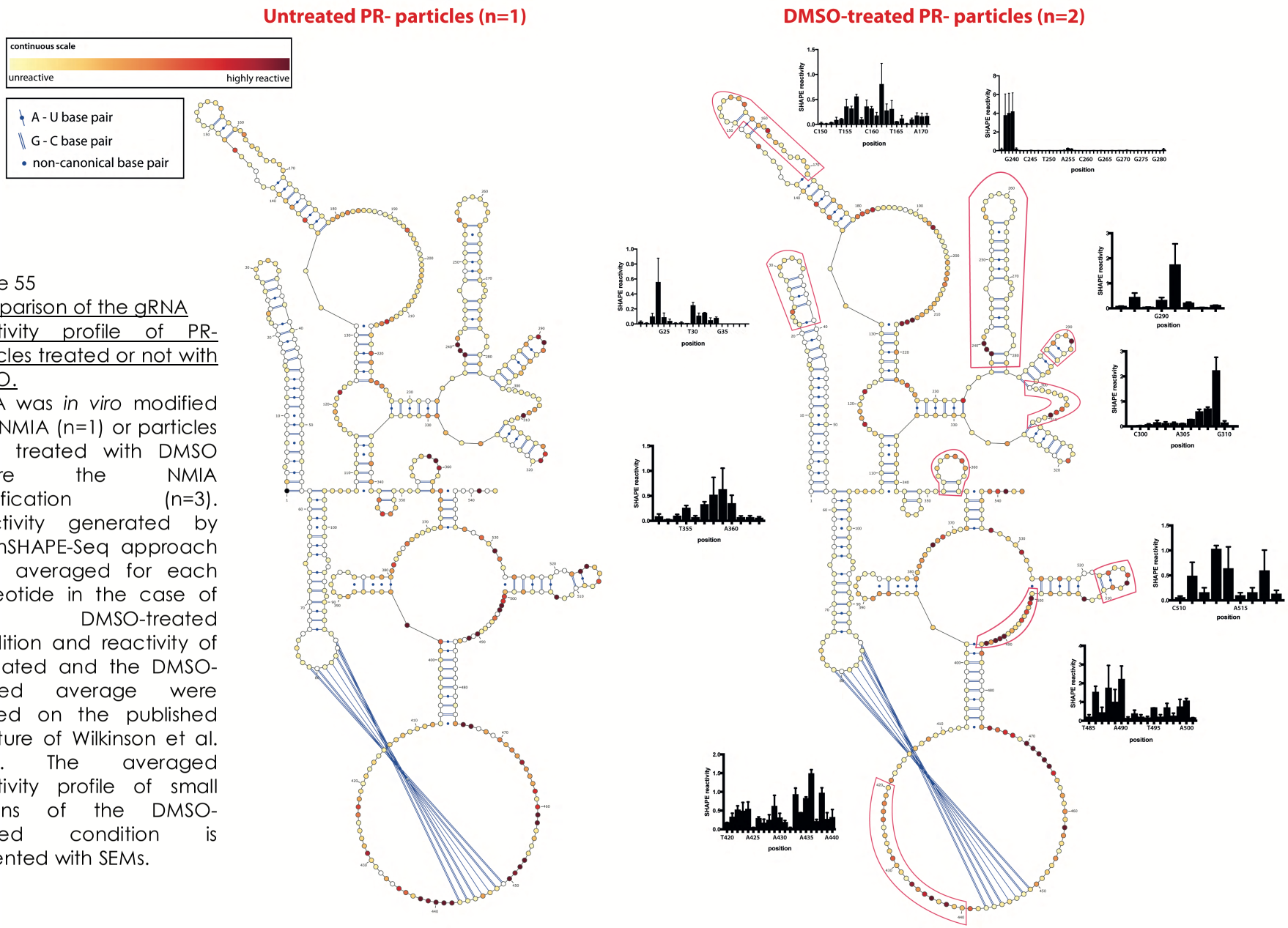


Figure 55
Comparison of the gRNA reactivity profile of PR-particles treated or not with DMSO.
gRNA was *in viro* modified with NMIA (n=1) or particles were treated with DMSO before the NMIA modification (n=3). Reactivity generated by the hSHAPE-Seq approach were averaged for each nucleotide in the case of the DMSO-treated condition and reactivity of untreated and the DMSO-treated average were placed on the published structure of Wilkinson et al. (426). The averaged reactivity profile of small regions of the DMSO-treated condition is presented with SEMs.

III.2 General overview of the structure of the 5'-UTR and the first 200 nucleotides of the gag coding region

A distinct difference in the levels of SHAPE reactivity is observed between the regulatory and coding regions with the inflection point coinciding with the AUG start codon of *gag* (**Figure 56 A**). The mean SHAPE reactivity is an indication of the amount of structure in the 5'-UTR and the first 200 nucleotides of the coding region. In the mature gRNA, the mean value is 0.12 for the 5'-UTR and 0.37 for the coding region (**Figure 56 B**). Thus, the hSHAPE-Seq approach shows the regulatory domain more structured compared to the coding region. This difference is clearly highlighted when using a 100-nucleotide rolling window, with a significant increase of the SHAPE reactivity from the AUG start codon. In the case of the immature PR- gRNA, the mean SHAPE reactivity is of 0.16 in the 5'-UTR and 0.33 in the coding region, consistent with observations made for the mature gRNA (**Figure 56 B**). However, the reactivity profile of the 5'-UTR is not as flat as in the mature gRNA with some flexible regions displaying higher reactivity (around nucleotides 200-250). Overall, a strong and significant difference in the amount of secondary structure is observed in these two regions and some differences in the level of reactivity are observed within the first 550 nucleotides of the gRNA in the mature and immature viral particles. The hSHAPE-Seq technique clearly distinguishes between regulatory and coding regions within the HIV-1 genome with the non-coding regulatory domain found to be more structured than the coding sequence. These results are consistent with previously published results of the first 900 nucleotides of the HIV-1 genome in the mature conformation from the team of K. Weeks (426). These results were obtained by hSHAPE using capillary sequencing, a conventional method.

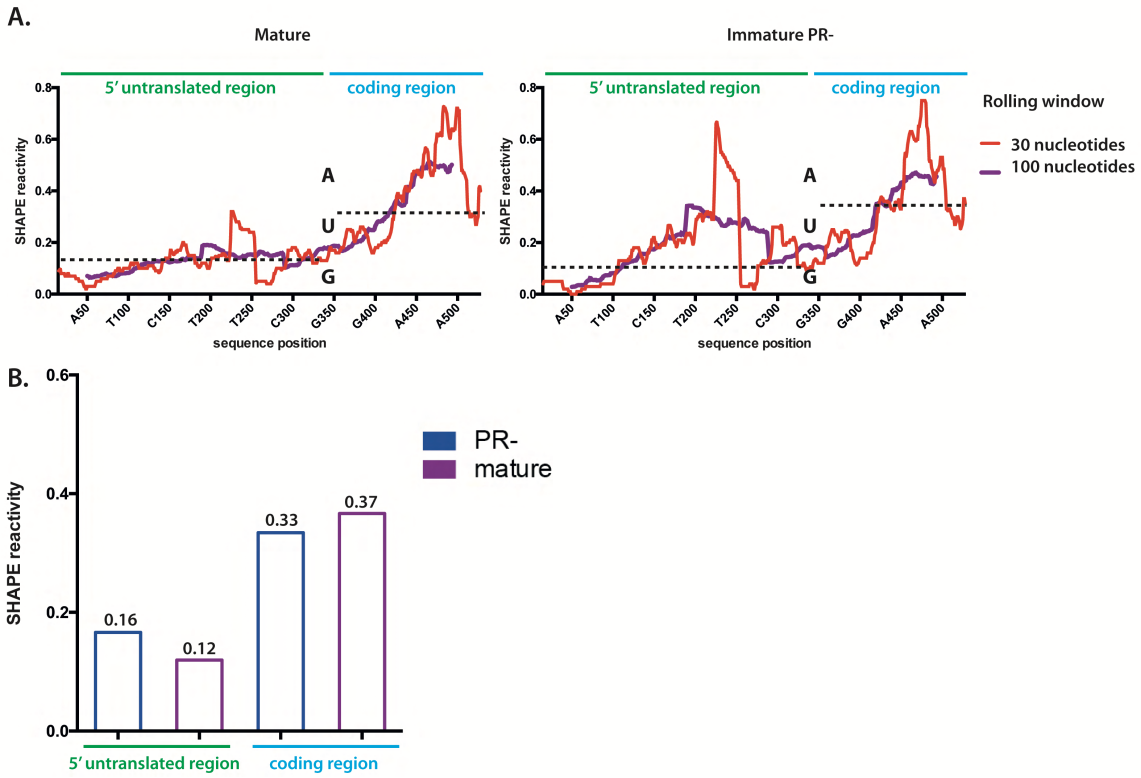


Figure 56
Global architecture of the first 550 nucleotides of mature and immature PR- gRNAs.
A. Rolling windows of 30 or 100 nucleotides show the differences in mean reactivity of the 5'-UTR and the beginning of the *gag* coding region in mature and immature gRNAs. The mean of each region is represented with dashed lines. **B**. The mean of nucleotide reactivity in these two regions is compared in mature and immature gRNAs

III.3 Evolution of the gRNA conformation during maturation of the viral particles

The evolution of the reactivity profile of the HIV-1 genome during the proteolytic processing of Pr55^{Gag} and the protein binding sites are detailed and discussed in this part of the manuscript. For the sake of simplicity of the figures, the overall structure has been split into small domains. Reactivity intensities obtained by hSHAPE-Seq have been placed on the secondary structure model published in 2008 by the team of K. Weeks (426) (**Figure 58**). Of note, more than 25 different secondary structure models of the HIV-1 gRNA 5' region have been proposed (28, 44, 91, 104, 175, 177, 191, 220, 269, 288, 319, 375, 383, 416, 422, 426).

The reactivity profile of the first 550 nucleotides of gRNA analysed by hSHAPE-Seq of untreated virus particles is presented in **Annexes 1** to **6**, DMSO-treated PR- particles in **Annexes 7** and **8**, the average of the two replicates of DMSO-treated PR- particles in **Annexes 9**, AT-2-treated PR- particles in **Annexe 10**, DMSO-treated mature particles in **Annexes 11** to **13,** the average of the three replicates of DMSO-treated mature particles in **Annexes 14** and AT-2-treated mature particles in **Annexe 15**.

Nucleotides are colour-coded according to their reactivity using a continuous scale (**Figure 57**). This scale goes from unreactive nucleotides (SHAPE reactivity of 0) represented in white to highly reactive positions coloured in dark red (SHAPE reactivity superior to 1).

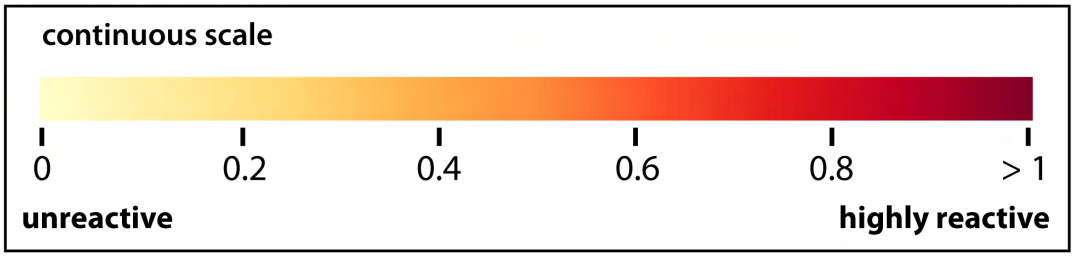


Figure 57
Description of the continuous scale representing nucleotide reactivity.
Nucleotides are colour-coded based on their SHAPE reactivity from not reactive in white to highly reactive in red.

The following pages are dedicated to the comparison of the gRNA reactivity profile between the mutants mimicking the Pr55^{Gag} proteolytic cascade, as well as PR- and mature samples. Similarities and major differences will be further discussed for each domain of the first 550 nucleotides of the HIV-1 gRNA. These domains are localised in **Figure 58**.

In addition, some results obtained by the *in vitro* approach are directly compared to *in viro* results (presented in black boxes). Indeed, the effect of the Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7 footprint to the 5' region of the HIV-1 gRNA is related to the mutants containing these intermediates of maturation.

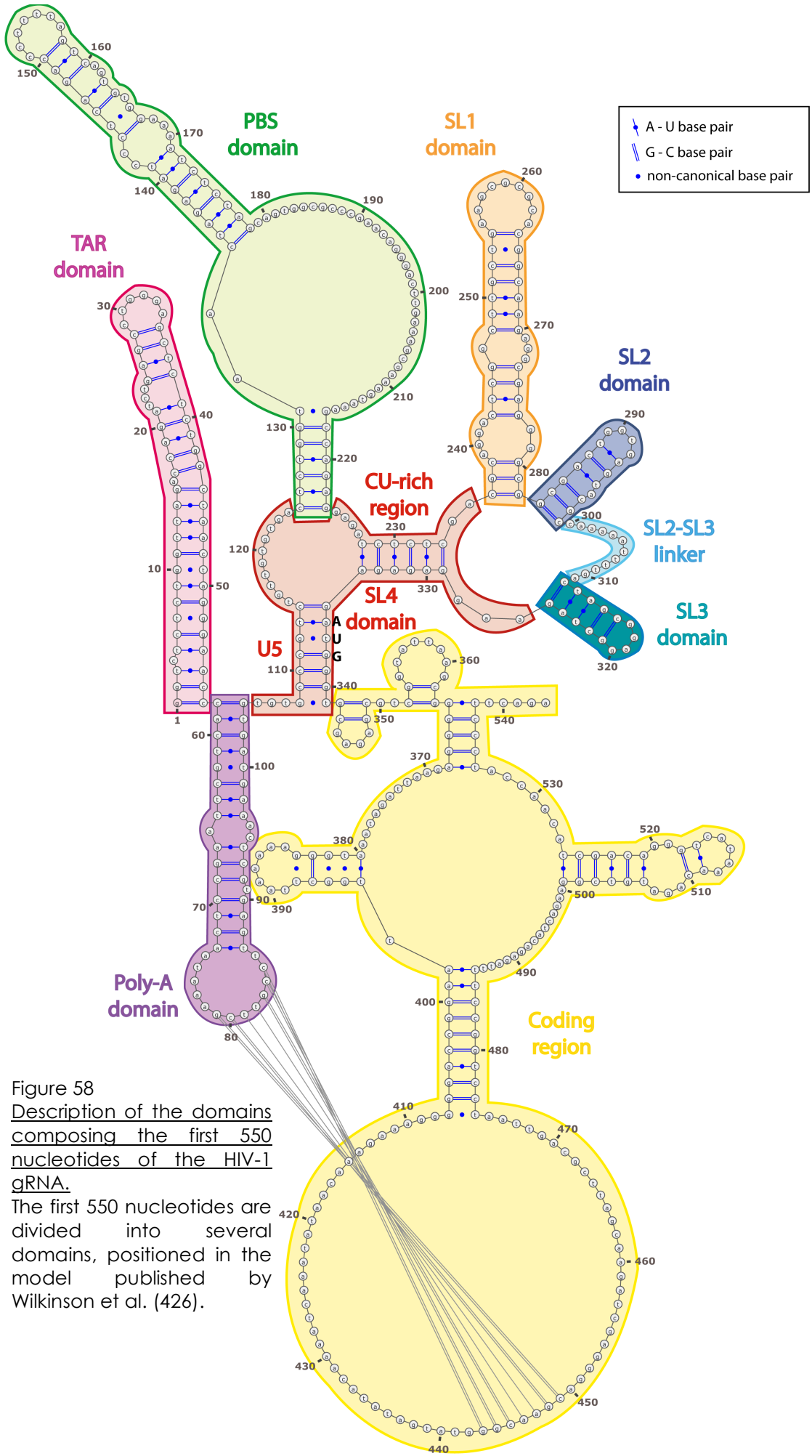


Figure 58
Description of the domains composing the first 550 nucleotides of the HIV-1 gRNA.
The first 550 nucleotides are divided into several domains, positioned in the model published by Wilkinson et al. (426).

III.3.1 The TAR domain

TAR is the first domain at the 5' end of the 5'-UTR and is 57 nucleotides long. At the mature stage, this domain forms a stem-loop with hSHAPE-Seq reactivity intensities consistent with the consensus secondary structure model for this region with two 1-nucleotide bulges (C4 and A16), the 3-pyrimidine bulge (U22-U24) and an apical loop containing 6 nucleotides (C29-A34) (**Figure 59 F**). Indeed, the mature structure of this stem-loop has been characterised *in vitro* (39, 42, 104), *ex viro* (383, 422), *in viro* (319, 426) and is constant in published models. The 3-nucleotide bulge has an intermediate reactivity and is consistent with the partially stacked nature of these nucleotides determined by nuclear magnetic resonance (NMR) (339). This structure of this stem-loop does not evolve during the cleavage cascade of Pr55^{Gag}, with similar reactivity profiles from the immature (**Figure 59 A**) to the mature stage (**Figure 59 F**).

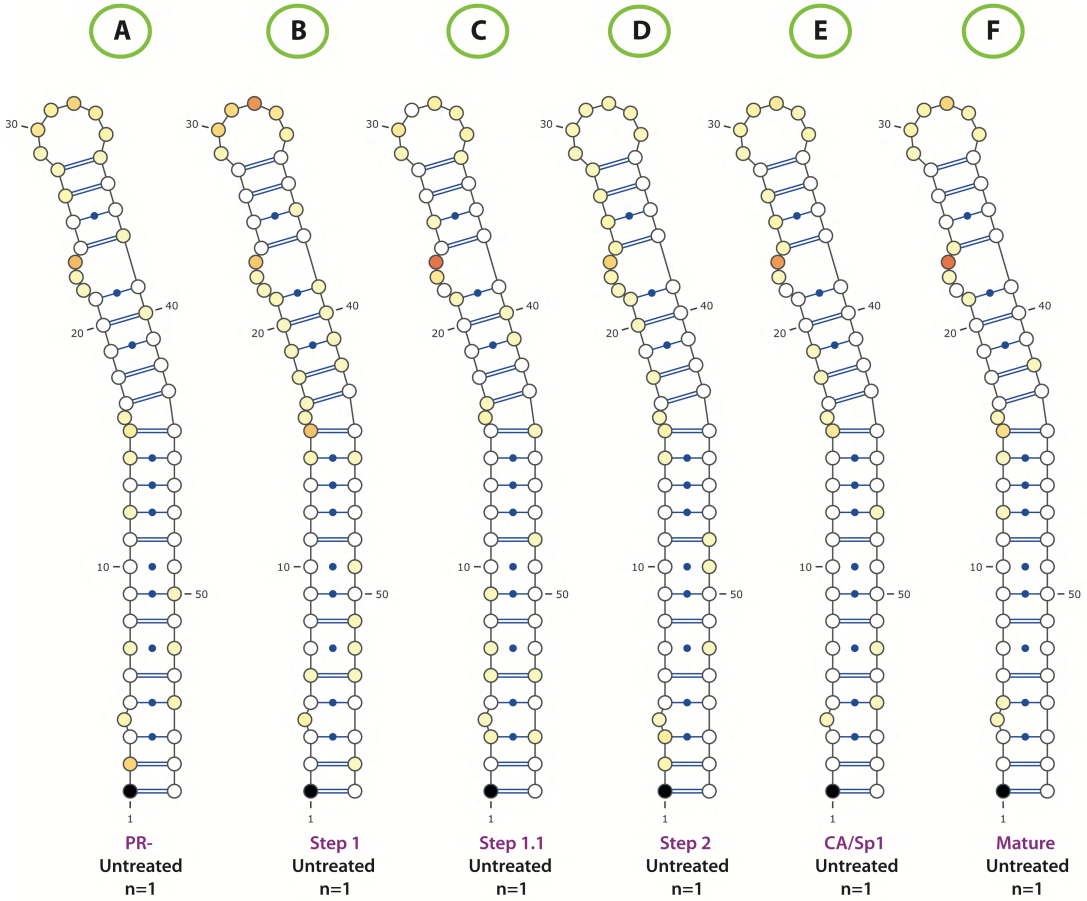


Figure 59
Evolution of the TAR domain of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivities, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples are placed on a published model (426).

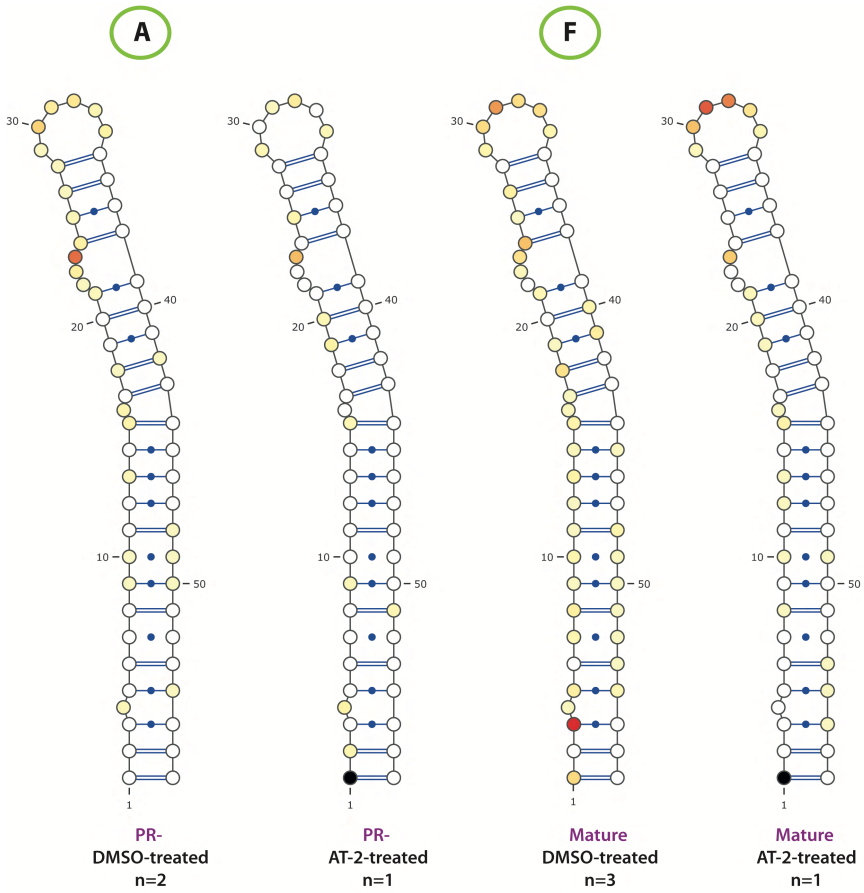


Figure 60
Effect of the AT-2 treatment on the TAR domain of HIV-1 gRNA from PR- and mature particles.
SHAPE reactivities, obtained by hSHAPE-Seq, of PR- and mature samples are placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

Regarding the apical loop, positions C29-A34 are poorly reactive in our hSHAPE-Seq experiments, independently of the maturation stage of the viral particle (**Figure 59 A** to F). In addition, the reactivity profile of nucleotides G31-G32 significantly increases upon AT-2 treatment in mature particles (**Figure 60 F**) whereas does not evolve in PR- particles (**Figure 60 A**). The apical loop has been proposed to be implicated in TAR-TAR kissing-loop inter-strand interaction. Indeed, several mutational analyses have suggested a base-pairing interaction between apical loops (19, 326, 389) but this interaction is not demonstrated *ex vivo* and no evidence for TAR-TAR interaction was found by mutational destabilisation and compensatory mutations (202). In addition, the noticed increase of reactivity in AT-2-treated mature particles is not in favour of a TAR-TAR interaction, which seems to be protected by a protein containing ZFs. The apical loop could thus be protected by NCp7 or a cellular factor in the mature gRNA conformation, which is not the case in PR- particles.

The tri-nucleotide bulge element is not evolving from the immature to the mature stage (**Figure 59 A to F**), with no influence as well of the AT-2 treatment (**Figure 60 A** and **F**). This bulge is described as the binding site of Tat (121, 122), with the basic domain of Tat mediating this interaction (339, 370). In addition, several cellular factors have been shown to interact with the 3-nucleotide bulge and/or the apical loop in order to trans-activate the 5'-LTR and thereby increase viral RNA transcription (126, 157, 176, 349, 380) and later to initiate the translation (reviewed in (314)).

However, these viral and cellular factors involved in transcription and translation do not seem to be incorporated inside virus particles and therefore they are not expected to impact the gRNA structure (reviewed in (66)) (80, 266).

III.3.2 The Poly-A domain

The Poly-A domain adopts a conserved stable secondary structure (41, 108, 236). Consistently, the 47 nucleotides composing the Poly-A stem-loop are poorly reactive in our hSHAPE-Seq experiments, independently of the maturation stage of the viral particles (**Figure 61 A to F**). This is consistent with published models except for positions A76-A78 reported as highly reactive (319, 375, 383, 422, 426). Positions A76-A78 are part of the AAUAAA polyadenylation signal localised in the apical loop (41).

This stem-loop has been proposed to be paired with the SL1 domain by the team of B. Berkhout. In this model, the 5'-UTR structure is an equilibrium between the LDI structure with the poly-A and SL1 domains paired and the BMH structure with the Poly-A and SL1 domains folded separately (191). This model has been further detailed in the review article "The Life-Cycle of the HIV-1 Gag–RNA Complex" published in "Viruses". Nevertheless, this LDI-BMH is not supported by *ex vivo* and *in vivo* chemical probing of HIV-1 gRNA (319, 383, 422, 426).

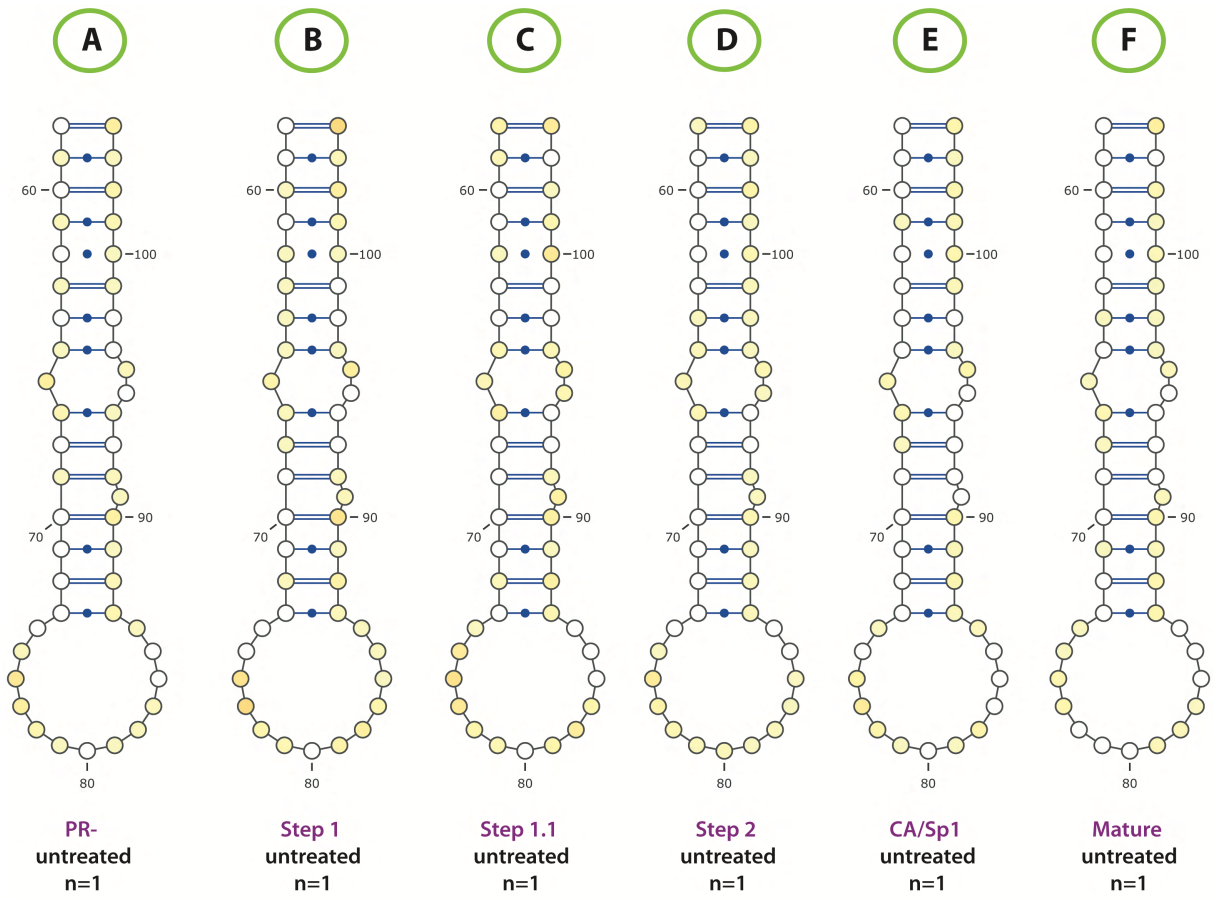


Figure 61
Evolution of the Poly-A domain of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
hSHAPE-Seq reactivity intensities, of PR-, intermediates and mature samples are placed on a published model (426).

Concerning the apical loop, positions G79-C85 and G443-C449 are poorly reactive (**Figure 61 A to C** and **62 A**). Nucleotides C445-A446 and C449 exhibit a higher reactivity from the Step 2 to the mature stage (**Figure 61 D to F** and **62 F**), consistent with data from Wilkinson et al. on the mature virus (426). Interestingly, the 4 nucleotides flanking the G443 to C449 region are highly reactive at all processing steps except Step 1 and Step 1.1 (**Figure 61 A to F**). Whether this fact is biologically relevant is difficult to assess with only one replicate. In addition, the reactivity profile of these two sites is different when comparing data from PR- particles treated or not with DMSO (**Figure 62 A** and **63 A**), with one of the two DMSO-treated PR- particles replicate consistent with untreated samples.

Interestingly, hSHAPE-Seq results correlate with the proposed involvement of the Poly-A stem-loop in a long-range interaction. This pseudoknot implicates the Poly-A apical loop (positions G79-C85 in NL43 isolate) and nucleotides in the matrix coding region (positions G443-C449 in NL-43 isolate). The heptanucleotide sequence has been identified *in vitro* by mobility-shift assay, enzymatic probing, ladder selection and antisense oligonucleotide mapping in the Mal isolate (324) and confirmed by chemical probing performed *ex viro* (383, 422) as well as *in viro* (426) in the NL-43 isolate.

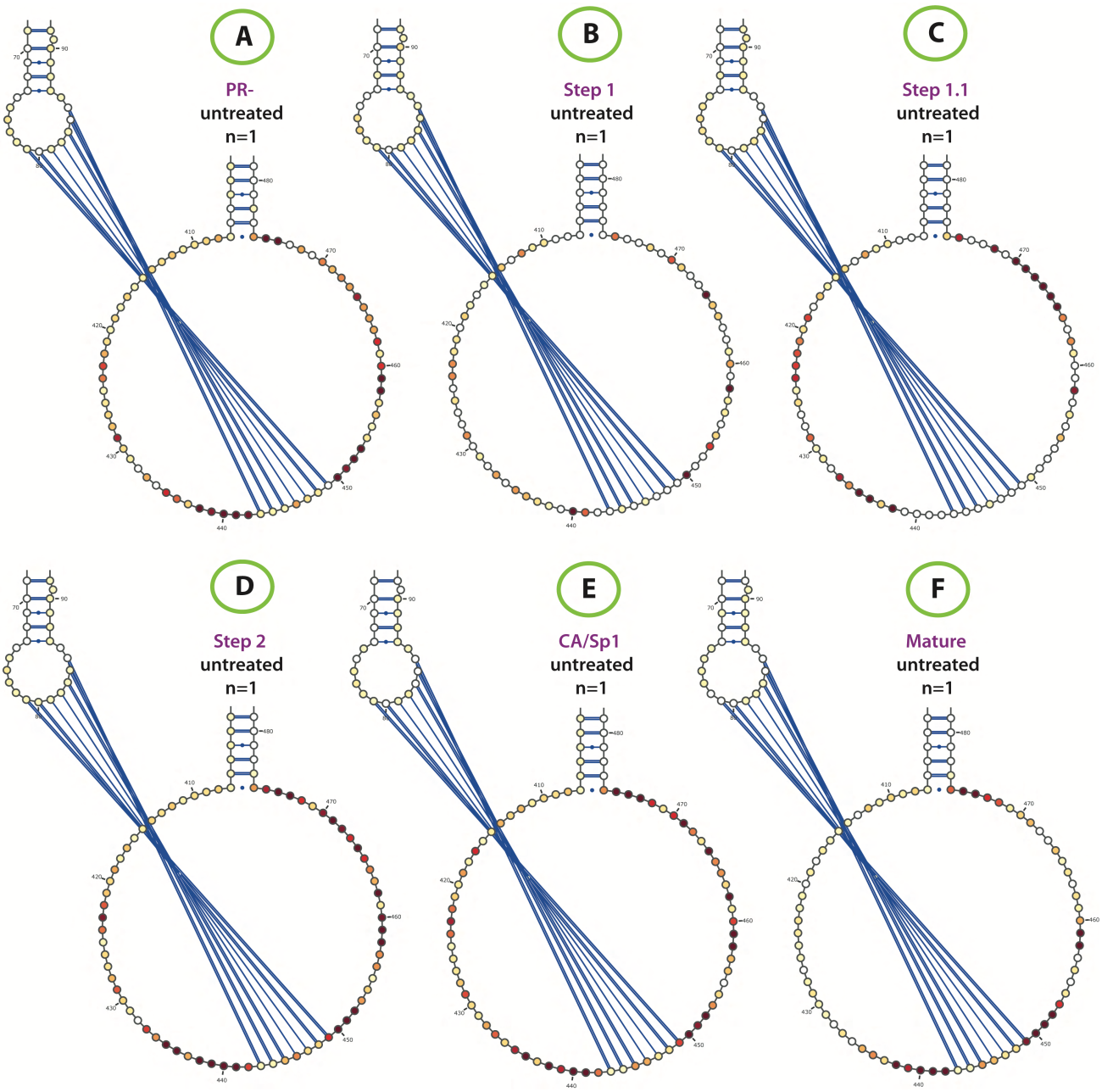


Figure 62
Evolution of the pseudoknot implicating positions G79-C85 and G443-C449 of HIV-1 gRNA during the proteolytic cascade of $Pr55^{Gag}$.
SHAPE reactivity intensities, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples are placed on a published model (426).

Regarding the AT-2 treatment, no difference was observed in the Poly-A domain between DMSO and AT-2 treated PR- and mature particles (**Figure 63 A** and **F**). These results are in contradiction with Wilkinson et al. reporting a structure-destabilising activity of NCp7 at positions U75-G79 (426) and Kenyon et al. reporting a $Pr55^{Gag}$ binding site (225).

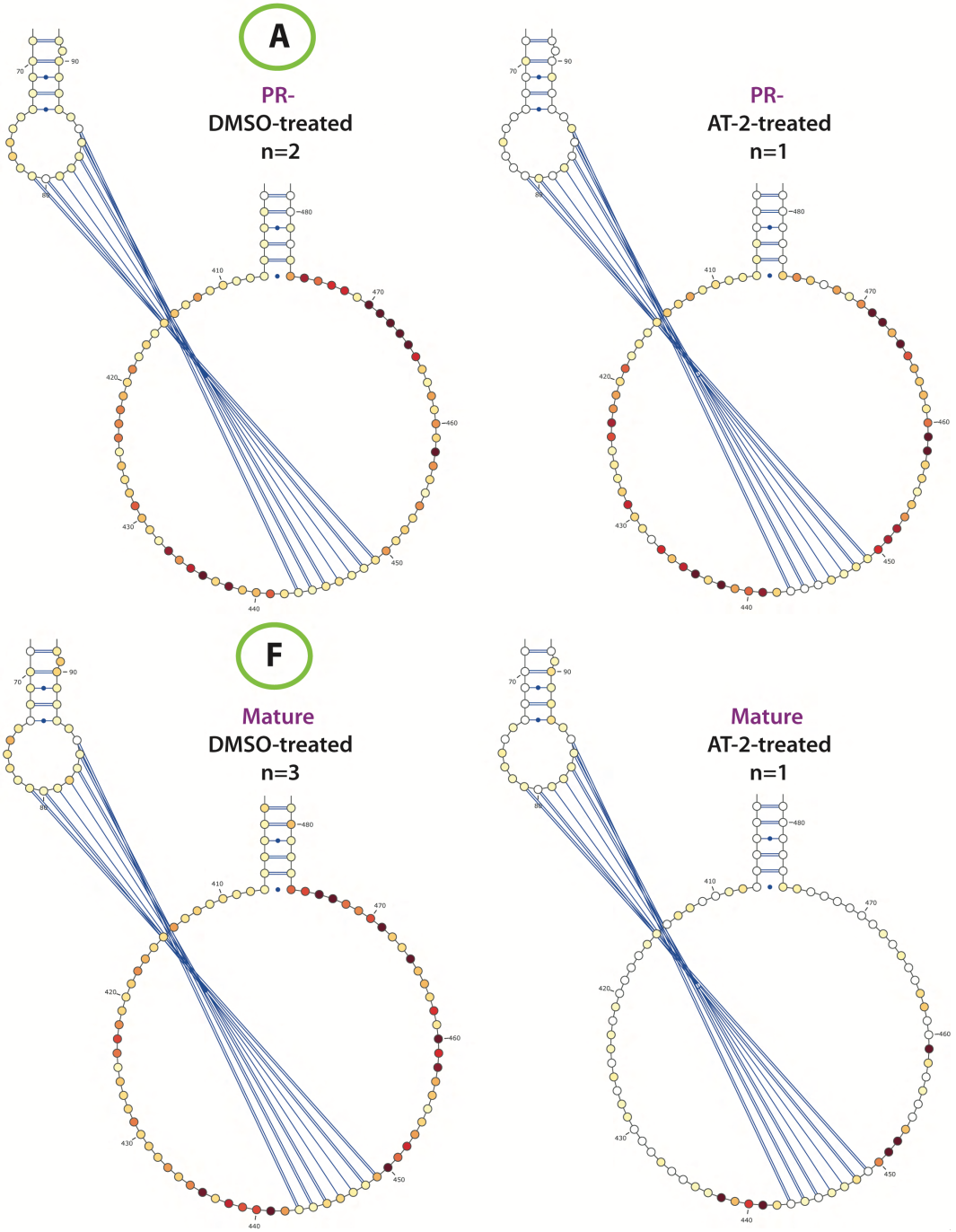


Figure 63
Effect of the AT-2 treatment on the pseudoknot implicating positions G79-C85 and G443-C449 of the HIV-1 gRNA from PR- and mature particles.
SHAPE reactivity obtained by hSHAPE-Seq, of PR- and mature samples is placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

The limited reactivity of the polyadenylation signal is in line with the recent finding by our laboratory that a functional 5' polyadenylation signal is required for optimal packaging of the gRNA (Smyth et al., in preparation). Indeed, it seems that recruitment of (at least part of) the cellular cleavage and polyadenylation machinery favours gRNA packaging and these proteins are therefore likely incorporated into viral particles, where they might protect the polyadenylation signal

from chemical modification. In addition, the absence of effect upon AT-2 treatment on the polyadenylation signal coincides with the hypothetical protection mediated by cleavage and polyadenylation specificity factor (CPSF), with no ZF structure (**Figure 63 A** and **F**).

III.3.3 The PBS domain

A decreased reactivity is observed for some of the 18 nucleotides (positions U182-C199) constituting the PBS during the maturation process. Nucleotides U182, G184, C193 and A194 are highly modified in PR- particles treated or not with DMSO (**Figure 64 A** and **65 A**) and become unreactive from the Step 1 stage to the completion of maturation (**Figure 64 B to F**). These results correlate with the formation of an internal stem-loop in the un-annealed PBS site, observed *in vitro* (4, 375) and *in viro*, in PR- particles (375) with G190-A194 corresponding to the apical loop. In mature particles (**Figure 64 F**), positions U182-C199 of the PBS involved in the annealing the 3' end of the $tRNA^{Lys,3}$ stem acceptor are unreactive, indicating the stable annealing of the $tRNA^{Lys,3}$ to these 18 nucleotides (262, 281). Localisation of the Acceptor stem of the $tRNA^{Lys,3}$ and the three additional sites interacting with the PBS domain is highlighted in **Figure 5** (in the Introduction part II.2.4.2 $tRNA^{Lys,3}$ annealing and packaging).

In addition, positions C179-U182 and A214-A216 located at each end of the PBS large loop are highly reactive in PR- treated or not with DMSO (**Figure 64 A** and **65 A**). The reactivity profile of these nucleotides is similar in processing intermediates (**Figure 64 B to E**) and significantly decreased in mature particles treated or not with DMSO (**Figure 64 F** and **65 F**).

The AT-2 treatment reveals several reactivity changes in the PBS loop (nucleotides C179-A216) with reduced reactivity at positions C179-U182, A214-A216, C193-A194 in PR- particles (**Figure 65 A**) whereas the reactivity profile at positions G181-U182 and A214 is increased in mature particles (**Figure 65 F**). These results indicate a conformational rearrangement of the PBS loop when ZF RNA-binding proteins, probably $Pr55^{Gag}$, are dissociated and a structural stabilisation of the PBS loop mediated by NCp7.

Regarding the other regions of the PBS domains interacting with $tRNA^{Lys,3}$, the C-rich (positions C142-A147) and A-rich (positions G167-A170) loops are unreactive in PR- mutant (**Figure 64 A** and **65 A**) and maturation intermediates (**Figure 64 B to E**). These regions are unreactive as well in mature particles (**Figure 64 F** and **65 F**), respectively because the variable loop of $tRNA^{Lys,3}$ interacts with the C-rich region (194, 196) and the anti-codon loop is complementary to the A-loop (160, 193, 194, 196).

Upon AT-2 treatment of PR- particles, the reactivity profile of the upper PBS stem decreases at positions A136, A138, A147, A157 and G162, indicating a conformational rearrangement of the upper stem when ZF RNA-binding proteins, probably $Pr55^{Gag}$, are dissociated (**Figure 65 A**). Interestingly, the reactivity profile of the C-rich (C142-A147) and the A-rich (G167-A170) loops is unchanged after AT-2 treatment and remains unreactive. This absence of reactivity is also noticed in maturation intermediates (**Figure 64 B to E**). In AT-2-treated mature particles, only positions G162 and G164, located in the loop 5' to the annealing site of the anti-codon loop of $tRNA^{Lys,3}$, are highly modified (**Figure 65 F**).

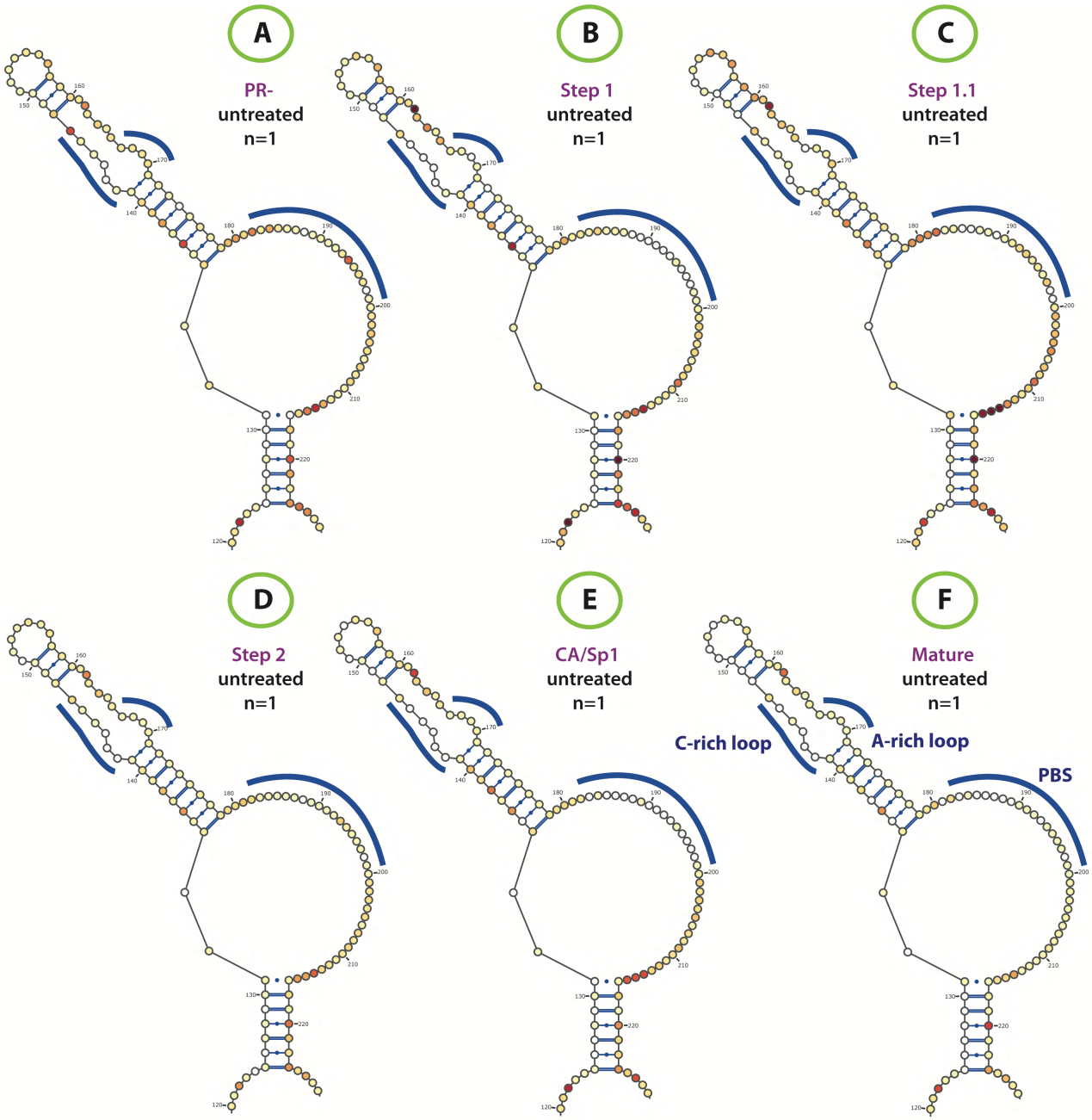


Figure 64
Evolution of the PBS domain of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples is placed on a published model (426).

In addition, a slight decrease of reactivity is noticed at positions U156 and A157 after AT-2 treatment of both PR- and mature particles implicating a destabilisation effect of Pr55^{Gag} and NCp7 at this site (**Figure 65 A** and **F**). This effect by NCp7 has been previously reported (426), strengthening these results. However, modification of U156 and A157 is weaker in untreated virus particles (**Figure 64 A to F**). Nucleotides G181, U182, G202, G208 and A214 have an increased reactivity as previously reported (375) and indicate NCp7 binding sites, not described in Wilkinson et al. (426).

To briefly summarise these results, the PBS is less tightly bound by the $tRNA^{Lys,3}$ stem acceptor in PR- particles and the optimal annealing of the PBS is achieved at the mature stage. These results correlate with multiple pieces of evidence suggesting that annealing can be achieved either by the NC domain within $Pr55^{Gag}$ (70, 138, 168, 186, 363, 364, 375) or by NCp7 (72, 115, 173, 375, 402, 426) and leading to an improved ability to induce RT with NCp7 (168, 364). These proteins probably help the annealing of the $tRNA^{Lys,3}$. In addition, the complementarity between the PBS sequence and the stem acceptor of $tRNA^{Lys,3}$ is not sufficient to efficiently initiate RTion and additional interactions between the $tRNA^{Lys,3}$ with the A-rich and C-rich loops and the HIV-1 gRNA being required (197, 263). Results obtained by the hSHAPE-Seq approach suggest that the variable loop and the anti-codon loop of $tRNA^{Lys,3}$ could potentially be annealed to the $tRNA^{Lys,3}$ immediately after gRNA packaging and viral assembly, with the stability of these interactions increasing during proteolytic processing of $Pr55^{Gag}$. Moreover, the destabilising activity of $Pr55^{Gag}$ on the upper stem seems required for the annealing of the variable loop and the anti-codon loop of $tRNA^{Lys,3}$. At the mature stage, NCp7 probably binds to positions G162 and G164 and increases binding stability of the variable loop and the anti-codon loop of $tRNA^{Lys,3}$. This structuration of the upper stem implicates the double-stranded region composed of nucleotides C134-U141 and A171-G178 which seems only perfectly stable at the final stage of maturation (**Figure 64 A to F** and **65 A** and **F**).

Our hSHAPE-Seq results show that nucleotides C218, A220, G221 and G223 are reactive in the PR- and maturation intermediate particles, (**Figure 64 A to E**). The reactivity of these nucleotides decreases from the Step 2 to the mature stage with only A220 remaining reactive (**Figure 64 F**). By contrast, nucleotides G123-G130 are not reactive in any maturation stage (**Figure 64 A to F**). Thus, nucleotides C218-G223 do not seem to be paired with positions C125-U131. Interestingly, AT-2-treated mature particles exhibit an increased reactivity of nucleotides A220, G221 and G223 suggesting the presence of an NCp7 binding site (**Figure 64 F**) whereas $Pr55^{Gag}$ does not seem to be present at this site (**Figure 64 A**). This absence of reactivity for positions C218-G223 in mature particles is thus misinterpreted when proposing a secondary structure model (319, 383, 422) and surprisingly NCp7 deprotection was not detected upon AT-2 treatment in the study by Wilkinson et al. (426).

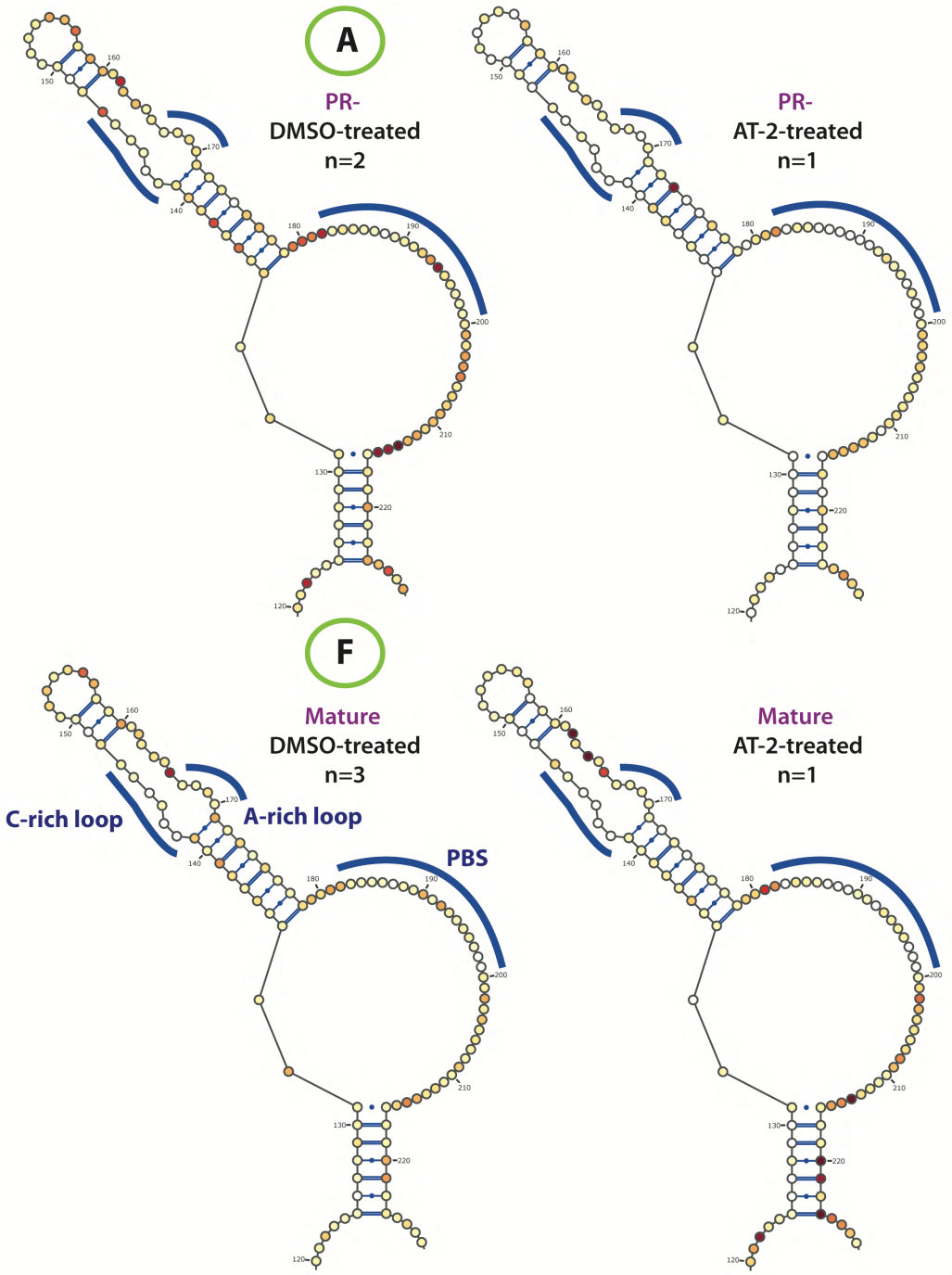


Figure 65
Effect of the AT-2 treatment on the PBS domain of HIV-1 gRNA from PR- and mature particles.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR- and mature samples is placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

The TΨC arm of $tRNA^{Lys,3}$, localised between the variable loop and the stem acceptor, has been proposed to interact with nucleotides G123-G130, an 8-nucleotide sequence in the U5 region termed the Primer Activation Signal (PAS). This interaction, proposed by the team of B. Berkhout (29–32), has been analysed by NMR (387), but its existence could not been conclusively demonstrated by probing experiments due to alternative nucleotide pairing possibilities (159, 160, 194). The team of B. Berkhout proposed a mechanism involving the PAS sequence to regulate

the RTion initiation (31). This sequence, initially base-paired in the U5 leader stem, anneals with the TΨC arm in the $tRNA^{Lys,3}$, stabilises primer annealing and stimulates RTion (2). Indeed, they describe this interaction as dynamic and favoured by NCp7 inducing a destabilisation of the hairpin. The timing of RTion initiation would therefore be controlled thanks to the proteolytic processing of $Pr55^{Gag}$ and the release of NCp7. In addition, the team of M. Summers shows by NMR that positions C127-G129 are paired with C219-G221 (220). However, our results suggest that the TΨC arm of the $tRNA^{Lys,3}$ already binds to the PAS in PR- immature particles. Thus, nucleotides C218-G223 do not seem to be paired with positions C125-U131.

Surprisingly, nucleotides C151-U154 located at the apical loop of the upper stem are noticed as unreactive independently of the maturation stage of the viral particles (**Figure 64 A to F**) as previously shown (319, 375, 383, 422, 426). Nevertheless, the reactivity profile of this potentially protein-binding site remains unchanged with an AT-2 treatment in the case of PR- and mature particles (**Figure 65 A** and **F**). Since $tRNA^{Lys,3}$ annealing has been suggested to be facilitated by the viral protein Vif (182, 341) and the cellular RNA helicase A (433), these proteins could potentially protect this apical loop.

III.3.4 The packaging signal

The packaging signal encompasses the SL1, SL2, SL3 and SL4 stem-loops (12, 43, 88, 91, 175, 185, 260, 320). The other domains composing the 5'-UTR have also been proposed for optimal gRNA packaging (90, 119, 179, 362).

III.3.4.1 SL1 domain

Two main higher-order structure models of the SL1 domain differing by their length have been proposed in published models of the 5' region of the HIV-1 gRNA, based on chemical probing studies:

- The short SL1 domain comprises positions C243-G277 (319, 368, 383). The short SL1 domain comprises the upper stem-loop containing the six nucleotide self-complementary sequence mediating gRNA dimerization (322, 386), an A271-G273 internal loop comprising nucleotides G247, A271, G272 and G273, and the in-between stem constituted by nucleotides C243-G246 base-paired to C274-G277. Smyth et al. study argues for the existence of a short SL1 domain, because the MIME technique suggests that $Pr55^{Gag}$ does not recognise the extended SL1 domain (388).

- The extended SL1 domain comprises positions C236-G282 (422, 426). The extended SL1 domain contains the short SL1 prolonged by an internal loop comprising nucleotides A239-A242 G278 and G279 and a lower stem containing nucleotides C236-C238 base-paired to G280-G282. An even longer SL1 domain comprising a few nucleotides of the SL2 domain (nucleotides G283-A286) has been proposed (174) but Sakaguchi et al. demonstrated that the 17-nucleotide SL2 structure is biologically functional, strongly suggesting that this longer SL1 structure does not exist (175, 365).

In mature particles treated or not with DMSO, the 6 self-complementary nucleotides mediating gRNA dimerisation are unreactive (**Figure 66 F** and **67 F**), which is expected since these nucleotides are involved in a kissing-loop interaction. This reactivity profile is similar in PR- viral particles treated or not with DMSO and in viral particles blocked at intermediate maturation stages (**Figure 66 A to F** and **67 A**). This

kissing-loop interaction is further supported by the absence of reactivity change upon AT-2 treatment (**Figure 67 A** and **F**).

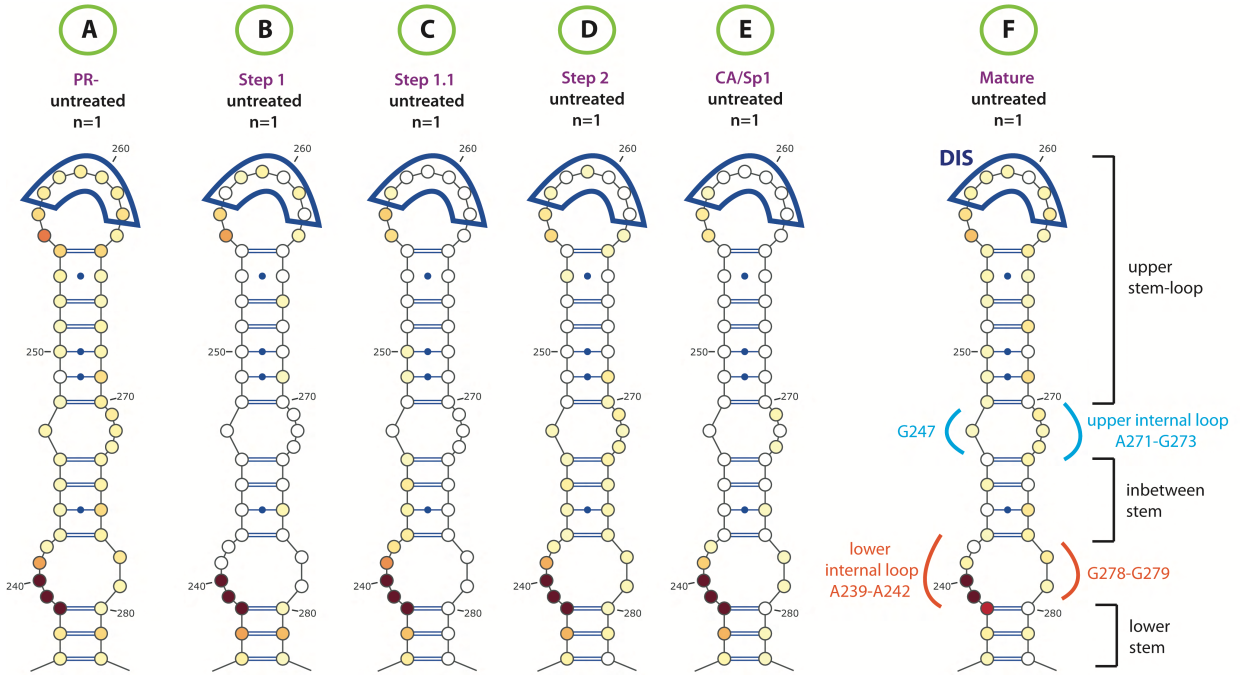


Figure 66
Evolution of the SL1 domain of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples is placed on a published model (426).

The self-complementary is preceded and followed by two and one purines, respectively A255-A256 and A263, in the apical loop. Positions A255 and A256 exhibit an intermediate reactivity from the immature to the mature stage of maturation (**Figure 66 A to F**) with a slightly higher reactivity profile for A255. Whereas AT-2 treatment does not impact the reactivity of A255 and A256 in PR- particles (**Figure 67 A**), an increased reactivity of A255 is noticed in mature particles (**Figure 67 F**). Nucleotide A263 appears unreactive from the immature to the fully mature stage (**Figure 66 A to F**), even after AT-2 treatment.

The lack of reactivity of A263 is in keeping with the fact that this purine is stacked inside the kissing-loop complex, as observed by X-ray crystallography (133). The reactivity profile of A255-A256 and A263 correlates well with published data obtained by chemical probing in solution (383, 422, 426), crystallography (133) and NMR (106, 220, 233, 234, 253, 301, 413), even if positions A255 and A256 have a higher reactivity. However, no NCp7 binding site was detected at these positions by Wilkinson et al. whereas hSHAPE-Seq data suggest NCp7 binding at this location (426). These three purine residues have been demonstrated to be required for efficient RNA dimerisation (325, 385).

In addition, our hSHAPE-Seq data show positions G254 and C264 weakly modified in PR- particles (**Figure 66 A**) whereas these nucleotides seem to be base-paired from the Step 1 stage (**Figure 67 B to F**). The team of Sakuragi recently proposed a larger hairpin loop with positions G254 and C264 unpaired. They postulated this extended apical loop to form thanks to the assistance of a chaperone, presumably NCp7 (368).

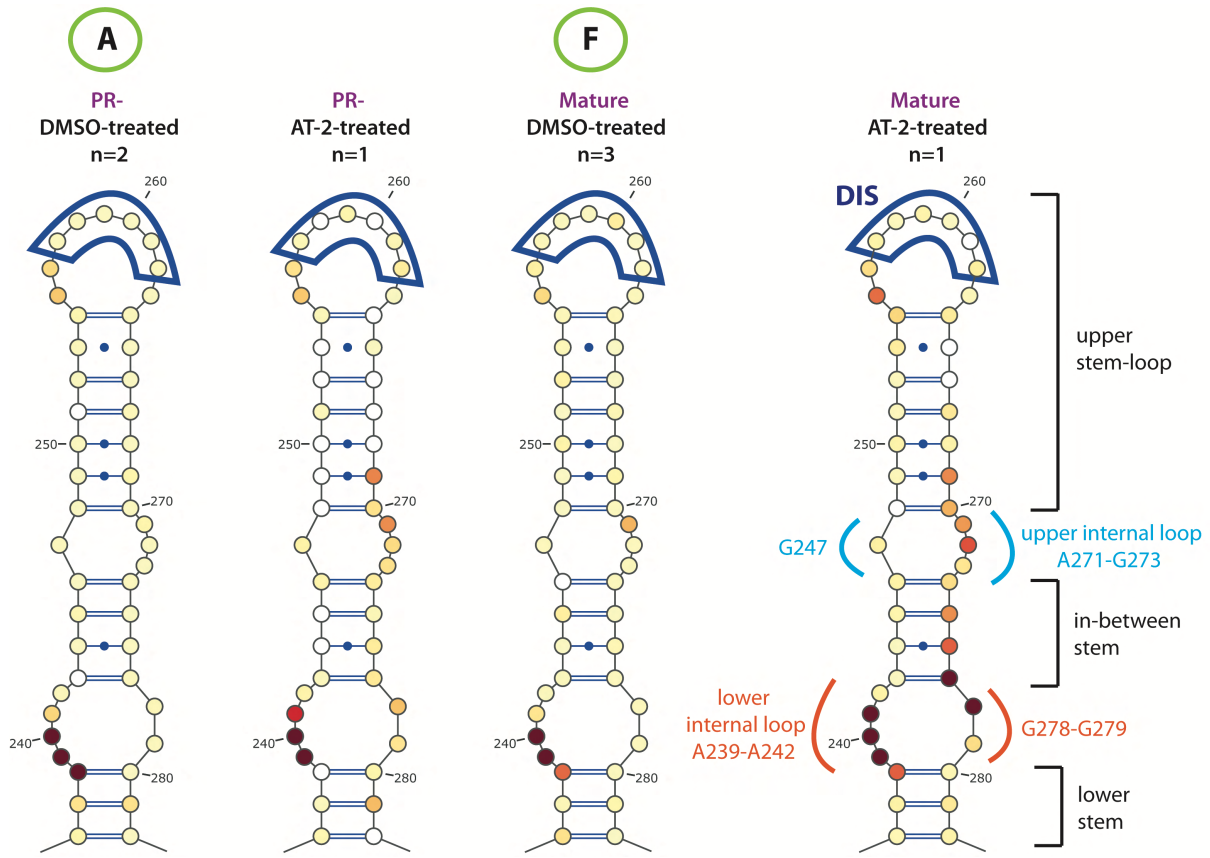


Figure 67
Effect of the AT-2 treatment on the SL1 domain of HIV-1 gRNA from PR- and mature particles.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR- and mature samples is placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

***In vitro* footprinting analysis of the U253-A263 region with Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7**

Nucleotides G257-C262, involved in gRNA dimerisation are not reactive with any of the different proteins and demonstrate that the gRNA is in the dimeric conformation (**Figure 68 A to D**). Pr55^{Gag} induces a strong protection of nucleotide A256 (**Figure 68 A**). This protection is not visualised *in viro* in PR- particles with no reactivity change of this position following an AT-2 treatment (**Figure 67 A**). Upon NCp15 and NCp9 binding, nucleotide A256 is still protected (**Figure 68 B** and **C**) and this protection is totally lost with NCp7 (**Figure 68 D**). These results correlate with the absence of reactivity change of nucleotide A256 *in viro* following AT-2 treatment of mature particles. In addition, nucleotide A255 is *in vitro* weakly protected by NCp7 and this protection correlates with the increased reactivity noticed *in viro* following AT-2 treatment. However, intensity of NCp7 protection is observed higher in viral particles (**Figure 67 F** and **68 D**). GagΔp6 induces a reduced protection of nucleotide A256 compared to Pr55^{Gag} and the protection of position A255 is lost (**Figure 68 A** and **E**). Taken together, *in vitro* footprint mediated by NCp15, NCp9 and NCp7 is representative of their behaviour in the viral particle with, however a slight variation of intensities, probably due to the difference of environment applied for *in vitro* and *in viro* experiments.

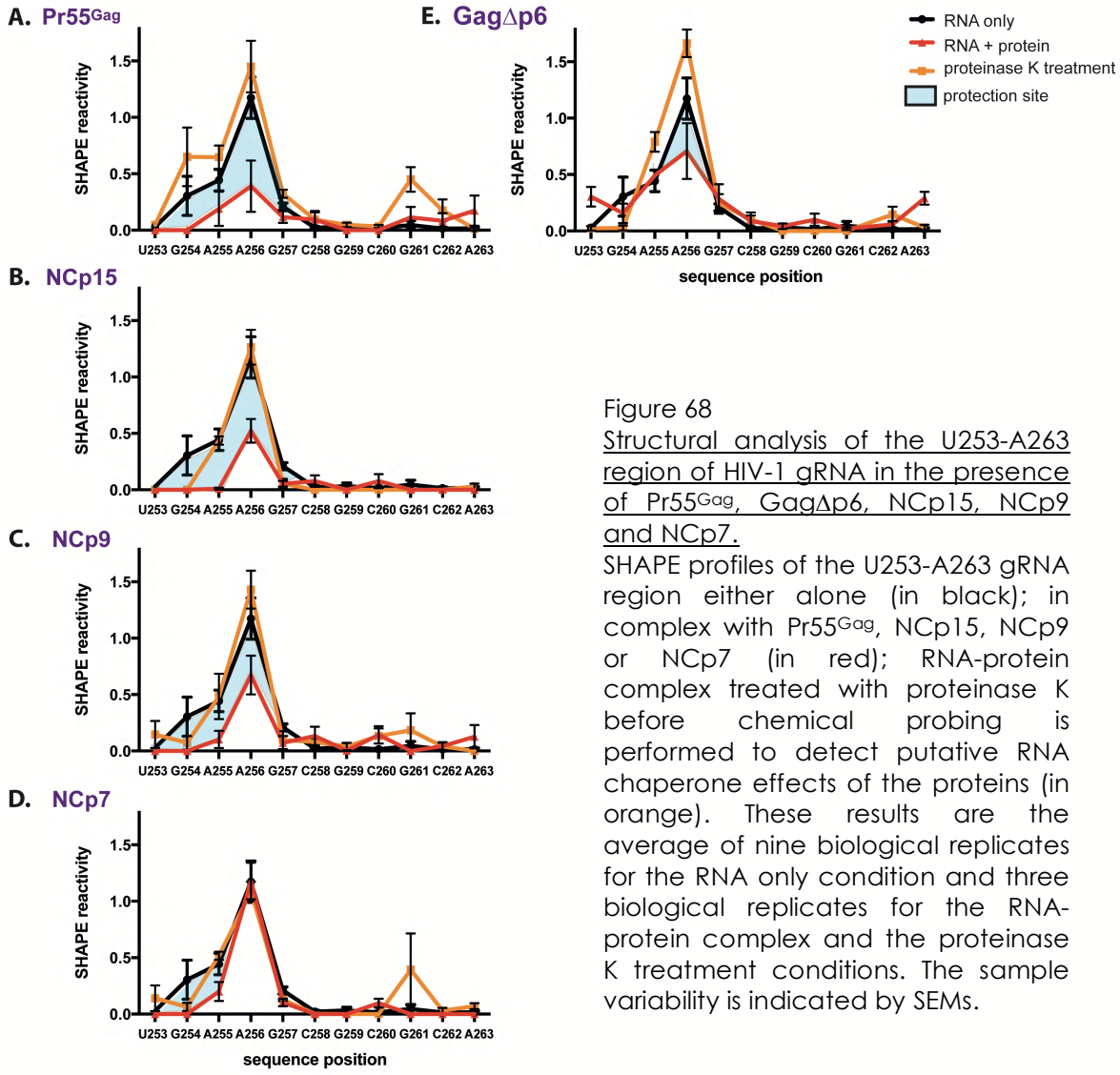


Figure 68
Structural analysis of the U253-A263 region of HIV-1 gRNA in the presence of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7.
SHAPE profiles of the U253-A263 gRNA region either alone (in black); in complex with Pr55^{Gag}, NCp15, NCp9 or NCp7 (in red); RNA-protein complex treated with proteinase K before chemical probing is performed to detect putative RNA chaperone effects of the proteins (in orange). These results are the average of nine biological replicates for the RNA only condition and three biological replicates for the RNA-protein complex and the proteinase K treatment conditions. The sample variability is indicated by SEMs.

Regarding the upper internal loop, positions A271-G273 are unreactive/weakly reactive independently of the maturation stage of the viral particle (**Figure 66 A to F**). Nucleotides A269-G272 becomes reactive after AT-2 treatment of PR- particles (**Figure 67 A**) and even more reactive in AT-2-treated mature particles (**Figure 67 F**). This deprotection induced by AT-2 was confirmed when SHAPE was analysed by PAGE on PR-, Step 1.1 and mature particles (**Figure 69 A, B** and **C**). These results highlight the presence of a ZF RNA binding protein, demonstrated to be $Pr55^{Gag}$ in immature PR- particles (4, 48, 388) and NCp7 (426) in mature particles. Although SL3 has historically being considered as the primary $Pr55^{Gag}$ binding site leading to gRNA packaging, our team determined that $Pr55^{Gag}$ primarily binds to the SL1 domain. A variety of biophysical, biochemical and footprinting assays has been used to demonstrate that $Pr55^{Gag}$ efficiently binds to the upper part of the SL1 domain, and in particular its internal loop (4, 48). The MIME technique confirmed that $Pr55^{Gag}$ binding was significantly impaired when point substitutions were introduced in the SL1 domain (388). Recently, Sakuragi et al. *ex vivo* demonstrated the importance of nucleotides G247 and G272-G273, in an exposed conformation, for both dimerisation and replication (368), thus reinforcing the existence of the primary $Pr55^{Gag}$ binding site in the upper internal loop of SL1.

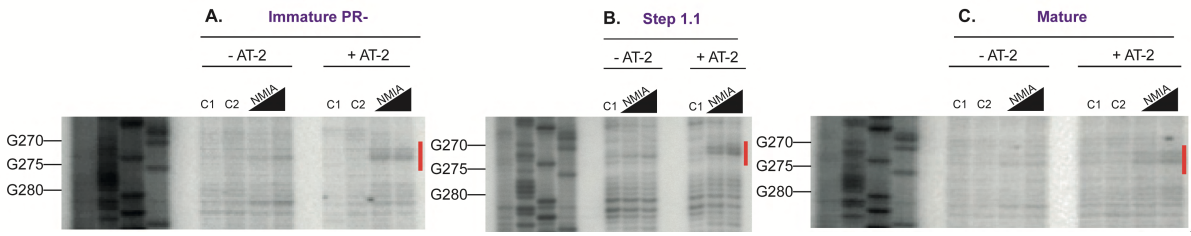


Figure 69

Probing of the HIV-1 gRNA inside PR-, Step 1.1 and mature particles, incubated or not with the AT-2 zinc ejector.

Primer extension (p365 Rv) of the untranslated gRNA region, modified in PR-, Step 1.1 and mature virus particles with NMIA, and analysed by denaturing PAGE. C1 control highlights natural RT stops whereas C2 control ensures that RNA modification took place inside the virus particle. NMIA concentrations of 3.9 and 7.8 mM were used. Position of modified nucleotides was determined thanks to the four sequencing reactions.

***In vitro* footprinting analysis of the G270-C274 region with Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7**

Pr55^{Gag} induces a strong protection of nucleotides A271-G273 (**Figure 70 A**). This protection is visualised *in viro* in PR- particles with a reactivity increase of nucleotides A269-A271 following AT-2 treatment (**Figure 67 A**). Upon binding of NCp15, NCp9 and NCp7, nucleotides A271-G273 are still strongly protected (**Figure 70 B, C** and **D**), in keeping with the absence of reactivity of this upper internal loop *in viro* in Step 1, Step 1.1, Step 2, CA/Sp1 and mature particles (**Figure 66 B to F**). Binding of NCp7 is also detected *in viro* with the reactivity profile of nucleotides A269-G272 increasing in AT-2 treated mature particles (**Figure 67 F**). Pr55^{Gag} and GagΔp6 induce a similar protection of nucleotides A271-G273 (**Figure 70 A** and **E**). Taken together, the pattern of protection induced by the different proteins is identical *in vitro* (**Figure 70 A to D**) and correlates with the protection of the upper lateral loop observed *in viro* (**Figure 66 A to F** and **70 A to D**).

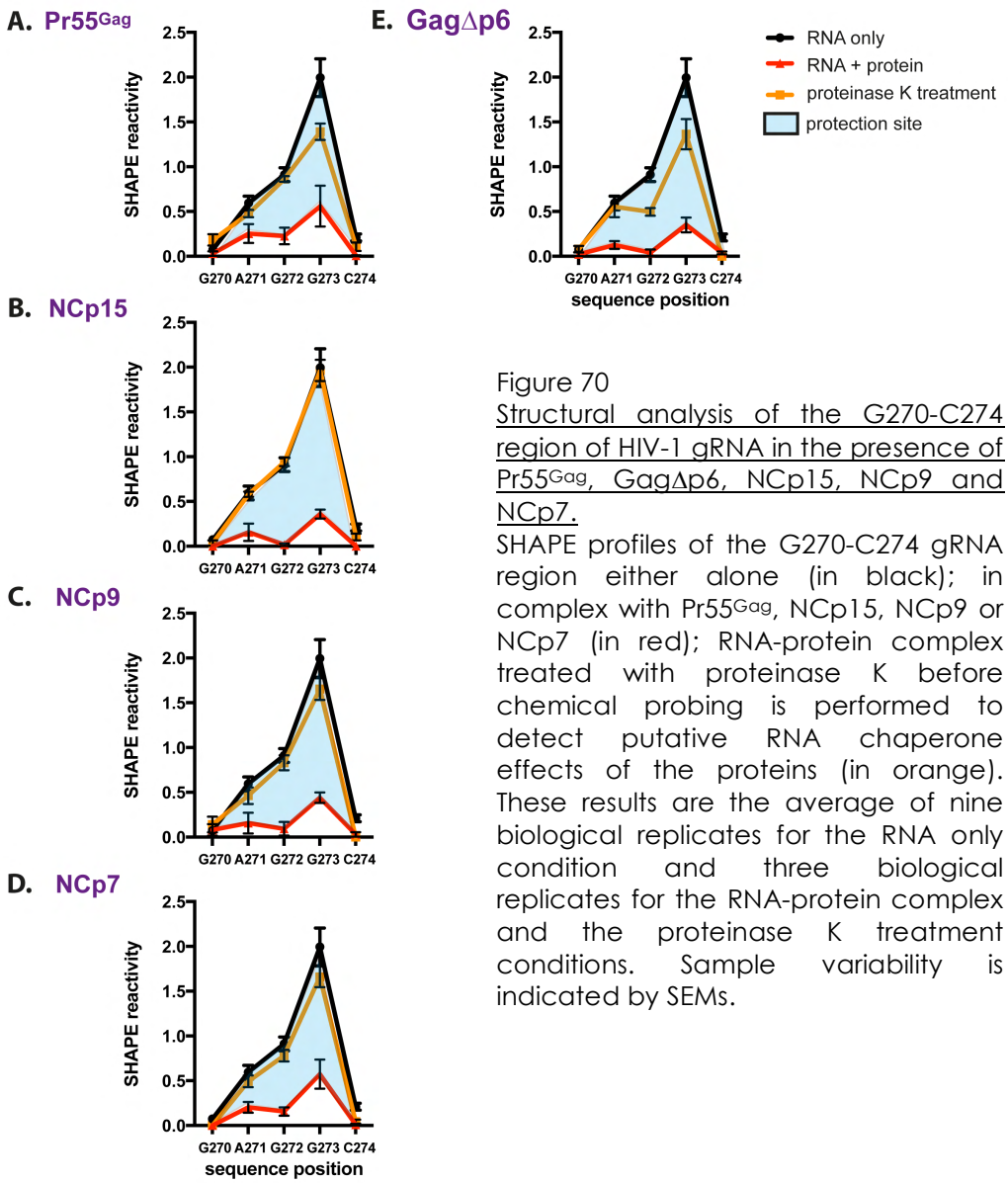


Figure 70
Structural analysis of the G270-C274 region of HIV-1 gRNA in the presence of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7.
SHAPE profiles of the G270-C274 gRNA region either alone (in black); in complex with Pr55^{Gag}, NCp15, NCp9 or NCp7 (in red); RNA-protein complex treated with proteinase K before chemical probing is performed to detect putative RNA chaperone effects of the proteins (in orange). These results are the average of nine biological replicates for the RNA only condition and three biological replicates for the RNA-protein complex and the proteinase K treatment conditions. Sample variability is indicated by SEMs.

Positions A239-A242 and G278-G279 are unpaired and form the lower internal loop in the extended SL1 model. This lower internal loop is closed by base-pairing between nucleotides C236-C238 and G280-G282 (422, 426). Nucleotides C238-G241 appear highly reactive in our hSHAPE-Seq data independently of the maturation stage of the viral particle (**Figure 66 A to F**) and have the highest hSHAPE-Seq reactivity in the first 550 nucleotides of the HIV-1 genome. Upon AT-2 treatment, reactivity becomes even higher in mature particles (**Figure 67 F** and **71**), likely indicating a NCp7 binding site. These data are coincident with a previous *in viro* study (426). However, this site does not appear to bind Pr55^{Gag} (**Figure 67 A** and **71**) as shown in Carlson et al. (67).

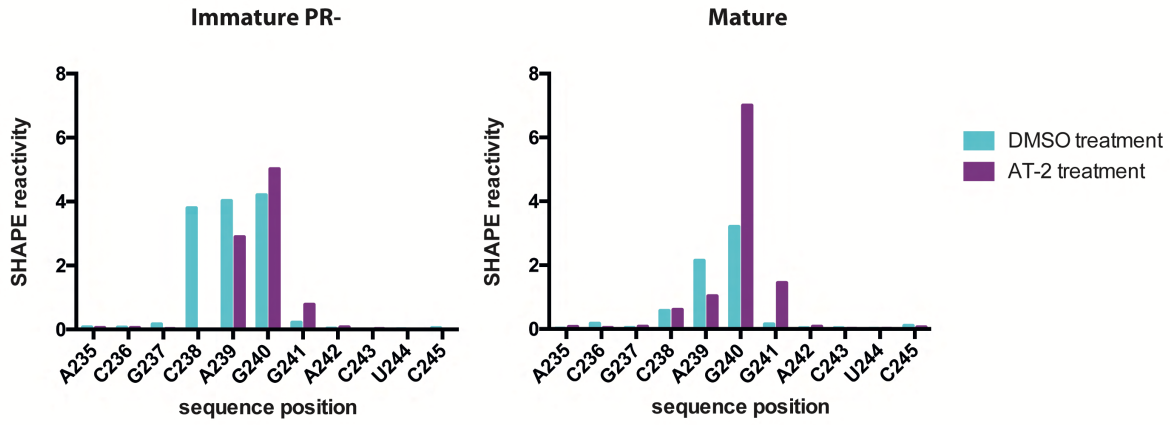


Figure 71
Comparison of SHAPE reactivity of positions A235-C245 in PR- and mature particles.
Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

The upper and lower internal loop consist of guanines and adenines exclusively. The NC domain of Pr55^{Gag} and NCp7 have been shown to preferentially bind to unpaired guanines and this may contribute to NCp7 binding to gRNA at these positions (141, 142, 412, 414, 418, 426). The NMR structure of NCp7 bound to SL2 and SL3 apical loops also highlights this preferential NCp7 interaction (16, 112).

In vitro footprinting analysis of the C236-U244 region with Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7

Pr55^{Gag} induces a strong protection of nucleotides G240-A242 *in vitro* (**Figure 72 A**). This protection is not as strong *in viro* in PR- particles (**Figure 67 A**). Upon NCp15 and NCp9 binding, nucleotides G241-A242 are still protected and nucleotides C238-A239 exhibit a slight increase of reactivity (**Figure 72 B** and **C**). Interestingly, nucleotides A239-G240 are significantly rearranged upon complex formation with NCp7, as shown by a 3-fold increase of SHAPE reactivity, while the increase in reactivity caused by GagΔp6, NCp15 and NCp9 is more limited (**Figure 72 D**). The reactivity of nucleotides G241-G242 is also increased by a 3-fold following AT-2 treatment of mature particles and could reflect a permanent structural rearrangement of these two nucleotides, even in absence of NCp7 (**Figure 67 F**), even if this is not visualised when RNA in complex with NCp7 is treated with proteinase K (**Figure 72 D**). Surprisingly, the footprint of Pr55^{Gag} and GagΔp6 proteins is not similar, with a structural rearrangement of nucleotides C238-A239 comparable to the one noticed with NCp15 and NCp9 (**Figure 72 C** and **E**).

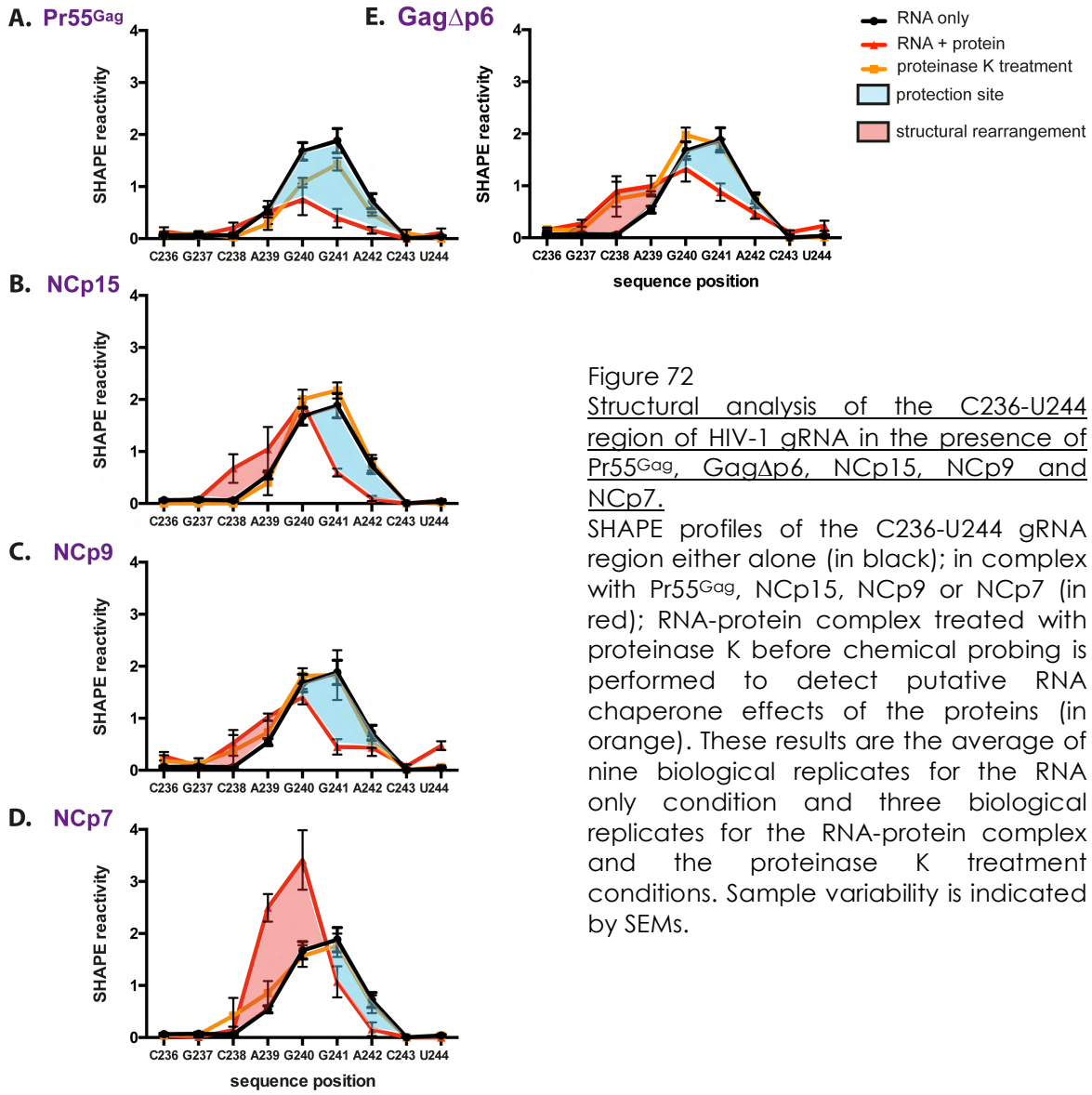


Figure 72
Structural analysis of the C236-U244 region of HIV-1 gRNA in the presence of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7.
SHAPE profiles of the C236-U244 gRNA region either alone (in black); in complex with Pr55^{Gag}, NCp15, NCp9 or NCp7 (in red); RNA-protein complex treated with proteinase K before chemical probing is performed to detect putative RNA chaperone effects of the proteins (in orange). These results are the average of nine biological replicates for the RNA only condition and three biological replicates for the RNA-protein complex and the proteinase K treatment conditions. Sample variability is indicated by SEMs.

Finally, the overall profile of base pairs constituting the SL1 domain are consistent with NMR results obtained in the context of the packaging signal (220, 253) or alone (106, 233, 234, 301, 413) and with chemical probing studies (319, 383, 422, 426). Positions C248-G254 and C264-G270 constituting the upper stem are unreactive, as well as C243-G246 and C274-G277 from the in-between stem (**Figure 66 A to F**). However, position C238 in the lower stem is detected as highly reactive independently on the maturation stage of the viral particle (**Figure 66 A to F**), suggesting that this position is unpaired. Position C238 is described as paired with nucleotide G280 in published models with an extended SL1 (422, 426) with, surprisingly, a reactivity below 0.5 in Wilkinson et al. (426).

Interestingly, the AT-2 treatment renders nucleotides G275-G278, implicated in the in-between stem and the lower internal bulge, highly reactive (**Figure 67**). This region is described as a stem in the mature conformation (319, 383, 422, 426) but the unreactive profile may have been misinterpreted. However, Wilkinson et al. did not detect a NCp7 binding site at this location (426). This increased reactivity profile after AT-2 treatment is not observed in PR- particles (**Figure 67 A**). This opened structure could be related to the formation of an extended duplex of the SL1 domain.

The initial kissing-loop dimer has been proposed to form an extended dimer duplex, resulting in the stabilisation of the SL1 domain by additional inter-molecular base-pairs. This extended dimer is promoted by NCp7 at physiological temperature or strongly facilitated *in vitro* by incubation at high temperature. Crystallography (24, 133, 134), NMR (158, 220, 231, 302), *in vitro* dimerisation (307, 308, 310) and single-molecule Förster resonance energy transfer (305) have provided insights of the extended duplex. However, these experiments were performed *in vitro* at non-physiological conditions. Moreover, the kissing-loop complex in the context of the entire 5'-region of the HIV-1 gRNA was not successfully converted into an extended duplex in other *in vitro* studies (323, 325). It is nevertheless important to mention the sole *in vivo* study addressing this problem and suggesting the existence of this extended duplex of the SL1 domain and its requirement for viral replication (416) and the recent hypothesis for a larger site called "dimer interactive site" and mediating dimerisation (368).

Regarding our hSHAPE-Seq data, the highly modified pattern of nucleotides G275-G278 in AT-2-treated mature particles indicates the region comprising A269-G278 to be in an opened conformation mediated by NCp7. This opened conformation could be due to the formation of an extended duplex involving only the upper SL1 stem (nucleotides C248-G270). The existence of the extended duplex cannot be inferred from the chemical probing studies. But this extended duplex questions the existence of the lower SL1 stem (positions C236-C238 and G280-G282) in the extended SL1 model due to steric constraints.

III.3.4.2 G283-G325 region comprising SL2 and SL3 domains

Two different secondary structures were proposed in the early 90's for SL2, with the apical loop comprising nucleotides G289-G292 (365) or A293-A296 (174). Compensatory mutants assayed in cell culture (175) indicated that the secondary structure of Sakaguchi et al. is functionally relevant *in vivo*, and NMR studies (16, 17) validated this model.

Our hSHAPE-Seq results are in agreement with an apical loop comprising positions G289-G292.

Nucleotides G290-G292 appear highly reactive in PR- particles and in the Step 1 intermediate (**Figure 73 A** and **B**). This region becomes poorly reactive from the Step 1.1 stage with position U291 maintaining a medium reactivity (**Figure 73 C to F**). In mature particles, the averaged reactivity of positions G290-G292 in DMSO treated samples (n=3) is lower than in the control and better corresponds to the reactivity profile of previous stages in the proteolytic cascade of Pr55^{Gag} (**Figure 73 F** and **74 F**).

Following AT-2 treatment of PR- particles, nucleotides U288 and G290-G292 are highly reactive, compared to the PR- particles treated or not with DMSO (**Figure 73 A** and **74 A**). This increased reactivity could reflect a Pr55^{Gag} binding site located at the apical loop and correlates with the reactivity pattern of positions U288 and G290-G292 in AT-2-treated mature particles (**Figure 73 F** and **74 F**). In addition, the study of Wilkinson et al. validated this potential NCp7 binding site (426). An increased affinity of Pr55^{Gag} processing intermediates and mature NCp7 is observed with an important decrease of the reactivity profile (16, 17). In addition, NCp7 binds to the SL2 domain with an affinity similar to that observed for the SL3 domain (89). The two ZFs bind to the exposed guanine nucleotides of the tetraloop with, however, other features significantly differing between the two complexes (16, 112). Moreover, the NMR structure of the SL2 domain in complex with NCp7 reveals that position G289 is stacked above the U288-A293 whereas the other nucleotides of the apical loop are available for potential interactions, perfectly correlating with hSHAPE-Seq data.

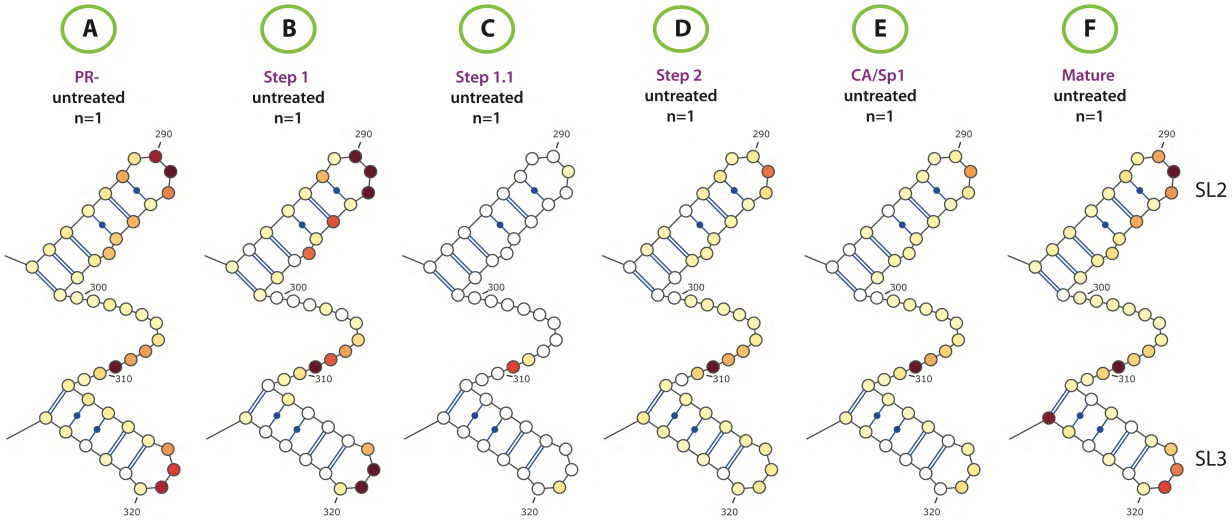


Figure 73
Evolution of the G283-G325 region of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples is placed on a published model (426).

In addition, positions G294-A296 become highly reactive in AT-2-treated PR- particles (**Figure 74 A**), whereas only the position G294 is unprotected in AT-2-treated mature particles (**Figure 74 F**). Positions U295 and A296 have been described as involved in a base-triple structure with position A286 (17), correlating with the reactivity profile found in mature particles. However, the increased reactivity profile after AT-2 treatment questions the existence of this base-triple structure at the immature stage. In addition, nucleotides C284-G285 become highly reactive in AT-2-treated mature particles besides position G294 (**Figure 74 F**). These results could indicate a stabilisation of the SL2 domain mediated by Pr55^{Gag} and NCp7.

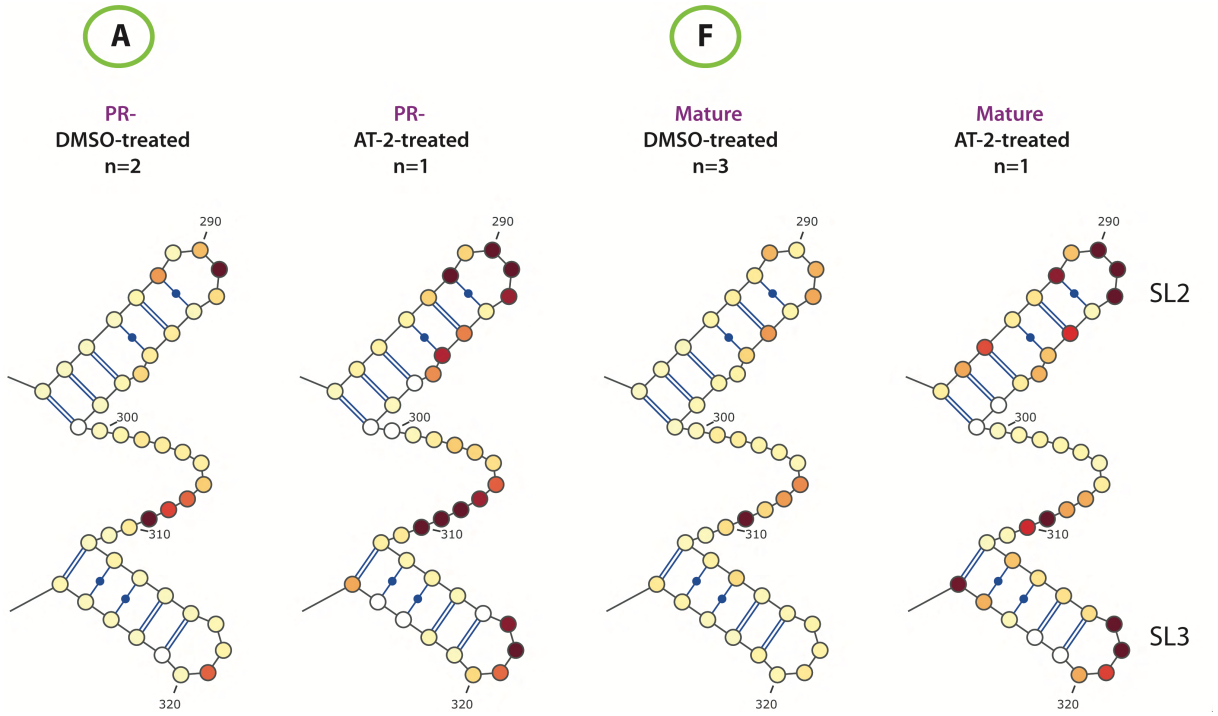


Figure 74
Effect of the AT-2 treatment on the G283-G325 region of HIV-1 gRNA from PR- and mature particles.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR- and mature samples is placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control, prior to NMIA modification.

The SL2 extended duplex model, proposed recently by the team of M. Summers, involves the formation of the U5-AUG interaction and the subsequent refolding of the SL2 domain when gRNA dimerisation occurs (**Figure 75**). A three-way junction is formed with positions G283-A286 paired to positions U232-A235, G289-U291 (involved in the apical loop in the "classical" model) paired to A227-C229, U295-C299 paired to G119-U122, with positions C287, U288 and G292-G294 remaining unpaired (220).

Whereas the SL2 apical loop is relatively unreactive in the mature gRNA, it does not mean that the SL2 domain is paired (**Figure 73 F** and **74 F**). In addition, the SL2 apical loop becomes highly modified upon AT-2 treatment (**Figure 74 F**). Thus, our hSHAPE-Seq data, as well as *ex viro* (383, 422) and *in viro* (319, 426) data do not correlate with this model. As noticed in the published paper, this study was performed in absence of RNA chaperone proteins, questioning the *in vivo* validity of the three-way junction model. In addition, the team of B. Berkhout performed a phylogenetic analysis of the three-way junction and the conventional SL2 domain and concludes that the observed sequence variation is highly compatible with the SL2 hairpin structure whereas only poorly compatible with the three-way junction folding (300).

Our hSHAPE-Seq data illustrate the strong structuration of the stem comprising positions C312-C316 paired with G321-G325, independently of the stage of maturation of the viral particle (**Figure 73 A to F**). This stem-loop correspond to the consensus structure for the SL3 domain, composed of 14 nucleotides with nucleotides G317-G320 forming the apical loop (28, 89, 105, 365). This 14-nucleotide

model has been further validated by NMR (112, 327) and correlates with recent *ex viro* (383, 422) and *in viro* (319, 426) chemical probing studies.

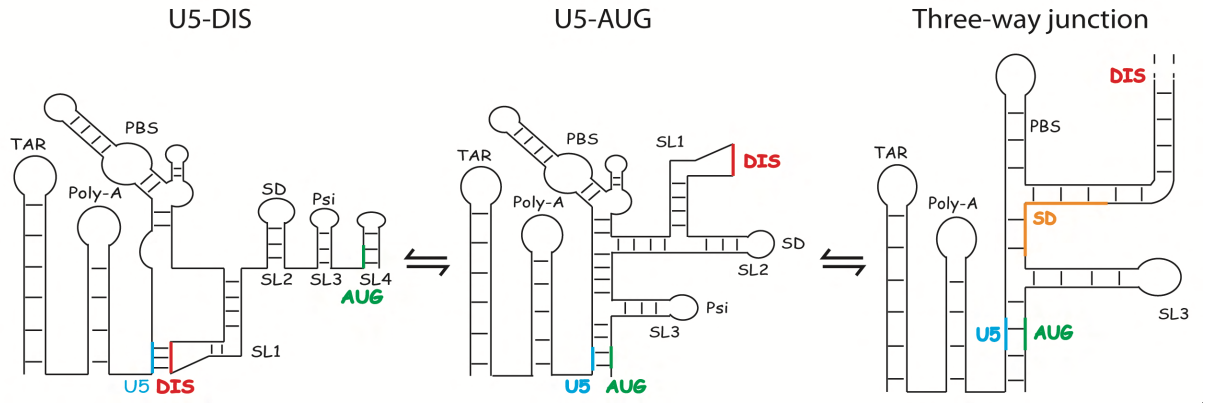


Figure 75
Description of RNA structures proposed by the team of M. Summers.
These structures are proposed to promote gRNA packaging thanks to an interaction between AUG and U5 region.

The residues G318 and A319, located in the apical loop, are highly reactive during the first steps of the maturation process (i.e. in PR- and Step 1 viral particles) (**Figure 73 A** and **B**). The apical loop becomes unreactive from Step 1.1 to the CA/Sp1 stage (**Figure 73 B to E**), whereas positions G318 and A319 are again highly reactive at the mature stage (**Figure 73 F**). This reactivity profile contrasts with the reactivity profile exhibited by these two residues in DMSO-treated mature particles (**Figure 74 F**). Nevertheless, despite these significant accessibility differences of the apical loop between PR- and mature particles treated or not with DMSO, the AT-2 treatment undeniably deprotects positions G317-A319, resulting in a pronounced increase of reactivity (**Figure 74 A** and **F**).

The SL3 domain has historically been considered as the main packaging signal based on mutagenesis studies (12, 91, 92, 260, 271, 288, 361). However, the exact deletion of the SL3 domain produce a more modest effect on packaging (175, 185) than previously observed with deletions of the SL3 domain plus flanking regions, which likely globally affect the structure of the 5' region of the HIV-1 gRNA. In addition, while the primary $Pr55^{Gag}$ binding site has been shown to reside in the SL1 domain, no $Pr55^{Gag}$ binding could be detected in the isolated SL3 domain by band-shift assay (4). We demonstrated as well that the SL1 domain is more important than the SL3 domain for encapsidation (185). Fluorescence spectroscopy further revealed that the SL3 domain displays a 3 to 8-fold lower $Pr55^{Gag}$ binding affinity compared to the SL1 domain (48). The lower binding of $Pr55^{Gag}$ to the SL3 domain is however not noticed in hSHAPE-Seq data.

The three-dimensional NMR structure of NCp7 bound to the SL3 domain revealed a flexible tetraloop, with positions G318 and G320 exposed to the solvent and involved in NCp7 binding (112, 327). In addition, NCp7 binding exposes A319, rendering it reactive (16). These results correlate with high reactivity of A319 in an *in vitro* $Pr55^{Gag}$ footprint assay of the gRNA (4), in *in viro* probing studies (319, 426).

***In vitro* footprinting analysis of the C312-G325 region with Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7**

Pr55^{Gag}, NCp15 and NCp7 protect nucleotides G317-G318 and G320 (**Figure 76 A, B** and **D**). Upon binding, these proteins expose residue A319 and increase its reactivity. In addition, this conformational rearrangement appears to be permanent, as proteinase K treatment does not reverse the increase of reactivity at this position (**Figure 76 A, B** and **D**). NCp9 and GagΔp6 bind to nucleotides G317-G320 but do not structurally rearrange residue A319. Protection of nucleotides G317-G318 and G320 is also observed *in viro* from the immature to the mature stage (**Figure 73 A to F**) with these positions highly reactive after AT-2 treatment of PR- and mature particles (**Figure 74 A** and **F**).

Position A319 is highly reactive in PR-, Step 1 and mature particles. These results correlate with the absence of conformational rearrangement induced by NCp9, present in Step 1.1 and Step 2 intermediate particles (**Figure 73 C** and **D**). However, the reason for the absence of reactivity of A319 at the CA/Sp1 stage, in the presence of NCp7, is unknown (**Figure 73 E**).

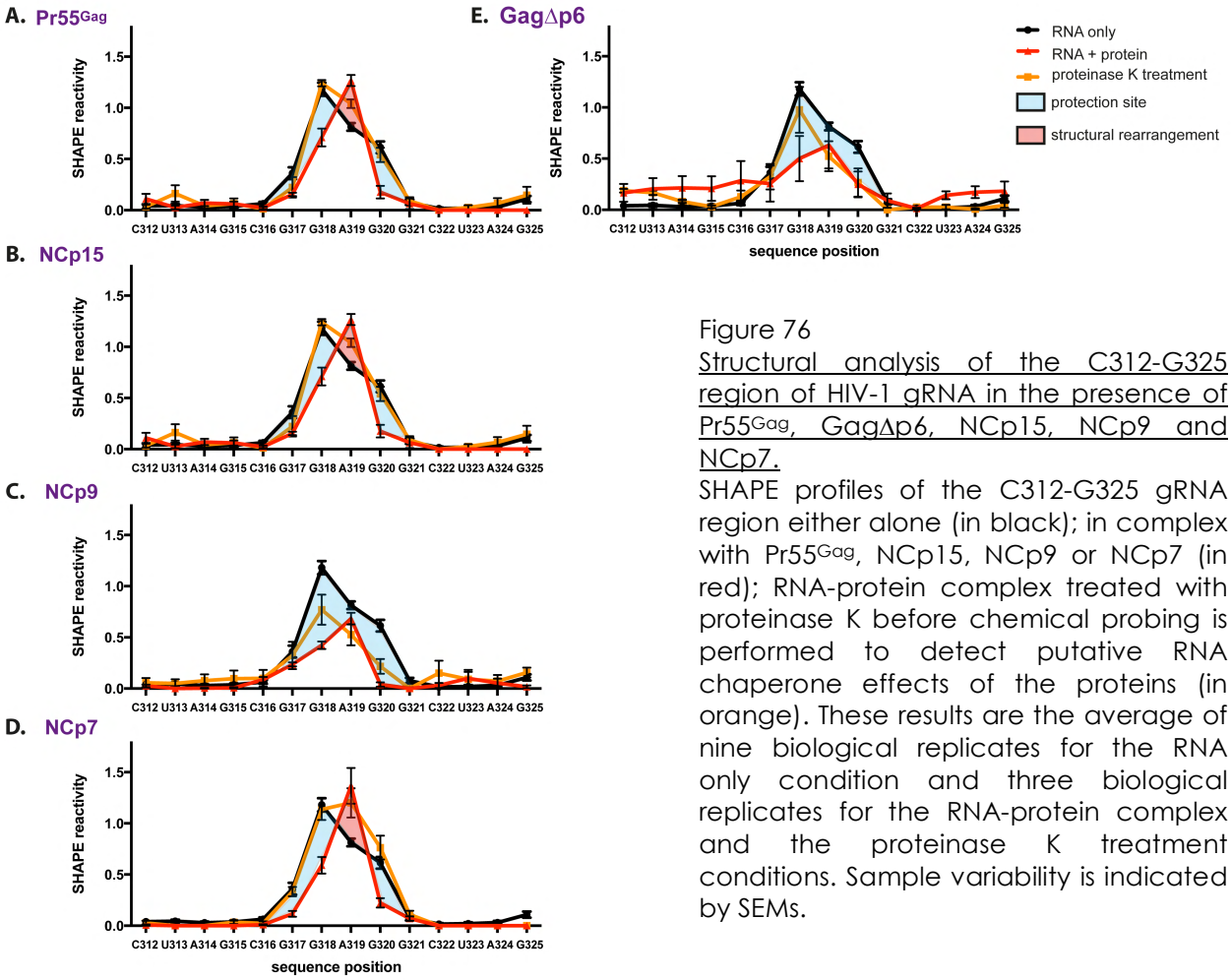


Figure 76
Structural analysis of the C312-G325 region of HIV-1 gRNA in the presence of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7.
SHAPE profiles of the C312-G325 gRNA region either alone (in black); in complex with Pr55^{Gag}, NCp15, NCp9 or NCp7 (in red); RNA-protein complex treated with proteinase K before chemical probing is performed to detect putative RNA chaperone effects of the proteins (in orange). These results are the average of nine biological replicates for the RNA only condition and three biological replicates for the RNA-protein complex and the proteinase K treatment conditions. Sample variability is indicated by SEMs.

The SL2-SL3 linker (positions C300-A311) is divided in two parts exhibiting a different reactivity profile. Nucleotides C300-U306 are unreactive independently of the maturation stage of the viral particle (**Figure 73 A to F**) and remain unreactive after AT-2 treatment (**Figure 74 A** and **F**). Positions A304-A305 have been previously reported as highly reactive *in vitro* (4, 104, 426), *ex viro* (383, 422) and *in viro* (319, 426). In addition, $Pr55^{Gag}$ binding induced a SHAPE reactivity enhancement at these two positions, implicating a $Pr55^{Gag}$ destabilising effect of positions A304-A305 (4). This reactivity profile is consistently not observed in the different samples composing the proteolytic cascade of $Pr55^{Gag}$.

In the second part of the linker (nucleotides U307-A311), nucleotides U307-U308 are reactive only in PR- and Step 1 particles (**Figure 73 A** and **B**) whereas position U309 is highly reactive from the immature to the mature stage (**Figure 73 A to F**). The reactivity of nucleotides U306-A311 becomes significantly higher following AT-2 treatment of PR- particles (**Figure 74 A**) whereas this effect is less apparent in mature particles, with a moderate increase of reactivity at positions U308-G310 (**Figure 74 F**). Reactivity of positions U307-U309 in PR- particles is consistent with the SHAPE reactivity enhancement observed in the presence of $Pr55^{Gag}$ and suggests a $Pr55^{Gag}$ destabilising effect of positions U307-U309 (4). Results obtained after AT-2 treatment suggest a $Pr55^{Gag}$ binding site at positions U306-A311, exposing residues U307-U309. This hypothesis is consistent with the moderate increase in reactivity of positions U308-G310 in AT-2-treated mature particles (**Figure 74 F**), reflecting a NCp7 binding site already reported by the team of K. Weeks (426). Reactivity of this site (nucleotides U307-A311) in hSHAPE-Seq experiments present an intermediate profile between a globally high reactivity pattern in some studies (383, 422, 426) and an absence of reactivity in another (319). The number of nucleotides with a high reactivity pattern and possibly implicated in ZF RNA binding proteins decreases during the proteolytic cascade with three nucleotides in PR-, Step 1 and Step 1.1 particles (U307-U309) (**Figure 73 A to C**) and only one from the Step 2 to the mature stage (position U309) (**Figure 73 D to F**). The $Pr55^{Gag}$ processing possibly explains these results with the production of smaller processing intermediates with different RNA binding parameters (**Figure 73 A to F**). In addition, position G310 is always unreactive, consistent with published models (4, 426).

In vitro footprinting analysis of the C300-U313 region with Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7

As noticed with the *in viro* results, the SL2-SL3 linker is divided in two parts with different reactivity profiles. Pr55^{Gag}, NCp15 and NCp9 induce a limited protection of nucleotides A301-A303 (**Figure 77 A, B and C**) and upon NCp7 binding, only nucleotide A301 is protected (**Figure 77 D**). However, this protection, mediated by Pr55^{Gag} and to a smaller extend by NCp7, is not detected *in viro* in PR- and mature particles (**Figure 74 A** and **F**). Interestingly, reactivity of nucleotides U307-U309 increases in presence of Pr55^{Gag}, NCp15 and NCp7 (**Figure 77 A, B** and **D**). This increase in reactivity is strong in the presence of Pr55^{Gag}, NCp15 and NCp7, with a significant variation between biological samples with Pr55^{Gag}. This reactivity profile suggests that Pr55^{Gag}, NCp15 and NCp7 bind to nucleotides A301-A303 and induce a conformational rearrangement of positions U307-U309. This conformational rearrangement is permanent with NCp15, semi-permanent with NCp7, whereas not permanent if Pr55^{Gag} is removed (**Figure 77 A, B** and **D**). Surprisingly, neither NCp9 nor GagΔp6 induce this conformational rearrangement (**Figure 77 C** and **E**). *In viro*, this second part of the linker is weakly modified at positions U307-U308 and highly modified at position U309 from the immature to the mature stage, (**Figure 73 A to F**), correlating with *in vitro* results. However, nucleotides U306-G310 become highly reactive in AT-2-treated PR- and mature particles, demonstrating the presence of a Pr55^{Gag} and NCp7 binding site at these positions (**Figure 74 A and F**) whereas *in vitro* results suggest a structural rearrangement (**Figure 77 A to D**).

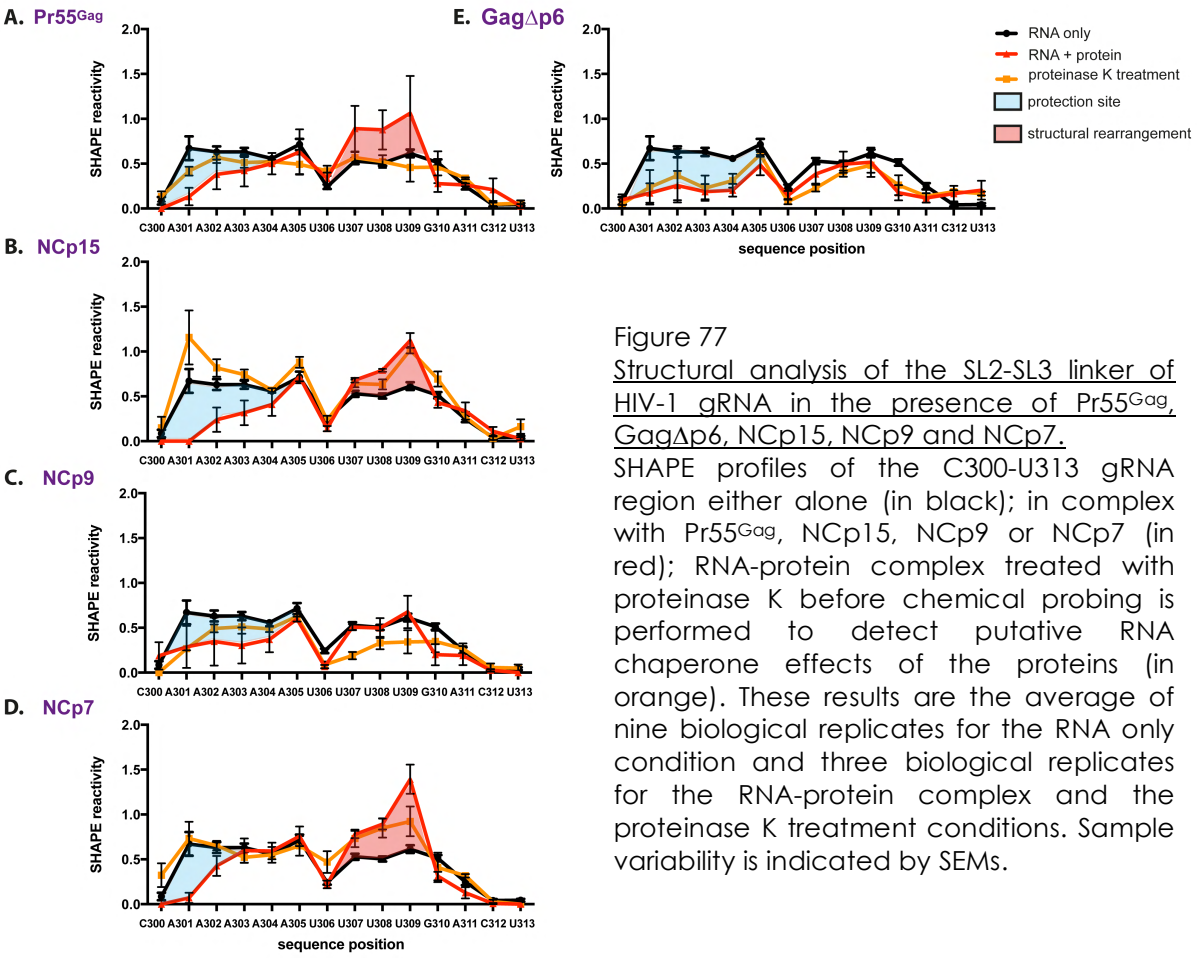


Figure 77
Structural analysis of the SL2-SL3 linker of HIV-1 gRNA in the presence of Pr55^{Gag}, GagΔp6, NCp15, NCp9 and NCp7.
SHAPE profiles of the C300-U313 gRNA region either alone (in black); in complex with Pr55^{Gag}, NCp15, NCp9 or NCp7 (in red); RNA-protein complex treated with proteinase K before chemical probing is performed to detect putative RNA chaperone effects of the proteins (in orange). These results are the average of nine biological replicates for the RNA only condition and three biological replicates for the RNA-protein complex and the proteinase K treatment conditions. Sample variability is indicated by SEMs.

III.3.4.3 Unique 5'-element, CU-rich region and the SL4 domain

In our hSHAPE-Seq results, positions U105-G116 in U5 are unreactive independently of the maturation stage of the viral particles (**Figure 78 A to F**) as well as after AT-2 treatment of PR- and mature particles (**Figure 79 A** and **F**). This absence of reactivity correlates with a base-pairing of the nucleotides U105-G116. This region has been proposed to base-pair with the SL4 domain by B. Berkhout (191) and M. Summers (220, 231, 269) labs with the LDI-BMH model and the tree-way junction, respectively.

The LDI-BMH model hypothesises the existence of a structural switch of the 5' region of HIV-1 gRNA, with an equilibrium between two conformations, termed LDI and BMH. In the LDI conformation, the poly-A and the SL1 domain are base-paired whereas in the BMH conformation, the PBS, SL1, SL2 and SL3 domains are independently folded and the SL4 domain (A334-G344) is paired to nucleotides U105-G116. For the sake of simplicity, this interaction is called the U5-AUG interaction. The BMH conformation is supported by an enzymatic probing study (104) but with less confidence for the ds region implicating positions G106-G108 and U341-C343. This pattern has been confirmed by the team of K. Weeks with a slightly smaller number of nucleotides implicated in the U5-AUG interaction. In this model, positions G108-C114 interact with G335-U341 (426).

The team of M. Summers also proposed a conformational RNA switch implicating the U5-AUG interaction, based on NMR results (269). In the monomeric conformation, the apical loop of the SL1 domain (and not the entire SL1 domain as in the LDI conformation) interacts with nucleotides U105-G116 and residues spanning the *gag* AUG form the SL4 hairpin. In the dimeric conformation, the AUG domain (nucleotides G333-G344) base pairs with nucleotides U105-G116 in U5 (220).

Nucleotides G333-G344, comprising the *gag* AUG, are mostly unreactive at the immature stage, except positions G338 and G340 appearing as weakly reactive (**Figure 78 A**). However, nucleotides G333-G344 become globally highly reactive following the AT-2 treatment of PR- particles (**Figure 79 A**). Interestingly, this increased reactivity pattern is even extended to a larger region than the SL4 region implicated in the U5-AUG interaction and extend to nucleotides A326-G346, with positions G329, G331, G335-A336 and G342 however remaining unreactive. $Pr55^{Gag}$ thus most likely protects this region.

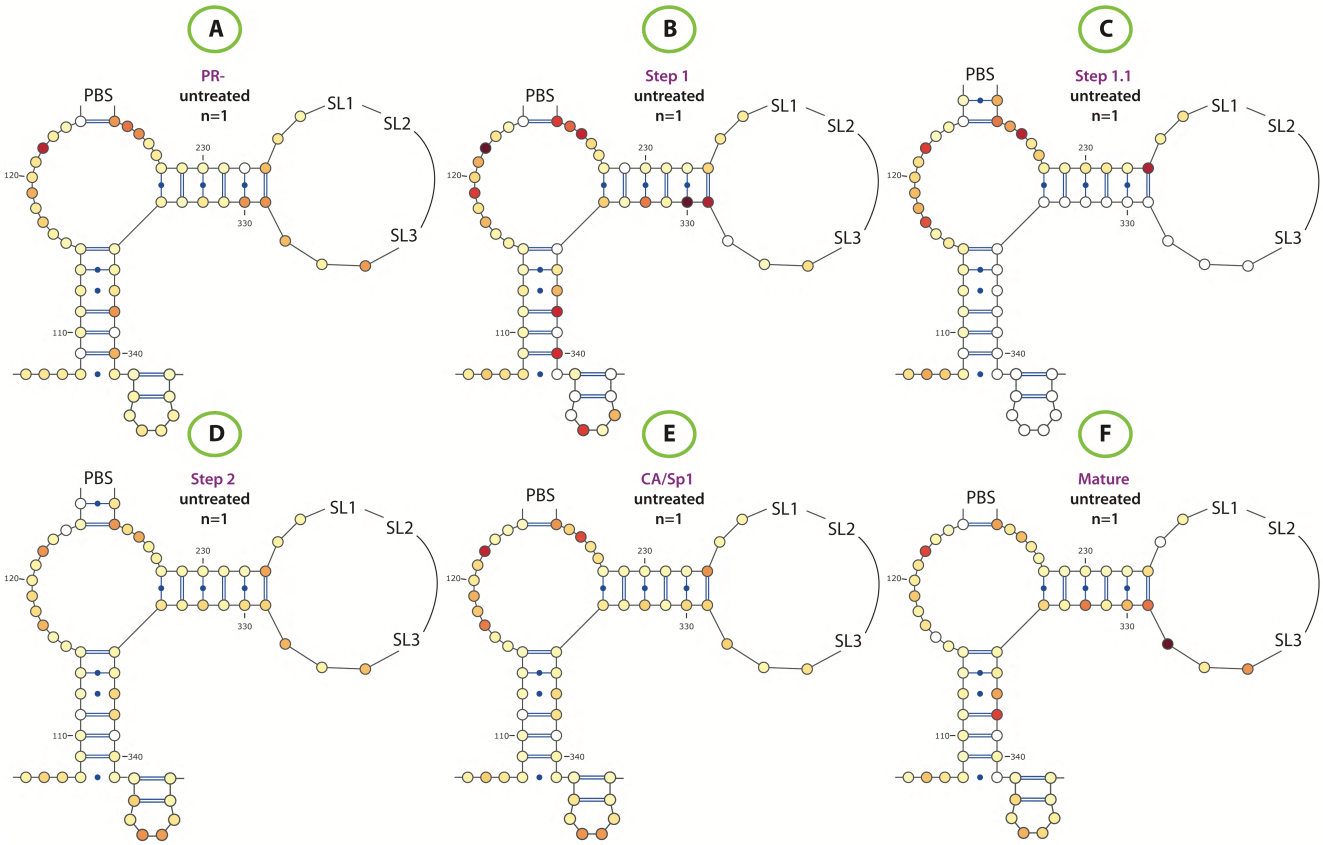


Figure 78
Evolution of the Unique-5' element, the CU-rich region and the SL4 domain of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples is placed on a published model (426).

The reactivity of nucleotides G329-A330, A332, G338, G340, A345 and A347 strongly increases when going from the immature stage to the Step 1 stage (**Figure 78 B**). This reactivity profile is reminiscent of the profiles observed with AT-2 treated mature, and to a lesser extend immature particles (**Figure 79**). This reactivity profile could indicate that NCp15 is unable to bind to this region once released from Pr55^{Gag}. From the Step 1.1 to the CA/Sp1 stage, nucleotides A326-G346 are unreactive (**Figure 78 C to E**), except positions A345-G346, which are weakly reactive at Step 2 and CA/Sp1 stages, potentially indicating a protection of this region by Pr55^{Gag} processing intermediates. The reactivity profile of this region for the Step 1.1 sample has also been obtained by PAGE (**Figure 80 B**) and confirms that positions A326-A327, G329-A330, A332, A334 and A336-A337 are strongly modified, similar to the mature sample (**Figure 80 C**), though reactivity are detected at these positions in the C1 lane. This Step 1.1 reactivity profile slightly differs from hSHAPE-Seq data (**Figure 80 C**).

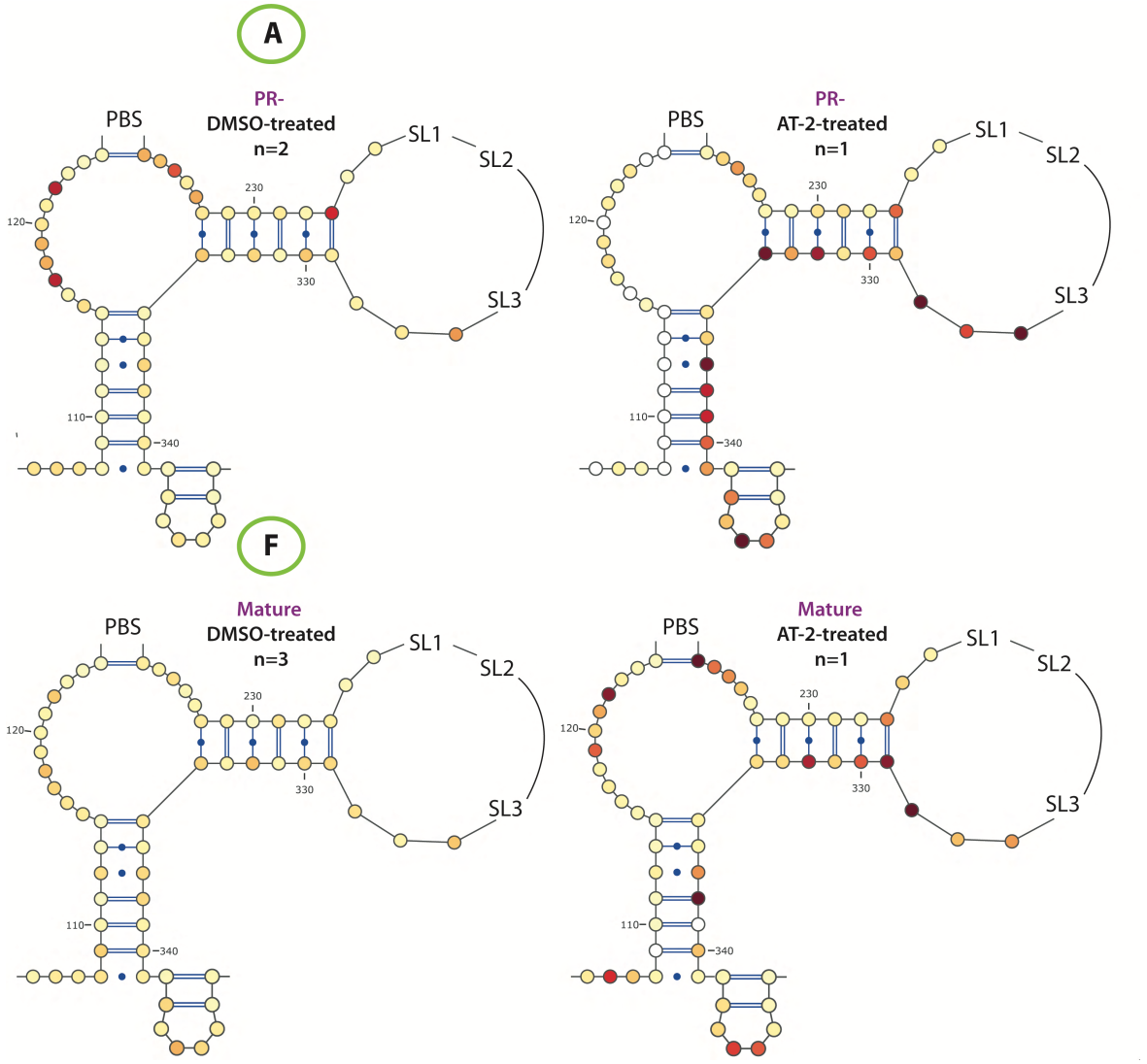


Figure 79
Effect of the AT-2 treatment on the Unique-5' element, the CU-rich region and the SL4 domain of HIV-1 gRNA from PR- and mature particles.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR- and mature samples is placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control prior to NMIA modification.

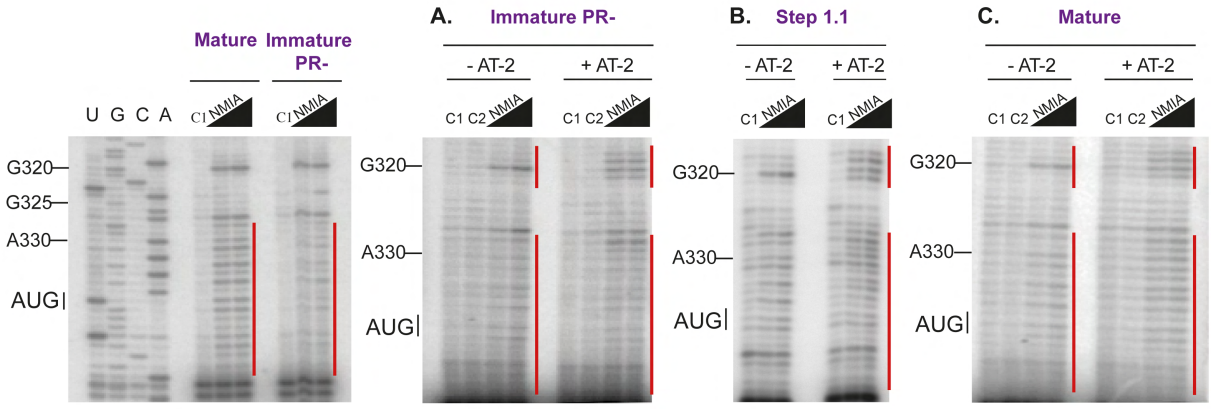


Figure 80
Probing of the HIV-1 gRNA inside PR-, Step 1.1 and mature particles, incubated or not with the AT-2 zinc ejector.
Primer extension (p365 Rv) of the 5' region of the HIV-1 gRNA region, modified *in viro* with NMIA in PR-, Step 1.1, and mature virus particles analysed by denaturing gel electrophoresis. C1 control reveals natural RT stops whereas C2 control ensures that RNA modification was performed inside the virus particle. NMIA concentrations of 3.9 and 7.8 mM were used. Position of modified nucleotides was determined thanks to the four sequencing reactions.

In the mature gRNA, the SL4 domain is highly modified upon AT-2 treatment and therefore highlights the impossibility for this region to be paired with positions U105-G116 in U5 (**Figure 79** and **81**). Some inconstancies, mostly present at nucleotides A326-G344, are revealed when comparing mature samples treated or not with DMSO. These nucleotides in the three DMSO-treated replicates are unreactive (**Figure 79 F**) whereas some are found moderately to highly reactive in the untreated replicate (**Figure 78 F**). Positions A326, G328-A330, A332 and U337-G338 shown to be reactive correspond to the positions becoming reactive following AT-2 treatment (**Figure 79 F**).

Thus, whether these positions are protected or not by NCp7, this region appears to be in a single-stranded conformation. This hypothesis is further supported by the reactivity profile of this region obtained by PAGE (**Figure 80 C**), which shows that positions A326-A327, G329-A330, A332, A334 and A336-A337 are strongly modified.

The structure of the SL4 domain, folded as an hairpin (28, 89, 105, 287), has been solved by NMR, describing the stem as unstable (226) and constituting a weak NCp7 binding site (18). The SL4 domain has also been proposed to be unfolded with positions A326-A327, A330, A332, A334 and A336 highly reactive (319), correlating with hSHAPE-Seq data.

Therefore, the reactivity of the SL4 domain indicates the impossibility for the U5-AUG interaction to exist in immature PR- and mature particles. Surprisingly, the team of K. Weeks does not observe a protection of the SL4 domain by NCp7 (426).

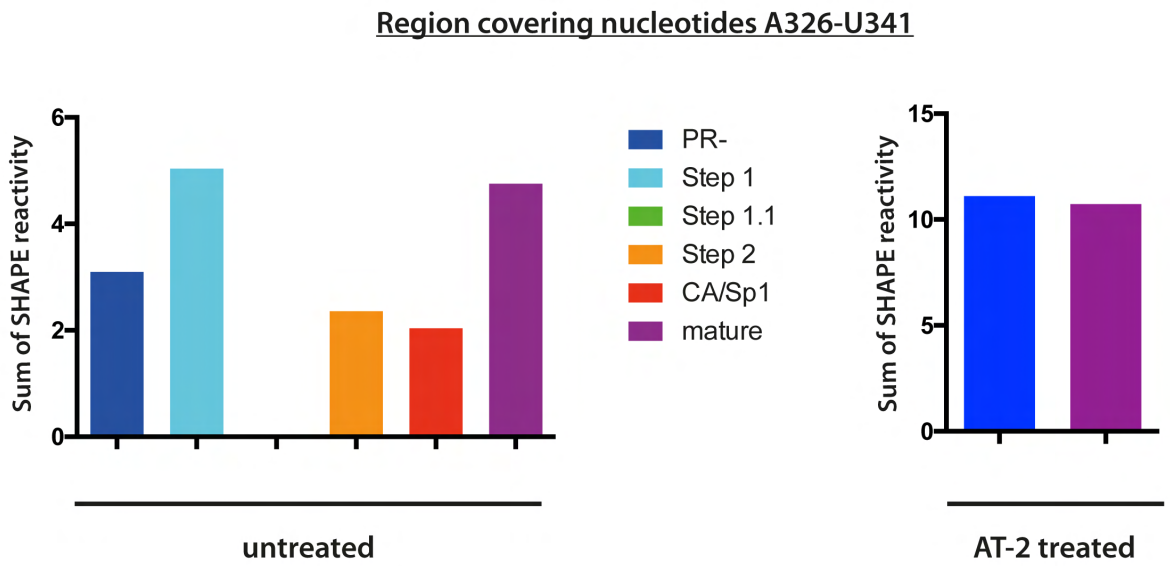


Figure 81
Evolution of the reactivity of the SL4 domain of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides A326-U341 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

During the Pr55^{Gag} proteolytic cascade, nucleotides involved in the CU-rich region (positions U228-C233) are not reactive (**Figure 78 A to F**), and their reactivity is not affected by the AT-2 treatment (**Figure 79 A** and **F**). The CU-rich region has first been proposed to interact with the AG-rich region (G329-A334) by the group of A. Lever (174). This long-range interaction is supported by one *in viro* study (426) and proposed in two *ex viro* studies (383, 422). Several other models of this CU-rich region have been proposed (3, 19, 105, 191, 220, 231, 361). Thus, the CU-rich region must be interacting elsewhere in the genome, since our hSHAPE-Seq data indicating it cannot interact with the AG-rich region. High reactivity of nucleotides in the AG-rich region was also obtained by *in viro* (319) and *ex viro* (383, 422) chemical probing.

In addition, nucleotides G223-A225, strongly modified up to the Step 1.1 stage (**Figure 78 A to C**), and weakly modified at Step 2 and CA/Sp1 stages (**Figure 78 D to E**), seem protected by NCp7 (**Figure 78 F**). These three nucleotides exhibit an increased reactivity following AT-2 treatment of mature particles (**Figure 79 F**). These results are consistent with a previously reported NCp7 binding site (426).

III.3.5 A336-A543 region: the first nucleotides of the gag coding region

Regarding positions G339-U341, three studies proposed that these nucleotides are paired with the U5 region (383, 422, 426). However, as discussed above, our hSHAPE-Seq data are not in favour of this U5-AUG interaction, as the reactivity pattern of positions A326-G338 increases in mature viral particles following AT-2 treatment (**Figure 82 F**). However, positions G339-U341 remain unmodified (**Figure 82 F**), thus suggesting these nucleotides are paired. A recently published model (116) propose that these nucleotides are involved in a stem composed of G339-C343 and G348-C352. This alternative structure is consistent with our hSHAPE-Seq results showing that nucleotides A345-G346, located in the apical loop in the model of Deforges et al., are highly modified following AT-2 treatment (**Figure 83 F**) whereas positions G348-C352 are unreactive (**Figure 82 F** and **83 F**). Following this G339-C352 stem-loop,

Deforges et al. proposed a long-distance interaction involving nucleotides G354-A359 and U572-C577, which is not consistent with the decreased reactivity of positions U358-A360 following AT-2 treatment (**Figure 82 F** and **83 F**). The region encompassing nucleotides G338-G363 is differently represented in every model with one (422) or two (426) stem-loops or as a single-strand (383). This highlights the existing inconsistency between the few available secondary structure models of the *gag* coding region, determined from *in vitro* (116, 353), *ex viro* (383, 422) or *in viro* (426) data.

There is an agreement that nucleotides G361-G367 form a long-distance interaction, but the interacting sequence is not the same in the different model. Nucleotides G364-A368 (426) or G363-A368 (116) are proposed to be paired with U534-U538 or U534-U539, or alternatively nucleotides G361-G365 (422) or G363-G367 (383) would be paired with C746-C750 or A745-C750. Depending on the study, positions A745-C750 are shown to be unreactive (426) or highly modified (116), and were not studied in our hSHAPE-Seq approach. The long-distance interaction with nucleotides U534-U538 is consistent with our hSHAPE-Seq results showing no reactivity of this region from the immature to the mature stage (**Figure 82 A to F**), even after AT-2 treatment of PR- and mature particles (**Figure 83 A** and **F**).

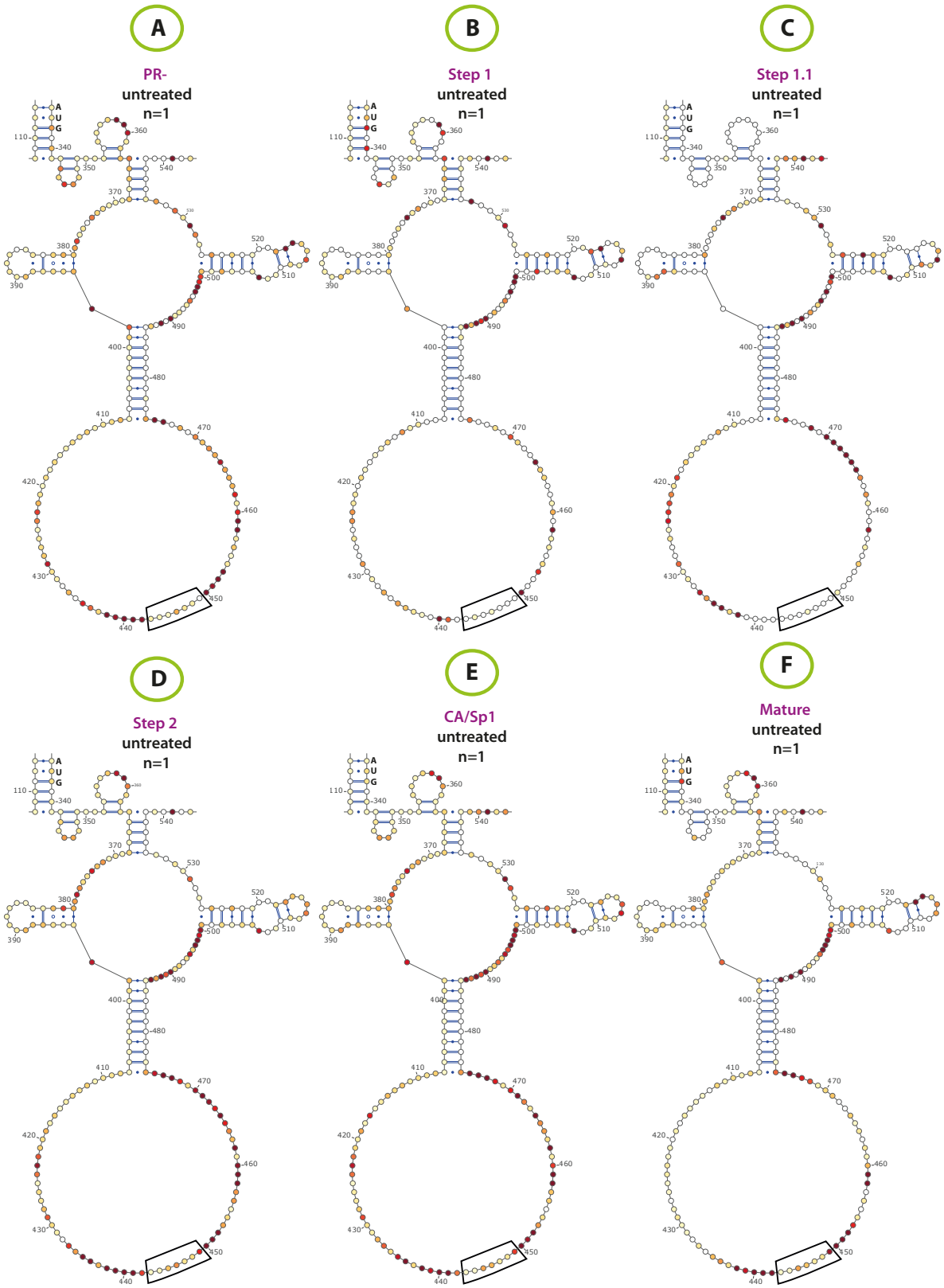


Figure 82
Evolution of the first 200 nucleotides of the *gag* coding region of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
SHAPE reactivity, obtained by hSHAPE-Seq, of PR-, intermediates and mature samples is represented on a published model (426). The heptanucleotide sequence implicated in the pseudoknot formed between the Poly-A apical loop (positions G79-C85) and nucleotides in the matrix coding region (positions G443-C449) is represented by the black box.

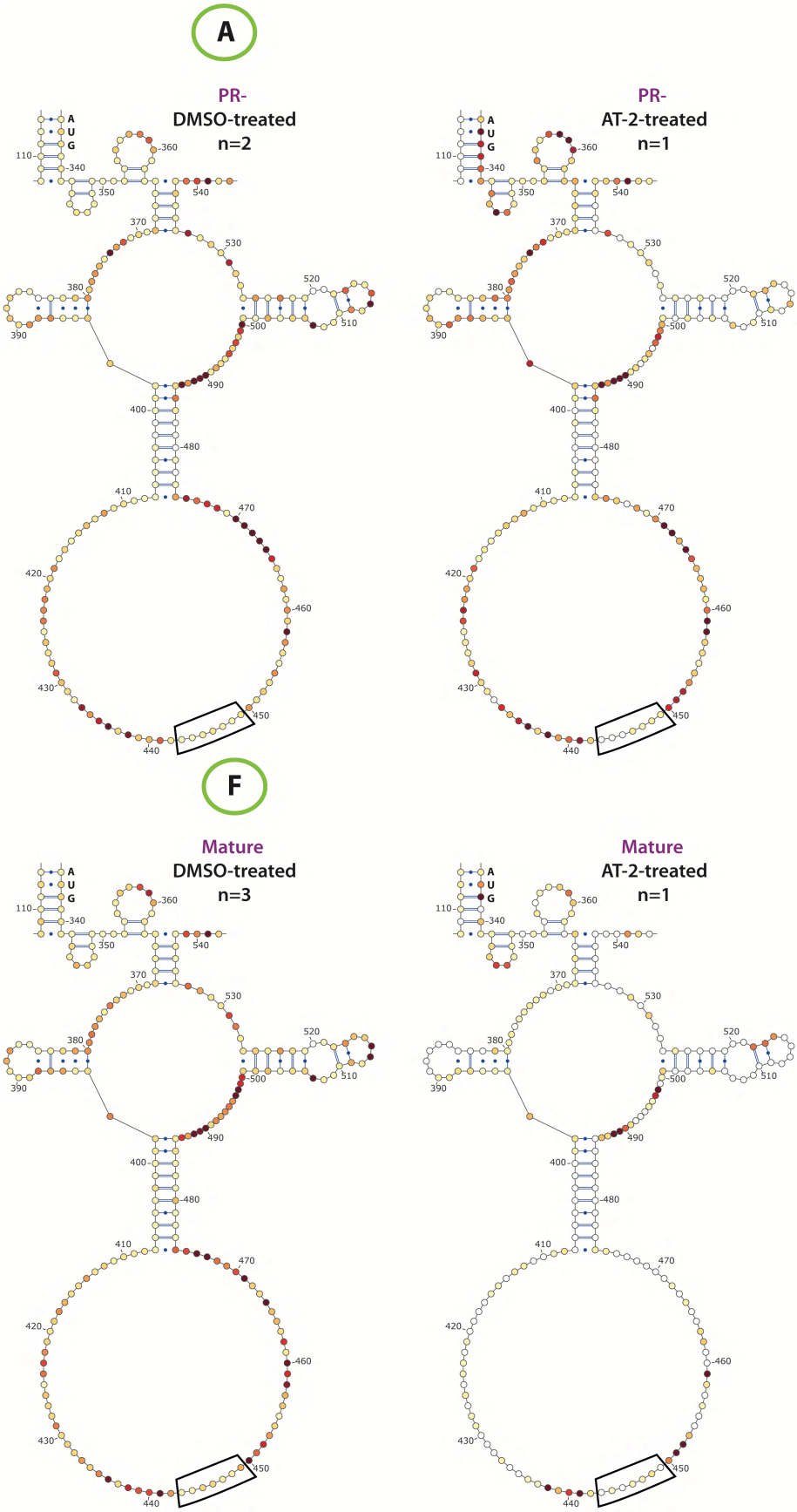


Figure 83

Effect of the AT-2 treatment on the first 200 nucleotides of the *gag* coding region of HIV-1 gRNA from PR- and mature particles.

SHAPE reactivity, obtained by hSHAPE-Seq, of PR- and mature samples are placed on a published model (426). Virus particles were treated with AT-2 or DMSO as a negative control prior to NMIA modification. The heptanucleotide sequence implicated in the pseudoknot formed between the Poly-A apical loop (positions G79-C85) and nucleotides in the matrix coding region (positions G443-C449) is represented by the black box.

The reactivity profile of nucleotides G369-A379 evolves during the Pr55^{Gag} proteolytic cascade (**Figure 84**). In PR- particles, positions U372-A374 and A378-A380 have a moderate reactivity (**Figure 82 A**) that increases for nucleotide A374 and is lost for positions A378-A379 at Step 1 and Step 1.1 stages (**Figure 82 B** and **C**). The AT-2 treatment does not change the reactivity profile of PR- particles, suggesting that nucleotides G369-A379 are unpaired (**Figure 83 A**). From Step 2 to the mature stage (Figure 35 D to F), positions U372-A374 and A378-A380 display an intermediate reactivity level, slightly higher in DMSO-treated mature samples (Figure 36 F) than in the untreated condition (**Figure 82 F**). These nucleotides become unmodified following AT-2 treatment of mature particles (**Figure 83 F**), suggesting a destabilisation effect of this region mediated by NCp7, but which was not detected by the team of K. Weeks (426). These results are in agreement with three published models describing this region as highly reactive and in a single-stranded conformation (383, 422, 426).

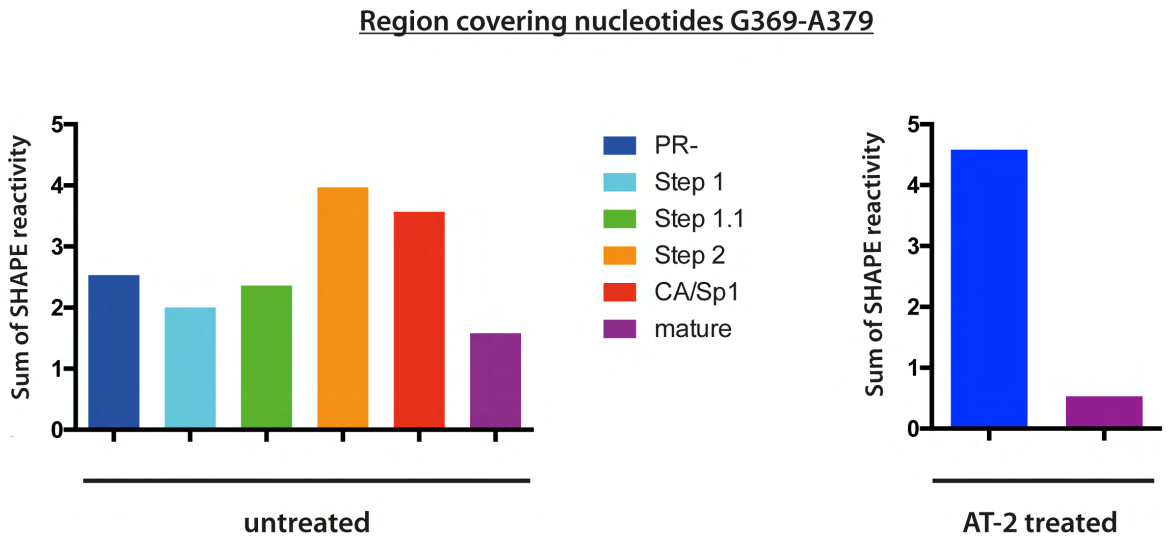


Figure 84
Evolution of the reactivity of nucleotides G369-A379 of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides G369-A379 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

In addition, nucleotides A387-U397, which are equally reactive in untreated PR- viral particles, treated with DMSO or AT-2 (**Figure 83 A**), and in Step 1 to CA/Sp1 viral particles (**Figure 82 B to E**) have a decreased reactivity profile following AT-2 treatment in mature particles (**Figure 83 F**). This region, which is involved in a stem-loop conformation (116, 353, 426) or a long-distance interaction (383, 422) in published models, could also be destabilised by NCp7.

The two GC-rich regions comprising nucleotides A398-G407 and U476-U485 are not reactive at any maturation stage (**Figure 82 A to F**) and their reactivity profile is unchanged following AT-2 treatment of PR- and mature particles (**Figure 83 A** and **F**). These two regions interact together in all existing HIV-1 secondary structure models (116, 353, 383, 422, 426), in agreement with our hSHAPE-Seq data. The region comprising positions G408-A475 is highly modified at all mature stages (**Figure 82 A to F**). The reactivity profile of DMSO-treated mature particles (**Figure 83 F**) is slightly higher than the untreated condition (**Figure 82 F**) and is in better agreement with results obtained for the processing intermediates and published data (116, 353, 383, 422, 426).

The significant decrease of reactivity at positions A414-U436 and A454-U476 in AT-2-treated mature particles suggests that these regions are destabilised by NCp7 (**Figure 82 F** and **85**). In addition, positions G443-C449, described to be involved in the pseudoknot with the Poly-A apical loop (G79-C85) are unmodified from the immature to the Step 1.1 stage (**Figure 82 A to C**). From the Step 2 stage to the completion of maturation, nucleotides C445-A446 becomes weakly modified and C449 highly reactive (**Figure 82 C to F**). This increased reactivity (**Figure 86**) profile is consistent with the models published by the team of K. Weeks (422, 426) and further questions the involvement of C449 in the pseudoknot interaction in the mature conformation. In addition, 4-5 highly reactive nucleotides flank the pseudoknot, except at Step 1 and Step 1.1 stages (**Figure 82 A to F** and **87**). Nucleotides U439-G442 and A450-G453 are highly accessible, probably due to the long-distance interaction of nucleotides G443-C449 with the Poly-A apical loop, imposing steric constraints.

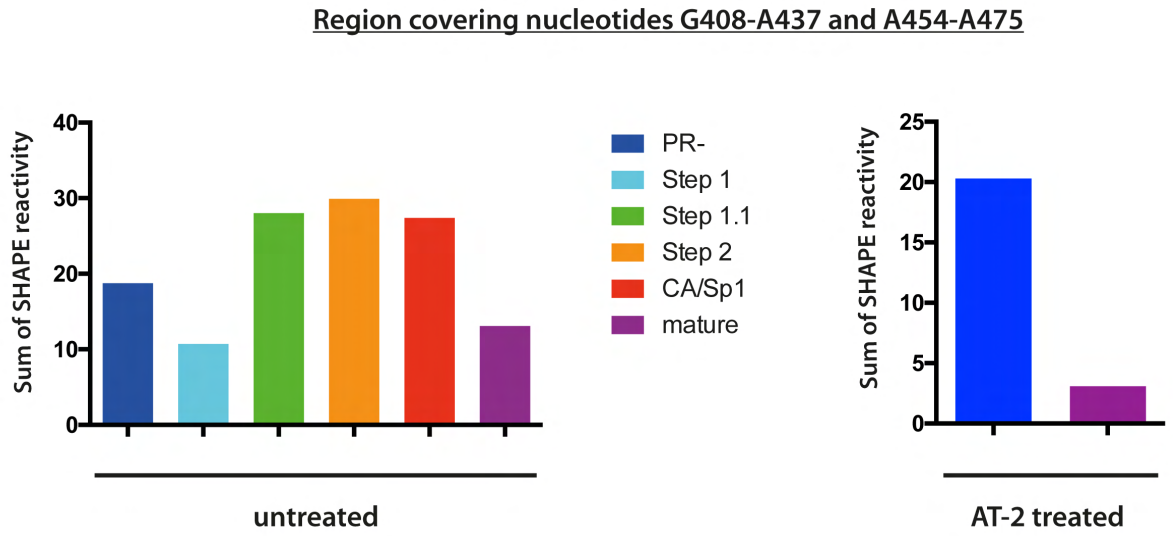


Figure 85
Evolution of the reactivity of nucleotides G408-A437 and A454-A475 of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides G408-A437 and A454-A475 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

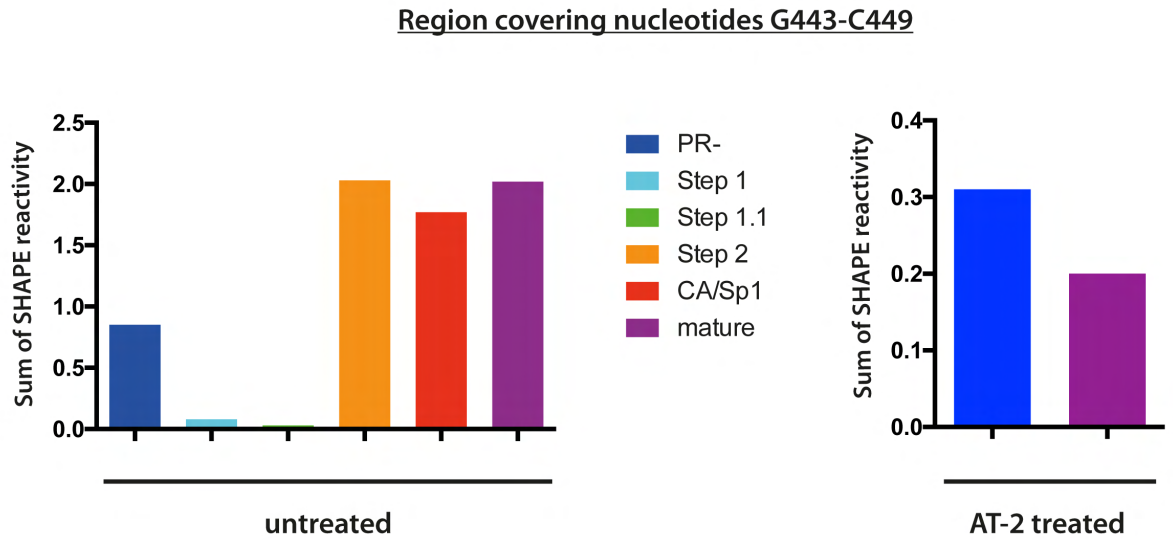


Figure 86
Evolution of the reactivity of nucleotides G443-C449 of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides G443-C449 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

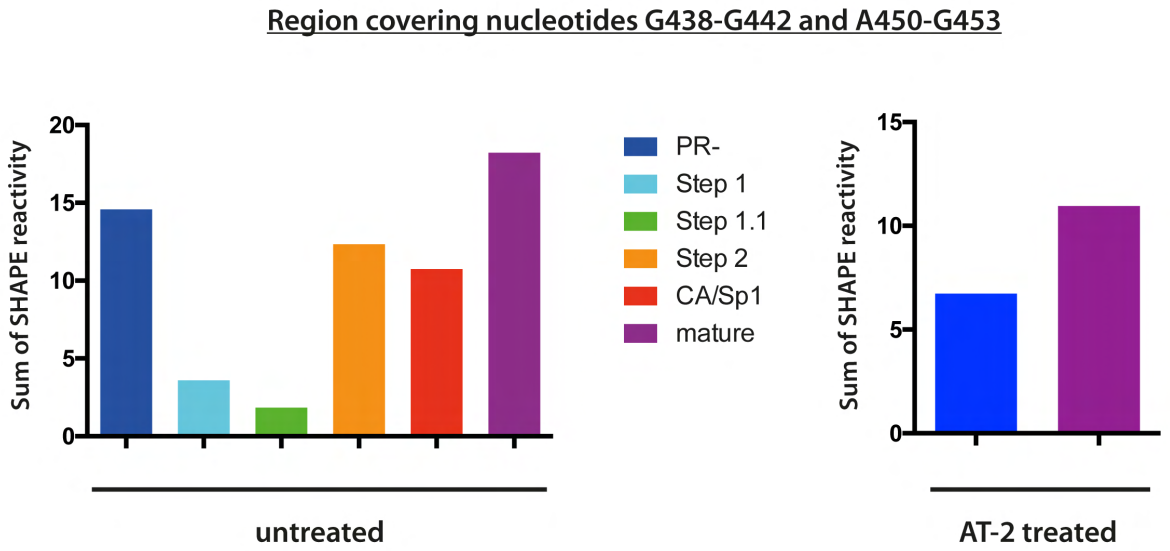


Figure 87
Evolution of nucleotides G438-G442 and A450-G453 of HIV-1 gRNA during the proteolytic cascade of Pr55^{Gag}.
The sum of SHAPE reactivity from nucleotides G443-C449 has been calculated for PR-, intermediate and mature samples in addition to AT-2-treated PR- and mature samples.

The G501-U526 region presents a moderate reactivity profile with positions A507, A513-U514 and A516-G519 highly reactive (**Figure 82 A to F**). Interestingly, this reactivity profile is significantly decreased following AT-2 treatment of PR- and mature particles (**Figure 83 A** and **F**); this could reflect destabilisation of this region by both Pr55^{Gag} and NCp7. Both the U486-A500 and A527-A533 regions are highly modified in PR- particles, without or with AT-2 treatment (**Figure 82 A** and **83 A**). Nucleotides U486-A500 remain highly reactive during the whole maturation process (**Figure 82 B to F** and **88**). The slight decrease of reactivity after AT-2 treatment of mature particles could reflect a NCp7 destabilising activity, which was however not previously reported (426). The reactivity profile of positions A527-A533 is increased in DMSO-treated mature particles (**Figure 83 F**) compared to the untreated condition (**Figure 82 F**), rendering the interpretation of AT-2-treated samples difficult.

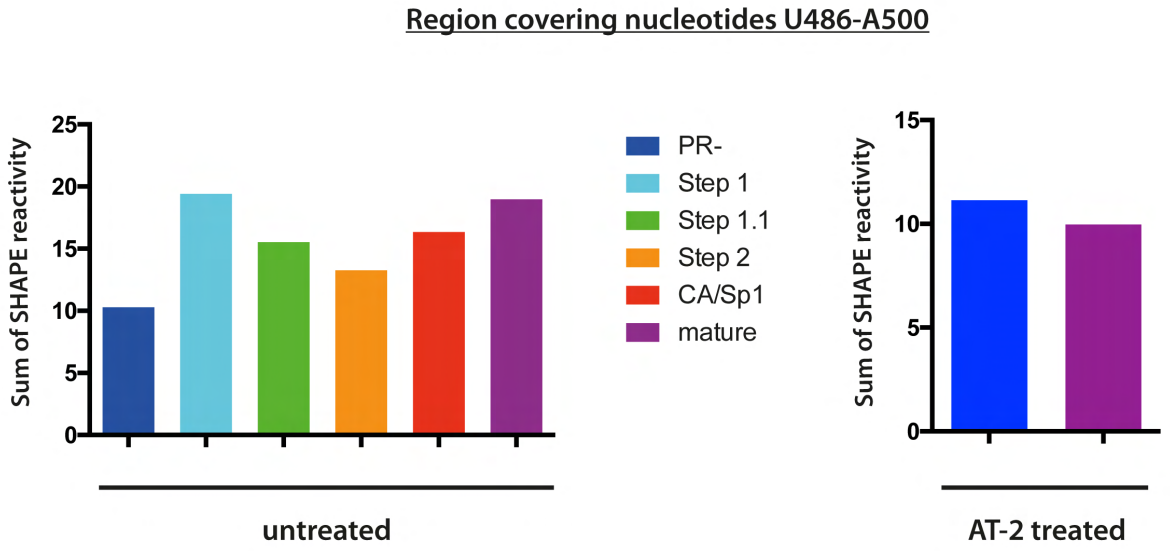


Figure 88
Evolution of the reactivity of nucleotides U486-A500 of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides U486-A500 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

IV. Mechanism of action of protease inhibitors

Another goal of my PhD project was to assess the impact of PIs on the structural maturation of the HIV-1 gRNA. Indeed, we hypothesise that PIs might inhibit viral replication by interfering with the genome structural maturation. In order to test this assumption, viral particles were produced by transfection of HEK-293T cells with PEI in the presence of PIs at the desired concentration (stock in DMSO). These pNL43 particles are pseudotyped with the protein G from the Stomatitis vesicular virus. Infectivity of virus particles in the presence of a concentration range of Lopinavir (LPV) or Atazanavir sulfate (ATVs), was assessed with TZM-BI reporter assay, as previously described (Experimental procedure III. Protease inhibitors). Mature and PR- viral particles in absence of PIs have been produced in parallel. Infectivity of viral particles produced in the presence of PIs is measured based on the luciferase reporter gene expression after a single round of virus infection (**Figure 89**). This reporter gene is under the control of the HIV-1 LTR, and thus requires infection, RTion, integration, transcription and synthesis of the Tat protein to be activated. A WB analysis of virus particles was conducted in parallel to visualise the proteolytic cascade of $Pr55^{Gag}$ and $Pr160^{GagPol}$ (**Figure 89**). The different characteristics of viral particles produced with LPV or ATVs are summarised in **Figure 90**.

I determined an inhibition of HIV infectivity with TZM-bl indicator cells at an IC_{50} of 7.2 nM for LPV and 2.1 nM for ATVs. The nanomolar range of IC_{50} for LPV and ATZs is consistent with other published reports (291, 295, 352, 376) with slight differences due to differences in host cells and experimental assays.

Regarding the extent of proteolytic cleavages in the presence of PIs, no detectable change was detected in overall at 6.6 nM of LPV and 2.6 nM ATVs, corresponding to the IC_{50}. An important inhibition of processing is noticed from 41 nM LPV and 16.4 nM ATVs. Importantly, virus infectivity is inhibited at much lower PI concentration than $Pr55^{Gag}$ and $Pr160^{GagPol}$ processing. At IC_{50}, protein precursors are indeed processed (almost) normally (**Figure 89**).

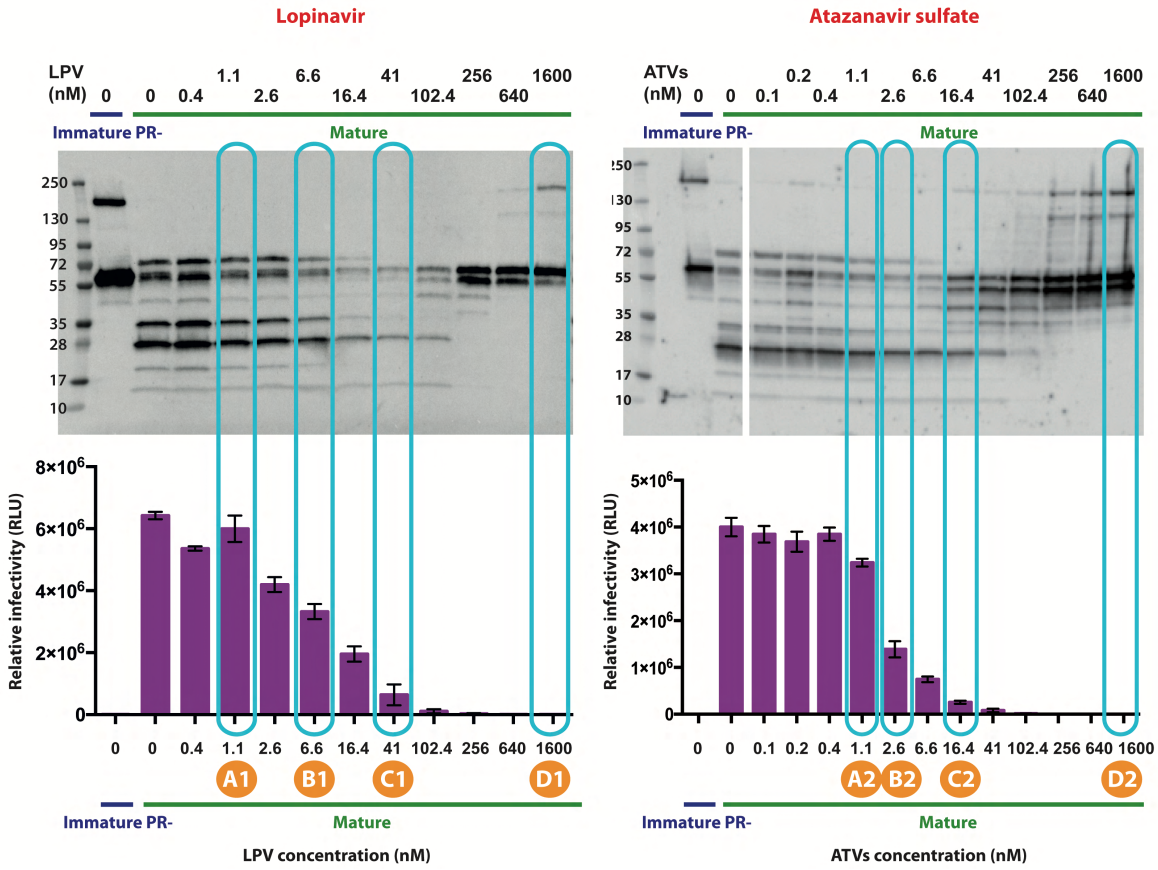


Figure 89
Influence of PIs on proteolytic processing and infectivity of viral particles.
Viral particles were produced in the presence of different concentrations of LPV (0.4-1600 nM) or ATVs (0.1-1600 nM) (PI in DMSO), in addition to PR- and mature virus particles in the presence of DMSO. LPV and ATV effect on $Pr55^{Gag}$ and $Pr160^{GagPol}$ processing was visualised by WB using a patient serum, and on infectivity by Luciferase assay. Infectivity values are mean values and SEMs from biological samples in triplicate, with a technical triplicate for each biological sample. PI concentrations of LPV and ATZs selected for hSHAPE-Seq analysis are highlighted (respectively A1 to D1 and A2 to D2).

The overall processing profile is similar to the mature sample at 41 nM of LPV and 16.4 nM of ATVs (**Figure 89**) even though these concentrations correspond to 5.7 and 7.8-fold the IC_{50} of these compounds. With a small influence of PIs on proteolytic processing at concentrations inhibiting at least 50% of viral particles, we hypothesised a potential effect of PIs on gRNA maturation, which might be affected by small amounts of intermediate products acting as trans-dominant inhibitors (304). I selected different concentrations of LPV and ATVs for further analysis by the hSHAPE-Seq i) with no impact on both viral infectivity and proteolytic processing (**Figures 89, 90 A1** and **A2**), ii) inhibiting viral infectivity with no visible effect on proteolytic processing (**Figures 89, 90 B1** and **B2**) and iii) impacting both viral infectivity and proteolytic processing with the presence of $Pr55^{Gag}$ processing intermediates (**Figures 89, 90 C1** and **C2**) or iv) with unprocessed $Pr55^{Gag}$ and $Pr160^{GagPol}$ (**Figures 89, 90 D1** and **D2**).

The reactivity profiles of the first 550 nucleotides of the HIV-1 gRNA of viral particles treated with increasing concentrations of LPV and ATVs obtained by hSHAPE-Seq are presented in **Annexes 16 to 19** for LPV and **Annexes 20 to 23** for ATVs.

VSV-G pseudotyped particles	PIs / DMSO	PI concentration (nM)	Overall processing profile of $Pr55^{Gag}$ and $Pr160^{GagPol}$	Infectivity	hSHAPE-Seq samples
'wild type' Δ*env*	DMSO		fully processed	> 90 %	
'wild type' Δ*env*	LPV ATVs	0.4, 1.1, 2.6 0.1, 0.2, 0.4, 1.1	fully processed	> 80 %	1.1 A1 1.1 A2
'wild type' Δ*env*	LPV ATVs	6.6 2.6	fully processed	~ 50 %	6.6 B1 2.6 B2
'wild type' Δ*env*	LPV ATVs	16.4 6.6	fully processed	< 30 %	
'wild type' Δ*env*	LPV ATVs	41, 102.4, 256 16.4, 41, 102.4, 256	*Presence of p41 intermediates *Disappearance of p32 IN, p17 MA, p24 CA	< 10 %	41 C1 16.4 C2
'wild type' Δ*env*	LPV ATVs	640, 1600 640, 1600	Not processed	0 %	1600 D1 1600 D2
PR-	DMSO		Not processed	0 %	

Figure 90
Influence of increasing concentrations of LPV and ATVs on proteolytic processing and infectivity of viral particles.
Virus particles were produced in presence of LPV (from 0.4 to 1600 nM) or ATVs (from 0.07 to 1600 nM). The effect of these PIs on $Pr55^{Gag}$ and $Pr160^{GagPol}$ on processing and infectivity was analysed by WB and Luciferase assay, respectively. PI concentrations selected for hSHAPE-Seq analysis are indicated.

The reactivity profiles obtained with LPV and ATVs are similar with only slight differences in the PBS loop, as positions C179-G181 and A214-A216 are highly modified in the presence de LPV but unreactive in the presence of ATVs (**Figure 91**). This difference is observed at all LPV concentrations (**Annexes 16 to 19**). In addition, the SL4 domain is moderately modified at 1.1 nM of LPV whereas this domain is not modified at 1.1 nM ATVs, neither at the other tested concentrations of PIs (**Annexes 16 to 23**). Moreover, the gRNA conformation does not evolve when increasing the PIs concentration of LPV or ATVs. The first 200 nucleotides of the coding region highlight the similar reactivity profiles of these nucleotides from 1.1 to 1600 nM final concentration of LPV (**Figure 92**). In addition, the gRNA profile obtained at 1.1 nM LPV (**Figure 92 A1**) is different from the immature and mature conformations (**Figure 92 A** and **F**). Thus, these results indicate the impossibility for gRNA maturation to further proceed to completion in the presence of PIs. Interestingly, the gRNA conformation at 1.1 nM LPV resembles the Step 2 and CA/Sp1 stages (**Figure 93 D** and **E**). These elements suggest that PIs induce a processing defect similar to the notice at Step 2 or CA/Sp1 stages.

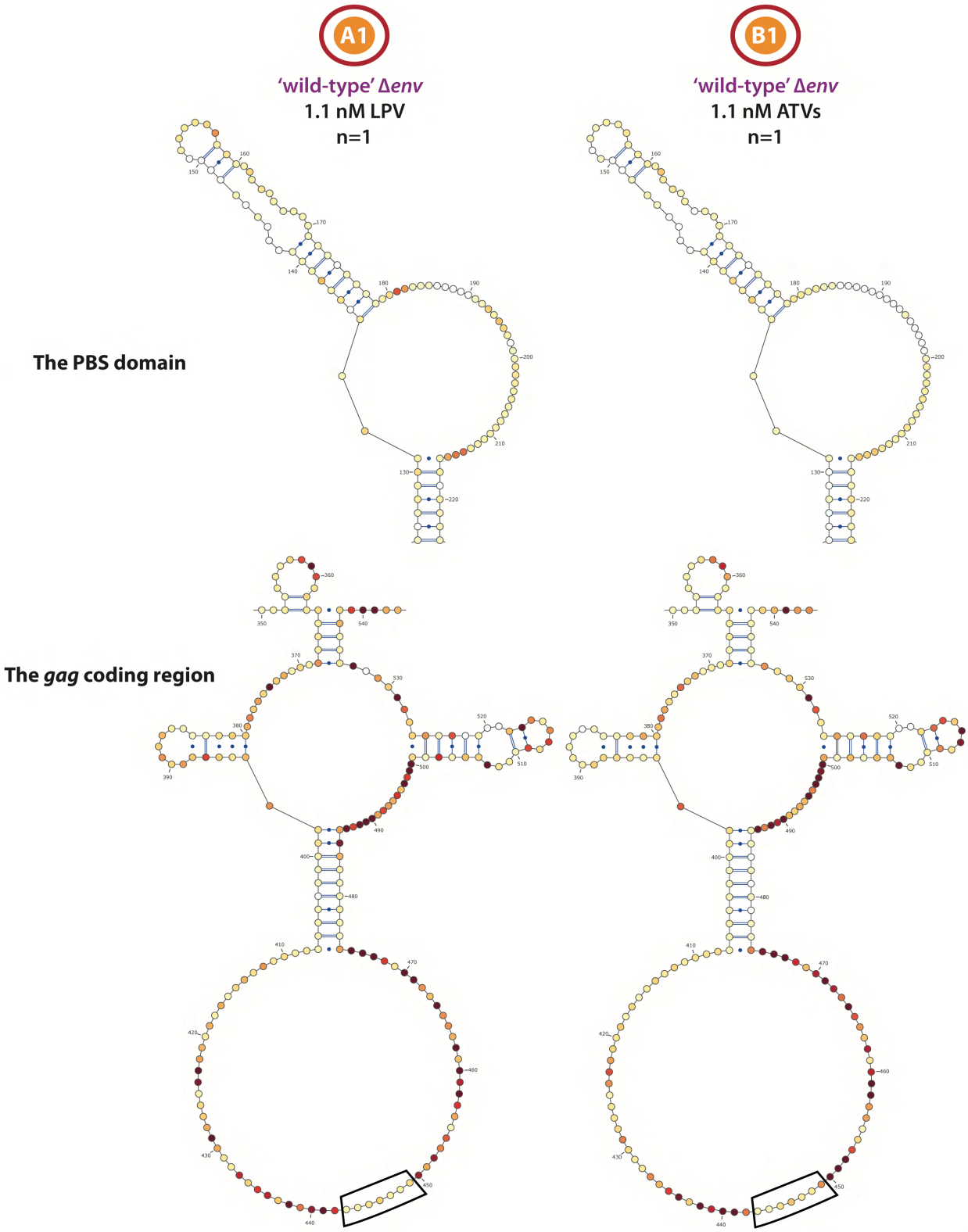


Figure 91
Comparison of the reactivity profile of gRNA with LPV and ATVs.
Cells were transfected with 'wild type' Δ*env* HIV-1 clones in the presence of 1.1 nM LPV or ATVs. SHAPE reactivity, obtained by hSHAPE-Seq, of viral particles produced in the presence of 1.1 nM LPV or ATVs LPV and ATVs-treated samples are represented on a published model (426). The reactivity profile of the PBS domain and positions G350-A543 at the beginning of the *gag* coding region with LPV (A1) and ATVs (B1) are compared.

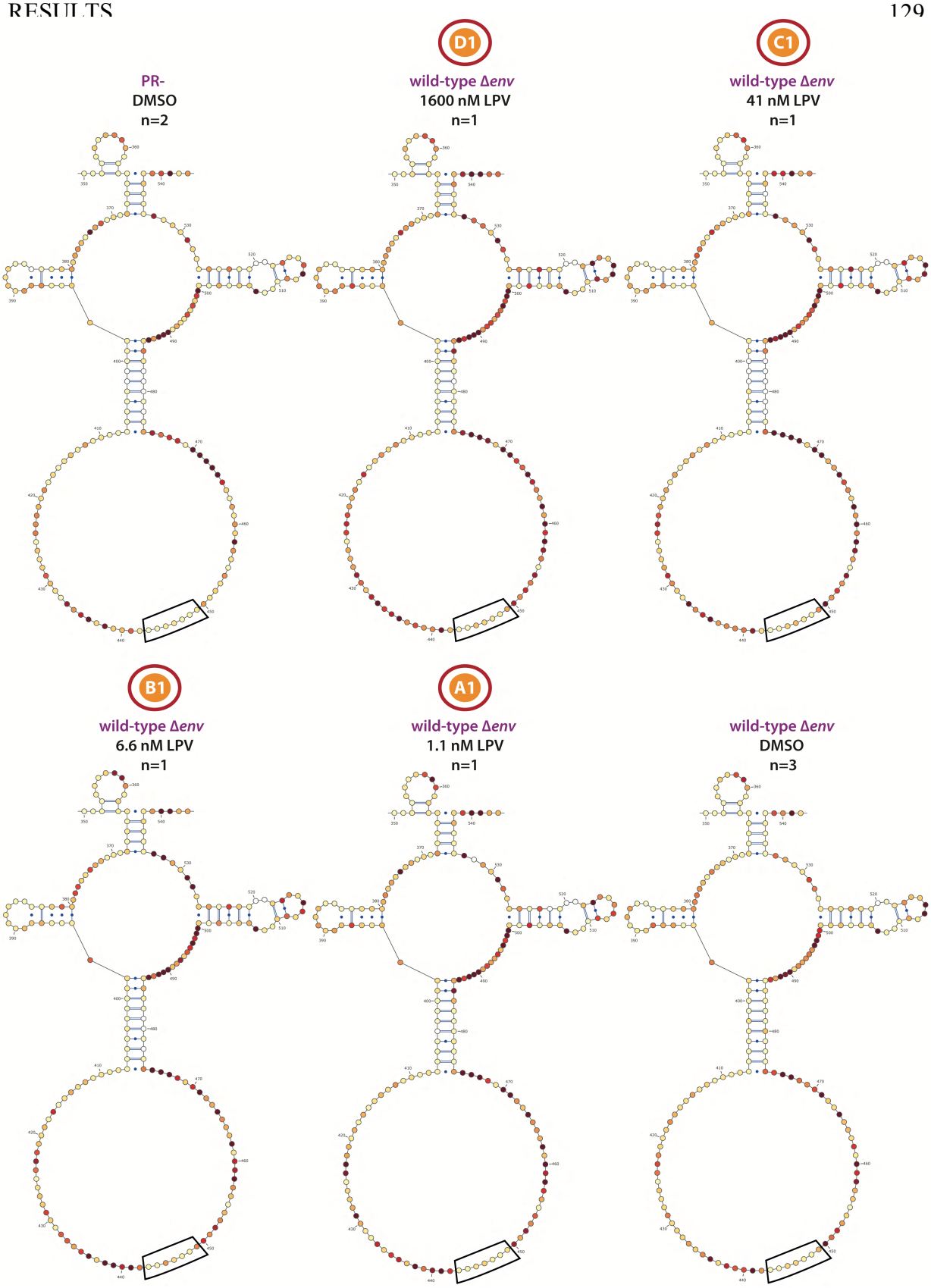


Figure 92

Evolution of the first 200 nucleotides of the *gag* coding region of HIV-1 gRNA in presence of an increasing concentrations of LPV.

PR- Δ*env* viral particles and 'wild type' Δenv viral particles were produced in the presence of 1.1 (A1), 6.6 (A2), 41 (A3), 1600 nM (A4) LPV or DMSO. SHAPE reactivity obtained by hSHAPE-Seq are presented on a published model (426). The reactivity profile of positions G350-A543 at the beginning of the *gag* coding region are compared.

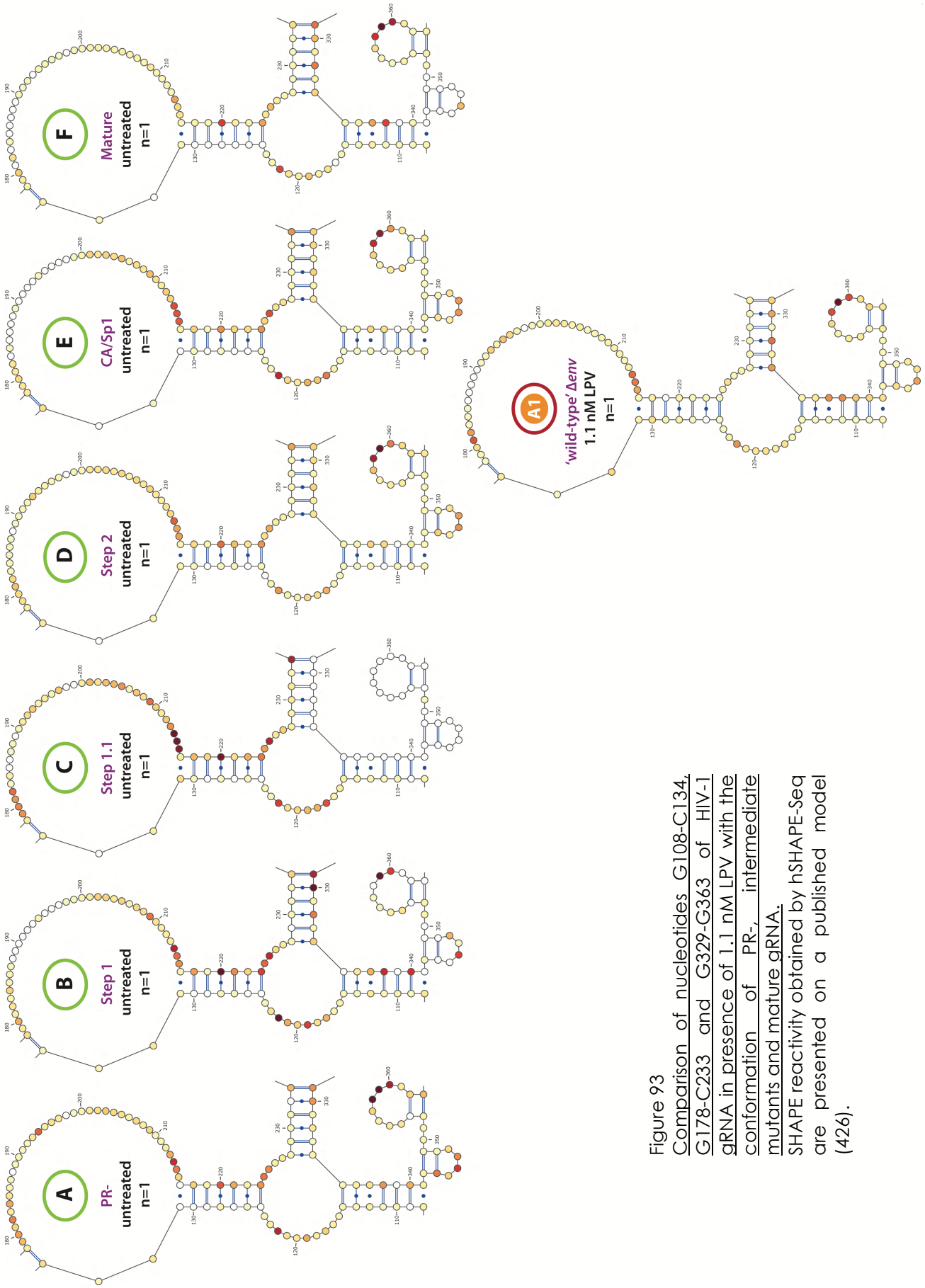


Figure 93
Comparison of nucleotides G108-C134, G178-C233 and G329-G363 of HIV-1 gRNA in presence of 1.1 nM LPV with the conformation of PR-, intermediate mutants and mature gRNA.
SHAPE reactivity obtained by hSHAPE-Seq are presented on a published model (426).

SUMMARY AND DISCUSSION

SUMMARY AND DISCUSSION

The purpose of my thesis was to determine the different steps leading to the formation of the mature conformation of the dimeric gRNA and to better understand the link between proteolytic processing and structural maturation of the genome. To this end, I developed an *in viro* chemical probing technique, termed hSHAPE-Seq. Using this approach, I analysed the first 550 nucleotides of the HIV-1 genome of the wild type virus and several mutants mimicking the sequential processing of $Pr55^{Gag}$. In the case of PR- and mature viral particles, the gRNA structure has also been studied after treatment with the AT-2 zinc ejector in order to determine $Pr55^{Gag}$ and NCp7 binding sites, as well as the gRNA destabilising activity of these proteins. In addition, the gRNA secondary structure has also been investigated *in vitro* in the presence and absence of $Pr55^{Gag}$, NCp15, NCp9, NCp7 and GagΔp6 proteins. Thanks to the hSHAPE-Seq approach, I also investigated the mechanism of action of PIs, which severely impact viral infectivity at concentrations that have only minor effects on proteolytic processing.

Development and validation of the hSHAPE-Seq methodology

The hSHAPE-Seq approach has been designed to determine the *in viro* accessibility of the HIV-1 genome and to follow the gRNA conformation during the proteolytic processing of $Pr55^{Gag}$. As shown in the manuscript, the hSHAPE-Seq approach indeed provided detailed structural information about the first 550 nucleotides of the HIV-1 genome inside viral particles. The hSHAPE-Seq protocol has been extensively optimised to generate the results presented in this manuscript. The main challenge was the very limited amount of material available for the non-replicative mutant virus. Therefore, several key steps of the protocol required time-consuming optimisation: i) large-scale production of PR- and intermediates viral particles, ii) AT-2 treatment, which has a dramatic effect on the RNA recovery yield, iii) RTion and iv) Illumina library preparation, including ligation of the adapters and normalisation of the samples.

The global gRNA reactivity profile of mature particles I obtained with hSHAPE-seq is globally highly consistent with the results previously published by Wilkinson et al. (426). In this study, gRNA was also modified *in viro*, but modifications were analysed by conventional hSHAPE, which uses capillary electrophoresis rather than next generation sequencing to identify their locations. Based on the SHAPE reactivity profile, we clearly showed that the 5' regulatory region is less reactive than the *gag* coding region (**Figure 94 A**).

However, unlike previous published works, our study provides a complete picture of the structural rearrangements of the 5' region of the gRNA during maturation of the viral particles. Indeed, our knowledge of the immature PR- gRNA conformation is very limited, with the analysis of the PBS domain only (375). Regarding $Pr55^{Gag}$ binding, some insights have been gained from *in vitro* studies (33, 48, 67, 127, 225, 354, 388). However, no similar *in viro* study has been performed on $Pr55^{Gag}$-containing

immature particles. Moreover, the gRNA conformation of mutants blocked at various steps of the maturation process (313) has not been determined previously. Thus, our results provide the first structural insight of the structural maturation of the gRNA that accompany the transformation of immature viral particles into infectious ones.

Two and three biological replicates were obtained for DMSO-treated immature and mature particles, respectively, and one untreated replicate has been produced for each stage. Comparing replicates, significant variability is observed in some regions even though the global reactivity profiles are consistent. One of the three DMSO-treated replicate of mature samples significantly differs from the two others and also appears inconsistent with the untreated sample; thus this sample should probably be removed from the analysis. This variability likely results from the fact that these gRNA samples were collected from different viral productions, and suggests that some regions of the gRNA structure might adopt alternative folds. Nevertheless, hSHAPE-Seq is overall highly reproducible, as highlighted by comparing the reactivity profiles obtained with the four PI concentrations (**Annexes 16 to 23**).

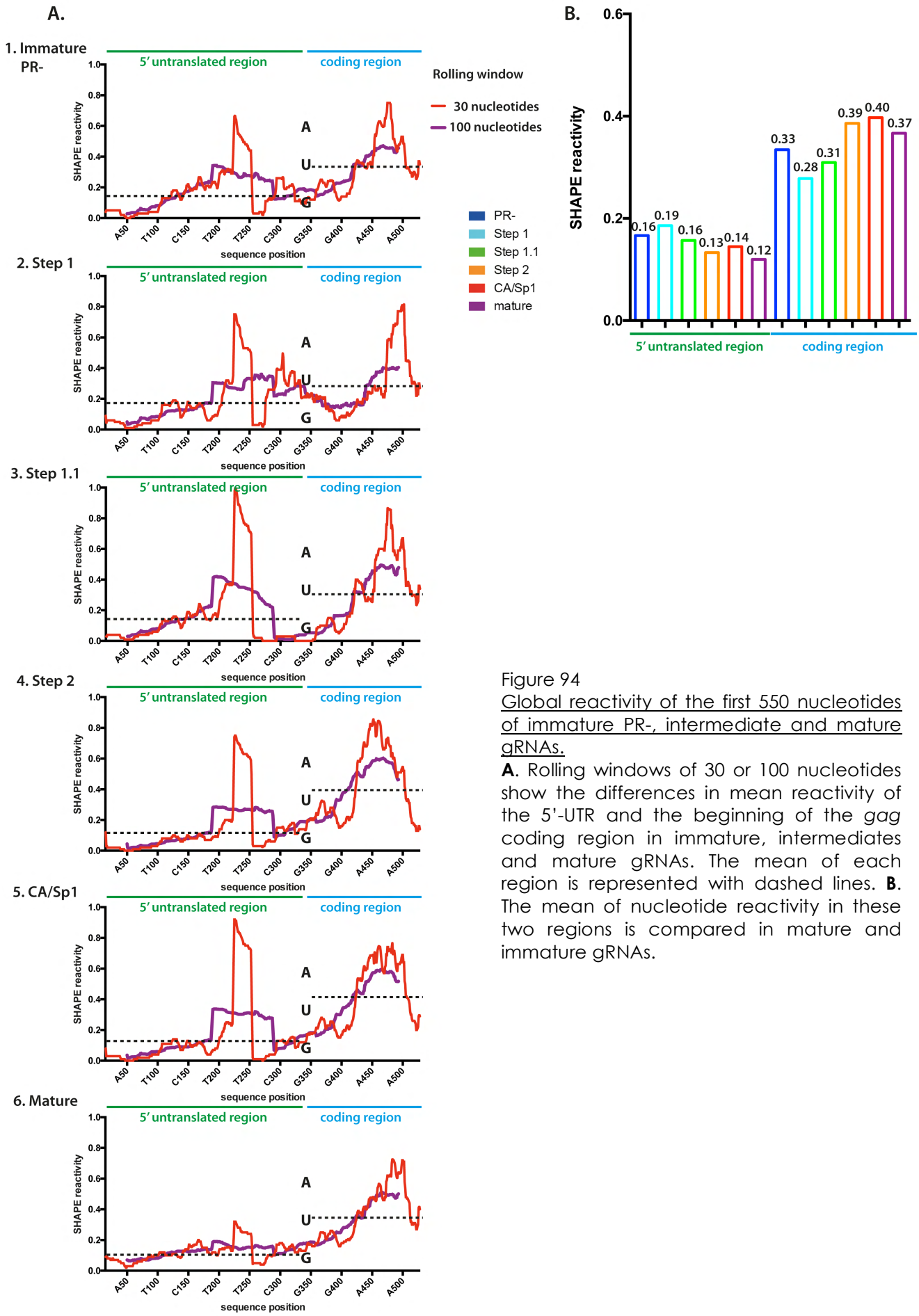


Figure 94
Global reactivity of the first 550 nucleotides of immature PR-, intermediate and mature gRNAs.
A. Rolling windows of 30 or 100 nucleotides show the differences in mean reactivity of the 5'-UTR and the beginning of the *gag* coding region in immature, intermediates and mature gRNAs. The mean of each region is represented with dashed lines. **B**. The mean of nucleotide reactivity in these two regions is compared in mature and immature gRNAs.

The HIV-1 genomic maturation process

I. Overall view of the *in viro* gRNA conformation of PR- and mature particles

The hSHAPE-Seq reactivity profiles are quite similar, but with some differences mostly located in the PBS domain and the *gag* coding region. However, AT-2 treatment reveals major differences between PR- and mature gRNAs, showing that the gRNA structure is indeed different between these two states and that $Pr55^{Gag}$ and NCp7 differently bind and affect it. Indeed, this treatment is of particular interest and allowed us to detect ZF-containing protein interactions with gRNA. Disrupting the RNA-ZF protein interactions results in both increases and decreases in reactivity. A difference plot comparing DMSO and AT-2-treated states is presented in **Figure 95** and highlights both binding sites (DMSO-treated < AT-2-treated nucleotide reactivity), and protein-induced RNA structure destabilisation (DMSO-treated > AT-2-treated nucleotide reactivity).

Upon AT-2 treatment, the reactivity of several regions evolves similarly in immature and mature viral particles (**Figure 95**, regions highlighted in red). These regions are mostly located in the packaging signal and protected by both $Pr55^{Gag}$ and NCp7. By contrast, $Pr55^{Gag}$ and NCp7 differently influence the gRNA conformation in the PBS domain and in the *gag* coding region (**Figure 95**, regions highlighted in green). The PBS domain becomes less accessible upon AT-2 treatment of the immature PR- viral particles, implying a destabilising effect of $Pr55^{Gag}$, while some regions are protected by NCp7 in the mature viral particles. The *gag* coding region is strongly destabilised by NCp7, while $Pr55^{Gag}$ binding has limited effect on this region.

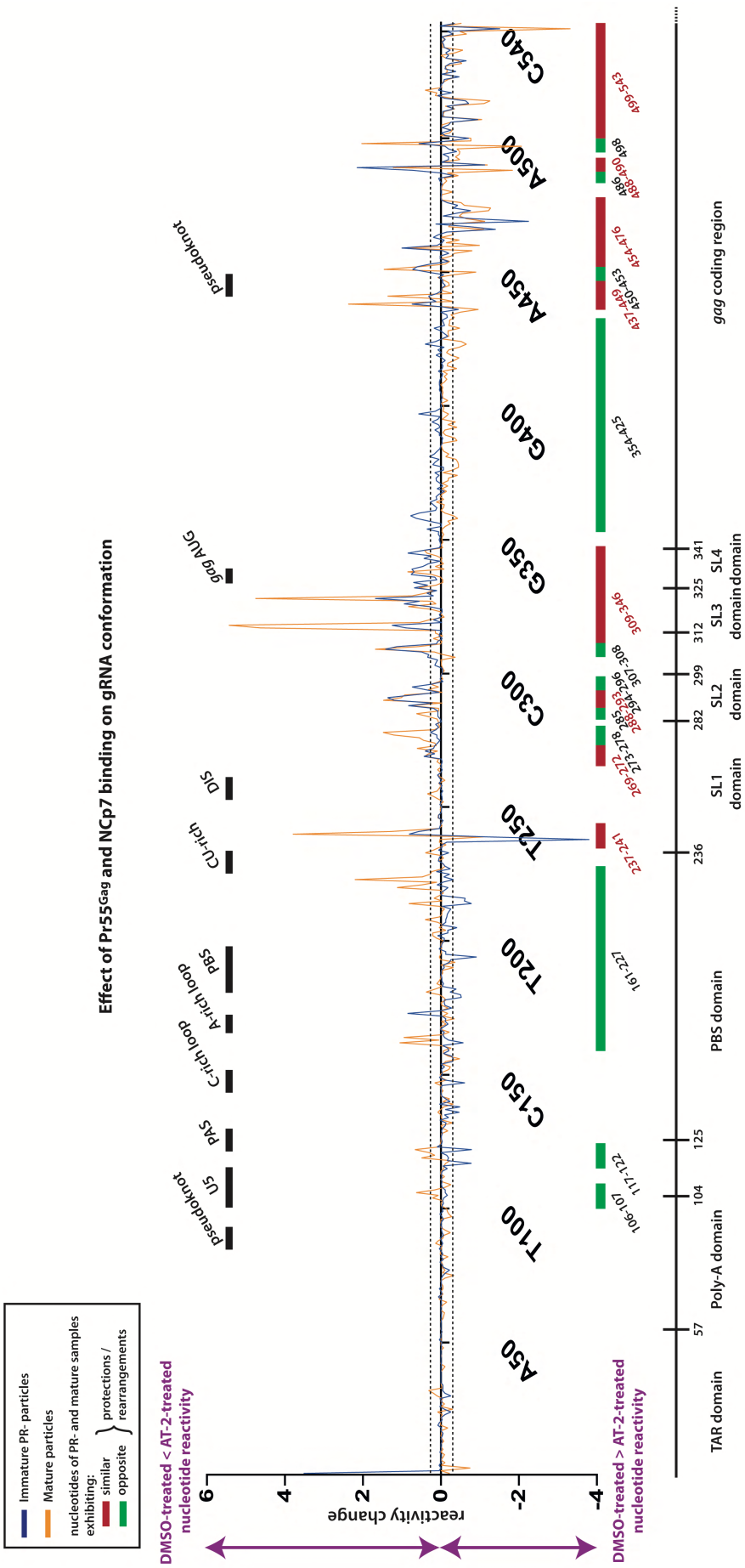


Figure 95
Effect of Pr55Gag and NCp7 binding on gRNA conformation.
Difference plot illustrating SHAPE reactivity changes upon AT-2 treatment, of PR- (blue) and mature (orange) particles, respectively. Reactivity change was calculated by subtracting the average of DMSO-treated samples from Positive peaks indicate protein binding sites, whereas reduced reactivity reflects destabilisation of the RNA structure by the protein. Sites with a similar reactivity change in immature and mature viral particles are highlighted in red and sites showing discordant changes are represented in green. A threshold at 0.2 and -0.2 is represented by a dashed line.

II. Do Pr55^{Gag} and NCp7 have a common consensus binding motif?

The nucleotide content of Pr55^{Gag} and NCp7 footprints has been compared thanks to the AT-2 treatment of PR- and mature particles (**Figure 96**). Ten Pr55^{Gag} binding footprints are present in the first 550 nucleotides of the HIV-1 genome; they correspond to a motif comprising approximately three nucleotides with a highly biased nucleotide composition at the third position, which might be either a G or an A residue. Sequence conservation at the first and second positions is poor with an equal relative frequency of A-, U- and G-residues. NCp7 footprints are one (12 sites), two (5 sites) or three (8 sites) nucleotide long. The predominance of A- and G-residues at the third position of the 3-nucleotide NCp7 footprint is similar that observed with Pr55^{Gag}. The predominance of G-residues at the end of the motif is also found in 1- and 2-nucleotide NCp7 footprints. The first and second positions of the 3-nucleotide NCp7 footprints are weakly conserved, but G-residues are relatively more frequent than in Pr55^{Gag} footprints. Thus, Pr55^{Gag} tightly binds A- and G-rich sites, whereas a higher proportion of G-residues is required for NCp7 binding.

The motifs of the HIV-1 gRNA that bind Pr55^{Gag} and NCp7 during packaging, assembly and maturation has been analysed by PAR-CLIP (246). Consistent with our data, Kutluay et al. showed that Pr55^{Gag} preferentially binds to A-rich motifs composed of five nucleotides with at least one G-residue. Pr55^{Gag} footprints determined by hSHAPE-Seq are smaller (three nucleotides) than the binding sites observed by PAR-CLIP but these differences are probably explained by the different experimental approaches, as hSHAPE-Seq only detects nucleotides that are protected by Pr55^{Gag}, while the binding sites may involve a larger region potentially including base-paired nucleotides. The G-rich consensus in NCp7 binding sites is in agreement with NMR structures obtained with NCp7 (16, 112, 413) as well as the *in viro* chemical probing data published by the team of K. Weeks (426). A NCp7 preference for GU-rich motifs is proposed in Kutluay et al. (2014) but, looking at their binding motifs, a predominance of G-residues is observed with one U-residue for two G-residues (246).

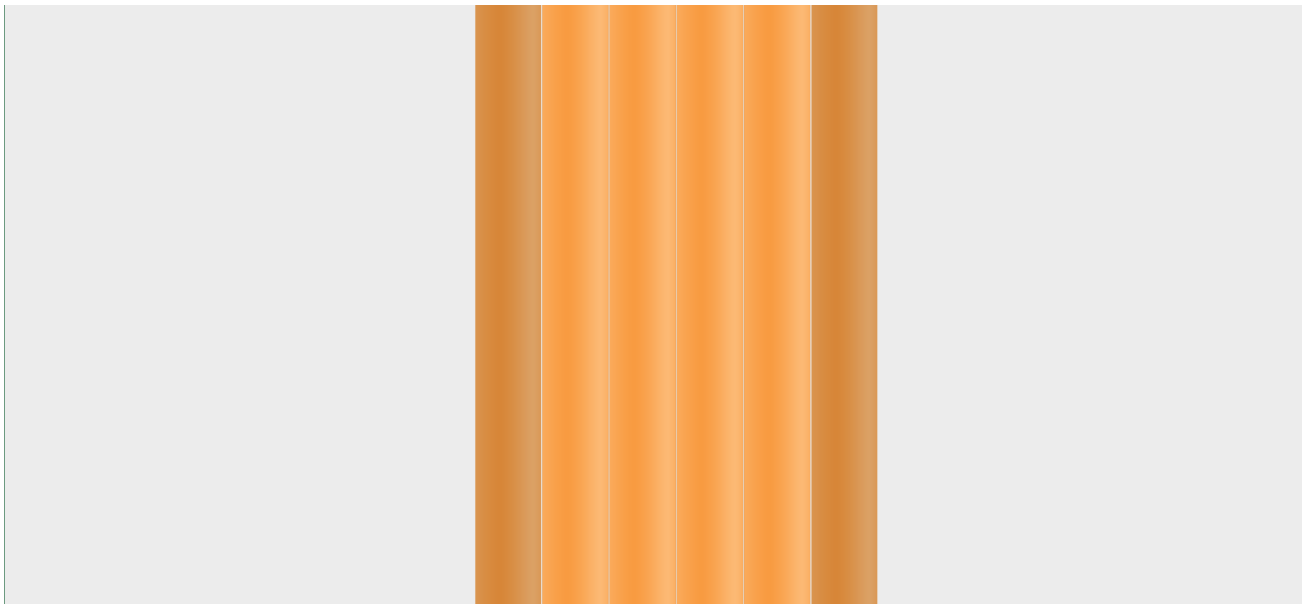


Figure 96
A sequence consensus for Pr55^{Gag} and NCp7 binding sites.
Logo plots of nucleotide sequences exhibiting an increased reactivity upon AT-2 treatment, in the case of PR- and mature particles, have been generated (http://weblogo.berkeley.edu/logo.cgi). The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of nucleotides at that position.

III. How does the gRNA structure evolve during Pr55^{Gag} processing? What is the impact of Pr55^{Gag} and its processing intermediates on gRNA maturation?

The four mutant viruses mimicking the proteolytic processing of Pr55^{Gag} display a coherent conformational evolution of the PBS domain, the packaging signal and the *gag* coding region. Moreover, this evolution correlates with the gRNA conformation in the immature PR- and mature viral particles. Step 1 and Step 1.1 stages exhibit the most important gRNA rearrangements. From Step 2 stage, this dynamic is lost and the gRNA conformation does not seem to be affected at the CA/Sp1 stage, as no clear changes are observed in our hSHAPE-Seq data. The final conformation is obtained at the mature stage, with the final rearrangements of the PBS and SL4 domains (**Figure 94 A**).

At Step 1, gRNA exhibits an increased nucleotide reactivity compared to the immature PR- stage (**Figure 94 A2**). In addition, the Step 1 reactivity profile is very similar to the AT-2-treated immature profile and thus suggests that NCp15 is unable to protect the first 550 nucleotides of the HIV-1 genome. Indeed, the packaging signal is highly reactive, including nucleotides composing the SL4 domain. These results are in agreement with the fact that NCp7 and NCp7-containing intermediates have different *in vitro* RNA binding and chaperone properties. NCp9 and NCp7 both promote the formation of large protein–RNA aggregates, whereas such aggregates were not detected with NCp15. This has been demonstrated by electron microscopy

of NC-DNA complexes (292, 293), biophysical and biochemical analysis of NC-RNA interactions (420, 432), and dimerisation of gRNA (201, 211, 313). Moreover, our hSHAPE-Seq data potentially explain the previously observed impairment of genome dimerisation (201, 211, 313), RT and integration (96) in mutant viral particles in which Pr55^{Gag} processing is stopped at the NCp15 stage (313).

Surprisingly, whereas NCp15 seems unable to protect gRNA as Pr55^{Gag} do, NCp15 destabilises the PBS domain as efficiently as Pr55^{Gag}, as reactivity in the upper PBS stem, the PBS loop and the sequence complementary to the PAS site is the same in PR- and Step 1 viral particles. Destabilisation of the PBS domain is lost upon Pr55^{Gag} removal by AT-2 treatment, indicating that the conformational rearrangement is not permanent at the immature PR- stage. As NCp15 only weakly interacts with gRNA, our results suggest that rearrangements of the PBS domain could become permanent at the Step 1 stage, after cleavage of Pr55^{Gag} at the Sp1/NC site. In addition, this increased reactivity is conserved at the Step 1.1 stage, which generates NCp9, and is progressively lost as maturation proceeds through the final stages, as reflected by the gradual protection of these sites.

Interestingly, nucleotides G257-G363 become completely unreactive at the Step 1.1 stage (**Figure 97**). This region comprises the packaging signal from the DIS and the first 20 nucleotides of the *gag* coding region. Besides this region, the PBS domain follows the same tendency as observed at the Step 1 stage and the strongest reactivity site, C238-A242, is even increased (**Figure 98**).

It would be interesting to test the AT-2 treatment at the Step 1.1 stage in order to determine if nucleotides G257-G363 are protected by NCp9 or if the gRNA conformation is rearranged. Changes in the reactivity profile of this region are observed when going from Step 1.1 to Step 2, even though maturation of Pr55^{Gag} is blocked at the NCp9 state in these two maturation intermediates (**Figure 97**). The only difference between these two stages is the MA/CA cleavage at Step 2, which releases the CA/Sp1 intermediate. This cleavage upstream of CA/Sp1 is required to initiate disassembly of the immature lattice in the disassembly-assembly model of the core formation (109, 222). Whether the disassembly of the immature lattice explains the conformational remodelling of this region is presently unknown.

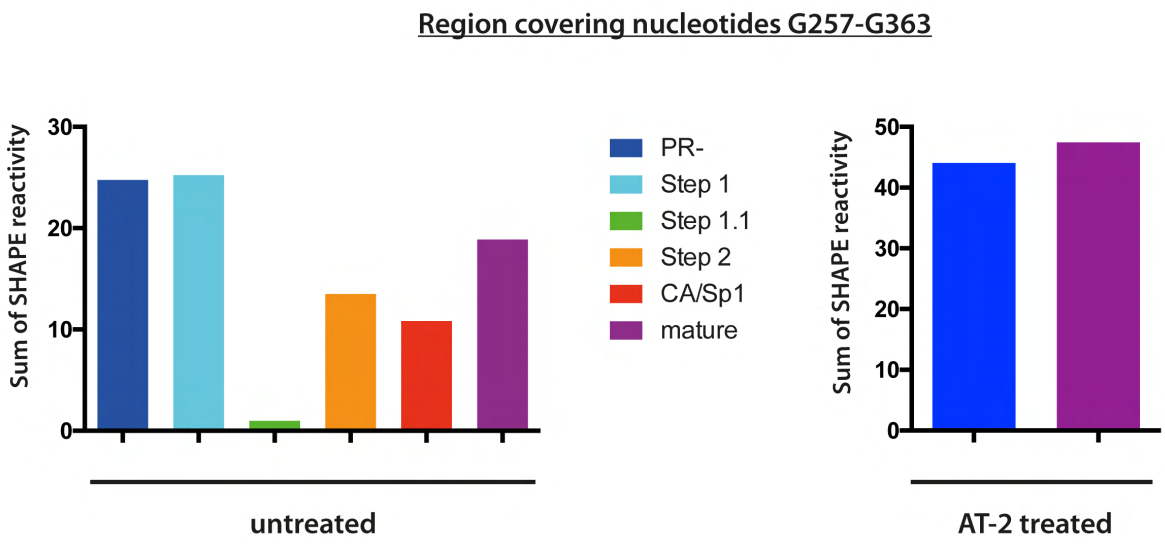


Figure 97
Evolution of the reactivity of nucleotides G257-G363 of HIV-1 gRNA during processing of Pr55^{Gag}. The sum of SHAPE reactivity of nucleotides G257-G263 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

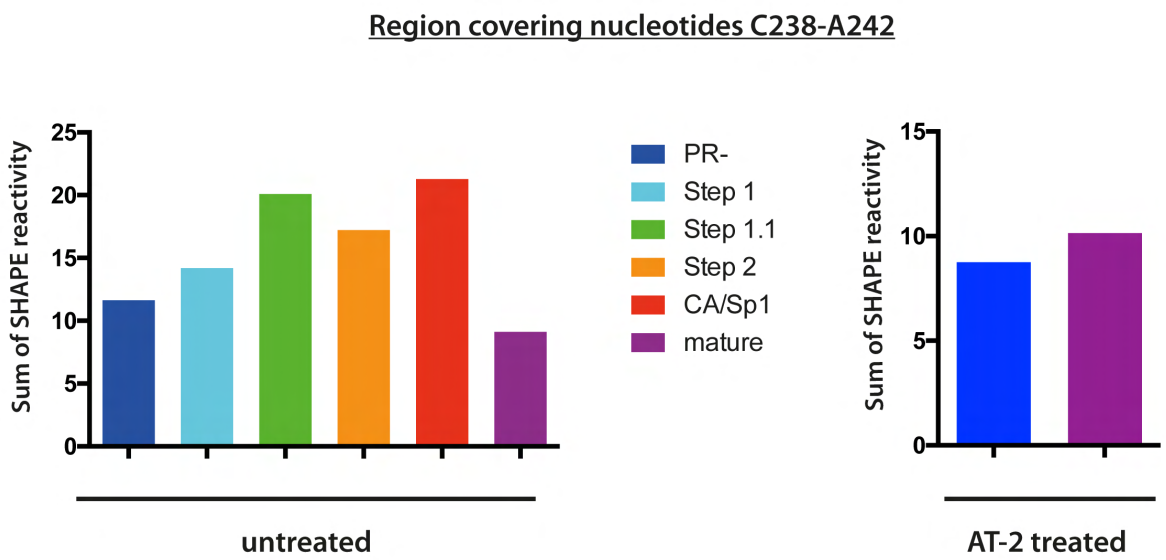


Figure 98
Evolution of the reactivity of nucleotides C238-A242 of HIV-1 gRNA during processing of Pr55^{Gag}.
The sum of SHAPE reactivity of nucleotides C238-A242 was calculated for PR-, intermediate and mature samples, as well as for AT-2-treated PR- and mature samples.

III.1 Evolution of the PBS domain

The PBS domain, which is destabilised by Pr55^{Gag} and in an "open" conformation, is still highly reactive at the Step 1 stage containing NCp15 and at the Step 1.1 and Step 2 stages containing NCp9. Reactivity decreases from the Step 2 stage, and the PBS domain is almost totally unreactive in the mature viral particles. The 18-nucleotide PBS is already annealed to the acceptor stem of tRNA^{Lys,3} at the immature PR- stage. AT-2 treatment of the immature PR- viral particles suggests that Pr55^{Gag} destabilises the gRNA secondary structure of the upper PBS stem (positions A136, A147, A138 and G162) and the PAS sequence (nucleotides A220, G221 and G223) proposed to interact tRNA^{Lys,3} with the anti-PAS so the additional interactions between tRNA^{Lys,3} and the PBS domain can already occur, at least transiently. Upon AT-2 treatment, reactivity of these regions decreases, reflecting a stabilisation of the intramolecular gRNA structure and indicating that interactions between the PBS domain and the variable loop, the anti-codon loop and the TΨC arm do not subsist in the absence of Pr55^{Gag}. However, the tRNA^{Lys,3} acceptor stem remains annealed, albeit imperfectly, with the 18 nucleotides of the PBS in the absence of Pr55^{Gag}, possibly due to the large interface of interaction and this annealing is improved at the mature stage.

Since NCp9 is present both at Step 1.1 and Step 2, the reactivity decrease of the PBS domain at the latter stage could be due to CA/Sp1 cleavage from the p41 intermediate (MA-CA-Sp1) and the subsequent initiation of core formation. This hypothesis is further supported by the unchanged reactivity profile of the PBS domain from the Step 2 to the CA/Sp1 stage, where CA/Sp1 is still present but NCp9 is processed into NCp7 and Sp2. The final structuration of the PBS domain is finally obtained through processing of the CA/Sp1 intermediate at the mature stage and the chaperone activity of NCp7. At this stage, interactions between the acceptor stem, the variable loop, the anti-codon loop and the TΨC arm of tRNA^{Lys,3} are all stable when NCp7 is dissociated by AT-2 treatment.

These observations are in agreement with the notion that $tRNA^{Lys,3}$ annealing is multi-step process, promoted firstly by $Pr55^{Gag}$, then by NCp7 after completion of proteolytic processing, and possibly also by NCp9 intermediates (71, 138, 168, 186, 375, 433). In addition, NCp7 may fine-tune the annealed $tRNA^{Lys,3}$ to the PBS domain to produce a fully functional initiation complex for RTion, which is probably not the case with NCp9. This hypothesis correlates with the infectivity of mutant CA/Sp1 viral particles, in contrast with Step 2 mutants, which are not infectious (313).

III.2 Evolution of the packaging signal

The SL1 domain is highly structured from Step 1 to the CA/Sp1 stage, in agreement with the reactivity profile of the immature PR- and mature stages. The 6-nucleotide palindromic sequence is consistently unreactive at each maturation stage and indicates that the gRNA dimerises via SL1. The lower internal loop is reactive at all maturation steps (**Figure 98**) and is the most reactive site of the first 550 nucleotides of the genome (**Figure 94 A**). At the immature stage, the upper internal loop is protected by $Pr55^{Gag}$, in keeping with the finding that this loop is the primary $Pr55^{Gag}$ binding site (4, 48, 388). Upon AT-2 treatment of the mature viral particles, not only does reactivity of the upper internal loop increase, reflecting a NCp7 binding site, but the entire region comprising positions A269-G278 becomes reactive. Nothing similar was observed in the immature viral particles. Taken together, our data strongly suggest that the SL1 intermediate stem, joining the two SL1 internal loops, does not exist at the mature stage. We suggest that this open conformation reflects the formation of an extended duplex involving the upper SL1 stem (nucleotides C248-G270). This extended duplex would induce structural constraints that would unfold the intermediate SL1 stem. This extended duplex also questions the existence of the lower SL1 stem in the extended SL1 model, due to steric constraint. Indeed, Mujeeb et al. showed that adding two base-pairs to the intermediate SL1 stem created a stable SL1 domain that did not form an extended duplex in the presence of NCp7 (303).

The apical loops of the SL2 and SL3 domains are protected by $Pr55^{Gag}$, NCp7 and presumably NCp9. The reactivity profile of the SL2 domain upon AT-2 treatment reveals the instability of the stem at the immature and the mature stage and the stabilisation of the SL2 conformation mediated by both $Pr55^{Gag}$ and NCp7.

The SL4 domain contains the *gag* start codon and is commonly drawn base-paired with nucleotides U105-G116 in U5 and the CU-rich region (positions U228-C233). This region is of particular interest because it has been proposed to regulate gRNA packaging thanks to a conformational switch of the gRNA 5'-UTR (1, 220, 269). Our hSHAPE-Seq data reveal a strong effect of $Pr55^{Gag}$ and NC-containing intermediates on reactivity of the SL4 domain rather than base-pairing of this domain. Indeed, $Pr55^{Gag}$ strongly protects the SL4 domain and this protection is lost upon the first $Pr55^{Gag}$ cleavage that releases NCp15. The SL4 domain is unreactive from the Step 1.1 to the CA/Sp1 stage, presumably protected by NCp9 and to a less extent by NCp7 at the mature stage, as shown by AT-2 treatment of the immature PR- and mature viral particles. Thus, while NCp15 is unable to efficiently bind the SL4 domain, this capacity is restored with NCp9 and NCp7. Deprotection of the SL4 domain upon dissociation of $Pr55^{Gag}$ and NCp7 precludes the interaction of this region with nucleotides U105-G116 in U5 and the CU-rich region at the immature stage and at least with the CU-rich region at the mature stage. Thus, our results are not consistent with the U5-AUG conformation proposed to be required for gRNA packaging and

published models based on results obtained *in viro* (426), *ex viro* (383, 422) and *in vitro* (191, 220). Of note, comparison of the gRNA structure without and with AT-2 treatment was instrumental for drawing our conclusions.

III.3 Evolution of the first 200 nucleotides of the gag coding region

The *gag* coding region is highly modified compared to the 5' regulatory region (**Figure 94 A**). The overall conformation is locked by two GC-rich regions, unreactive at all maturation steps, without and with AT-2 treatment, and thus most likely base-paired. Several domains are highly reactive and $Pr55^{Gag}$ does not strongly influence this region, neither by protecting nor by destabilising it. However, the reactivity of positions G369-U397 and U486-A533 is lost upon AT-2 treatment of mature particles. Thus, NCp7, and potentially NCp9, strongly destabilise these nucleotides. Regarding nucleotides G408-A475, their reactivity increases during the $Pr55^{Gag}$ proteolytic processing, except at Step 1. In addition, our hSHAPE-Seq results demonstrate a strong destabilising effect of NCp7 at the beginning of the *gag* coding region, whereas the team of K. Weeks suggested that NCp7 has limited destabilising activity on the *gag* coding region (426).

IV. Are results obtained by *in vitro* and *in viro* footprinting consistent?

In vitro footprinting assays with $Pr55^{Gag}$, GagΔp6, NCp15, NCp9 and NCp7 were performed in order to further validate hSHAPE-Seq results and to compare the properties of each protein. These results were compared with the corresponding viral mutant(s) of the $Pr55^{Gag}$ processing cascade and *in vitro* and *in viro* results were largely similar. Some discrepancies regarding for example the SL2-SL3 linker have been found but it is therefore important to keep in mind that *in vitro* conditions represent a simplified system compared to the *in viro* environment. In addition, *in vitro* data are compatible with the existence of the U5-AUG interaction as well as the interaction between the AG-rich and the CU-rich regions (results not shown). Of note, *in vitro* footprinting experiments were performed in the absence of $tRNA^{Lys,3}$. We showed *in viro* that the final structuration of the PBS domain is obtained through processing of the CA/Sp1 intermediate and the chaperone activity of NCp7, stabilising $tRNA^{Lys,3}$ annealing. Thus, the influence of the annealed $tRNA^{Lys,3}$ onto the final structuration of the PBS domain potentially explains the discrepancy between *in viro* and *in vitro* results.

Our results correlate with an enhancement of the chaperone activity of the NC domain-containing proteins during the processing cascade, which is consistent with *in viro* results and with the idea that the NCp7 release increases its flexibility and influences its nucleic acid binding/dissociation properties. Regarding the NCp15 intermediate, no strong difference of behaviour is noticed *in vitro* and *in viro*.

In this study, we also tested the GagΔp6 protein, an intermediate which does not exist during the natural maturation process, because this protein is often used as a substitute for $Pr55^{Gag}$ in *in vitro* studies (110, 111, 206, 225, 306, 423, 431). Surprisingly, $Pr55^{Gag}$ and GagΔp6 have quite different characteristics, as clearly highlighted at

the SL1 and SL3 apical loops and at the lower internal loop of SL1. Interestingly, the reactivity profiles obtained upon binding of GagΔp6 and NCp9, which both lack the p6 domain, are very similar. Unlike $Pr55^{Gag}$, NCp15 and NCp7, these two proteins did not completely dissociate from the RNA matrix in the absence of SDS. These results highlight the aggregative properties of NCp9 and GagΔp6 and suggest that Sp2 and/or p6 affect the binding mode of these proteins. Interestingly, it has been proposed that the NC domain transiently interacts with p6 (101, 292, 293, 420). GagΔp6 has been often used as a substitute for $Pr55^{Gag}$ in *in vitro* studies because full-length Gag is sensitive to proteolytic cleavage during expression and purification (65). Our laboratory recently addressed the impact of the p6 domain on binding specificity of $Pr55^{Gag}$ to viral RNAs (Dubois et al. unpublished results). Surprisingly, this comparison reveals that GagΔp6 binds gRNA, spliced viral RNAs and cellular RNAs with very similar affinity. In addition, the SL1 domain does not constitute the major recognition element of GagΔp6, on the contrary to $Pr55^{Gag}$. Taken together, these findings demonstrated that the p6 domain plays a crucial role in the selective binding of $Pr55^{Gag}$ to the HIV-1 gRNA. Thus, $Pr55^{Gag}$ and GagΔp6 cannot be considered to be equivalent with respect of their RNA binding properties.

V. Do PIs impact the gRNA conformation?

Low concentrations of PIs achieved in clinical settings efficiently inhibit viral replication while producing only subtle or no detectable defects in the processing of the $Pr55^{Gag}$ and $Pr160^{GagPol}$ precursors. In addition, PIs affect the process of RTion but not the enzymatic activity of RT itself (304), and an accumulation of NCp9 processing intermediates have been reported (304). Based on these facts, we hypothesised that the activity of PIs might be linked to the processing of the NC domain and to defects in the structural maturation of the gRNA.

Comparison of the gRNA conformation at the four different concentrations of LPV and ATV did not reveal any evolution correlating with the increase of PI concentration. In addition, the gRNA profile obtained at 1.1 nM LPV is different from the immature PR- and mature conformations. Thus, our results indicate the gRNA structural maturation is blocked at an intermediate stage in the presence of PIs.

Indeed, the gRNA conformation at 1.1 nM LPV resembles the Step 2 and CA/Sp1 stages, as no clear difference is observed between these two reactivity profiles. In addition, LPV has been reported to affect processing of the NC/Sp2 cleavage site (304), possibly reflecting a processing defect similar to the one observed in the Step 2 mutant. The NCp9 intermediate is already present at the Step 1.1 stage, where CA/Sp1 is still attached to MA and thus linked to the viral membrane. However, the reactivity profile of PI-treated samples is different from samples blocked at the Step 1.1 stage, possibly pointing at an influence of the core formation on the structural rearrangement of the genome. Moore et al. further investigated the stability of gRNA in the presence of PI at IC_{50} or IC_{90} (295). Northern blot analysis showed only a slight impact of PIs on the stability of the dimeric gRNA, while approximately 40 % of viral particles displayed an aberrant morphology with an eccentric RNP complex. Thus, these authors concluded that stabilisation of the dimeric gRNA is independent from the viral core formation.

Detectable changes in the HIV-1 particle morphology are noticed in the presence of low PI doses. A gradual increase of aberrant particles with diffuse or empty cores, as

well as an accumulation of electron-dense eccentric material was observed (215, 295, 304). According to Müller et al., the proportion of particles exhibiting these defects is too low to explain the infectivity loss. However, the proportion of morphologically mature particles at a LPV IC_{50} concentration is diminished by approximately 40 %, with 20 % of particles exhibiting a maturation defect and 30 % of immature particles (304). This is consistent with Moore et al. showing approximately 25 % of particles with an eccentric electron-dense region and 15 % of immature particles in the presence of LPV and ATVs IC_{50} concentrations (296). The proportion of viral particles with a mature morphology decreases with increasing concentrations of PIs, with a remaining 20-30 % at the PI IC_{90}. In addition, the degree of $Pr55^{Gag}$ processing impairment and the reduction of infectivity at low PIs concentrations is not directly and quantitatively correlated.

The mislocalised RNP complex observed in the presence of PIs is similar to the morphological phenotype observed with NCINIs (144, 210, 228). In the case of NCINIs, the phenotype has been proposed to be due to IN multimerisation that would disrupt the IN-RNP interaction and thus initiation of the core morphogenesis. It is thus tempting to link the action of PIs to the core formation.

VI. Secondary structure model of the 5' region of the HIV-1 genome at the mature stage

Based on our hSHAPE-Seq data obtained upon AT-2 treatment of mature particles, we propose a new secondary structure model of the first 550 nucleotides of the HIV-1 genome (**Figure 99**). The proposed NCp7 binding sites and the sites destabilised by NCp7 are indicated on the model.

In this model, nucleotides U105-G116 of U5 interact with nucleotides G361-G365 of the *gag* coding region, instead of the U5-AUG interaction (191, 220, 383, 422, 426) (**Figure 99 A**). Moreover, the CU-rich region is still proposed to base pair with the SL4 domain, but in a shifted way, with nucleotides G331-A336 interacting with the CU-rich region instead of the commonly proposed nucleotides C329-A334 (**Figure 99 B**).

The overall structuration of the PBS domain is similar to the consensus structure found in published models, except for the PAS (**Figure 99 C**). Indeed, our results indicate that the PAS is annealed to $tRNA^{Lys,3}$. The anti-PAS is protected by NCp7 and is thus proposed to be unpaired. The PBS loop comprises a smaller number of nucleotides as two additional interactions are proposed (positions U131-A133 with positions U200-G202, and positions U115-U118 with positions A205-G208).

The SL1 domain is drawn as an irregular intramolecular stem-loop, however, the apical loop and stem probably form an extended duplex. The intermediate stem, which is predicted by the folding algorithm (RNA structure software (346)) even when the SHAPE reactivity are introduced as constraints does probably not exist: it might unfold under the geometrical constraints imposed by the extended duplex. The SL2 and SL3 domains plus the SL2-SL3 linker adopt the widely accepted structure (**Figure 99 D**).

The first 200 nucleotides of the *gag* coding region are highly structured, but this region is destabilised by NCp7, as shown by the decreased reactivity observed upon AT-2 treatment (**Figure 99 E**).

We also tried to model the 5' region of the HIV-1 genome at the immature stage, using AT-2-treated hSHAPE-Seq results. However, the strong destabilising effect of $Pr55^{Gag}$ on the PBS domain makes the modelling of the PBS domain difficult, and hence of the whole 5' region of the HIV-1 genome.

It will be interesting to build a 3D structure model of the mature HIV-1 RNA structure. This will be an important test for our secondary structure model, as it will indicate whether it is topologically possible. The same strategy could be applied to discriminate between several possible secondary structure models of the immature RNA conformation.

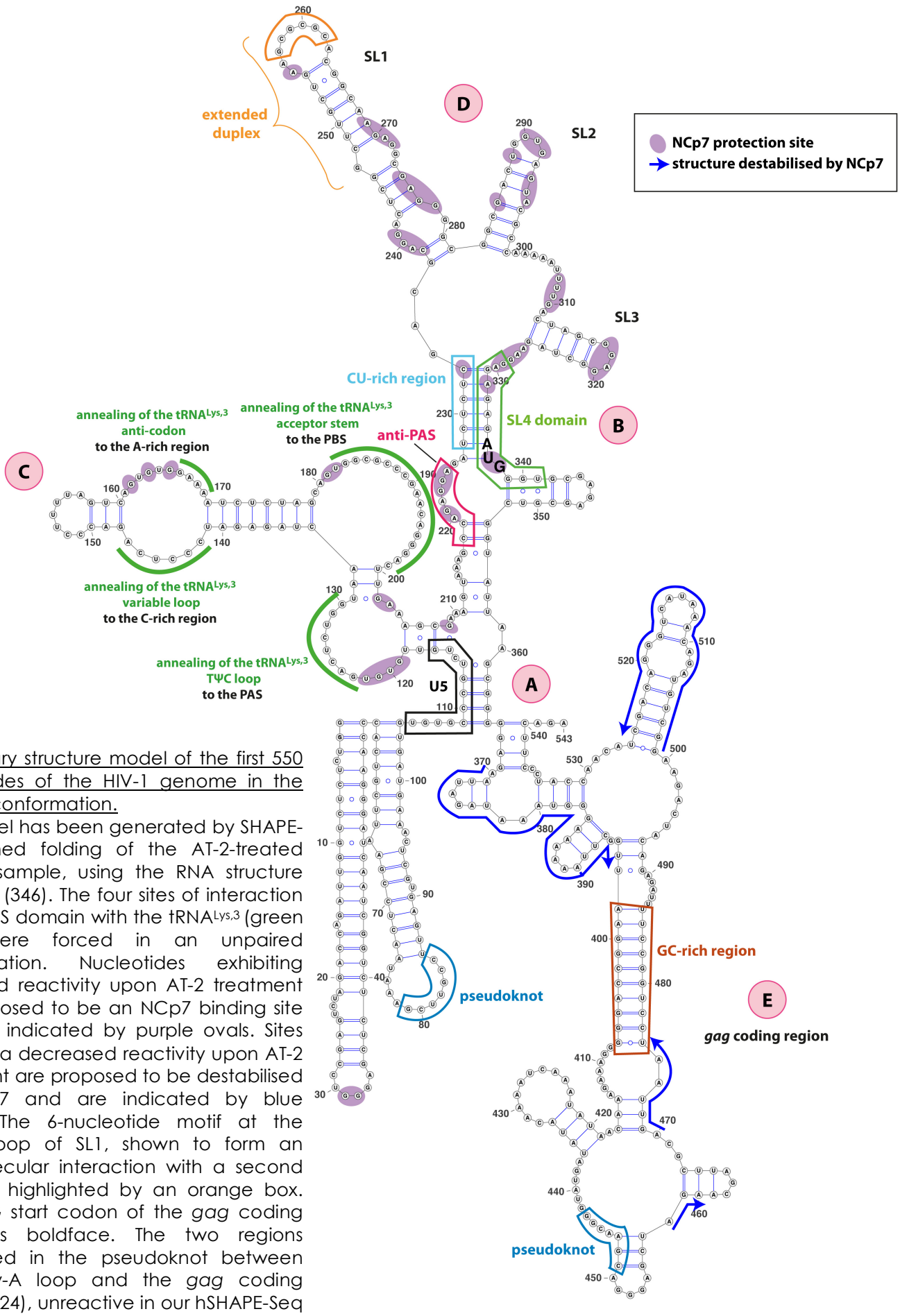


Figure 99
Secondary structure model of the first 550 nucleotides of the HIV-1 genome in the mature conformation.
This model has been generated by SHAPE-constrained folding of the AT-2-treated mature sample, using the RNA structure software (346). The four sites of interaction of the PBS domain with the tRNA^{Lys,3} (green lines) were forced in an unpaired conformation. Nucleotides exhibiting increased reactivity upon AT-2 treatment are proposed to be an NCp7 binding site and are indicated by purple ovals. Sites showing a decreased reactivity upon AT-2 treatment are proposed to be destabilised by NCp7 and are indicated by blue arrows. The 6-nucleotide motif at the apical loop of SL1, shown to form an intermolecular interaction with a second gRNA, is highlighted by an orange box. The AUG start codon of the *gag* coding region is boldface. The two regions implicated in the pseudoknot between the poly-A loop and the *gag* coding region (324), unreactive in our hSHAPE-Seq data, are indicated by blue boxes.

PERSPECTIVES

PERSPECTIVES

Short term perspectives

I. Analysis of AT-2-treated intermediates and confirmation of the results obtained by hSHAPE-Seq

The results we obtained thanks to the hSHAPE-Seq approach are very promising and demonstrated that the the experimental procedure is powerful enough to detect conformational rearrangements mediated by $Pr55^{Gag}$ and NC-containing intermediates. The impact of NCp15 and NCp9 on the gRNA conformation is deduced indirectly thanks to the AT-2 treated immature and mature particles. However, it would be interesting to test the gRNA structure of AT-2 treated Step 1, Step 1.1, Step 2 and CA/Sp1 viral particles in order to confirm our analysis. Moreover, obtaining one additional replicate for each intermediate and for the untreated mature and immature PR- particles would strengthen our results.

II. Addition of the *ex viro* condition

The AT-2 treatment is very useful to determine ZF protein binding sites but is not informative regarding the influence of other RNA-binding proteins. An *ex viro* RNA condition could be produced with viral gRNA gently extracted from viral particles, treated with proteinase K (and SDS if required) to remove interacting proteins, and refolded before chemical modification and analysis with the hSHAPE-Seq methodology. The folding of this protein-free *ex viro* RNA is supposed to be strongly influenced by the authentic virus environment and simultaneously, this condition lacks the complex influence of interacting proteins. Thus, the *ex viro* condition should allow us to determine protein-binding site at "physiological conditions" by comparing it to the *in viro* condition. Such *ex viro* conditions have been widely utilised by Kevin Weeks and co-workers (383, 422, 426), but it is difficult to ensure that extraction of the gRNA from the viral particles and degradation of the proteins does not affect RNA structure. Comparison of the gRNA structure *ex viro* and in AT-2 treated viral particles would be important in that respect.

III. Development of a genome-wide hSHAPE-Seq technique

We demonstrated that the hSHAPE-Seq methodology is reproducible and powerful enough to follow the conformational rearrangements of the 5' region of the HIV-1 gRNA. Now, we would like to extend this analysis to the entire genome in order to investigate the putative formation of additional intermolecular contacts and structural rearrangements throughout the entire HIV-1genome.

To that goal, the hSHAPE-Seq protocol needs to be adapted for genome-wide analysis. The genome-wide hSHAPE-Seq approach, which is similar but not identical to icSHAPE (392), is developed together with Dr Valérie Vivet-Boudou, who

synthesised NMIA derivatives such the NMIA-N_3 characterised by the presence of an azide group. This feature is important for the genome wide adaptation of the protocol.

Following chemical modification and extraction from the viral particles, RNA is randomly fragmented. Experimental conditions for this step have already been set up and allow production of ~200-nucleotide long fragments. A selection step is introduced to discard non-modified fragments. Indeed, these fragments are uninformative and bias results since the RT stop at the 5' end of the RNA fragment will be mistakenly considered as a chemical modification. To this aim, the azide function of NMIA-N_3 is coupled to sulfodibenzocyclooctine biotine by copper-free « click chemistry ». Thanks to the strong interaction between the biotin and the streptavidin, modified fragments are retained on streptavidin-coated magnetic beads, then selectively eluted. The "click chemistry" reaction has already been optimised on *in vitro* produced RNA fragments. The selection step with streptavidin beads has also been validated and the selection step needs now to be tested on small amounts of viral RNAs. Prior to RTion, the population of RNA fragments will be poly-adenylated in a controlled manner resulting in the addition of approximately 50 A-residues. This step will allow the RTion reaction to be performed with a single poly-T primer. Regarding the Illumina library generation, the smaller and more homogenous size of the cDNAs should facilitate preparation of the Illumina library, compared to the hSHAPE-Seq methodology applied to the 5' region of the genome.

Long term perspectives

Thanks to the genome-wide hSHAPE-Seq methodology, we will identify regions undergoing conformational rearrangements during maturation of the viral particles and further focus our attention on these regions. An exhaustive bio-informatic analysis will be required to model possible intra- and inter-molecular interactions. Classical bio-informatics approaches such as covariation analysis will be used to assess the conservation of the proposed interactions between the different HIV-1 clades. We will also evaluate if the sites where conformational rearrangements take place are hotspots for recombination.

Based on these results, a small number of regions undergoing a structural rearrangement during the $Pr55^{Gag}$ proteolytic processing will be mutated. Mutations will be designed in order to hinder or to facilitate structural rearrangements. The hSHAPE-Seq data presented in this thesis already hint at the U5 region, SL1 and SL4 domains as strong candidates.

Thanks to these mutations, we will be able to test the impact of these structural rearrangements on viral infectivity thanks to functionality tests assessing viral entry, RTion, decapsidation, integration, packaging and maturation.

BIBLIOGRAPHY

1. Abbink TEM, Berkhout B. 2003. A Novel Long Distance Base-pairing Interaction in Human Immunodeficiency Virus Type 1 RNA Occludes the Gag Start Codon. *J. Biol. Chem.* 278(13):11601–11
2. Abbink TEM, Berkhout B. 2008. HIV-1 reverse transcription initiation: a potential target for novel antivirals? *Virus Res.* 134(1-2):4–18
3. Abbink TEM, Ooms M, Haasnoot PCJ, Berkhout B. 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry*. 44(25):9058–66
4. Abd El-Wahab EW, Smyth RP, Mailler E, Bernacchi S, Vivet-Boudou V, Hijnen M, Jossinet F, Mak J, Paillart J-C, Marquet R. 2014. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5:4304
5. Abdurahman S, Youssefi M, Höglund S, Vahlne A. 2007. Characterization of the invariable residue 51 mutations of human immunodeficiency virus type 1 capsid protein on in vitro CA assembly and infectivity. *Retrovirology*. 4(1):69
6. Abrahamyan LG, Chatel-Chaix L, Ajamian L, Milev MP, Monette A, Clément J-F, Song R, Lehmann M, DesGroseillers L, Laughrea M, Boccaccio G, Mouland AJ. 2010. Novel Staufen1 ribonucleoproteins prevent formation of stress granules but favour encapsidation of HIV-1 genomic RNA. *J. Cell Sci.* 123(Pt 3):369–83
7. Accola MA, Strack B, Göttlinger HG. 2000. Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J. Virol.* 74(12):5395–5402
8. Ahmad N, Maitra RK, Venkatesan S. 1989. Rev-induced modulation of Nef protein underlies temporal regulation of human immunodeficiency virus replication. *Proc. Natl. Acad. Sci. U. S. A.* 86(16):6111–15
9. Ajamian L, Abel K, Rao S, Vyboh K, García-de-Gracia F, Soto-Rifo R, Kulozik AE, Gehring NH, Mouland AJ. 2015. HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA. *Biomolecules*. 5(4):2808–39
10. Ajamian L, Abrahamyan L, Milev M, Ivanov P V, Kulozik AE, Gehring NH, Mouland AJ. 2008. Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation. *RNA*. 14(5):914–27
11. Ako-Adjei D, Johnson MC, Vogt VM. 2005. The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles. *J. Virol.* 79(21):13463–72
12. Aldovini A, Young RA. 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* 64(5):1920–26
13. Alfadhli A, Barklis E. 2014. The roles of lipids and nucleic acids in HIV-1 assembly. *Front. Microbiol.* 5:
14. Alfadhli A, Mack A, Ritchie C, Cylinder I, Harper L, Tedbury PR, Freed EO, Barklis E. 2016. Trimer Enhancement Mutation Effects on HIV-1 Matrix Protein Binding Activities. *J. Virol.* 90(12):5657–64
15. Allain B, Lapadat-Tapolsky M, Berlioz C, Darlix JL. 1994. Transactivation of the minus-strand DNA transfer by nucleocapsid protein during reverse transcription of the retroviral genome. *EMBO J.* 13(4):973–81
16. Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF. 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* 301(2):491–511
17. Amarasinghe GK, De Guzman RN, Turner RB, Summers MF. 2000. NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J. Mol. Biol.* 299(1):145–56
18. Amarasinghe GK, Zhou J, Miskimon M, Chancellor KJ, McDonald JA, Matthews AG, Miller RR, Rouse MD, Summers MF. 2001. Stem-loop SL4 of the HIV-1 Ψ RNA packaging signal exhibits weak affinity for the nucleocapsid protein. structural studies and implications for genome recognition1. *J. Mol. Biol.* 314(5):961–70

19. Andersen ES, Contera SA, Knudsen B, Damgaard CK, Besenbacher F, Kjems J. 2004. Role of the trans-activation response element in dimerization of HIV-1 RNA. *J. Biol. Chem.* 279(21):22243–49
20. Arhel N, Genovesio A, Kim K-A, Miko S, Perret E, Olivo-Marin J-C, Shorte S, Charneau P. 2006. Quantitative four-dimensional tracking of cytoplasmic and nuclear HIV-1 complexes. *Nat. Methods*. 3(10):817–24
21. Ashe MP, Furger A, Proudfoot NJ. 2000. Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. *RNA*. 6(2):170–77
22. Aviran S, Trapnell C, Lucks JB, Mortimer SA, Luo S, Schroth GP, Doudna JA, Arkin AP, Pachter L. 2011. Modeling and automation of sequencing-based characterization of RNA structure. *Proc. Natl. Acad. Sci. U. S. A.* 108(27):11069–74
23. Azoulay J, Clamme JP, Darlix JL, Roques BP, Mély Y. 2003. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J. Mol. Biol.* 326(3):691–700
24. Baba S, Takahashi K, Noguchi S, Takaku H, Koyanagi Y, Yamamoto N, Kawai G. 2005. Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J. Biochem.* 138(5):583–92
25. Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, Niedziela-Majka A, Stray KM, Sakowicz R, Cihlar T. 2013. Non-catalytic site HIV-1 integrase inhibitors disrupt core maturation and induce a reverse transcription block in target cells. *PLoS One*. 8(9):e74163
26. Ballandras-Colas A, Maskell DP, Serrao E, Locke J, Swuec P, Jónsson SR, Kotecha A, Cook NJ, Pye VE, Taylor IA, Andrésdóttir V, Engelman AN, Costa A, Cherepanov P. 2017. A supramolecular assembly mediates lentiviral DNA integration. *Science*. 355(6320):93–95
27. Banerjee A, Benjamin R, Balakrishnan K, Ghosh P, Banerjee S. 2014. Human protein Staufen-2 promotes HIV-1 proliferation by positively regulating RNA export activity of viral protein Rev. *Retrovirology*. 11(1):18
28. Baudin F, Marquet R, Isel C, Darlix JL, Ehresmann B, Ehresmann C. 1993. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* 229(2):382–97
29. Beerens N, Berkhout B. 2002. 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.* 76(5):2329–39
30. Beerens N, Groot F, Berkhout B. 2001. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.* 276(33):31247–56
31. Beerens N, Jepsen MDE, Nechyporuk-Zloy V, Krüger AC, Darlix J-L, Kjems J, Birkedal V. 2013. Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA*. 19(4):517–26
32. Beerens N, Klaver B, Berkhout B. 2000. A structured RNA motif is involved in correct placement of the tRNA(3)(Lys) primer onto the human immunodeficiency virus genome. *J. Virol.* 74(5):2227–38
33. Bell NM, Kenyon JC, Balasubramanian S, Lever AML. 2012. Comparative structural effects of HIV-1 Gag and nucleocapsid proteins in binding to and unwinding of the viral RNA packaging signal. *Biochemistry*. 51(15):3162–69
34. Beltz H, Azoulay J, Bernacchi S, Clamme J-P, Ficheux D, Roques B, Darlix J-L, Mély Y. 2003. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* 328(1):95–108
35. Beltz H, Clauss C, Piémont E, Ficheux D, Gorelick RJ, Roques B, Gabus C, Darlix J-L, de Rocquigny H, Mély 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(5):1113–26
36. Bender W, Chien YH, Chattopadhyay S, Vogt PK, Gardner MB, Davidson N. 1978. High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis

virus all have similar dimer structures. *J. Virol.* 25(3):888–96
37. Bender W, Davidson N. 1976. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell*. 7(4):595–607
38. Benjamin J, Ganser-Pornillos BK, Tivol WF, Sundquist WI, Jensen GJ. 2005. Three-dimensional Structure of HIV-1 Virus-like Particles by Electron Cryotomography. *J. Mol. Biol.* 346(2):577–88
39. Berkhout B. 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res.* 20(1):27–31
40. Berkhout B, Essink BB, Schoneveld I. 1993. In vitro dimerization of HIV-2 leader RNA in the absence of PuGGAPuA motifs. *FASEB J.* 7(1):181–87
41. Berkhout B, Klaver B, Das AT. 1995. A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses. *Virology*. 207(1):276–81
42. Berkhout B, Silverman RH, Jeang KT. 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell*. 59(2):273–82
43. Berkhout B, van Wamel JL. 1996. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J. Virol.* 70(10):6723–32
44. Berkhout B, van Wamel JL. 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA*. 6(2):282–95
45. Berkowitz R, Fisher J, Goff SP. 1996. RNA packaging. *Curr. Top. Microbiol. Immunol.* 214:177–218
46. Berkowitz RD, Luban J, Goff SP. 1993. Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* 67(12):7190–7200
47. Berman PW, Nunes WM, Haffar OK. 1988. Expression of membrane-associated and secreted variants of gp160 of human immunodeficiency virus type 1 in vitro and in continuous cell lines. *J. Virol.* 62(9):3135–42
48. Bernacchi S, Abd El-Wahab EW, Dubois N, Hijnen M, Smyth RP, Mak J, Marquet R, Paillart J-C. 2017. HIV-1 Pr55(Gag) binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol.* 14(1):90–103
49. Bernacchi S, Stoylov S, Piémont E, Ficheux D, Roques BP, Darlix JL, Mély Y. 2002. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* 317(3):385–99
50. Bhattacharya A, Alam SL, Fricke T, Zadrozny K, Sedzicki J, Taylor AB, Demeler B, Pornillos O, Ganser-Pornillos BK, Diaz-Griffero F, Ivanov DN, Yeager M. 2014. Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U. S. A.* 111(52):18625–30
51. Bichel K, Price AJ, Schaller T, Towers GJ, Freund SM V, James LC. 2013. HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. *Retrovirology*. 10(1):81
52. Bonnet-Mathonière B, Girard PM, Muriaux D, Paoletti J. 1996. Nucleocapsid protein 10 activates dimerization of the RNA of Moloney murine leukaemia virus in vitro. *Eur. J. Biochem.* 238(1):129–35
53. Bouchat S, Van Driessche B, Van Lint C. 2014. Transcription (Initiation, Regulation, Elongation). In *Encyclopedia of AIDS*, pp. 1–13. New York, NY: Springer New York
54. Bourbigot S, Ramalanjaona N, Boudier C, Salgado GFJ, Roques BP, Mély Y, Bouaziz S, Morellet N. 2008. How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *J. Mol. Biol.* 383(5):1112–28
55. Briggs J a G, Kräusslich H-G. 2011. The molecular architecture of HIV. *J. Mol. Biol.* 410(4):491–500
56. Briggs JAG, Grünewald K, Glass B, Förster F, Kräusslich H-G, Fuller SD. 2006. The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure*. 14(1):15–20
57. Briggs JAG, Riches JD, Glass B, Bartonova V, Zanetti G, Kräusslich H-G. 2009. Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. U. S. A.* 106(27):11090–95

58. Briggs JAG, Simon MN, Gross I, Kräusslich H-G, Fuller SD, Vogt VM, Johnson MC. 2004. The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* 11(7):672–75
59. Briggs JAG, Wilk T, Welker R, Kräusslich H-G, Fuller SD. 2003. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* 22(7):1707–15
60. Brogna S, Wen J. 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* 16(2):107–13
61. Brügger B, Glass B, Haberkant P, Leibrecht I, Wieland FT, Kräusslich H-G. 2006. The HIV lipidome: a raft with an unusual composition. *Proc. Natl. Acad. Sci. U. S. A.* 103(8):2641–46
62. Buckman JS, Bosche WJ, Gorelick RJ. 2003. Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J. Virol.* 77(2):1469–80
63. Bukrinskaya a, Brichacek B, Mann A, Stevenson M. 1998. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *J. Exp. Med.* 188(11):2113–25
64. Campbell EM, Hope TJ. 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* 13(8):471–83
65. Campbell S, Rein A. 1999. In Vitro Assembly Properties of Human Immunodeficiency Virus Type 1 Gag Protein Lacking the p6 Domain. *J. Virol.* 73(3):2270–79
66. Cantin R, Méthot S, Tremblay MJ. 2005. Plunder and stowaways: incorporation of cellular proteins by enveloped viruses. *J. Virol.* 79(11):6577–87
67. Carlson L-A, Bai Y, Keane SC, Doudna JA, Hurley JH. 2016. Reconstitution of selective HIV-1 RNA packaging in vitro by membrane-bound Gag assemblies. *Elife*. 5:1–19
68. Carlson L-A, Briggs JAG, Glass B, Riches JD, Simon MN, Johnson MC, Müller B, Grünewald K, Kräusslich H-G. 2008. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe*. 4(6):592–99
69. Cassan E, Arigon-Chifolleau A-M, Mesnard J-M, Gross A, Gascuel O. 2016. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. *Proc. Natl. Acad. Sci. U. S. A.* 113(41):11537–42
70. Cen S, Huang Y, Khorchid A, Darlix JL, Wainberg MA, Kleiman L. 1999. The role of Pr55(gag) in the annealing of tRNA3Lys to human immunodeficiency virus type 1 genomic RNA. *J. Virol.* 73(5):4485–88
71. Cen S, Khorchid A, Gabor J, Rong L, Wainberg MA, Kleiman L. 2000. Genomic Placement and the Initiation Step of Reverse Transcription in Human Immunodeficiency Virus Type 1. . 74(22):10796–800
72. Chan B, Weidemaier K, Yip WT, Barbara PF, Musier-Forsyth K. 1999. Intra-tRNA distance measurements for nucleocapsid proteindependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* 96(2):459–64
73. Charneau P, Alizon M, Clavel F. 1992. A second origin of DNA plus-strand synthesis is required for optimal human A Second Origin of DNA Plus-Strand Synthesis Is Required for Optimal Human Immunodeficiency Virus Replication. *J. Virol.* 66(5):2814–20
74. Charneau P, Clavel F. 1991. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. *J. Virol.* 65(5):2415–21
75. Chatel-Chaix L, Abrahamyan L, Fréchina C, Mouland AJ, DesGroseillers L. 2007. The host protein Staufen1 participates in human immunodeficiency virus type 1 assembly in live cells by influencing pr55Gag multimerization. *J. Virol.* 81(12):6216–30
76. Chatel-Chaix L, Boulay K, Mouland AJ, Desgroseillers L. 2008. The host protein Staufen1 interacts with the Pr55Gag zinc fingers and regulates HIV-1 assembly via its N-terminus. *Retrovirology*. 5:41
77. Chatel-Chaix L, Clément J-F, Martel C, Bériault V, Gatignol A, DesGroseillers L, Mouland AJ. 2004. Identification of Staufen in the Human Immunodeficiency Virus Type 1 Gag Ribonucleoprotein Complex and a Role in Generating Infectious Viral Particles. *Mol. Cell. Biol.*

24(7):2637–48
78. Checkley MA, Luttge BG, Freed EO. 2011. HIV-1 Envelope Glycoprotein Biosynthesis, Trafficking, and Incorporation. *J. Mol. Biol.* 410(4):582–608
79. Checkley MA, Luttge BG, Soheilian F, Nagashima K, Freed EO. 2010. The capsid-spacer peptide 1 Gag processing intermediate is a dominant-negative inhibitor of HIV-1 maturation. *Virology*. 400(1):137–44
80. Chertova E, Chertov O, Coren L V., Roser JD, Trubey CM, Bess JW, Sowder RC, Barsov E, Hood BL, Fisher RJ, Nagashima K, Conrads TP, Veenstra TD, Lifson JD, Ott DE. 2006. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J. Virol.* 80(18):9039–52
81. Chertova E, Crise BJ, Morcock DR, Bess JW, Henderson LE, Lifson JD. 2003. Sites, mechanism of action and lack of reversibility of primate lentivirus inactivation by preferential covalent modification of virion internal proteins. *Curr. Mol. Med.* 3(3):265–72
82. Cheung K, Smith RE, Stone MP, Joklik WK. 1972. Comparison of immature (rapid harvest) and mature Rous sarcoma virus particles. *Virology*. 50(3):851–64
83. Chojnacki J, Müller B. 2013. Investigation of HIV-1 assembly and release using modern fluorescence imaging techniques. *Traffic*. 14(1):15–24
84. Chojnacki J, Staudt T, Glass B, Bingen P, Engelhardt J, Anders M, Schneider J, Müller B, Hell SW, Kräusslich H-G. 2012. Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science*. 338(6106):524–28
85. Chukkapalli V, Inlora J, Todd GC, Ono A. 2013. Evidence in support of RNA-mediated inhibition of phosphatidylserine-dependent HIV-1 Gag membrane binding in cells. *J. Virol.* 87(12):7155–59
86. Chukkapalli V, Ono A. 2011. Molecular determinants that regulate plasma membrane association of HIV-1 Gag. *J. Mol. Biol.* 410(4):512–24
87. Cimarelli A, Sandin S, Höglund S, Luban J. 2000. Basic Residues in Human Immunodeficiency Virus Type 1 Nucleocapsid Promote Virion Assembly via Interaction with RNA. *J. Virol.* 74(7):3046–57
88. Clavel F, Orenstein JM. 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* 64(10):5230–34
89. Clever J, Sassetti C, Parslow TG. 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* 69(4):2101–9
90. Clever JL, Miranda D, Parslow TG. 2002. RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J. Virol.* 76(23):12381–87
91. Clever JL, Parslow TG. 1997. Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* 71(5):3407–14
92. Clever JL, Taplitz R a, Lochrie M a, Polisky B, Parslow TG. 2000. A heterologous, high-affinity RNA ligand for human immunodeficiency virus Gag protein has RNA packaging activity. *J. Virol.* 74(1):541–46
93. Cocchi F, DeVico AL, Garzino-Demo A, Arya SK, Gallo RC, Lusso P. 1995. Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science*. 270(5243):1811–15
94. Coffin JM. 1979. Structure, Replication, and Recombination of Retrovirus Genomes: Some Unifying Hypotheses. *J. Gen. Virol.* 42(1):1–26
95. Coffin JM. 1992. Structure and Classification of Retroviruses. In *The Retroviridae*, pp. 19–49. Boston, MA: Springer US
96. Coren L V, Thomas JA, Chertova E, Sowder RC, Gagliardi TD, Gorelick RJ, Ott DE. 2007. Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *J. Virol.* 81(18):10047–54
97. Cosnefroy O, Murray PJ, Bishop KN. 2016. HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*. 13(1):58

98. Craigie R, Bushman FD. 2012. HIV DNA integration. *Cold Spring Harb. Perspect. Med.* 2(7):a006890
99. Crist RM, Datta SAK, Stephen AG, Soheilian F, Mirro J, Fisher RJ, Nagashima K, Rein A. 2009. Assembly properties of human immunodeficiency virus type 1 Gag-leucine zipper chimeras: implications for retrovirus assembly. *J. Virol.* 83(5):2216–25
100. Cruceanu M, Gorelick RJ, Musier-Forsyth K, Rouzina I, Williams MC. 2006. Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J. Mol. Biol.* 363(5):867–77
101. Cruceanu M, Urbaneja MA, Hixson C V., Johnson DG, Datta SA, Fivash MJ, Stephen AG, Fisher RJ, Gorelick RJ, Casas-Finet JR, Rein A, Rouzina I, Williams MC. 2006. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.* 34(2):593–605
102. Cullen BR. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell*. 46(7):973–82
103. Dam E, Quercia R, Glass B, Descamps D, Launay O, Duval X, Kräusslich H-G, Hance AJ, Clavel F, ANRS 109 Study Group. 2009. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog.* 5(3):e1000345
104. Damgaard CK, Andersen ES, Knudsen B, Gorodkin J, Kjems J. 2004. RNA interactions in the 5' region of the HIV-1 genome. *J. Mol. Biol.* 336(2):369–79
105. Damgaard CK, Dyhr-Mikkelsen H, Kjems J. 1998. Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res.* 26(16):3667–76
106. Dardel F, Marquet R, Ehresmann C, Ehresmann B, Blanquet S. 1998. Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res.* 26(15):3567–71
107. Darlix JL, Gabus C, Nugeyre MT, Clavel F, Barré-Sinoussi F. 1990. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* 216(3):689–99
108. Das AT, Klaver B, Klasens BI, van Wamel JL, Berkhout B. 1997. A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J. Virol.* 71(3):2346–56
109. Datta S a K, Temeselew LG, Crist RM, Soheilian F, Kamata A, Mirro J, Harvin D, Nagashima K, Cachau RE, Rein A. 2011. On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J. Virol.* 85(9):4111–21
110. Datta SAK, Curtis JE, Ratcliff W, Clark PK, Crist RM, Lebowitz J, Krueger S, Rein A. 2007. Conformation of the HIV-1 Gag protein in solution. *J. Mol. Biol.* 365(3):812–24
111. Datta SAK, Heinrich F, Raghunandan S, Krueger S, Curtis JE, Rein A, Nanda H. 2011. HIV-1 Gag Extension: Conformational Changes Require Simultaneous Interaction with Membrane and Nucleic Acid. *J. Mol. Biol.* 406(2):205–14
112. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*. 279(5349):384–88
113. de Marco A, Heuser A-M, Glass B, Kräusslich H-G, Müller B, Briggs JAG. 2012. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J. Virol.* 86(24):13708–16
114. de Marco A, Müller B, Glass B, Riches JD, Kräusslich H-G, Briggs JAG. 2010. Structural Analysis of HIV-1 Maturation Using Cryo-Electron Tomography. *PLoS Pathog.* 6(11):
115. De Rocquigny H, Gabus C, Vincent A, Fournié-Zaluski MC, Roques B, Darlix JL. 1992. Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. U. S. A.* 89(14):6472–76
116. Deforges J, de Breyne S, Ameur M, Ulryck N, Chamond N, Saadi A, Ponty Y, Ohlmann T, Sargueil B. 2017. Two ribosome recruitment sites direct multiple translation events within

HIV1 Gag open reading frame. *Nucleic Acids Res.*, pp. 7382–7400
117. Desfarges S, Ciuffi A. 2010. Retroviral integration site selection. *Viruses*. 2(1):111–30
118. Deshmukh L, Ghirlando R, Clore GM. 2015. Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 112(11):3374–79
119. Didierlaurent L, Racine PJ, Houzet L, Chamontin C, Berkhout B, Mougel M. 2011. Role of HIV-1 RNA and protein determinants for the selective packaging of spliced and unspliced viral RNA and host U6 and 7SL RNA in virus particles. *Nucleic Acids Res.* 39(20):8915–27
120. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. 2014. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 505(7485):696–700
121. Dingwall C, Ernberg I, Gait MJ, Green SM, Heaphy S, Karn J, Lowe AD, Singh M, Skinner MA, Valerio R. 1989. Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 86(18):6925–29
122. Dingwall C, Ernberg I, Gait MJ, Green SM, Heaphy S, Karn J, Lowe a D, Singh M, Skinner M a. 1990. HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. *EMBO J.* 9(12):4145–53
123. Dugré-Brisson S, Elvira G, Boulay K, Chatel-Chaix L, Mouland AJ, DesGroseillers L. 2005. Interaction of Staufen1 with the 5' end of mRNA facilitates translation of these RNAs. *Nucleic Acids Res.* 33(15):4797–4812
124. Eckwahl MJ, Arnion H, Kharytonchyk S, Zang T, Bieniasz PD, Telesnitsky A, Wolin SL. 2016. Analysis of the human immunodeficiency virus-1 RNA packageome. *RNA*
125. Eckwahl MJ, Telesnitsky A, Wolin SL. 2016. Host RNA Packaging by Retroviruses: A Newly Synthesized Story. *MBio*. 7(1):
126. Eilebrecht S, Wilhelm E, Benecke B-J, Bell B, Benecke AG. 2013. HMGA1 directly interacts with TAR to modulate basal and Tat-dependent HIV transcription. *RNA Biol.* 10(3):436–44
127. El-Wahab EWA, Smyth RP, Mailler E, Bernacchi S, Vivet-Boudou V, Hijnen M, Jossinet F, Mak J, Paillart J-C, Marquet R. 2014. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat. Commun.* 5:
128. Emery A, Zhou S, Pollom E, Swanstrom R. 2017. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* 91(6):1–21
129. Engelman a, Englund G, Orenstein JM, Martin M a, Craigie R. 1995. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J. Virol.* 69(5):2729–36
130. Engelman A. 1999. In vivo analysis of retroviral integrase structure and function. *Adv. Virus Res.* 52:411–26
131. Engelman A, Cherepanov P. 2008. The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication. *PLoS Pathog.* 4(3):e1000046
132. Engelman A, Mizuuchi K, Craigie R. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*. 67(6):1211–21
133. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. 2001. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* 8(12):1064–68
134. Ennifar E, Yusupov M, Walter P, Marquet R, Ehresmann B, Ehresmann C, Dumas P. 1999. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*. 7(11):1439–49
135. Erickson-Viitanen S, Manfredi J, Viitanen P, Tribe DE, Tritch R, Hutchison CA, Loeb DD, Swanstrom R. 1989. Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res. Hum. Retroviruses*. 5Erickson-(6):577–91
136. Erkelenz S, Theiss S, Otte M, Widera M, Peter JO, Schaal H. 2014. Genomic HEXploring allows landscaping of novel potential splicing regulatory elements. *Nucleic Acids Res.* 42(16):10681–97

137. Feng Y, Broder CC, Kennedy PE, Berger EA. 1996. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science*. 272(5263):872–77
138. Feng YX, Campbell S, Harvin D, Ehresmann B, Ehresmann C, Rein A. 1999. The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol.* 73(5):4251–56
139. Feng YX, Copeland TD, Henderson LE, Gorelick RJ, Bosche WJ, Levin JG, Rein A. 1996. HIV-1 nucleocapsid protein induces “maturation” of dimeric retroviral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 93(15):7577–81
140. Ferrer M, Clerté C, Chamontin C, Basyuk E, Lainé S, Hottin J, Bertrand E, Margeat E, Mougel M. 2016. Imaging HIV-1 RNA dimerization in cells by multicolor super-resolution and fluctuation microscopies. *Nucleic Acids Res.*, p. gkw511
141. Fisher RJ, Fivash MJ, Stephen AG, Hagan NA, Shenoy SR, Medaglia M V., Smith LR, Worthy KM, Simpson JT, Shoemaker R, McNitt KL, Johnson DG, Hixson C V., Gorelick RJ, Fabris D, Henderson LE, Rein A. 2006. Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. *Nucleic Acids Res.* 34(2):472–84
142. Fisher RJ, Rein A, Fivash M, Urbaneja MA, Casas-Finet JR, Medaglia M, Henderson LE. 1998. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.* 72(3):1902–9
143. Fitzon T, Leschonsky B, Bieler K, Paulus C, Schröder J, Wolf H, Wagner R. 2000. Proline residues in the HIV-1 NH2-terminal capsid domain: structure determinants for proper core assembly and subsequent steps of early replication. *Virology*. 268(2):294–307
144. Fontana J, Jurado KA, Cheng N, Ly NL, Fuchs JR, Gorelick RJ, Engelman AN, Steven AC. 2015. Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-inhibited Virions: A Role for Integrase in Maturation. *J. Virol.*, pp. JVI.01522–15
145. Fontana J, Keller PW, Urano E, Ablan SD, Steven AC, Freed EO. 2016. Identification of an HIV-1 Mutation in Spacer Peptide 1 That Stabilizes the Immature CA-SP1 Lattice. *J. Virol.* 90(2):972–78
146. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. 2002. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.* 76(11):5667–77
147. Franke EK, Yuan HEH, Luban J. 1994. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature*. 372(6504):359–62
148. Freed EO. 2002. Viral late domains. *J. Virol.* 76(10):4679–87
149. Freed EO. 2015. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* 13(8):484–96
150. Fu W, Dang Q, Nagashima K, Freed EO, Pathak VK, Hu W-S. 2006. Effects of Gag Mutation and Processing on Retroviral Dimeric RNA Maturation. *J. Virol.* 80(3):1242–49
151. Fu W, Gorelick RJ, Rein A. 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol.* 68(8):5013–18
152. Fu W, Rein A. 1993. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J. Virol.* 67(9):5443–49
153. Galetto R, Giacomoni V, Véron M, Negroni M. 2006. Dissection of a circumscribed recombination hot spot in HIV-1 after a single infectious cycle. *J. Biol. Chem.* 281(5):2711–20
154. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP. 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*. 87(7):1285–94
155. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI. 1999. Assembly and Analysis of Conical Models for the HIV-1 Core. *Science (80-.).* 283(5398):80–83
156. Garcia EL, Onafuwa-Nuga A, Sim S, King SR, Wolin SL, Telesnitsky A. 2009. Packaging of host mY RNAs by murine leukemia virus may occur early in Y RNA biogenesis. *J. Virol.* 83(23):12526–34
157. Gatignol A, Buckler-White A, Berkhout B, Jeang KT. 1991. Characterization of a human TAR

RNA-binding protein that activates the HIV-1 LTR. *Science*. 251(5001):1597–1600
158. Girard F, Barbault F, Gouyette C, Huynh-Dinh T, Paoletti J, Lancelot G. 1999. Dimer initiation sequence of HIV-1Lai genomic RNA: NMR solution structure of the extended duplex. *J. Biomol. Struct. Dyn.* 16(6):1145–57
159. Goldschmidt V, Ehresmann C, Ehresmann B, Marquet R. 2003. Does the HIV-1 primer activation signal interact with tRNA3(Lys) during the initiation of reverse transcription? *Nucleic Acids Res.* 31(3):850–59
160. Goldschmidt V, Paillart J-C, Rigourd M, Ehresmann B, Aubertin A-M, Ehresmann C, Marquet R. 2004. Structural variability of the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* 279(34):35923–31
161. Gorelick RJ, Nigida SM, Bess JW, Arthur LO, Henderson LE, Rein A. 1990. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* 64(7):3207–11
162. Göttlinger HG, Dorfman T, Sodroski JG, Haseltine W a. 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. U. S. A.* 88(8):3195–99
163. Göttlinger HG, Sodroski JG, Haseltine W a. 1989. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A.* 86(15):5781–85
164. Grawenhoff J, Engelman AN. 2017. Retroviral integrase protein and intasome nucleoprotein complex structures. *World J. Biol. Chem.* 8(1):32–44
165. Gres AT, Kirby KA, KewalRamani VN, Tanner JJ, Pornillos O, Sarafianos SG. 2015. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science*. 349(6243):99–103
166. Grohman JK, Gorelick RJ, Lickwar CR, Lieb JD, Bower BD, Znosko BM, Weeks KM. 2013. A Guanosine-Centric Mechanism for RNA Chaperone Function. *Science*. 340(6129):190–95
167. Grohmann D, Godet J, Mély Y, Darlix J-L, Restle T. 2008. HIV-1 nucleocapsid traps reverse transcriptase on nucleic acid substrates. *Biochemistry*. 47(46):12230–40
168. Guo F, Saadatmand J, Niu M, Kleiman L. 2009. Roles of Gag and NCp7 in facilitating tRNA(Lys)(3) Annealing to viral RNA in human immunodeficiency virus type 1. *J. Virol.* 83(16):8099–8107
169. Guo J, Wu T, Anderson J, Kane BF, Johnson DG, Gorelick RJ, Henderson LE, Levin JG. 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* 74(19):8980–88
170. Haffar OK, Dowbenko DJ, Berman PW. 1988. Topogenic analysis of the human immunodeficiency virus type 1 envelope glycoprotein, gp160, in microsomal membranes. *J. Cell Biol.* 107(5):1677–87
171. Hahn F, Schmalen A, Setz C, Friedrich M, Schlößer S, Kölle J, Spranger R, Rauch P, Fraedrich K, Reif T, Karius-Fischer J, Balasubramanyam A, Henklein P, Fossen T, Schubert U. 2017. Proteolysis of mature HIV-1 p6 Gag protein by the insulin-degrading enzyme (IDE) regulates virus replication in an Env-dependent manner. *PLoS One*. 12(4):e0174254
172. Hanne J, Göttfert F, Schimer J, Anders-Össwein M, Konvalinka J, Engelhardt J, Müller B, Hell SW, Kräusslich H-G. 2016. Stimulated Emission Depletion Nanoscopy Reveals Time-Course of Human Immunodeficiency Virus Proteolytic Maturation. *ACS Nano*. 10(9):8215–22
173. Hargittai MRS, Gorelick RJ, Rouzina I, 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(4):951–68
174. Harrison GP, Lever AM. 1992. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* 66(7):4144–53
175. Harrison GP, Miele G, Hunter E, Lever AM. 1998. Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J. Virol.* 72(7):5886–96

176. Hart CE, Saltrelli MJ, Galphin JC, Schochetman G. 1995. A human chromosome 12-associated 83-kilodalton cellular protein specifically binds to the loop region of human immunodeficiency virus type 1 trans-activation response element RNA. *J. Virol.* 69(10):6593–99
177. Hayashi T, Ueno Y, Okamoto T. 1993. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS Lett.* 327(2):213–18
178. Heilman-Miller SL, Wu T, Levin JG. 2004. Alteration of nucleic acid structure and stability modulates the efficiency of minus-strand transfer mediated by the HIV-1 nucleocapsid protein. *J. Biol. Chem.* 279(42):44154–65
179. Helga-Maria C, Hammarskjöld M-L, Rekosh D. 1999. An Intact TAR Element and Cytoplasmic Localization Are Necessary for Efficient Packaging of Human Immunodeficiency Virus Type 1 Genomic RNA. *J. Virol.* 73(5):4127–35
180. Hemelaar J. 2012. The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18(3):182–92
181. Hendrix J, Baumgärtel V, Schrimpf W, Ivanchenko S, Digman MA, Gratton E, Kräusslich H-G, Müller B, Lamb DC. 2015. Live-cell observation of cytosolic HIV-1 assembly onset reveals RNA-interacting Gag oligomers. *J. Cell Biol.* 210(4):629–46
182. Henriet S, Sinck L, Bec G, Gorelick RJ, Marquet R, Paillart J-C. 2007. Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res.* 35(15):5141–53
183. Hill MK, Shehu-Xhilaga M, Campbell SM, Poumbourios P, Crowe SM, Mak J. 2003. The Dimer Initiation Sequence Stem-Loop of Human Immunodeficiency Virus Type 1 Is Dispensable for Viral Replication in Peripheral Blood Mononuclear Cells. *J. Virol.* 77(15):8329–35
184. Hill MK, Shehu-Xhilaga M, Crowe SM, Mak J. 2002. Proline Residues within Spacer Peptide p1 Are Important for Human Immunodeficiency Virus Type 1 Infectivity, Protein Processing, and Genomic RNA Dimer Stability. *J. Virol.* 76(22):11245–53
185. Houzet L, Paillart JC, Smagulova F, Maurel S, Morichaud Z, Marquet R, Mougel M. 2007. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res.* 35(8):2695–2704
186. Huang Y, Khorchid A, Gabor J, Wang J, Li X, Darlix JL, Wainberg MA, Kleiman L. 1998. The role of nucleocapsid and U5 stem/A-rich loop sequences in tRNA(3Lys) genomic placement and initiation of reverse transcription in human immunodeficiency virus type 1. *J. Virol.* 72(5):3907–15
187. Huang Y, Mak J, Cao Q, Li Z, Wainberg MA, Kleiman L. 1994. Incorporation of excess wild-type and mutant tRNA(3Lys) into human immunodeficiency virus type 1. *J. Virol.* 68(12):7676–83
188. Hulme AE, Perez O, Hope TJ. 2011. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U. S. A.* 108(24):9975–80
189. Hunter E. 1997. *Viral Entry and Receptors*, Vol. 1087. 135–58 pp.
190. Hurley JH. 2015. ESCRTs are everywhere. *EMBO J.* 34(19):2398–2407
191. Huthoff H, Berkhout B. 2001. Two alternating structures of the HIV-1 leader RNA. *RNA*. 7(1):143–57
192. Incarnato D, Neri F, Anselmi F, Oliviero S. 2014. Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15(10):491
193. Isel C, Ehresmann C, Ehresmann B, Marquet R. 1996. Determining the conformation of RNAs in solution. Application to a retroviral system: structure of the HIV-1 primer binding site region and effect of tRNA(3Lys) binding. *Pharm. Acta Helv.* 71(1):11–19
194. Isel C, Ehresmann C, Keith G, Ehresmann B, Marquet R. 1995. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.* 247(2):236–50
195. Isel C, Ehresmann C, Marquet R. 2010. Initiation of HIV Reverse Transcription. *Viruses*. 2(1):213–43
196. Isel C, Marquet R, Keith G, Ehresmann C, Ehresmann B. 1993. Modified nucleotides of

tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.* 268(34):25269–72
197. Isel C, Westhof E, Massire C, Le Grice SFJ, Ehresmann B, Ehresmann C, Marquet R. 1999. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J.* 18(4):1038–48
198. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*. 331(6153):280–83
199. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P. 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90(13):6320–24
200. Jacques DA, McEwan WA, Hilditch L, Price AJ, Towers GJ, James LC. 2016. HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*. 536(7616):349–53
201. Jalalirad M, Laughrea M. 2010. 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*. 407(2):225–36
202. Jalalirad M, Saadatmand J, Laughrea M. 2012. Dominant role of the 5' TAR bulge in dimerization of HIV-1 genomic RNA, but no evidence of TAR-TAR kissing during in vivo virus assembly. *Biochemistry*. 51(18):3744–58
203. Ji X, Klarmann GJ, Preston BD. 1996. Effect of human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein on HIV-1 reverse transcriptase activity in vitro. *Biochemistry*. 35(1):132–43
204. Jiang J, Ablan SD, Derebail S, Hercík K, Soheilian F, Thomas JA, Tang S, Hewlett I, Nagashima K, Gorelick RJ, Freed EO, Levin JG. 2011. The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability. *Virology*. 421(2):253–65
205. Jiang M, Mak J, Ladha A, Cohen E, Klein M, Rovinski B, Kleiman L. 1993. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. *J. Virol.* 67(6):3246–53
206. Jones CP, Datta S a K, Rein A, Rouzina I, Musier-Forsyth K. 2011. Matrix domain modulates HIV-1 Gag's nucleic acid chaperone activity via inositol phosphate binding. *J. Virol.* 85(4):1594–1603
207. Jouvenet N. 2012. Dynamics of ESCRT proteins. *Cell. Mol. Life Sci.* 69(24):4121–33
208. Jouvenet N, Bieniasz PD, Simon SM. 2008. Imaging the biogenesis of individual HIV-1 virions in live cells. *Nature*. 454(7201):236–40
209. Julien J-P, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse P-J, Burton DR, Sanders RW, Moore JP, Ward AB, Wilson IA. 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science*. 342(6165):1477–83
210. Jurado KA, Wang H, Slaughter A, Feng L, Kessl JJ, Koh Y, Wang W, Ballandras-Colas A, Patel PA, Fuchs JR, Kvaratskhelia M, Engelman A. 2013. Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation. *Proc. Natl. Acad. Sci. U. S. A.* 110(21):8690–95
211. Kafaie J, Dolatshahi M, Ajamian L, Song R, Mouland AJ, Rouiller I, Laughrea M. 2009. Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization. *Virology*. 385(1):233–44
212. Kafaie J, Song R, Abrahamyan L, Mouland AJ, Laughrea M. 2008. Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging. *Virology*. 375(2):592–610
213. Kaminska M, Shalak V, Francin M, Mirande M. 2007. Viral hijacking of mitochondrial lysyl-tRNA synthetase. *J. Virol.* 81(1):68–73
214. Kaplan AH, Manchester M, Swanstrom R. 1994. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the

release of viral proteins and is required for release to occur with maximum efficiency. *J. Virol.* 68(10):6782–86
215. Kaplan AH, Zack JA, Knigge M, Paul DA, Kempf DJ, Norbeck DW, Swanstrom R. 1993. Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *J. Virol.* 67(7):4050–55
216. Karabiber F, McGinnis JL, Favorov O V, Weeks KM. 2013. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA*. 19(1):63–73
217. Karn J, Stoltzfus CM. 2012. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med.* 2(2):a006916
218. Karsten C, Pöhlmann S. 2013. Attachment/Binding. In *Encyclopedia of AIDS*, pp. 1–10. New York, NY: Springer New York
219. Kaye JF, Lever a M. 1998. Nonreciprocal packaging of human immunodeficiency virus type 1 and type 2 RNA: a possible role for the p2 domain of Gag in RNA encapsidation. *J. Virol.* 72(7):5877–85
220. Keane SC, Heng X, Lu K, Kharytonchyk S, Ramakrishnan V, Carter G, Barton S, Hosic A, Florwick A, Santos J, Bolden NC, McCowin S, Case DA, Johnson B, Salemi M, Telesnitsky A, Summers MF. 2015. Structure of the HIV-1 RNA Packaging Signal. *Science*. 348(6237):917–21
221. Keene SE, King SR, Telesnitsky A. 2010. 7SL RNA is retained in HIV-1 minimal virus-like particles as an S-domain fragment. *J. Virol.* 84(18):9070–77
222. Keller PW, Huang RK, England MR, Waki K, Cheng N, Heymann JB, Craven RC, Freed EO, Steven AC. 2013. A two-pronged structural analysis of retroviral maturation indicates that core formation proceeds by a disassembly-reassembly pathway rather than a displacive transition. *J. Virol.* 87(24):13655–64
223. Kemler I, Meehan A, Poeschla EM. 2010. Live-cell coimaging of the genomic RNAs and Gag proteins of two lentiviruses. *J. Virol.* 84(13):6352–66
224. Kennedy EM, Bogerd HP, Kornepati AVR, Kang D, Ghoshal D, Marshall JB, Poling BC, Tsai K, Gokhale NS, Horner SM, Cullen BR. 2016. Posttranscriptional m(6)A Editing of HIV-1 mRNAs Enhances Viral Gene Expression. *Cell Host Microbe*. 19(5):675–85
225. Kenyon JC, Prestwood LJ, Lever AML. 2015. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci. Rep.* 5:
226. Kerwood DJ, Cavaluzzi MJ, Borer PN. 2001. Structure of SL4 RNA from the HIV-1 packaging signal. *Biochemistry*. 40(48):14518–29
227. Kessl JJ, Kutluay SB, Townsend D, Rebensburg S, Slaughter A, Larue RC, Shkriabai N, Bakouche N, Fuchs JR, Bieniasz PD, Kvaratskhelia M. 2016. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. 166(5):1257–68.e12
228. Kessl JJ, Kutluay SB, Townsend D, Rebensburg S, Slaughter A, Larue RC, Shkriabai N, Bakouche N, Fuchs JR, Bieniasz PD, Kvaratskhelia M. 2016. HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis. *Cell*. 166(5):1257–68.e12
229. Khan M a, Goila-Gaur R, Opi S, Miyagi E, Takeuchi H, Kao S, Strebel K. 2007. Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. *Retrovirology*. 4:48
230. Khan R, Giedroc DP. 1994. Nucleic acid binding properties of recombinant Zn2 HIV-1 nucleocapsid protein are modulated by COOH-terminal processing. *J. Biol. Chem.* 269(36):22538–46
231. Kharytonchyk S, Monti S, Smaldino PJ, Van V, Bolden NC, Brown JD, Russo E, Swanson C, Shuey A, Telesnitsky A, Summers MF. 2016. Transcriptional start site heterogeneity modulates the structure and function of the HIV-1 genome. *Proc. Natl. Acad. Sci. U. S. A.* 113(47):13378–83
232. Khorchid a, Javanbakht H, Wise S, Halwani R, Parniak M a, Wainberg M a, Kleiman L. 2000. Sequences within Pr160gag-pol affecting the selective packaging of primer tRNA(Lys3) into

HIV-1. *J. Mol. Biol.* 299(1):17–26
233. Kieken F, Arnoult E, Barbault F, Paquet F, Huynh-Dinh T, Paoletti J, Genest D, Lancelot G. 2002. HIV-1(Lai) genomic RNA: combined used of NMR and molecular dynamics simulation for studying the structure and internal dynamics of a mutated SL1 hairpin. *Eur. Biophys. J.* 31(7):521–31
234. Kieken F, Paquet F, Brulé F, Paoletti J, Lancelot G. 2006. A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. *Nucleic Acids Res.* 34(1):343–52
235. King AM. 1976. High molecular weight RNAs from Rous sarcoma virus and Moloney murine leukemia virus contain two subunits. *J. Biol. Chem.* 251(1):141–49
236. Klasens BIF, Das AT, Berkhout B. 1998. Inhibition of polyadenylation by stable RNA secondary structure. *Nucleic Acids Res.* 26(8):1870–76
237. Kleiman L, Jones CP, Musier-Forsyth K. 2010. Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett.* 584(2):359–65
238. Klein KC, Reed JC, Tanaka M, Nguyen VT, Giri S, Lingappa JR. 2011. HIV Gag-Leucine Zipper Chimeras Form ABCE1-Containing Intermediates and RNase-Resistant Immature Capsids Similar to Those Formed by Wild-Type HIV-1 Gag ▿. *J. Virol.* 85(14):7419–35
239. Knipe D, Howley P. 2013. Fields Virology. Wolters Kluwer. 6th editio ed.
240. Kobbi L, Octobre G, Dias J, Comisso M, Mirande M. 2011. Association of mitochondrial Lysyl-tRNA synthetase with HIV-1 GagPol involves catalytic domain of the synthetase and transframe and integrase domains of Pol. *J. Mol. Biol.* 410(5):875–86
241. Kondo E, Göttlinger HG. 1996. A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *J. Virol.* 70(1):159–64
242. Kondo E, Mammano F, Cohen E a, Göttlinger HG. 1995. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *J. Virol.* 69(5):2759–64
243. Konvalinka J, Kräusslich H-G, Müller B. 2015. Retroviral proteases and their roles in virion maturation. *Virology*. 479-480:403–17
244. Kräusslich HG. 1992. Specific inhibitor of human immunodeficiency virus proteinase prevents the cytotoxic effects of a single-chain proteinase dimer and restores particle formation. *J. Virol.* 66(1):567–72
245. Kräusslich HG, Fäcke M, Heuser a M, Konvalinka J, Zentgraf H. 1995. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J. Virol.* 69(6):3407–19
246. Kutluay SB, Zang T, Blanco-Melo D, Powell C, Jannain D, Errando M, Bieniasz PD. 2014. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell*. 159(5):1096–1109
247. L'Hernault A, Greatorex JS, Crowther RA, Lever AM. 2007. Dimerisation of HIV-2 genomic RNA is linked to efficient RNA packaging, normal particle maturation and viral infectivity. *Retrovirology*. 4:90
248. L'Hernault A, Weiss EU, Greatorex JS, Lever AM. 2012. HIV-2 Genome Dimerization Is Required for the Correct Processing of Gag: a Second-Site Reversion in Matrix Can Restore Both Processes in Dimerization-Impaired Mutant Viruses. *J. Virol.* 86(10):5867–76
249. Lanchy J-M, Ivanovitch JD, Lodmell JS. 2003. A structural linkage between the dimerization and encapsidation signals in HIV-2 leader RNA. *RNA*. 9(8):1007–18
250. Lanman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, Prevelige PE. 2004. Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat. Struct. Mol. Biol.* 11(7):676–77
251. Laspia MF, Rice AP, Mathews MB. 1989. HIV-1 Tat protein increases transcriptional initiation and stabilizes elongation. *Cell*. 59(2):283–92
252. Laughrea M, Jetté L. 1994. A 19-nucleotide sequence upstream of the 5' major splice donor is part of the dimerization domain of human immunodeficiency virus 1 genomic RNA.

Biochemistry. 33(45):13464–74
253. Lawrence DC, Stover CC, Noznitsky J, Wu Z, Summers MF. 2003. Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J. Mol. Biol.* 326(2):529–42
254. Lee K, Ambrose Z, Martin TD, Oztop I, Mulky A, Julias JG, Vandegraaff N, Baumann JG, Wang R, Yuen W, Takemura T, Shelton K, Taniuchi I, Li Y, Sodroski J, Littman DR, Coffin JM, Hughes SH, Unutmaz D, Engelman A, KewalRamani VN. 2010. Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*. 7(3):221–33
255. Lee S-K, Harris J, Swanstrom R. 2009. A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *J. Virol.* 83(17):8536–43
256. Lee S-K, Potempa M, Kolli M, Özen A, Schiffer CA, Swanstrom R. 2012. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J. Biol. Chem.* 287(16):13279–90
257. Lelek M, Di Nunzio F, Henriques R, Charneau P, Arhel N, Zimmer C. 2012. Superresolution imaging of HIV in infected cells with FlAsH-PALM. *Proc. Natl. Acad. Sci. U. S. A.* 109(22):8564–69
258. Lener D, Tanchou V, Roques BP, Le Grice SF, Darlix JL. 1998. Involvement of HIV-I nucleocapsid protein in the recruitment of reverse transcriptase into nucleoprotein complexes formed in vitro. *J. Biol. Chem.* 273(50):33781–86
259. Leschonsky B, Ludwig C, Bieler K, Wagner R. 2007. Capsid stability and replication of human immunodeficiency virus type 1 are influenced critically by charge and size of Gag residue 183. *J. Gen. Virol.* 88(Pt 1):207–16
260. Lever A, Gottlinger H, Haseltine W, Sodroski J. 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* 63(9):4085–87
261. Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* 7(6):754–74
262. Li X, Mak J, Arts EJ, Gu Z, Kleiman L, Wainberg MA, Parniak MA. 1994. Effects of alterations of primer-binding site sequences on human immunodeficiency virus type 1 replication. *J. Virol.* 68(10):6198–6206
263. Liang C, Li X, Rong L, Inouye P, Quan Y, Kleiman L, Wainberg MA. 1997. The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. *J. Virol.* 71(8):5750–57
264. Liang C, Rong L, Cherry E, Kleiman L, Laughrea M, Wainberg MA. 1999. Deletion Mutagenesis within the Dimerization Initiation Site of Human Immunodeficiency Virus Type 1 Results in Delayed Processing of the p2 Peptide from Precursor Proteins. *J. Virol.* 73(7):6147–51
265. Liang C, Rong L, Laughrea M, Kleiman L, Wainberg M a. 1998. Compensatory point mutations in the human immunodeficiency virus type 1 Gag region that are distal from deletion mutations in the dimerization initiation site can restore viral replication. *J. Virol.* 72(8):6629–36
266. Linde ME, Colquhoun DR, Ubaida Mohien C, Kole T, Aquino V, Cotter R, Edwards N, Hildreth JEK, Graham DR. 2013. The conserved set of host proteins incorporated into HIV-1 virions suggests a common egress pathway in multiple cell types. *J. Proteome Res.* 12(5):2045–54
267. Liu H, Dow EC, Arora R, Kimata JT, Bull LM, Arduino RC, Rice AP. 2006. Integration of human immunodeficiency virus type 1 in untreated infection occurs preferentially within genes. *J. Virol.* 80(15):7765–68
268. Loughrey D, Watters KE, Settle AH, Lucks JB. 2014. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* 42(21):gku909
269. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hiremath A, Divakaruni SS, LaCotti C, Barton S, Tummillo D, Hosic A, Edme K, Albrecht S,

Telesnitsky A, Summers MF. 2011. NMR Detection of Structures in the HIV-1 5´-Leader RNA that Regulate Genome Packaging. *Science*. 334(6053):242–45
270. Luban J, Bossolt KL, Franke EK, Kalpana G V, Goff SP. 1993. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell*. 73(6):1067–78
271. Luban J, Goff SP. 1994. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* 68(6):3784–93
272. Lucks JB, Mortimer SA, Trapnell C, Luo S, Aviran S, Schroth GP, Pachter L, Doudna JA, Arkin AP. 2011. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U. S. A.* 108(27):11063–68
273. Lusic M, Siliciano RF. 2017. Nuclear landscape of HIV-1 infection and integration. *Nat. Rev. Microbiol.* 15(2):69–82
274. Lyumkis D, Julien J, de Val N, Cupo A, Potter CS, Klasse P, Burton DR, Sanders RW, Moore JP, Carragher B, Wilson IA, Ward AB. 2013. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science*. 342(6165):1484–90
275. Madison MK, Lawson DQ, Elliott J, Ozantürk AN, Koneru PC, Townsend D, Errando M, Kvaratskhelia M, Kutluay SB. 2017. Allosteric HIV-1 Integrase Inhibitors Lead to Premature Degradation of the Viral RNA Genome and Integrase in Target Cells. *J. Virol.*, pp. JVI.00821–17
276. Malim MH, Hauber J, Le SY, Maizel J V., Cullen BR. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature*. 338(6212):254–57
277. Mammano F, Ohagen A, Höglund S, Göttlinger HG. 1994. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J. Virol.* 68(8):4927–36
278. Mangel WF, Delius H, Duesberg PH. 1974. Structure and molecular weight of the 60-70S RNA and the 30-40S RNA of the Rous sarcoma virus. *Proc. Natl. Acad. Sci. U. S. A.* 71(11):4541–45
279. Maquat LE. 1995. When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA*. 1(5):453–65
280. Marquet R, Baudin F, Gabus C, Darlix JL, Mougel M, Ehresmann C, Ehresmann B. 1991. Dimerization of human immunodeficiency virus (type 1) RNA: stimulation by cations and possible mechanism. *Nucleic Acids Res.* 19(9):2349–57
281. Marquet R, Isel C, Ehresmann C, Ehresmann B. 1995. tRNAs as primer of reverse transcriptases. *Biochimie*. 77(1-2):113–24
282. Matreyek KA, Engelman A. 2013. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses*. 5(10):2483–2511
283. Matreyek KA, Yücel SS, Li X, Engelman A. 2013. Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog.* 9(10):e1003693
284. Mattei S, Anders M, Konvalinka J, Kräusslich H-G, Briggs JAG, Müller B. 2014. Induced maturation of human immunodeficiency virus. *J. Virol.* 88(23):13722–31
285. Mattei S, Flemming A, Anders-Össwein M, Kräusslich H-G, Briggs JAG, Müller B. 2015. RNA and Nucleocapsid Are Dispensable for Mature HIV-1 Capsid Assembly. *J. Virol.* 89(19):9739–47
286. Mattei S, Glass B, Hagen WJH, Kräusslich H, Briggs JAG. 2016. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science*. 354(6318):1434–37
287. McBride MS, Panganiban AT. 1996. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* 70(5):2963–73
288. McBride MS, Schwartz MD, Panganiban AT. 1997. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J. Virol.* 71(6):4544–54
289. McKinstry WJ, Hijnen M, Tanwar HS, Sparrow LG, Nagarajan S, Pham ST, Mak J. 2014. Expression and purification of soluble recombinant full length HIV-1 Pr55(Gag) protein in Escherichia coli. *Protein Expr. Purif.* 100(2014):10–18

290. McLaughlin S, Aderem A. 1995. The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* 20(7):272–76
291. Mendonça LM, Poeys SC, Abreu CM, Tanuri A, Costa LJ. 2014. HIV-1 Nef inhibits Protease activity and its absence alters protein content of mature viral particles. *PLoS One*. 9(4):e95352
292. Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Justome A, Delain E, Gorelick RJ, Le Cam E. 2006. 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.* 364(3):496–511
293. Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Borde I, Reboud-Ravaux M, Restle T, Gorelick RJ, Le Cam E. 2007. HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. *PLoS One*. 2(7):e669
294. Mirambeau G, Lyonnais S, Gorelick RJ. 2010. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol.* 7(6):724–34
295. Moore MD, Fu W, Soheilian F, Nagashima K, Ptak RG, Pathak VK, Hu W-S. 2008. Suboptimal inhibition of protease activity in human immunodeficiency virus type 1: effects on virion morphogenesis and RNA maturation. *Virology*. 379(1):152–60
296. Moore MD, Nikolaitchik OA, Chen J, Hammarskjöld M-L, Rekosh D, Hu W-S. 2009. Probing the HIV-1 Genomic RNA Trafficking Pathway and Dimerization by Genetic Recombination and Single Virion Analyses. *PLoS Pathog.* 5(10):
297. Morellet N, Druillennec S, Lenoir C, Bouaziz S, Roques BP. 2005. Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging. *Protein Sci.* 14(2):375–86
298. Morellet N, Jullian N, De Rocquigny H, Maigret B, Darlix JL, Roques BP. 1992. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *EMBO J.* 11(8):3059–65
299. Mortimer SA, Trapnell C, Aviran S, Pachter L, Lucks JB. 2009. SHAPE–Seq: High-Throughput RNA Structure Analysis. John Wiley & Sons, Inc.
300. Mueller N, Das AT, Berkhout B. 2016. A Phylogenetic Survey on the Structure of the HIV-1 Leader RNA Domain That Encodes the Splice Donor Signal. *Viruses*. 8(7):200
301. Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG. 1998. Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat. Struct. Biol.* 5(6):432–36
302. Mujeeb A, Parslow TG, Zarrinpar A, Das C, James TL. 1999. NMR structure of the mature dimer initiation complex of HIV-1 genomic RNA. *FEBS Lett.* 458(3):387–92
303. Mujeeb A, Ulyanov NB, Georgantis S, Smirnov I, Chung J, Parslow TG, James TL. 2007. Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: sequence effects and mechanism of RNA refolding. *Nucleic Acids Res.* 35(6):2026–34
304. Müller B, Anders M, Akiyama H, Welsch S, Glass B, Nikovics K, Clavel F, Tervo H-M, Keppler OT, Kräusslich H-G. 2009. HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity. *J. Biol. Chem.* 284(43):29692–703
305. Mundigala H, Michaux JB, Feig AL, Ennifar E, Rueda D. 2014. HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. *Nucleic Acids Res.* 42(11):7281–89
306. Munro JB, Nath A, Färber M, Datta SAK, Rein A, Rhoades E, Mothes W. 2014. A conformational transition observed in single HIV-1 Gag molecules during in vitro assembly of virus-like particles. *J. Virol.* 88(6):3577–85
307. Muriaux D, Fossé P, Paoletti J. 1996. A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry*. 35(15):5075–82
308. Muriaux D, Girard PM, Bonnet-Mathonière B, Paoletti J. 1995. Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5’ leader region is evidenced by an antisense oligonucleotide. *J. Biol. Chem.* 270(14):8209–16

309. Muriaux D, Mirro J, Harvin D, Rein A. 2001. RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci. U. S. A.* 98(9):5246–51
310. Muriaux D, Rocquigny H De, Roques B, Paoletti J. 1996. Molecular Genetics : NCp7 Activates HIV-1 Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer NCp7 Activates HIV-1 Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer *. *J Biol Chem.* 271(52):33686–92
311. Murti KG, Bondurant M, Tereba A. 1981. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol.* 37(1):411–19
312. Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. 2012. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* 40(20):10345–55
313. Ohishi M, Nakano T, Sakuragi S, Shioda T, Sano K, Sakuragi J. 2011. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res.* 39(8):3404–17
314. Ohlmann T, Mengardi C, López-Lastra M. 2014. Translation initiation of the HIV-1 mRNA. *Transl. (Austin, Tex.)*. 2(2):e960242
315. Onafuwa-Nuga A, Telesnitsky A. 2009. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol. Mol. Biol. Rev.* 73(3):451–80, Table of Contents
316. Onafuwa-Nuga AA, Telesnitsky A, King SR. 2006. 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA*. 12(4):542–46
317. Ott DE. 2008. Cellular proteins detected in HIV-1. *Rev. Med. Virol.* 18(3):159–75
318. Ott DE, Coren L V, Shatzer T. 2009. The nucleocapsid region of human immunodeficiency virus type 1 Gag assists in the coordination of assembly and Gag processing: role for RNA-Gag binding in the early stages of assembly. *J. Virol.* 83(15):7718–27
319. Paillart J-C, Dettenhofer M, Yu X-F, Ehresmann C, Ehresmann B, Marquet R. 2004. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* 279(46):48397–403
320. Paillart JC, Berthoux L, Ottmann M, Darlix JL, Marquet R, Ehresmann B, Ehresmann C. 1996. 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.* 70(12):8348–54
321. Paillart JC, Göttlinger HG. 1999. Opposing effects of human immunodeficiency virus type 1 matrix mutations support a myristyl switch model of gag membrane targeting. *J. Virol.* 73(4):2604–12
322. Paillart JC, Marquet R, Skripkin E, Ehresmann B, Ehresmann C. 1994. Mutational analysis of the bipartite dimer linkage structure of human immunodeficiency virus type 1 genomic RNA. *J. Biol. Chem.* 269(44):27486–93
323. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 1996. 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.* 93(11):5572–77
324. Paillart JC, Skripkin E, Ehresmann B, Ehresmann C, Marquet R. 2002. In Vitro evidence for a long range pseudoknot in the 5’-untranslated and matrix coding regions of HIV-1 genomic RNA. *J. Biol. Chem.* 277(8):5995–6004
325. Paillart JC, Westhof E, Ehresmann C, Ehresmann B, Marquet R. 1997. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.* 270(1):36–49
326. Pallesen J. 2011. Structure of the HIV-1 5’ untranslated region dimer alone and in complex with gold nanocolloids: support of a TAR-TAR-containing 5' dimer linkage site (DLS) and a 3' DIS-DIS-containing DLS. *Biochemistry*. 50(28):6170–77
327. Pappalardo L, Kerwood DJ, Pelczer I, Borer PN. 1998. Three-dimensional folding of an RNA

hairpin required for packaging HIV-1. *J. Mol. Biol.* 282(4):801–18
328. Passos DO, Li M, Yang R, Rebensburg S V, Ghirlando R, Jeon Y, Shkriabai N, Kvaratskhelia M, Craigie R, Lyumkis D. 2017. Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science*. 355(6320):89–92
329. Pavon-Eternod M, Wei M, Pan T, Kleiman L. 2010. Profiling non-lysyl tRNAs in HIV-1. *RNA*. 16(2):267–73
330. Paxton W, Connor RI, Landau NR. 1993. Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *J. Virol.* 67(12):7229–37
331. Pettit SC, Everitt LE, Choudhury S, Dunn BM, Kaplan AH. 2004. Initial cleavage of the human immunodeficiency virus type 1 GagPol precursor by its activated protease occurs by an intramolecular mechanism. *J. Virol.* 78(16):8477–85
332. Pettit SC, Gulnik S, Everitt L, Kaplan AH. 2003. The dimer interfaces of protease and extra-protease domains influence the activation of protease and the specificity of GagPol cleavage. *J. Virol.* 77(1):366–74
333. Pettit SC, Henderson GJ, Schiffer CA, Swanstrom R. 2002. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease. *J. Virol.* 76(20):10226–33
334. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. 2005. 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*. 2:66
335. Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet P V, Klein CA, Swanstrom R. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* 68(12):8017–27
336. Poljak L, Batson SM, Ficheux D, Roques BP, Darlix J-L, Käs E. 2003. Analysis of NCp7-dependent activation of HIV-1 cDNA integration and its conservation among retroviral nucleocapsid proteins. *J. Mol. Biol.* 329(3):411–21
337. Potempa M, Nalivaika E, Ragland D, Lee S-K, Schiffer CA, Swanstrom R. 2015. A Direct Interaction with RNA Dramatically Enhances the Catalytic Activity of the HIV-1 Protease In Vitro. *J. Mol. Biol.* 427(14):2360–78
338. Poulsen LD, Kielpinski LJ, Salama SR, Krogh A, Vinther J. 2015. SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA*. 21(5):1042–52
339. Puglisi JD, Tan R, Calnan BJ, Frankel a D, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*. 257(5066):76–80
340. Purcell DF, Martin MA. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* 67(11):6365–78
341. Racine P-J, Chamontin C, de Rocquigny H, Bernacchi S, Paillart J-C, Mougel M. 2016. Requirements for nucleocapsid-mediated regulation of reverse transcription during the late steps of HIV-1 assembly. *Sci. Rep.* 6(January 2016):27536
342. Rankovic S, Varadarajan J, Ramalho R, Aiken C, Rousso I. 2017. Reverse transcription mechanically initiates HIV-1 capsid disassembly. *J. Virol.*
343. Rein A. 2010. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol.* 7(6):700–705
344. Rein A, Datta SAK, Jones CP, Musier-Forsyth K. 2011. DIVERSE INTERACTIONS OF RETROVIRAL GAG PROTEINS WITH RNAs. *Trends Biochem. Sci.* 36(7):373–80
345. Rein A, Henderson LE, Levin JG. 1998. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem. Sci.* 23(8):297–301
346. Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*. 11(1):129
347. Rice AP. 2015. Roles of microRNAs and long-noncoding RNAs in human immunodeficiency virus replication. *Wiley Interdiscip. Rev. RNA*. 6(6):661–70
348. Rice WG, Supko JG, Malspeis L, Buckheit RW, Clanton D, Bu M, Graham L, Schaeffer C a,

Turpin J a, Domagala J, Gogliotti R, Bader JP, Halliday SM, Coren L, Sowder RC, Arthur LO, Henderson LE. 1995. Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science*. 270(5239):1194–97

349. Richter S, Ping Y-H, Rana TM. 2002. TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. *Proc. Natl. Acad. Sci. U. S. A.* 99(12):7928–33
350. Riggin CH, Bondurant M, Mitchell WM. 1975. Physical properties of moloney murine leukemia virus high-molecular-weight RNA: a two subunit structure. *J. Virol.* 16(6):1528–35
351. Robinson BA, Reed JC, Geary CD, Swain JV, Lingappa JR. 2014. A temporospatial map that defines specific steps at which critical surfaces in the Gag MA and CA domains act during immature HIV-1 capsid assembly in cells. *J. Virol.* 88(10):5718–41
352. Robinson BS, Riccardi KA, Gong YF, Guo Q, Stock DA, Blair WS, Terry BJ, Deminie CA, Djang F, Colonno RJ, Lin PF. 2000. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.* 44(8):2093–99
353. Roldan A, Liang C, Wainberg MA. 2011. Structural changes in the SL5 and SL6 leader sequences of HIV-1 RNA following interactions with the viral mGag protein. *Virus Res.* 155(1):98–105
354. Roldan A, Russell RS, Marchand B, Götte M, Liang C, Wainberg MA. 2004. In vitro identification and characterization of an early complex linking HIV-1 genomic RNA recognition and Pr55Gag multimerization. *J. Biol. Chem.* 279(38):39886–94
355. Rosa A, Chande A, Ziglio S, De Sanctis V, Bertorelli R, Goh SL, McCauley SM, Nowosielska A, Antonarakis SE, Luban J, Santoni FA, Pizzato M. 2015. HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*. 526(7572):212–17
356. Rossio JL, Esser MT, Suryanarayana K, Schneider DK, Bess JW, Vasquez GM, Wiltrout TA, Chertova E, Grimes MK, Sattentau Q, Arthur LO, Henderson LE, Lifson AJD. 1998. Inactivation of Human Immunodeficiency Virus Type 1 Infectivity with Preservation of Conformational and Functional Integrity of Virion Surface Proteins. *J. Virol.* 72(10):7992–8001
357. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. 2014. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 505(7485):701–5
358. Roy C, Tounekti N, Mougel M, Darlix JL, Paoletti C, Ehresmann C, Ehresmann B, Paoletti J. 1990. An analytical study of the dimerization of in vitro generated RNA of Moloney murine leukemia virus MoMuLV. *Nucleic Acids Res.* 18(24):7287–92
359. Roy NH, Chan J, Lambelé M, Thali M. 2013. Clustering and mobility of HIV-1 Env at viral assembly sites predict its propensity to induce cell-cell fusion. *J. Virol.* 87(13):7516–25
360. Rulli SJ, Hibbert CS, Mirro J, Pederson T, Biswal S, Rein A. 2007. Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* 81(12):6623–31
361. Russell RS, Hu J, Bériault V, Mouland AJ, Laughrea M, Kleiman L, Wainberg MA, Liang C. 2003. Sequences downstream of the 5' splice donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J. Virol.* 77(1):84–96
362. Russell RS, Hu J, Laughrea M, Wainberg MA, Liang C. 2002. Deficient Dimerization of Human Immunodeficiency Virus Type 1 RNA Caused by Mutations of the U5 RNA Sequences. *Virology*. 303(1):152–63
363. Saadatmand J, Guo F, Cen S, Niu M, Kleiman L. 2008. Interactions of reverse transcriptase sequences in Pol with Gag and LysRS in the HIV-1 tRNALys3 packaging/annealing complex. *Virology*. 380(1):109–17
364. Saadatmand J, Niu M, Kleiman L, Guo F. 2009. The contribution of the primer activation signal to differences between Gag- and NCp7-facilitated tRNA(Lys3) annealing in HIV-1. *Virology*. 391(2):334–41
365. Sakaguchi K, Zambrano N, Baldwin ET, Shapiro B a, Erickson JW, Omichinski JG, Clore GM, Gronenborn a M, Appella E. 1993. Identification of a binding site for the human immunodeficiency virus type 1 nucleocapsid protein. *Proc. Natl. Acad. Sci. U. S. A.* 90(11):5219–23

366. Sakuragi J-I, Ueda S, Iwamoto A, Shioda T. 2003. Possible Role of Dimerization in Human Immunodeficiency Virus Type 1 Genome RNA Packaging. *J. Virol.* 77(7):4060–69
367. Sakuragi JI, Panganiban AT. 1997. Human immunodeficiency virus type 1 RNA outside the primary encapsidation and dimer linkage region affects RNA dimer stability in vivo. *J. Virol.* 71(4):3250–54
368. Sakuragi S, Yokoyama M, Shioda T, Sato H, Sakuragi J. 2016. SL1 revisited: functional analysis of the structure and conformation of HIV-1 genome RNA. *Retrovirology*. 13(1):79
369. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E. 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* 385(3):693–713
370. Schneeberger E-M, Breuker K. 2017. Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type tat Peptide for Binding Site Mapping. *Angew. Chem. Int. Ed. Engl.* 56(5):1254–58
371. Schopman NCT, van Montfort T, Willemsen M, Knoepfel SA, Pollakis G, van Kampen A, Sanders RW, Haasnoot J, Berkhout B. 2012. Selective packaging of cellular miRNAs in HIV-1 particles. *Virus Res.* 169(2):438–47
372. Schur FKM, Hagen WJH, Rumlová M, Ruml T, Müller B, Kräusslich H-G, Briggs JAG. 2015. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature*. 517(7535):505–8
373. Schwartz MD, Geraghty RJ, Panganiban AT. 1996. HIV-1 particle release mediated by Vpu is distinct from that mediated by p6. *Virology*. 224(1):302–9
374. Seetin MG, Kladwang W, Bida JP, Das R. 2014. Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. , pp. 95–117. Springer
375. Seif E, Niu M, Kleiman L. 2015. In virio SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology*. 12:40
376. Sham HL, Kempf DJ, Molla A, Marsh KC, Kumar GN, Chen CM, Kati W, Stewart K, Lal R, Hsu A, Betebenner D, Korneyeva M, Vasavanonda S, McDonald E, Saldivar A, Wideburg N, Chen X, Niu P, Park C, Jayanti V, Grabowski B, Granneman GR, Sun E, Japour AJ, Leonard JM, Plattner JJ, Norbeck DW. 1998. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* 42(12):3218–24
377. Shehu-Xhilaga M, Crowe SM, Mak J. 2001. Maintenance of the Gag/Gag-Pol Ratio Is Important for Human Immunodeficiency Virus Type 1 RNA Dimerization and Viral Infectivity. *J. Virol.* 75(4):1834–41
378. Shehu-Xhilaga M, Hill M, Marshall JA, Kappes J, Crowe SM, Mak J. 2002. The Conformation of the Mature Dimeric Human Immunodeficiency Virus Type 1 RNA Genome Requires Packaging of Pol Protein. *J. Virol.* 76(9):4331–40
379. Shehu-Xhilaga M, Kraeusslich HG, Pettit S, Swanstrom R, Lee JY, Marshall JA, Crowe SM, Mak J. 2001. Proteolytic Processing of the P2/Nucleocapsid Cleavage Site Is Critical for Human Immunodeficiency Virus Type 1 RNA Dimer Maturation. *J. Virol.* 75(19):9156–64
380. Sheline CT, Milocco LH, Jones KA. 1991. Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. *Genes Dev.* 5(12B):2508–20
381. Sheng N, Pettit SC, Tritch RJ, Ozturk DH, Rayner MM, Swanstrom R, Erickson-Viitanen S. 1997. Determinants of the human immunodeficiency virus type 1 p15NC-RNA interaction that affect enhanced cleavage by the viral protease. *J. Virol.* 71(8):5723–32
382. Shoemaker CJ, Green R. 2012. Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* 19(6):594–601
383. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM. 2014. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods*. 11(9):959–65
384. Sinck L, Richer D, Howard J, Alexander M, Purcell DFJ, Marquet R, Paillart J. 2007. In vitro dimerization of human immunodeficiency virus type 1 (HIV-1) spliced RNAs. *RNA*. 13(12):2141–50
385. Skripkin E, Paillart JC, Marquet R, Blumenfeld M, Ehresmann B, Ehresmann C. 1996.

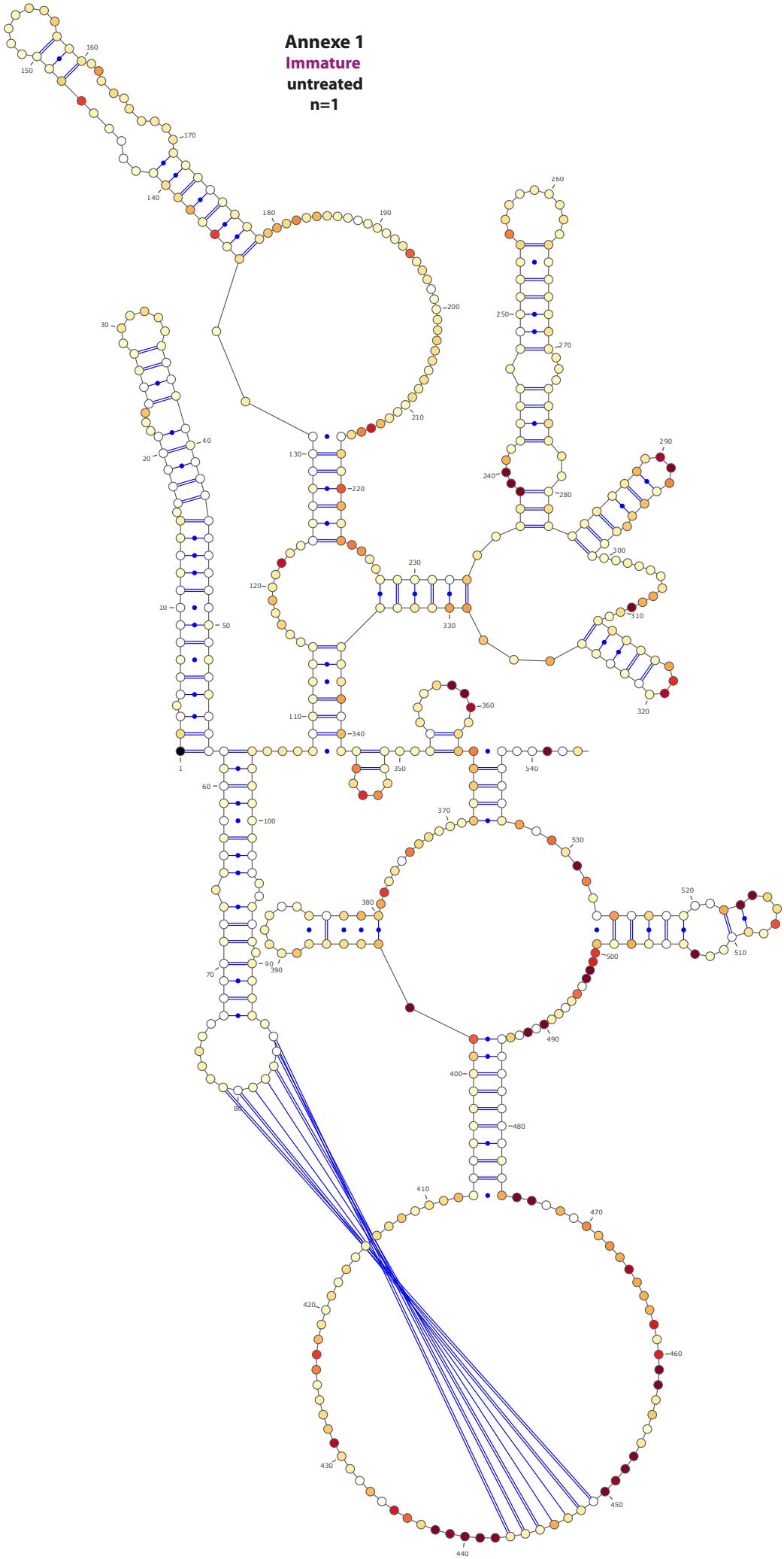
Mechanisms of inhibition of in vitro dimerization of HIV type I RNA by sense and antisense oligonucleotides. *J. Biol. Chem.* 271(46):28812–17
386. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. 1994. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 91(11):4945–49
387. Sleiman D, Barraud P, Brachet F, Tisne C. 2013. The Interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. *PLoS One*. 8(6):e64700
388. Smyth RP, Despons L, Huili G, Bernacchi S, Hijnen M, Mak J, Jossinet F, Weixi L, Paillart J-C, von Kleist M, Marquet R. 2015. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat. Methods*. 12(9):866–72
389. Song R, Kafaie J, Laughrea M. 2008. Role of the 5′ TAR Stem−Loop and the U5-AUG Duplex in Dimerization of HIV-1 Genomic RNA. *Biochemistry*. 47(10):3283–93
390. Song R, Kafaie J, Yang L, Laughrea M. 2007. HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. *J. Mol. Biol.* 371(4):1084–98
391. Soto-Rifo R, Rubilar PS, Ohlmann T. 2013. The DEAD-box helicase DDX3 substitutes for the cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. *Nucleic Acids Res.* 41(12):6286–99
392. Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung J-W, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, Chang HY. 2015. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*. advance on:
393. Stewart-Maynard KM, Cruceanu M, Wang F, Vo M-N, Gorelick RJ, Williams MC, Rouzina I, Musier-Forsyth K. 2008. Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J. Virol.* 82(20):10129–42
394. Stoltzfus CM, Snyder PN. 1975. Structure of B77 sarcoma virus RNA: stabilization of RNA after packaging. *J. Virol.* 16(5):1161–70
395. Stoylov SP, Vuilleumier C, Stoylova E, De Rocquigny H, Roques BP, Gérard D, Mély Y. 1997. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*. 41(3):301–12
396. Summers MF, Henderson LE, Chance MR, Bess JW, South TL, Blake PR, Sagi I, Perez-Alvarado G, Sowder RC, Hare DR. 1992. 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.* 1(5):563–74
397. Sundquist WI, Heaphy S. 1993. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA. *Proc. Natl. Acad. Sci. U. S. A.* 90(8):3393–97
398. Sundquist WI, Kräusslich H-G. 2012. HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* 2(7):a006924
399. Suzuki Y, Craigie R. 2007. The road to chromatin - nuclear entry of retroviruses. *Nat. Rev. Microbiol.* 5(3):187–96
400. Swanstrom R, Wills J. 1997. *Synthesis, Assembly, and Processing of Viral Proteins*
401. Talkish J, May G, Lin Y, Jr JLW, Mcmanus CJ. 2014. Mod-seq : high-throughput sequencing for chemical probing of RNA structure Mod-seq : high-throughput sequencing for chemical probing of RNA structure. *RNA*. 20:0–8
402. Tanchou V, Gabus C, Rogemond V, Darlix JL. 1995. Formation of stable and functional HIV-1 nucleoprotein complexes in vitro. *J. Mol. Biol.* 252(5):563–71
403. Tang S, Murakami T, Agresta BE, Campbell S, Freed EO, Levin JG. 2001. Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. *J. Virol.* 75(19):9357–66
404. Tang Y, Bouvier E, Kwok CK, Ding Y, Nekrutenko A, Bevilacqua PC, Assmann SM. 2015. StructureFold: Genome-wide RNA secondary structure mapping and reconstruction in vivo.

Bioinformatics. 31(16):2668–75
405. Tedbury PR, Novikova M, Ablan SD, Freed EO. 2016. Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. *Proc. Natl. Acad. Sci. U. S. A.* 113(2):E182–90
406. Telesnitsky A, Goff S. 1997. *Reverse Transcriptase and the Generation of Retroviral DNA*
407. Telesnitsky A, Wolin SL. 2016. The Host RNAs in Retroviral Particles. *Viruses*. 8(8):235
408. Terwilliger EF, Cohen EA, Lu YC, Sodroski JG, Haseltine WA. 1989. Functional role of human immunodeficiency virus type 1 vpu. *Proc. Natl. Acad. Sci. U. S. A.* 86(13):5163–67
409. Thierry S, Thierry E, Subra F, Deprez E, Leh H, Bury-Moné S, Delelis O. 2016. Opposite transcriptional regulation of integrated vs unintegrated HIV genomes by the NF-κB pathway. *Sci. Rep.* 6(April):25678
410. Tözsér J, Bláha I, Copeland TD, Wondrak EM, Oroszlan S. 1991. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett.* 281(1-2):77–80
411. Troutt AB, McHeyzer-Williams MG, Pulendran B, Nossal GJ. 1992. Ligation-anchored PCR: a simple amplification technique with single-sided specificity. *Proc. Natl. Acad. Sci. U. S. A.* 89(20):9823–25
412. Turner KB, Hagan NA, Fabris D. 2007. Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein. *J. Mol. Biol.* 369(3):812–28
413. Ulyanov NB, Mujeeb A, Du Z, Tonelli M, Parslow TG, James TL. 2006. NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J. Biol. Chem.* 281(23):16168–77
414. Urbaneja MA, Wu M, Casas-Finet JR, Karpel RL. 2002. HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. *J. Mol. Biol.* 318(3):749–64
415. Usami Y, Wu Y, Göttlinger HG. 2015. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature*. 526(7572):218–23
416. van Bel N, Das AT, Cornelissen MTE, Abbink TEM, Berkhout B. 2014. A Short Sequence Motif in the 5' Leader of the HIV-1 Genome Modulates Extended RNA Dimer Formation and Virus Replication. *J. Biol. Chem.*
417. Votteler J, Sundquist WI. 2013. Virus Budding and the ESCRT Pathway. *Cell Host Microbe*. 14(3):
418. Vuilleumier C, Bombarda E, Morellet N, Gérard D, Roques BP, Mély Y. 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry*. 38(51):16816–25
419. Wang S-W, Aldovini A. 2002. RNA incorporation is critical for retroviral particle integrity after cell membrane assembly of Gag complexes. *J. Virol.* 76(23):11853–65
420. Wang W, Naiyer N, Mitra M, Li J, Williams MC, Rouzina I, Gorelick RJ, Wu Z, Musier-Forsyth K. 2014. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res.* 42(11):7145–59
421. Watters KE, Abbott TR, Lucks JB. 2016. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.* 44(2):e12
422. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. *Nature*. 460(7256):711–16
423. Webb J a, Jones CP, Parent LJ, Rouzina I, Musier-Forsyth K. 2013. Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: implications for viral genomic RNA packaging. *RNA*. 19(8):1078–88
424. Whitney JB, Wainberg MA. 2006. Impaired RNA incorporation and dimerization in live attenuated leader-variants of SIVmac239. *Retrovirology*. 3(1):96
425. Wiegers K, Rutter G, Kottler H, Tessmer U, Hohenberg H, Kräusslich H-GG. 1998. Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of

individual Gag polyprotein cleavage sites. *J. Virol.* 72(4):2846–54
426. Wilkinson K a, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* 6(4):e96
427. Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* 1(3):1610–16
428. Woodward CL, Cheng SN, Jensen GJ. 2014. Electron Cryotomography Studies of Maturing HIV-1 Particles Reveal the Assembly Pathway of the Viral Core. *J. Virol.* 89(2):1267–77
429. Wright ER, Schooler JB, Ding HJ, Kieffer C, Fillmore C, Sundquist WI, Jensen GJ. 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* 26(8):2218–26
430. Wu H, Mitra M, Naufer MN, McCauley MJ, Gorelick RJ, Rouzina I, Musier-Forsyth K, Williams MC. 2014. Differential contribution of basic residues to HIV-1 nucleocapsid protein's nucleic acid chaperone function and retroviral replication. *Nucleic Acids Res.* 42(4):2525–37
431. Wu T, Datta SAK, Mitra M, Gorelick RJ, Rein A, Levin JG. 2010. Fundamental differences between the nucleic acid chaperone activities of HIV-1 nucleocapsid protein and Gag or Gag-derived proteins: biological implications. *Virology*. 405(2):556–67
432. Wu T, Gorelick RJ, Levin JG. 2014. Selection of fully processed HIV-1 nucleocapsid protein is required for optimal nucleic acid chaperone activity in reverse transcription. *Virus Res.* 193:52–64
433. Xing L, Liang C, Kleiman L. 2011. Coordinate roles of Gag and RNA helicase A in promoting the annealing of formula to HIV-1 RNA. *J. Virol.* 85(4):1847–60
434. Xu H, Franks T, Gibson G, Huber K, Rahm N, Strambio De Castillia C, Luban J, Aiken C, Watkins S, Sluis-Cremer N, Ambrose Z. 2013. Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. *Retrovirology*. 10(1):70
435. Yamashita M, Perez O, Hope TJ, Emerman M. 2007. Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS Pathog.* 3(10):1502–10
436. Yedavalli VSRK, Jeang K-T. 2010. Trimethylguanosine capping selectively promotes expression of Rev-dependent HIV-1 RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 107(33):14787–92
437. Zaitseva L, Myers R, Fassati A. 2006. tRNAs promote nuclear import of HIV-1 intracellular reverse transcription complexes. *PLoS Biol.* 4(10):e332
438. Zennou V, Petit C, Guetard D, Nerhbass U, Montagnier L, Charneau P. 2000. HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell*. 101(2):173–85
439. Zhang Y, Barklis E. 1997. Effects of nucleocapsid mutations on human immunodeficiency virus assembly and RNA encapsidation. *J. Virol.* 71(9):6765–76
440. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, Zhang P. 2013. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*. 497(7451):643–46
441. Zhou W, Parent LJ, Wills JW, Resh MD. 1994. Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* 68(4):2556–69
442. Zhu P, Liu J, Bess J, Chertova E, Lifson JD, Grisé H, Ofek G a, Taylor K a, Roux KH. 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature*. 441(7095):847–52
443. Zubradt M, Gupta P, Persad S, Lambowitz AM, Weissman JS, Rouskin S. 2017. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*. 14(1):75–82

ANNEXES

Annexe 1
Immature
untreated
n=1

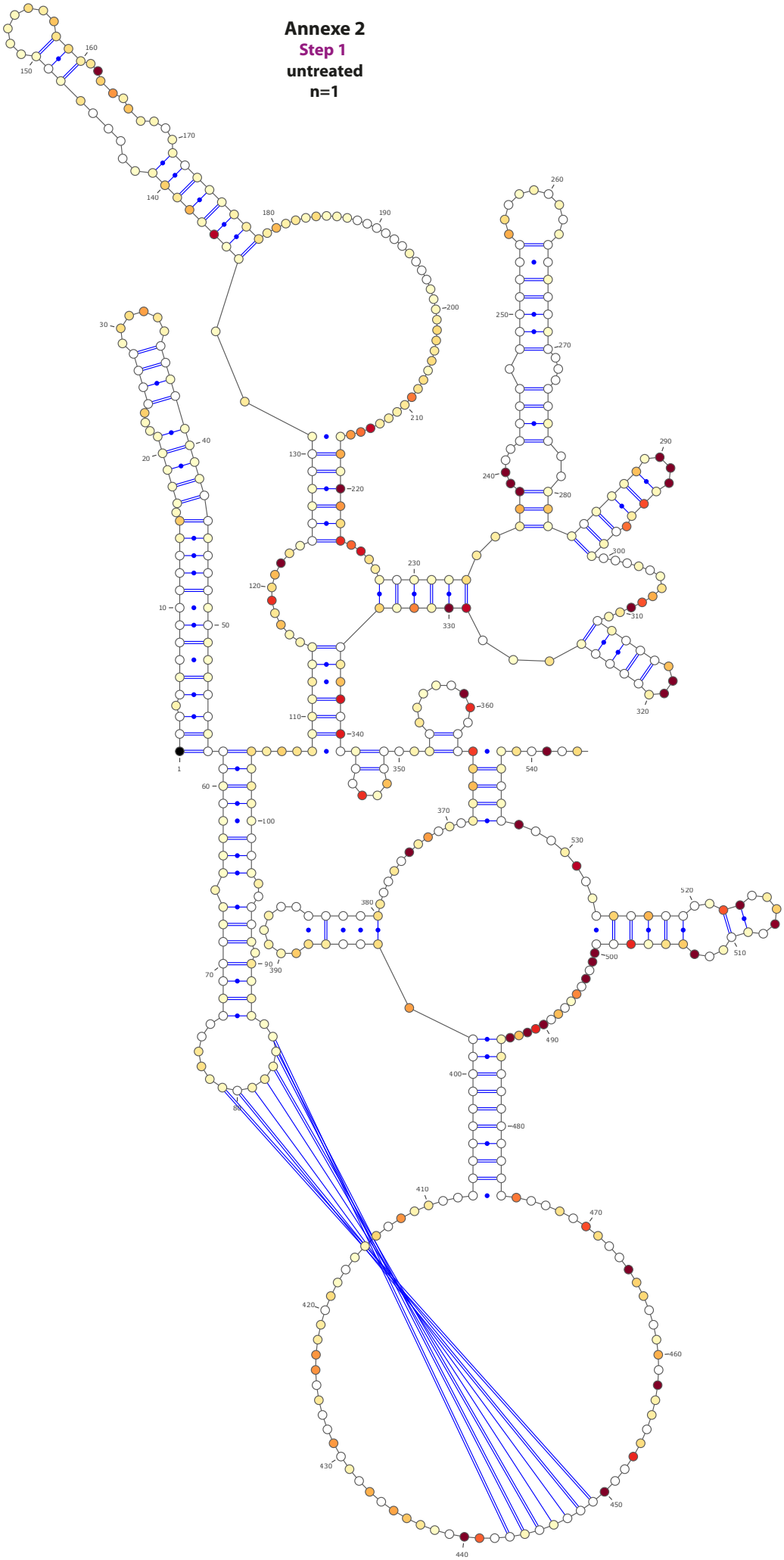


Annexe 2

Step 1

untreated

n=1

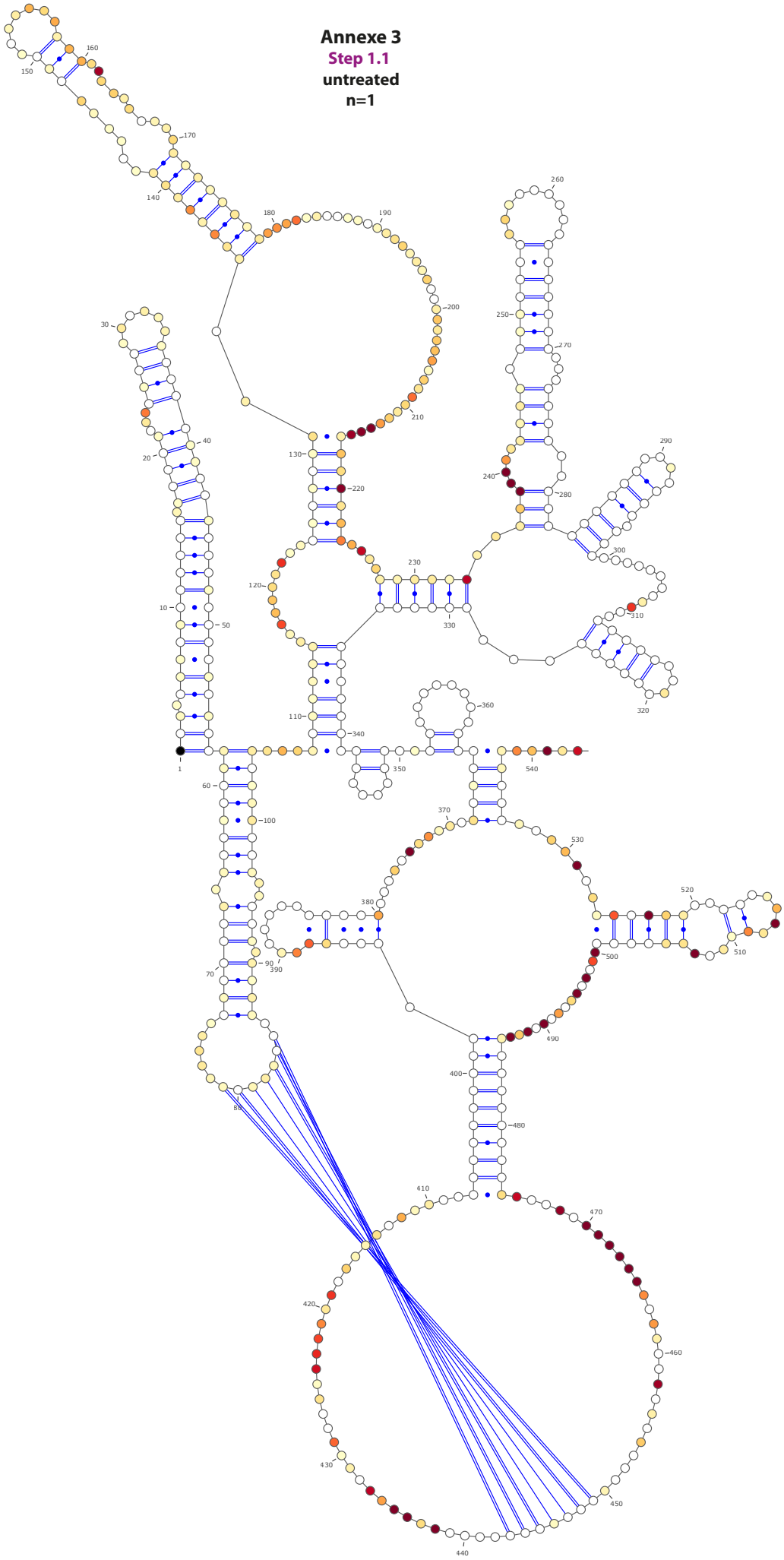


Annexe 3

Step 1.1

untreated

n=1

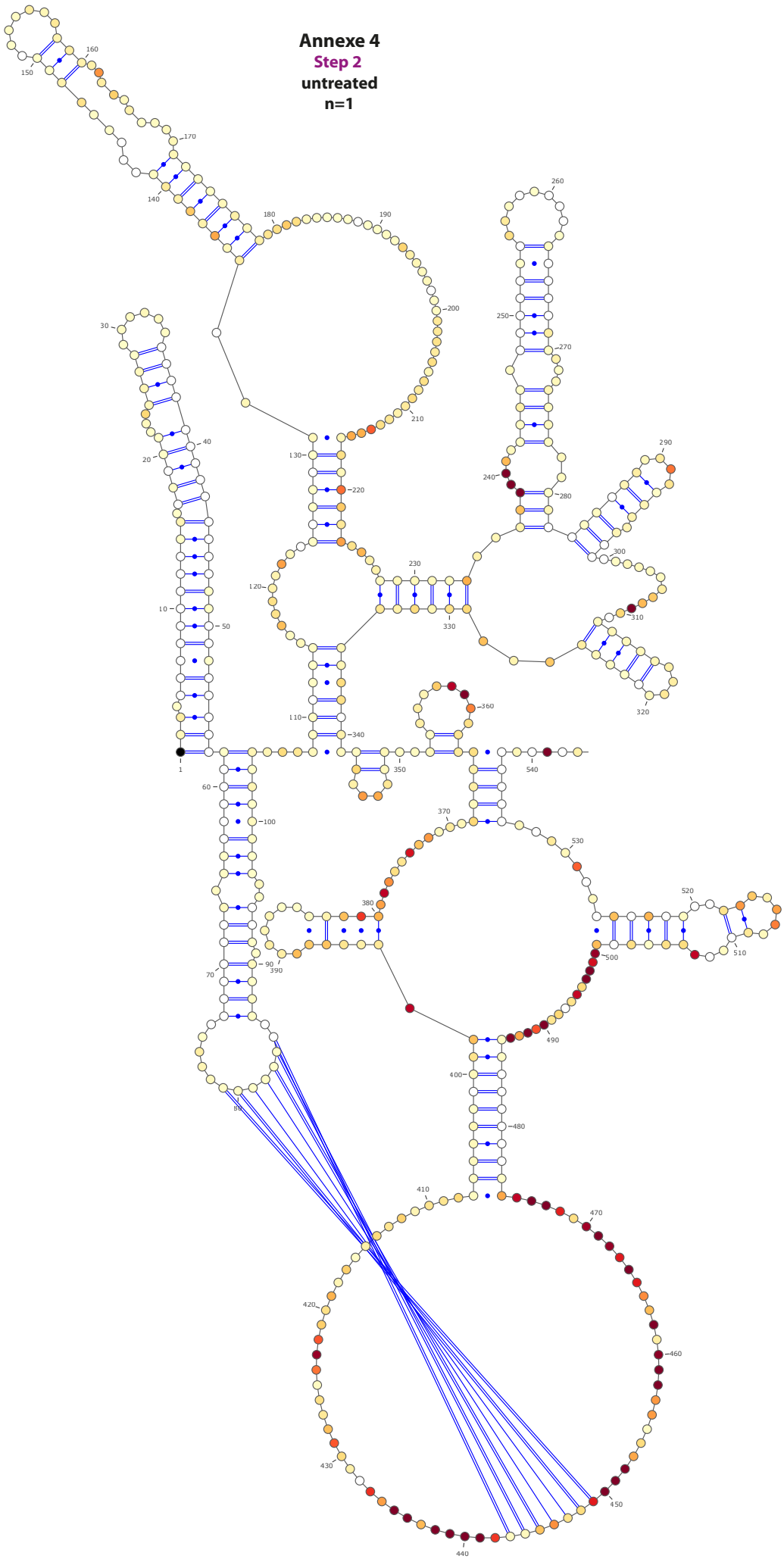


Annexe 4

Step 2

untreated

n=1

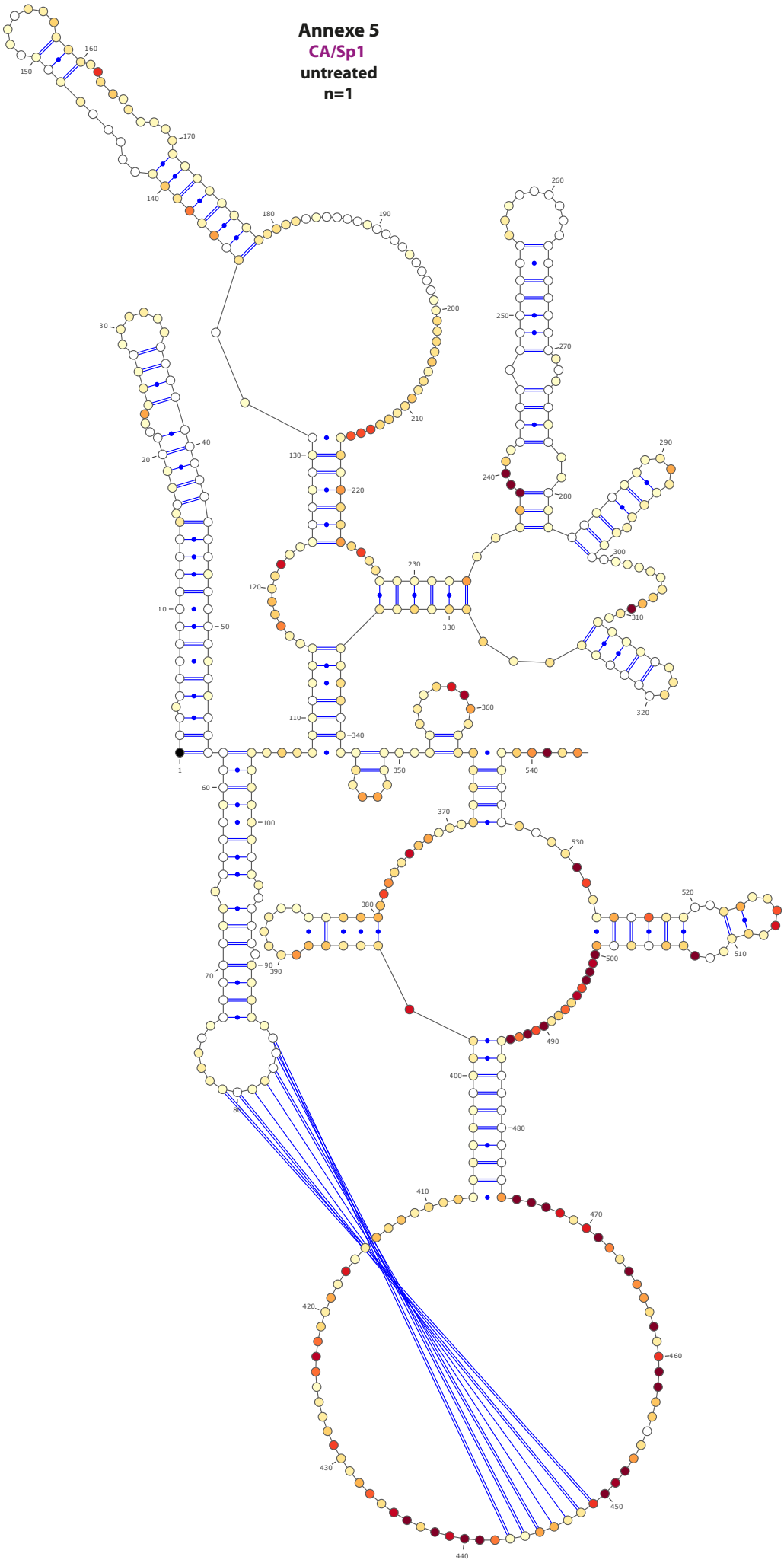


Annexe 5

CA/Sp1

untreated

n=1

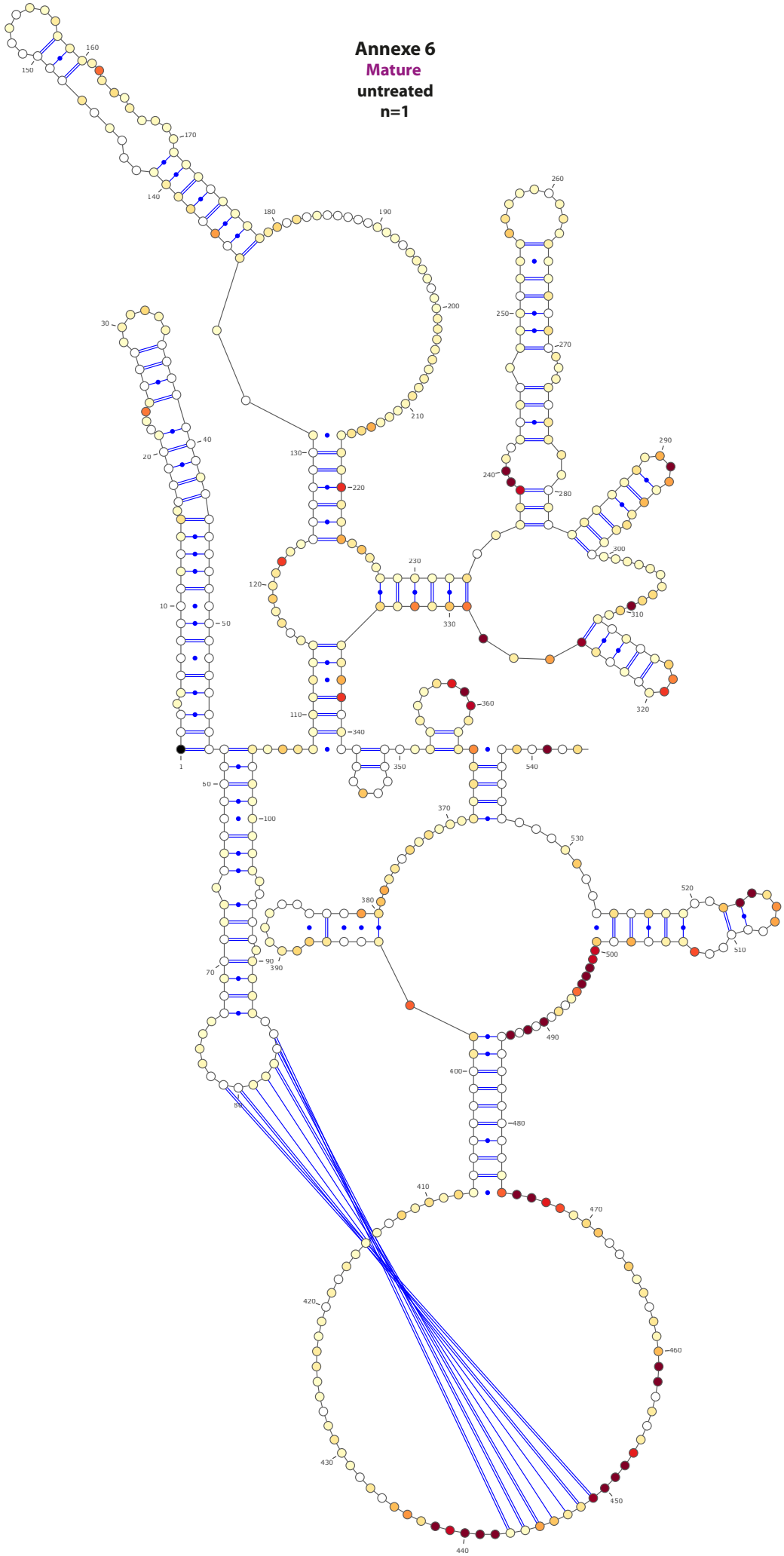


Annexe 6

Mature

untreated

n=1

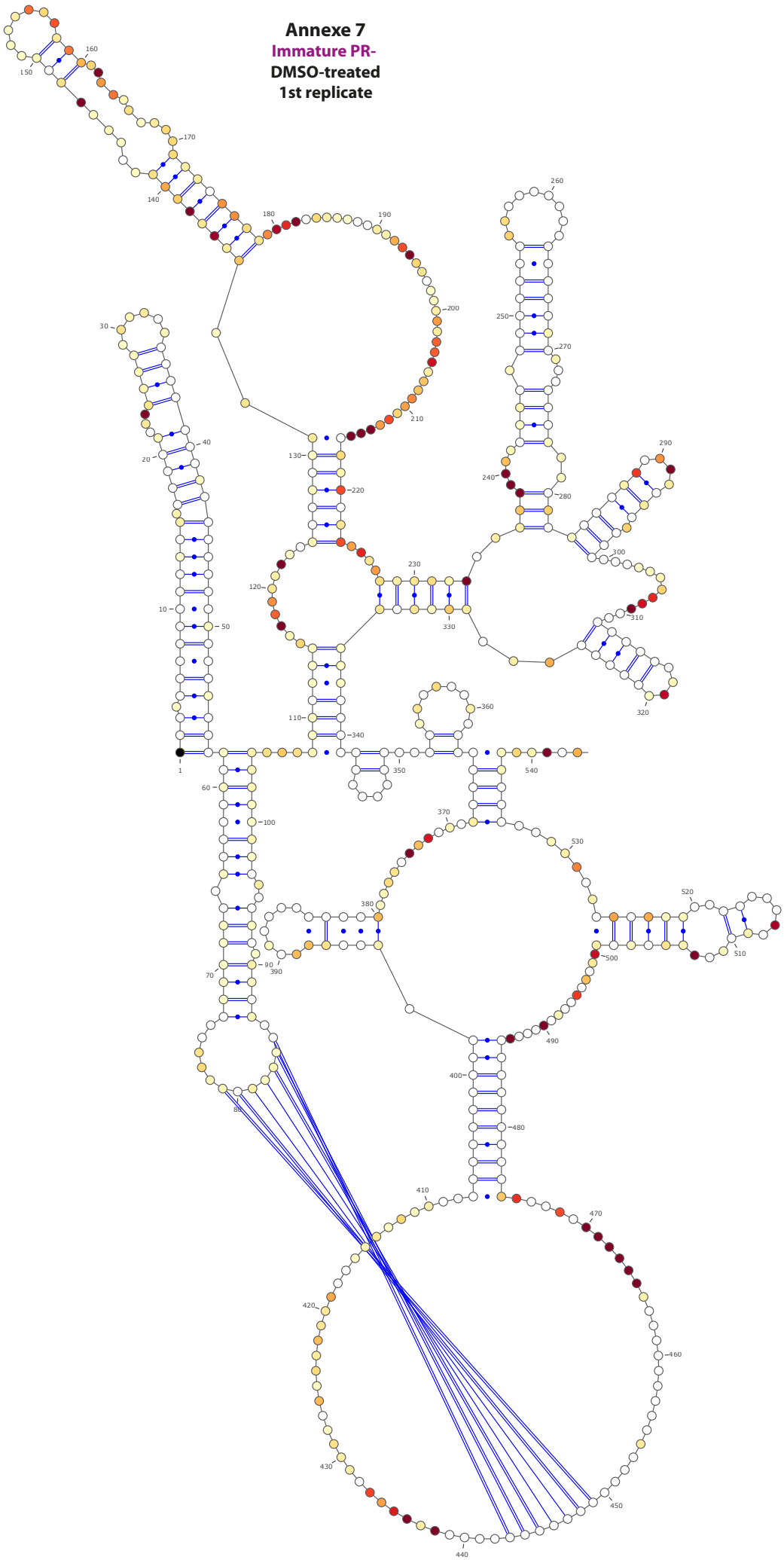


Annexe 7

Immature PR-

DMSO-treated

1st replicate

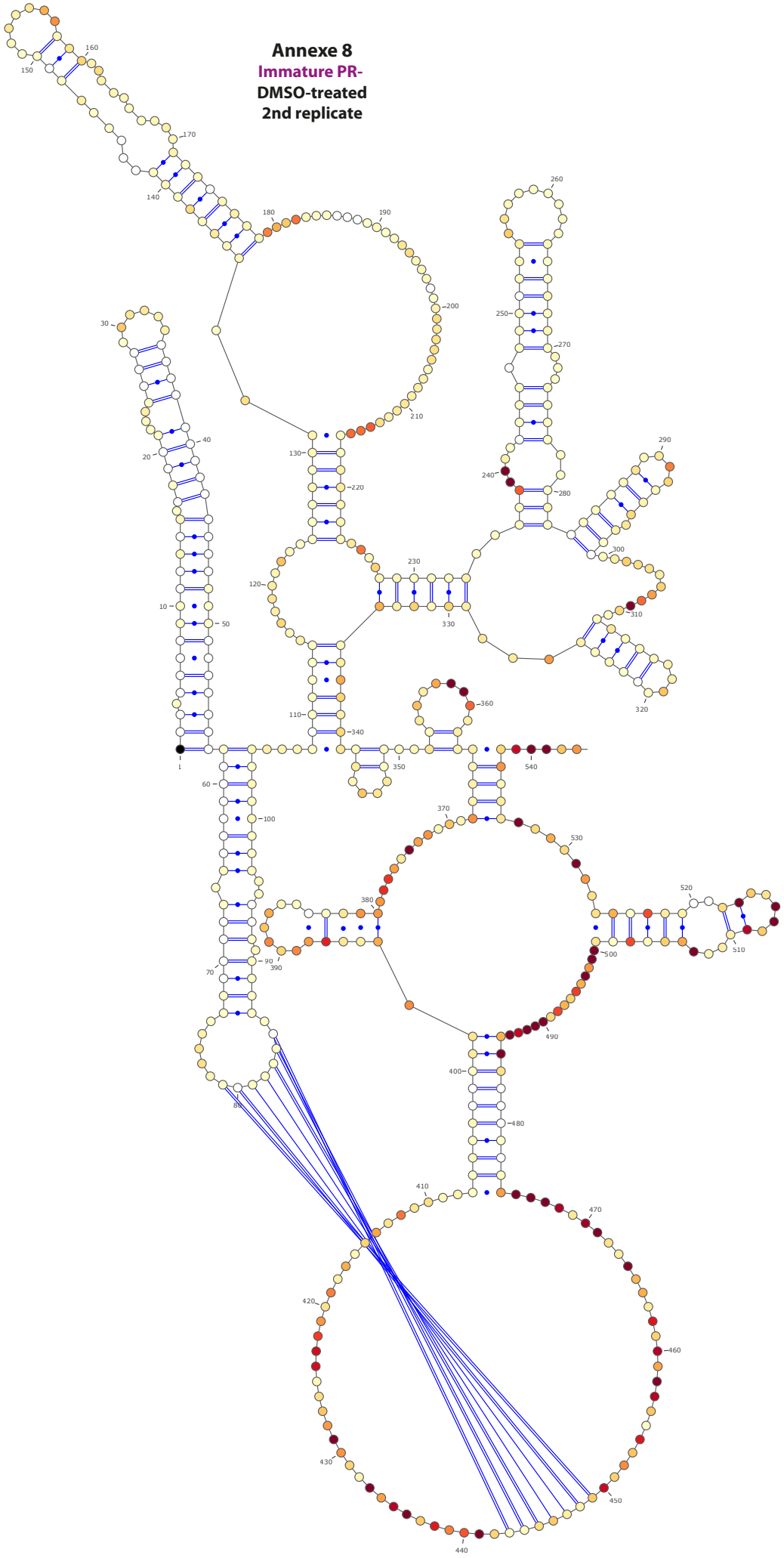


Annexe 8

Immature PR-

DMSO-treated

2nd replicate

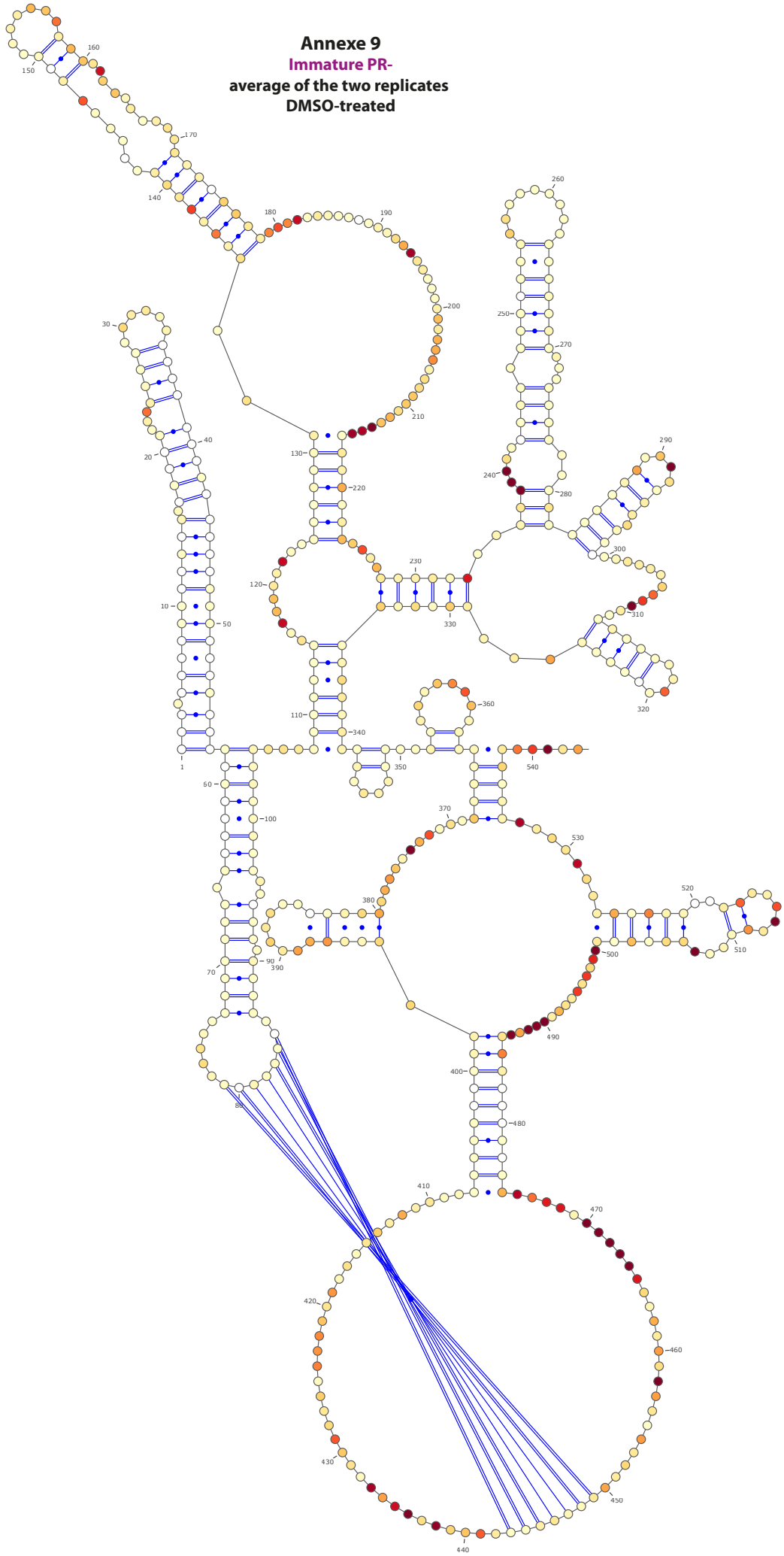


Annexe 9

Immature PR-

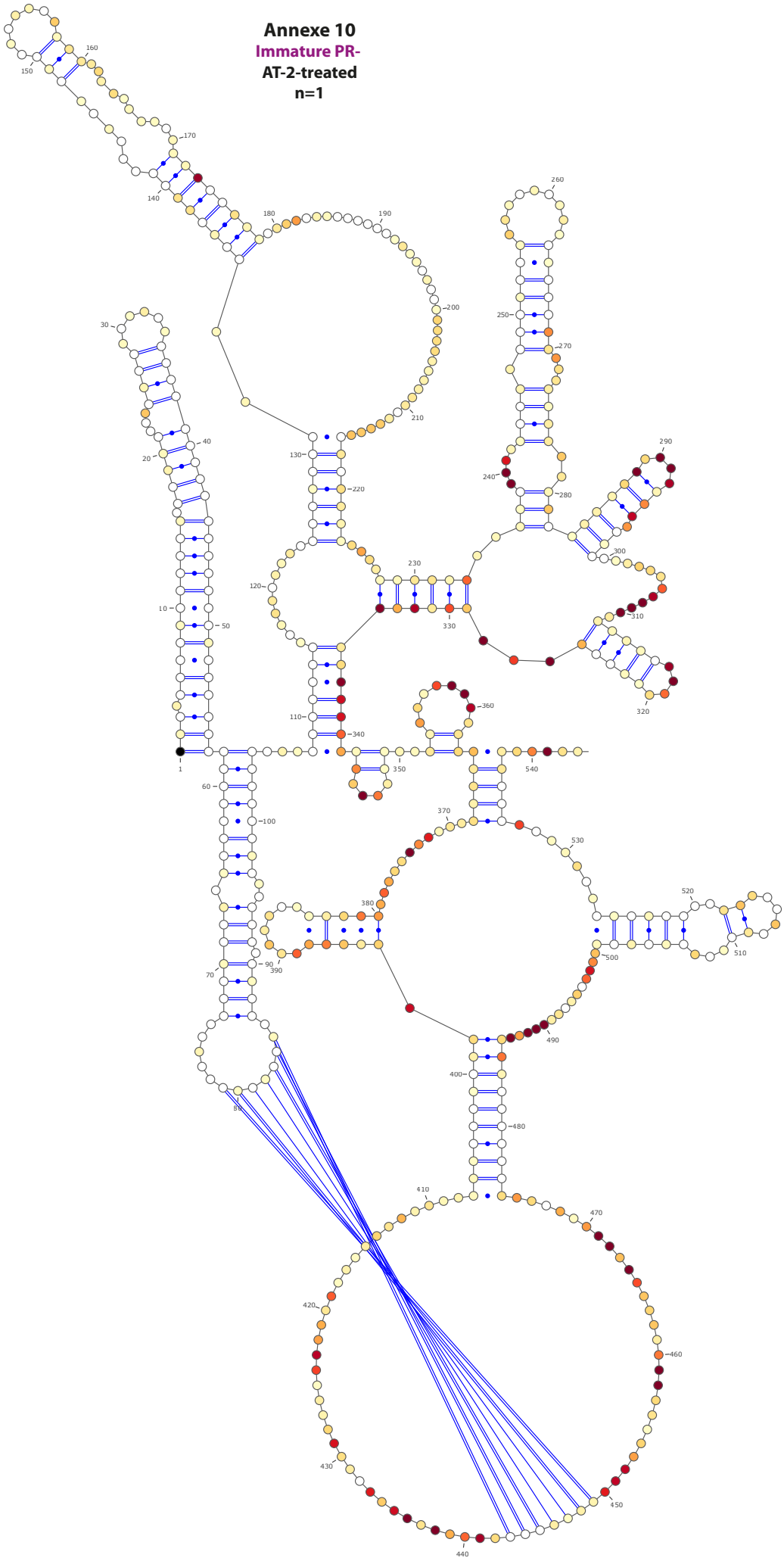
average of the two replicates

DMSO-treated



Annexe 10

Immature PR-AT-2-treated n=1

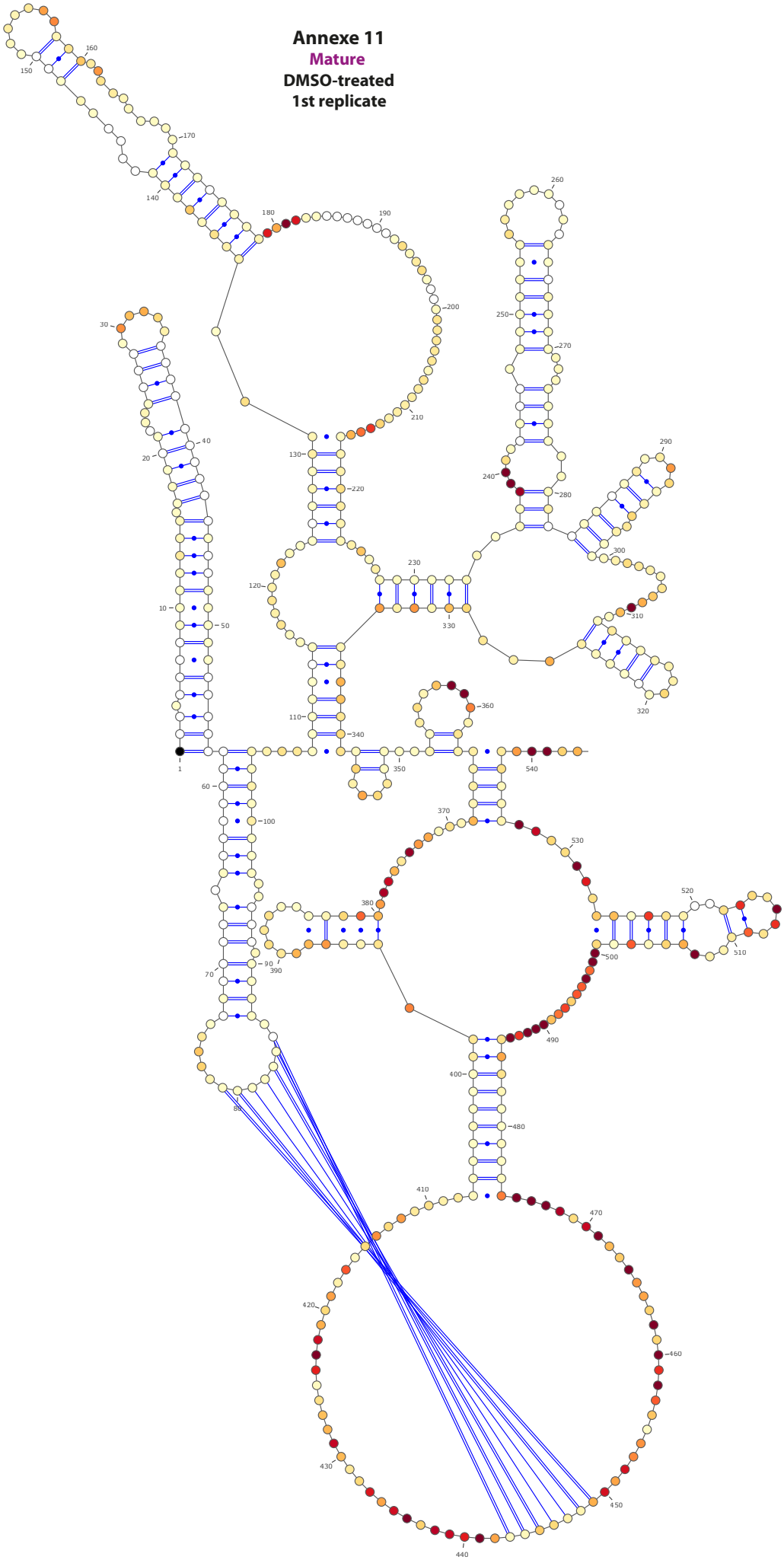


Annexe 11

Mature

DMSO-treated

1st replicate

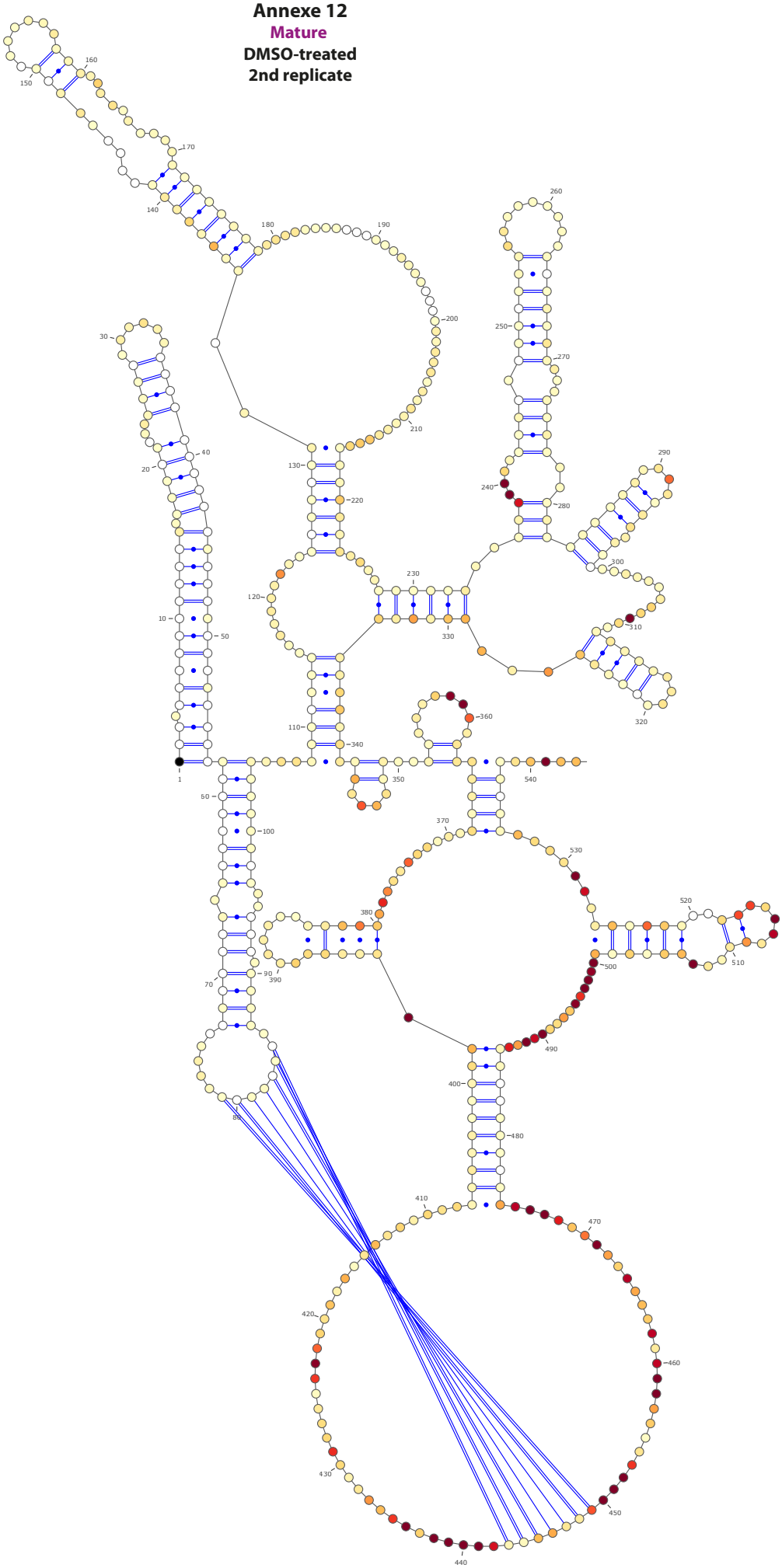


Annexe 12

Mature

DMSO-treated

2nd replicate

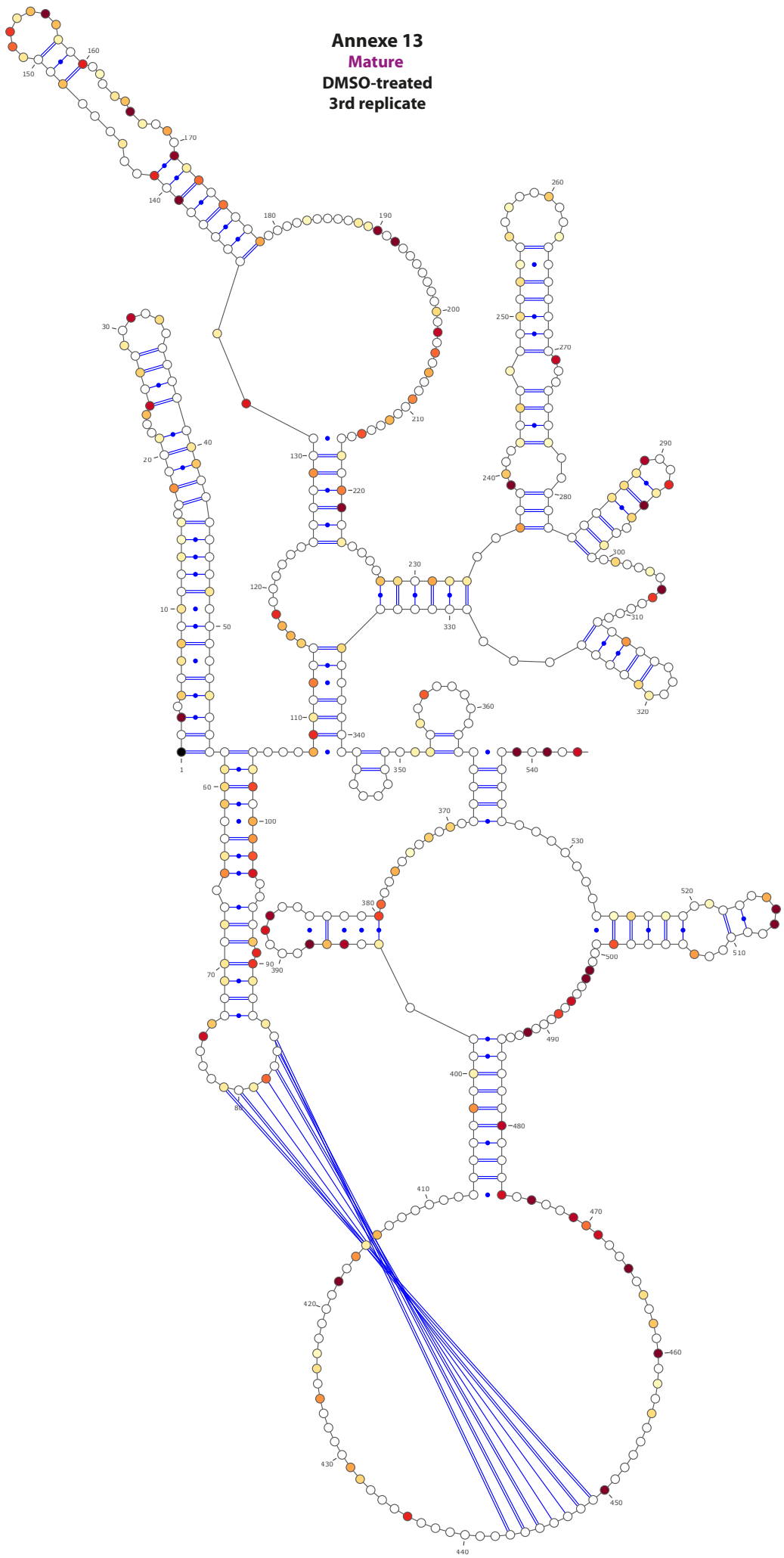


Annexe 13

Mature

DMSO-treated

3rd replicate

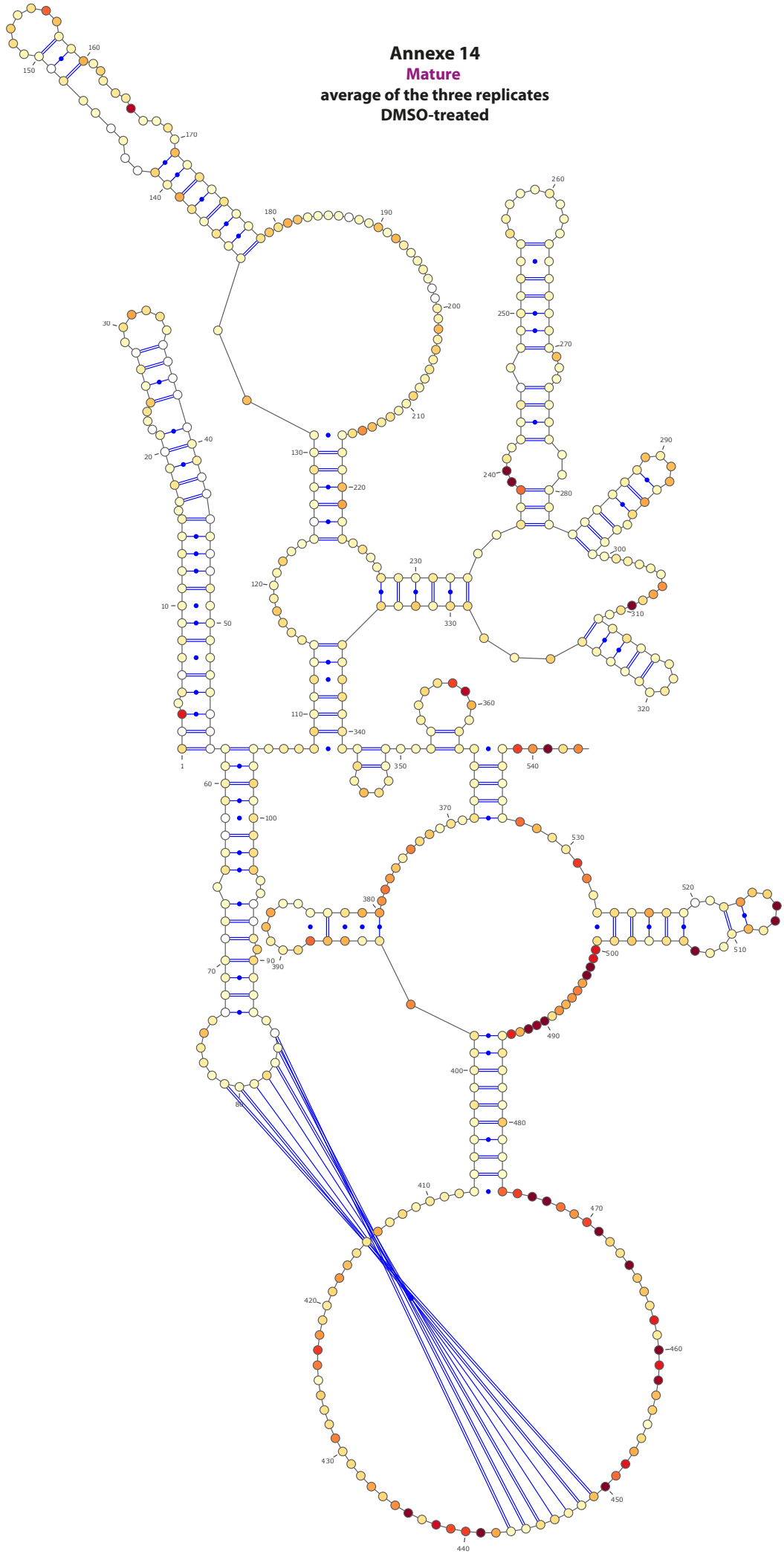


Annexe 14

Mature

average of the three replicates

DMSO-treated

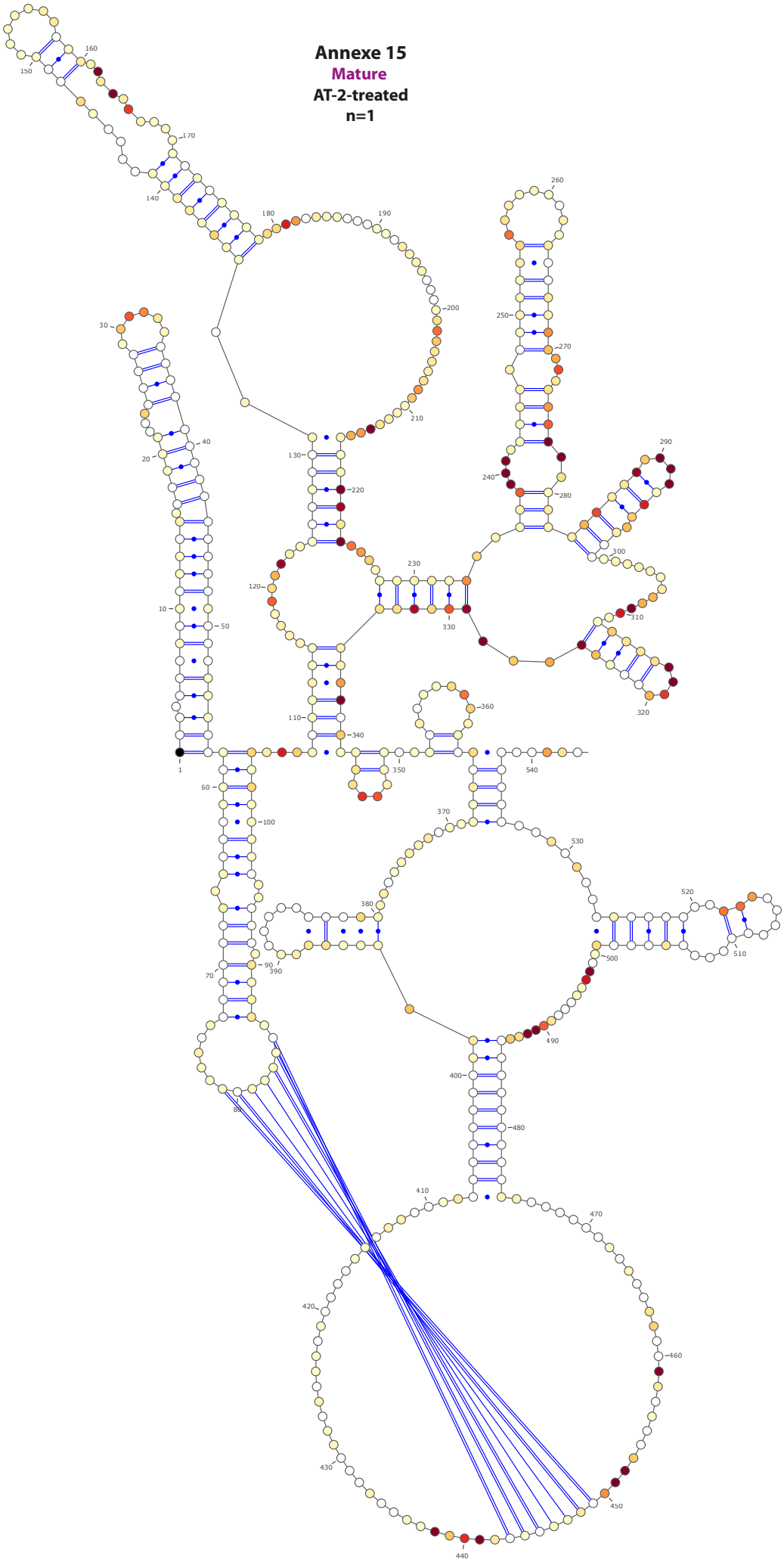


Annexe 15

Mature

AT-2-treated

n=1

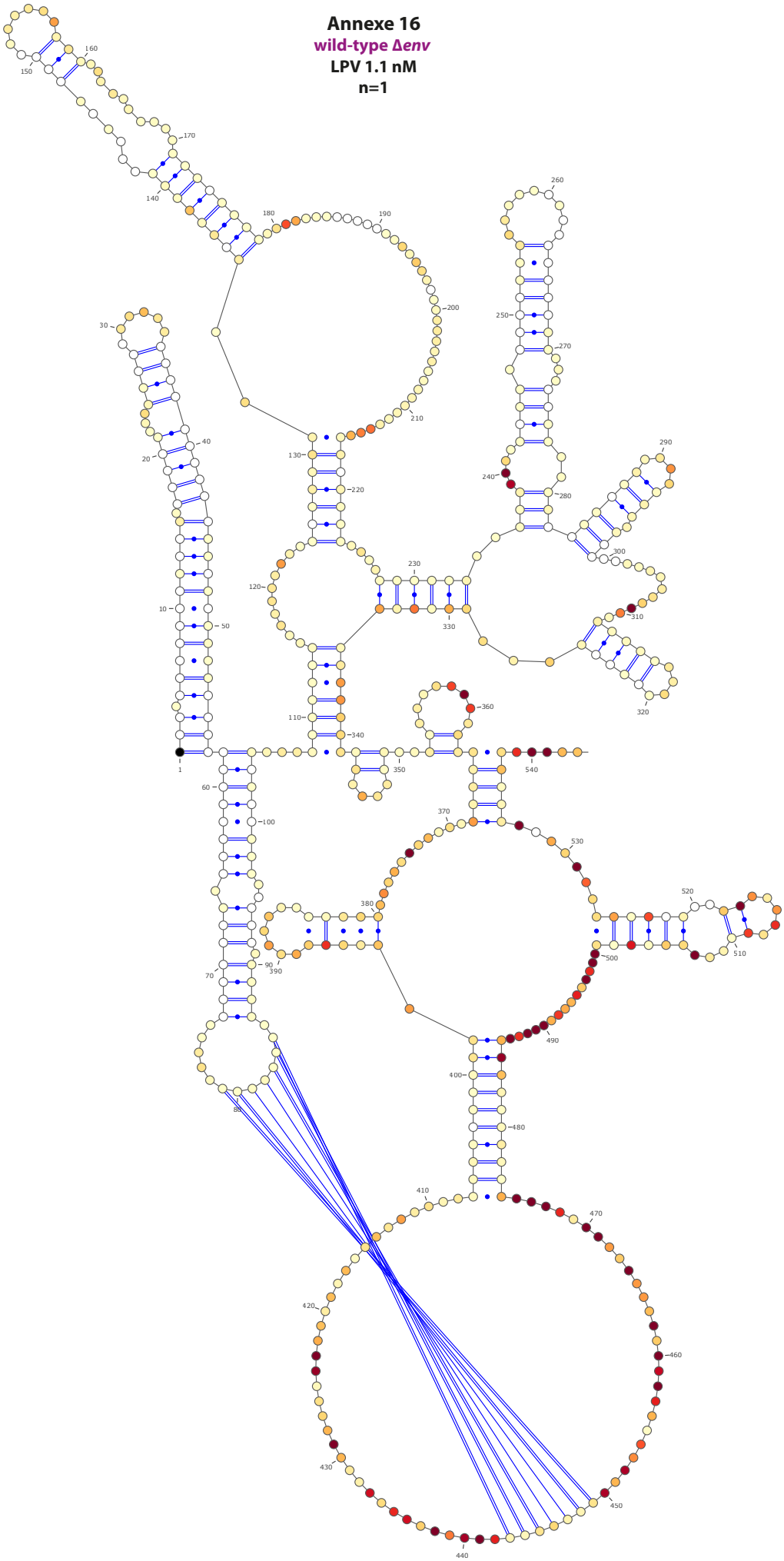


Annexe 16

wild-type Δ*env*

LPV 1.1 nM

n=1

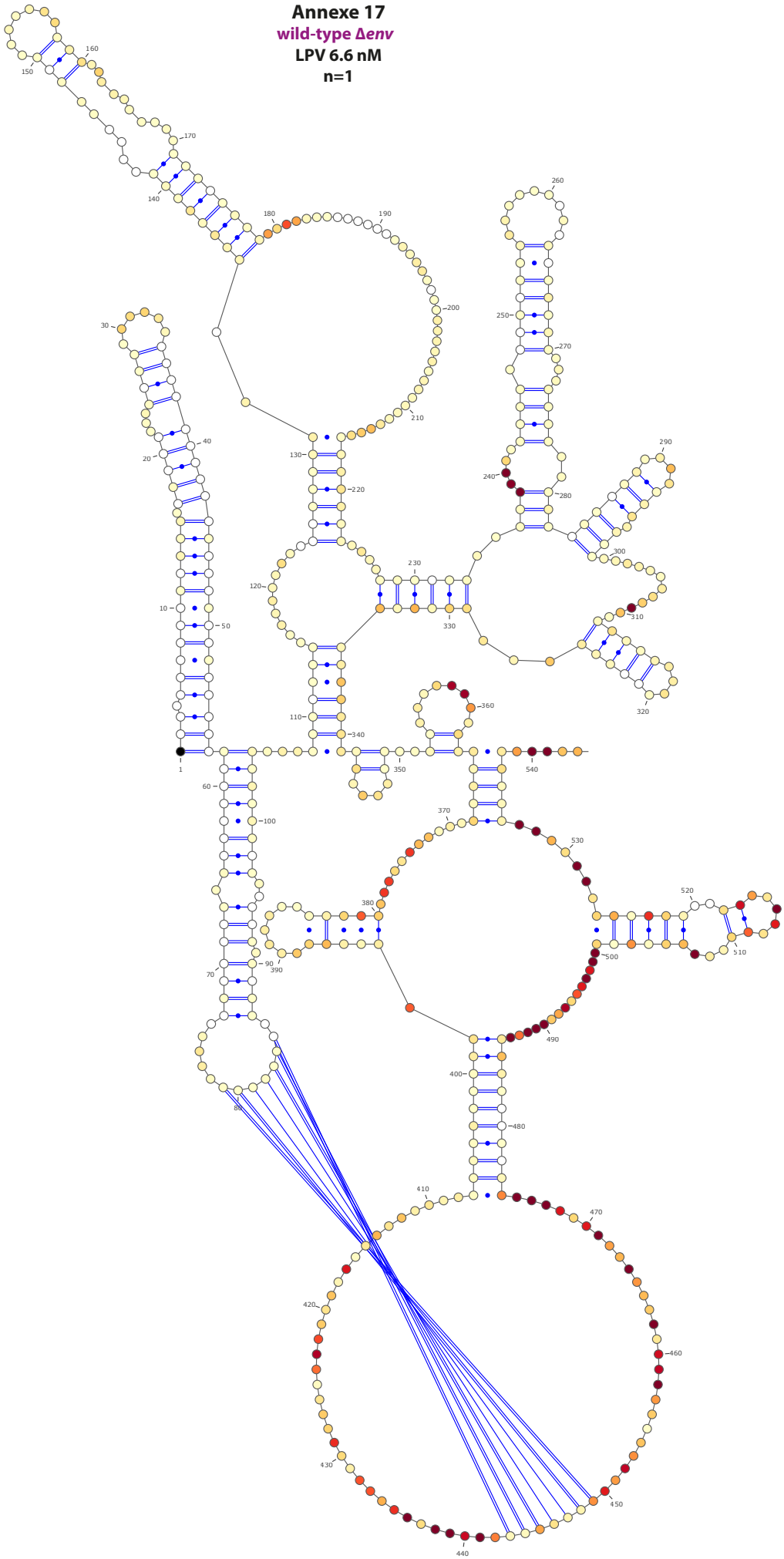


Annexe 17

wild-type Δ*env*

LPV 6.6 nM

n=1

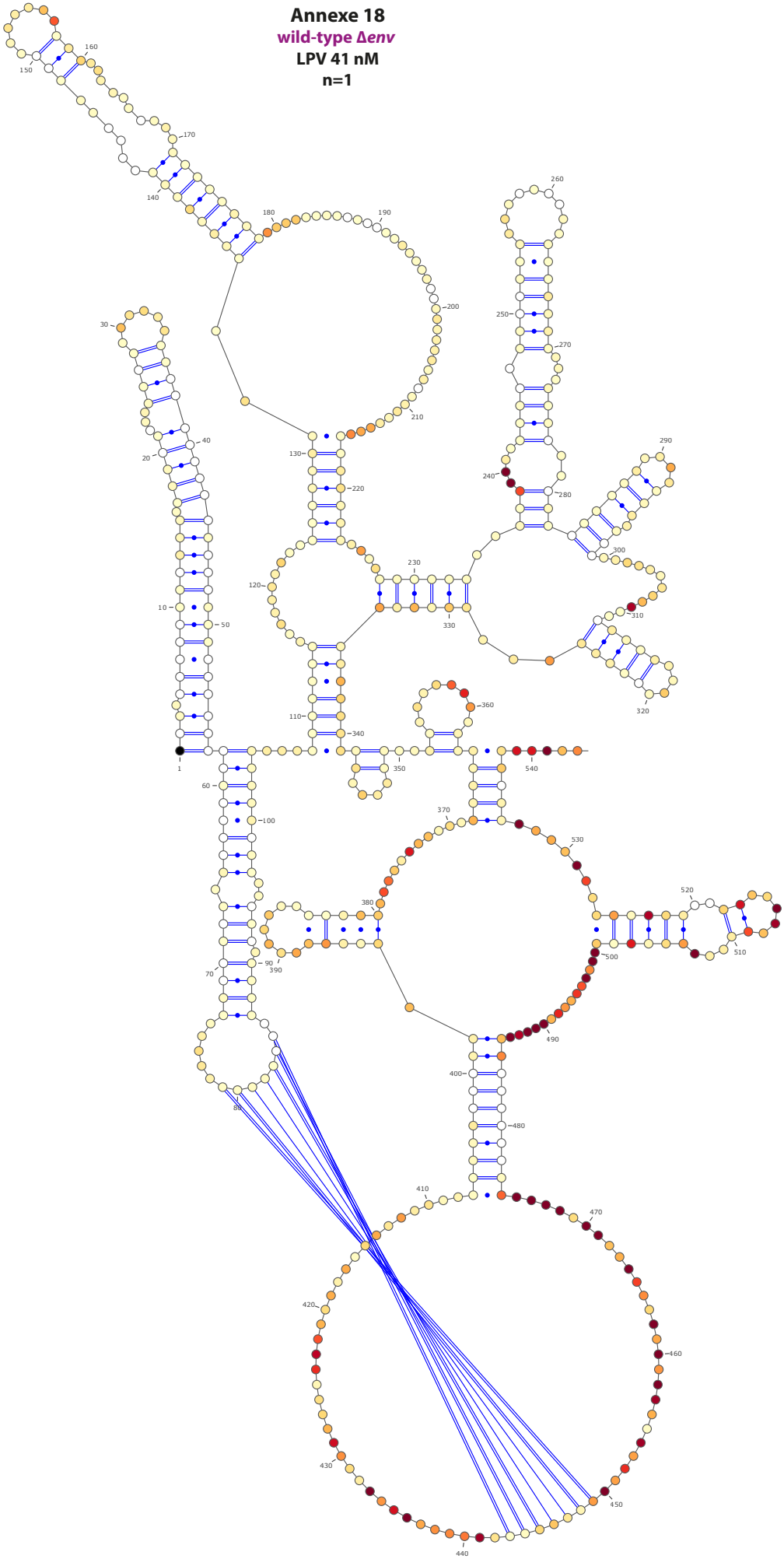


Annexe 18

wild-type Δ*env*

LPV 41 nM

n=1

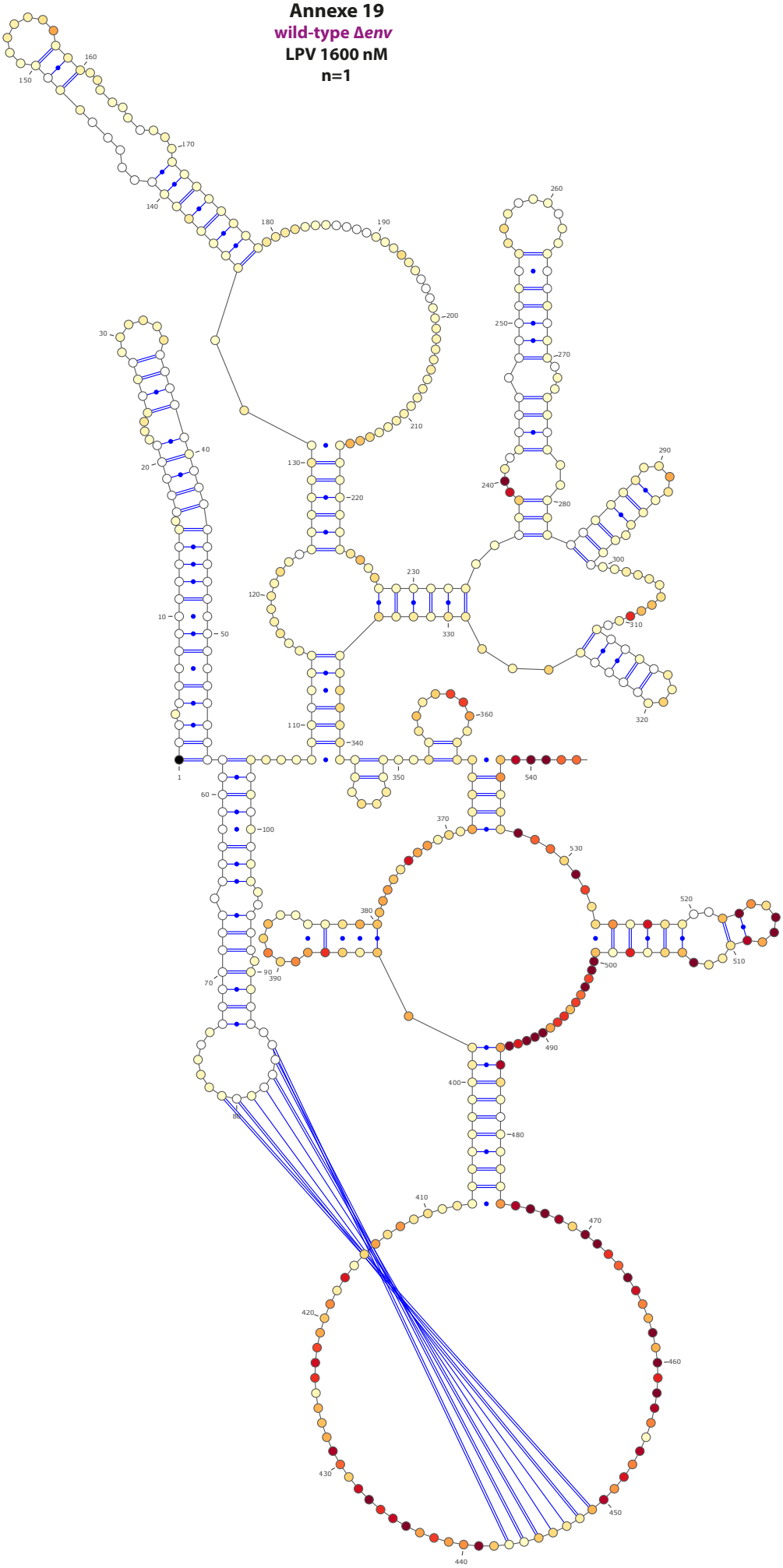


Annexe 19

wild-type Δ*env*

LPV 1600 nM

n=1

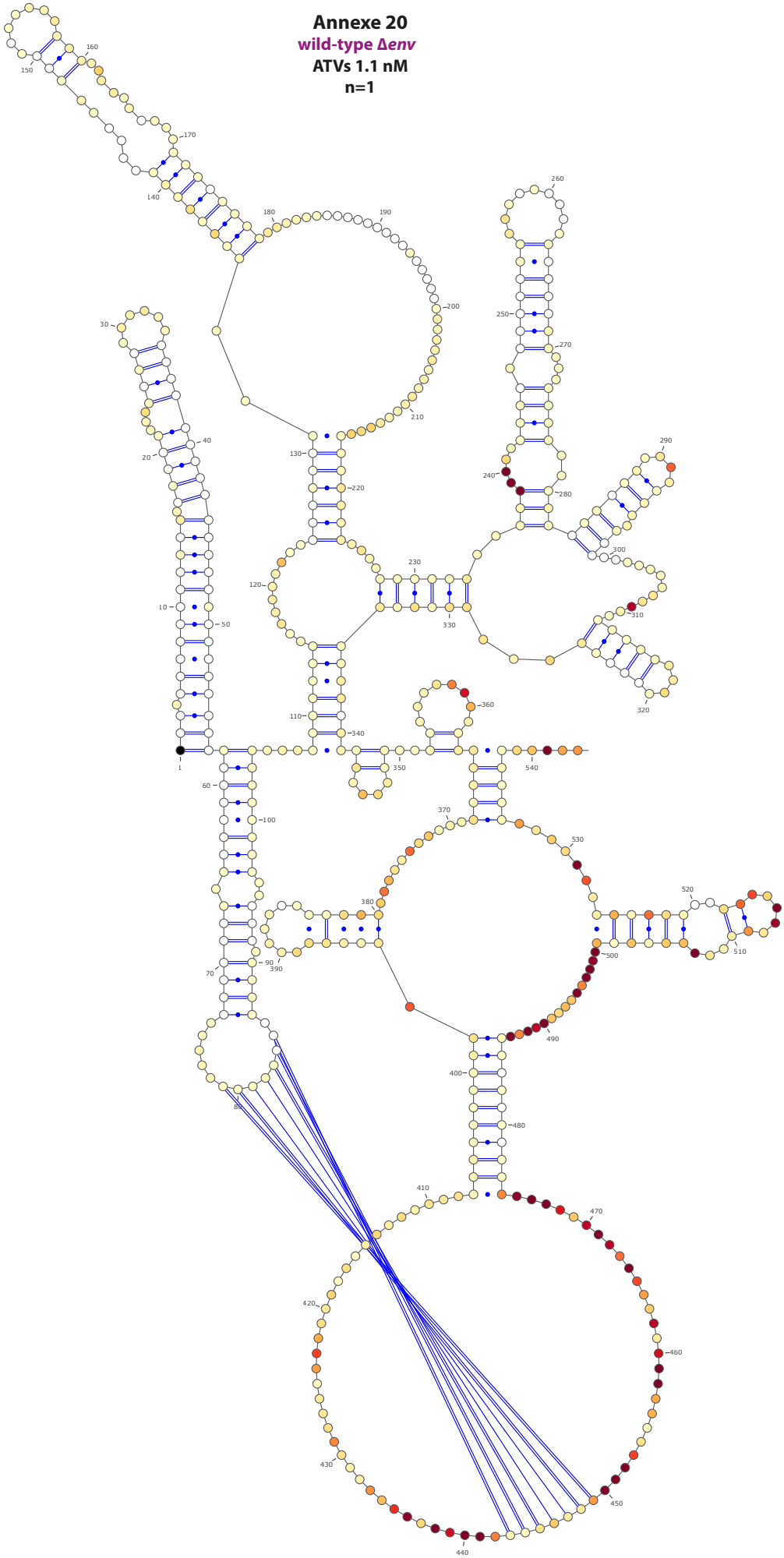


Annexe 20

wild-type Δ*env*

ATVs 1.1 nM

n=1

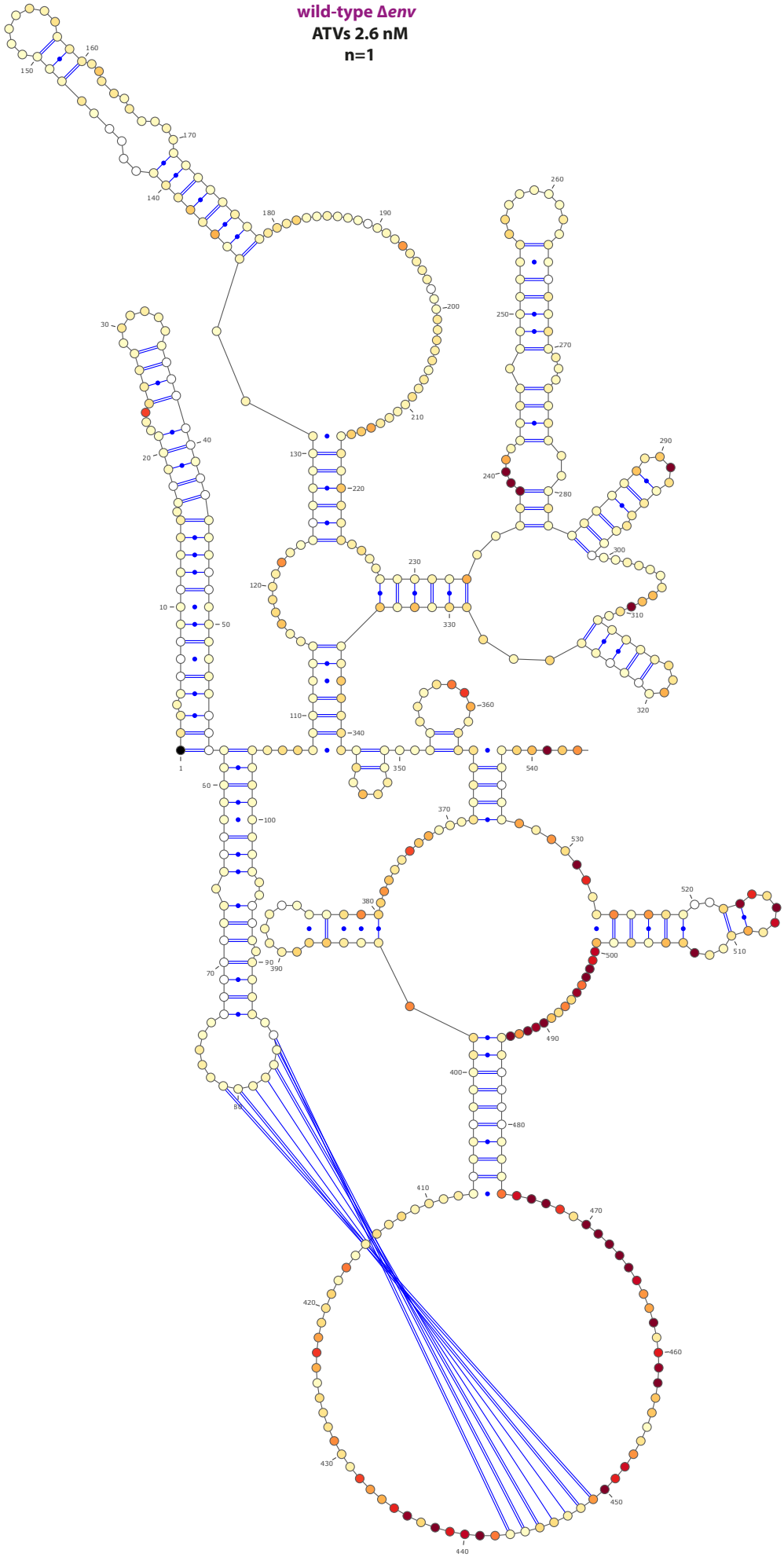


Annexe 21

wild-type Δ*env*

ATVs 2.6 nM

n=1

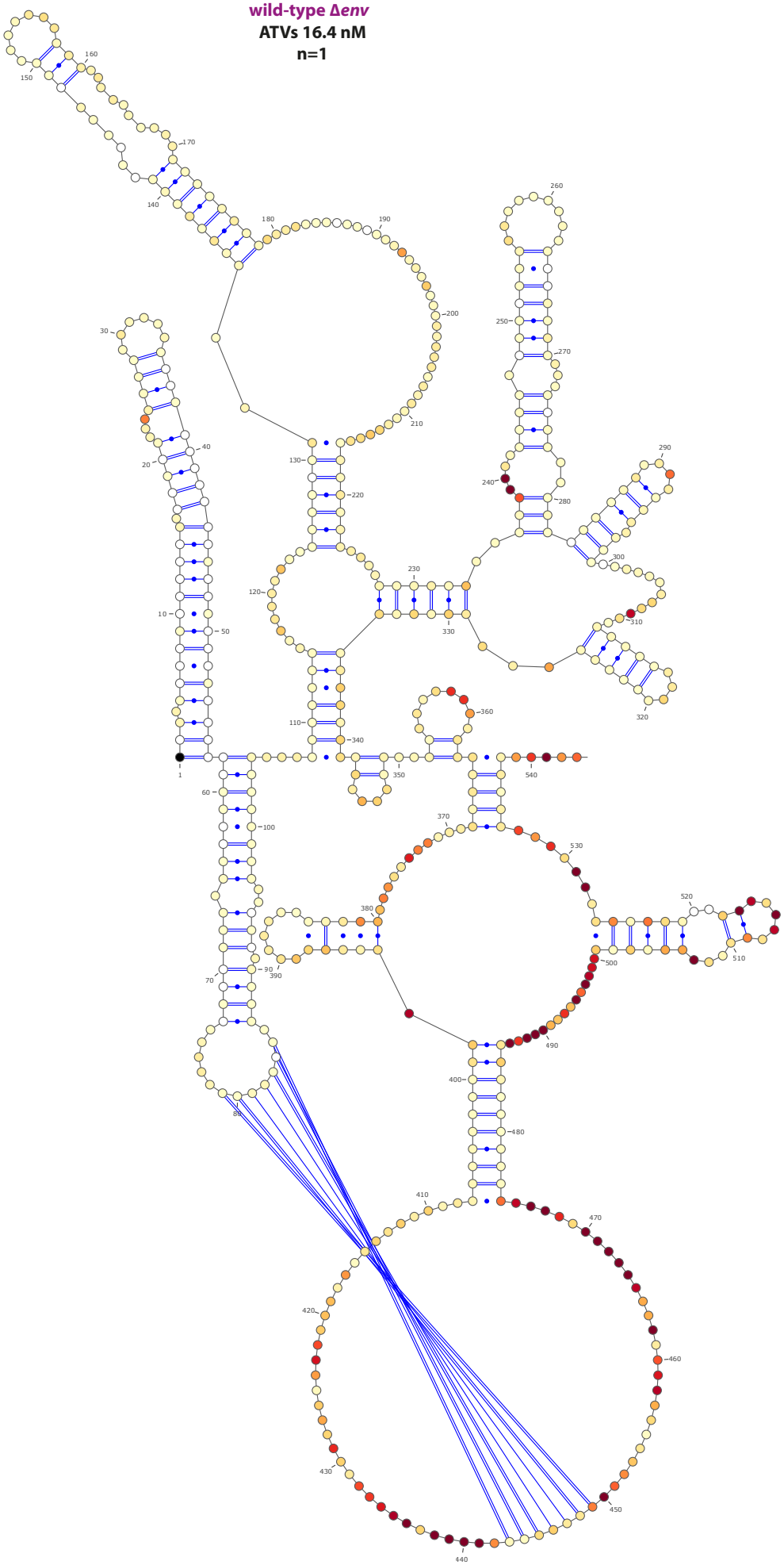


Annexe 22

wild-type Δ*env*

ATVs 16.4 nM

n=1

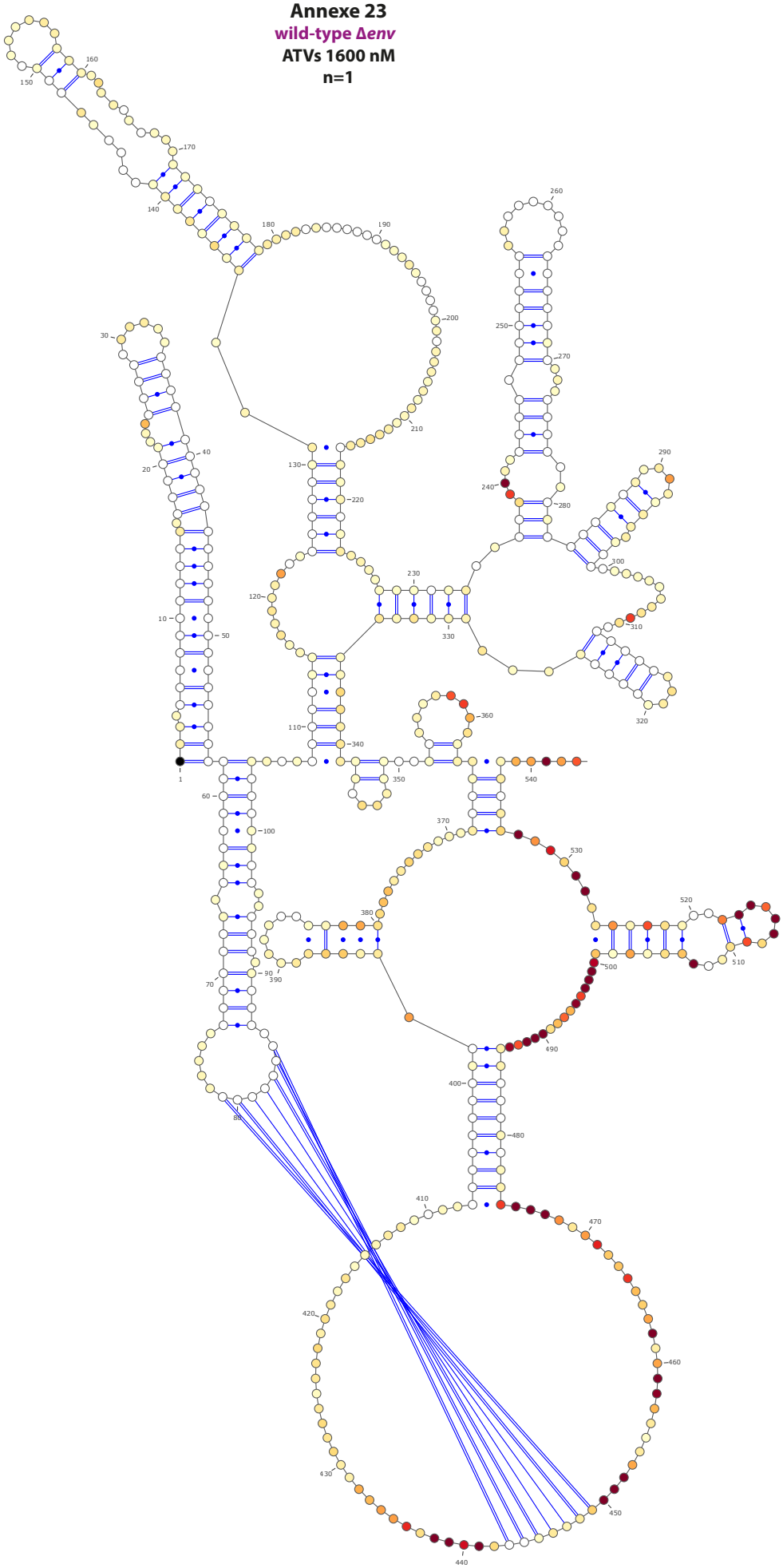


Annexe 23

wild-type Δ*env*

ATVs 1600 nM

n=1



Annexe 24: hSHAPE-Seq procedure

1. ***in viro* SHAPE protocol**

NMIA modification of viral particles was performed in NMIA reaction buffer for 50 min at 37°C.

2. **RNA extraction and DNase treatment**

An RNA extraction was performed using with the « NucleoSpin virus » kit followed by ethanol precipitation (standard protocol with 3 volumes of ethanol 100%, 0.3 M sodium acetate and 1 μg glycogen). After centrifugation at 11,000 g, pellets were washed twice with 80% ethanol, vacuum dried and resuspended in 20 μl milliQ H_2O. RNA was incubated with 2 U of DNase I, 1x DNase buffer and milliQ H_2O in a final volume of 200 μl for 30 min at 37°C. An additional 2 U of DNase I was added at 30 min and incubated for an extra 30 min.

3. **Optimised T-GIRT III RT protocol**

The RT reaction was performed in a T100 thermal PCR cycler with a heated lid set at 105°C. Ethanol precipitated viral RNA was dissolved in 15.1 μl milliQ H_2O and subjected to primer annealing with 1 μl of 1 μM p628 Rv stock in addition to 100 mM Tris-HCl pH 7.5 (at 20°C) and 0.1 mM EDTA. The reaction mix was incubated at 82°C, then cooled at 25°C. The annealed RNA-primer was then pre-incubated for 30 min at room temperature with 300 U of T-GIRT III (Ingex) and 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris–HCl pH 7.5, 5 mM DTT and milliQ H_2O to a final volume of 32.9 μl. This pre-incubation step allowed the tight binding of T-GIRT III with the RNA template. Following addition of dNTPs at a final concentration of 1.5 mM (dNTPs mix is an equimolar mixture of dATP, dCTP, dGTP, and dTTP), the RTion reaction was initiated at 60°C for 1 h in a final reaction volume of 35 μl.

4. **Treatment of DNA samples following T-GIRT III RT**

T-GIRT III RT reaction was ended by adding 50 U of RNase A into the 35 μl RT mix. Samples were incubated at 60°C for 15 min. RNase A was degraded by a proteinase K treatment, performed at 37°C for 30 min. A phenol-chloroform followed by a chloroform extraction were performed prior to ethanol precipitation.

5. **Illumina library generation**

-ligation of the first Illumina adapter

ssDNA samples were incubated at 90°C for 2 min with the A1 primer in order to denature both DNA molecules. Then, 1x T4 RNA ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 1 mM DTT), 25% PEG 8000, 1 mM ATP, 1 mM DTT (in addition to the 1 mM from the commercial buffer) and 10 U T4 RNA ligase 1 were added and samples incubated at 25°C for 16 h. Following ligation, samples were ethanol precipitated and resuspended in 10 μl milliQ H_2O.

-selection of the ligated DNAs with magnetic beads

The required amount of beads was calculated based on their binding capacity, transferred into a new tube and washed according to the supplier protocol. The tubes (5 μl for each sample) were placed on a magnet for 2 min followed by aspiration of the supernatant while the tubes were on the magnet. Beads were resuspended with 200 μl of 1x washing buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl). Following 3 washes, beads were resuspended with 2x washing buffer to a final concentration of 5

µg/µl (twice the original volume). Ligated samples were precipitated and resuspended in 20 µl milliQ H_2O. An equal volume of the 2x washing buffer was added and samples were incubated for 15 min at room temperature using gentle agitation. The biotinylated ssDNA coated beads were separated with a magnet for 3 min then washed 3 times with 150 µl of 1x washing buffer. An additional wash was performed with 150 µl milliQ H_2O. Beads with the immobilized ssDNA molecules were resuspended with 15 µl milliQ H_2O.

-addition of the second Illumina adapter by PCR

A mix of 25 µl containing 1 µl of magnetic beads coated with biotinylated ssDNA samples was prepared with 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of PCR1.0 primer and 0.5 µM of A2PCR primer, 0.02 U Phusion polymerase and milliQ H_2O. Samples were denatured at 98°C for 3 min followed by five PCR cycles composed of 15 s at 98°C (denaturation), 15 s at 61.2°C (hybridization) and 2 min at 72°C (elongation).

-normalisation of the different samples and Indexing

***Evagreen based qPCR (normalisation)**

A mix of 25 µl containing 1 µl of PCR PCR1.0-A2PCR was prepared with 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of PCR1.0 and Index1 primers, 0.5x Evagreen dye, 0.02 U Phusion polymerase and milliQ H_2O. Samples were denatured at 98°C for 3 min and 29 PCR cycles composed of 15 s at 98°C, 15 s at 56°C and 2 min at 72°C were performed.

***Phusion PCR protocol**

A mix of 50 µl containing DNA, 1x Phusion HF buffer, 200 nM dNTPs, 0.25 µM of each primer (Index and PCR1.0), 0.01 U Phusion (Thermo Scientific) and milliQ H_2O was prepared. Samples were denatured at 98°C for 3 min followed by a PCR cycle composed of 15 s at 98°C (denaturation), 15 s at 56°C (hybridization) and 2 min at 72°C (elongation). Based on the qPCR quantification, samples were normalised by performing a specific number of PCR1.0-Index PCR cycles.

Half of PCR1.0-Index PCR reaction from each sample were pooled, ethanol precipitated and resuspended in 30 µl milliQ H_2O to obtain the Illumina library.

Annexe 25: Synthèse en français

I. Contexte et objectifs de ce projet

Au cours du cycle de réplicatif, les particules du VIH-1 bourgeonnent de la cellule infectée sous forme immature et doivent subir le processus de maturation pour acquérir leur pouvoir infectieux. Le traitement protéolytique de $Pr55^{Gag}$ déclenche des réarrangements morphologiques du virus afin de former une particule mature possédant un cone de forme conique contenant le dimère d'ARN génomique (ARNg). En parallèle, le dimère d'ARNg subit également des remaniements et devient plus stable et plus compact. Alors que la maturation de $Pr55^{Gag}$ est bien appréhendée, la maturation structurale de l'ARNg reste mal comprise. Les réarrangements structuraux du génome sont facilités par l'activité chaperonne de la protéine NCp7. De plus, les produits intermédiaires contenant le domaine NC sont cruciaux pour l'infectivité virale, bien que produits de manière transitoire lors de la maturation.

Malgré les nombreuses études accomplies pour comprendre le processus de maturation, plusieurs questions demeurent non résolues. Le déclencheur exact de l'activation de la protéase virale reste inconnu ainsi que le moment précis de l'initiation, même s'il est communément admis que la maturation virale démarre immédiatement après la libération des particules.

En outre, la transition morphologique du stade immature au stade mature reste incertaine, impliquant probablement des interactions inter- et intra-moléculaires, ainsi que la relation entre la maturation protéique et la maturation génomique.

Plusieurs groupes ont analysé la structure de l'ARNg en couvrant l'ensemble du génome mais ces études ont été réalisées après l'extraction de l'ARNg de particules virales (383, 422). Une étude similaire a été réalisée dans les particules virales mais n'a ciblé que les 900 premiers nucléotides (427). Néanmoins, il est important de souligner que seule la conformation de l'ARNg mature a été étudiée alors que la compréhension de la maturation du génome du VIH-1 nécessite d'étudier l'ensemble des processus de manière séquentielle.

Depuis plusieurs années, le groupe de J.C. Paillart et R. Marquet, avec pour thématiques les complexes ribonucléoprotéiques, l'encapsidation du génome viral et l'assemblage de la particule virale, a réalisé plusieurs découvertes dans le domaine du VIH-1. En ce qui concerne l'encapsidation du génome du VIH-1, l'équipe a étudié le mécanisme de dimérisation de l'ARNg et identifié les 6 nucléotides impliqués dans la dimérisation de l'ARN. Cette interaction intermoléculaire est la seule démontrée à l'heure actuelle entre les deux copies de l'ARNg du VIH-1. L'importance des trois purines de part et d'autre du DIS pour le

processus de dimérisation a également été démontrée. La fixation de $Pr55^{Gag}$ à de nombreux mutants de l'ARNg a de plus permis de déterminer *in vitro* et *in viro* que SL1 est le site de liaison majeur de $Pr55^{Gag}$. En outre, l'équipe possède une solide expertise dans l'interrogation de la structure de l'ARN en utilisant des approches de cartographie chimique en solution, à la fois *in vitro* et *in vivo*.

Dans ce contexte, mon projet de doctorat a été axé sur l'étape de maturation. J'ai initié ce projet dans lequel nous souhaitons déterminer les différentes étapes menant à la formation d'un dimère d'ARNg stable et mature. Nous souhaitons également mieux comprendre le lien entre le traitement protéolytique et la maturation de l'ARNg. Dans ce but, j'ai analysé la structure secondaire de la région 5' du génome du VIH-1 (comprenant les 550 premiers nucléotides) en combinant des approches *in vitro* et *in viro*. La structure de cette région a été évaluée par cartographie chimique :

1. *In vitro* en présence ou en l'absence de $Pr55^{Gag}$, GagΔp6, intermédiaires de maturation contenant le domaine NC (NCp15 et NCp9) et NCp7. Cette comparaison nous permet de déterminer ces sites de fixation de ces protéines à l'ARNg et si ces sites de protection sont conservés pendant la cascade protéolytique. Deux conditions ont été testées en parallèle, à savoir l'ARN seul et le complexe ARNg-protéines traité avec la protéinase K avant modification chimique. Ces conditions permettent de déterminer si les différents réarrangements structuraux induits par les protéines sont permanents au niveau de l'ARNg.
2. *In viro* au sein de particules virales délétées de la protéase virale (PR), de mutants mimant la cascade protéolytique $Pr55^{Gag}$ et de particules virales matures. Ces particules ont également été traitées avec un éjecteur de zinc (AT-2) qui déstabilise les doigts de zinc des intermédiaires de maturation contenant le domaine NC. La structure de l'ARNg, chimiquement modifiée *in viro*, a été déterminée par l'approche hSHAPE-Seq. Cette technique a été développée pour étudier la maturation structurale de l'ARNg à haute résolution grâce au couplage de la cartographie chimique au séquençage à haut débit.

La structuration de cette région a également été étudiée au sein de particules virales traitées avec deux inhibiteurs de protéase différents afin de mieux comprendre le mécanisme d'action de ces antiviraux et d'évaluer leurs effets sur la maturation de l'ARN génomique. En effet, nous proposons l'hypothèse d'un défaut de maturation causé par les inhibiteurs de protéase avec l'intermédiaire de maturation NC/Sp2 agissant comme dominant négatif. L'impossibilité de maturer totalement le domaine NC pourrait en effet bloquer le réarrangement structural de l'ARNg et inhiberait la réplication virale, compte tenu du rôle crucial de la chaperonne d'ARN NCp7.

Ce projet vise donc à répondre aux questions suivantes :
- Comment la conformation structurale des premiers 550 nucléotides évolue-t-elle pendant la cascade protéolytique de $Pr55^{Gag}$?

- Ces réarrangements sont-ils en corrélation avec la cascade protéolytique de $Pr55^{Gag}$? Où sont-ils exactement localisés et quelle est leur nature (interaction inter- et/ou intra-moléculaire ?
- Quel est l'impact de $Pr55^{Gag}$ et de ses intermédiaires de maturation sur la maturation génomique ? Où sont localisés les différents sites de fixation protéique ? Est-ce que $Pr55^{Gag}$ et NCp7 ont un motif commun de liaison ?
- Quel est le mécanisme d'action des inhibiteurs de protéase sur la maturation des particules virales ? Ce mécanisme pourrait-il être lié à la maturation des ARNg ?

Mon projet de doctorat a nécessité un effort technique important, avec le développement d'une nouvelle technique pour analyser la structure de l'ARNg à l'intérieur de la particule virale, car aucun protocole approprié n'était disponible au début de mon doctorat. De plus, afin de mettre en lumière le lien entre maturation protéique et maturation génomique, des mutants bloqués aux différentes étapes de la maturation protéique ont été conçus. Il a été impossible de produire ces particules virales par cycle réplicatif, car ces mutants ne se répliquent pas. Ainsi, la quantité d'ARNg produite a été limitée et a donc exigé une analyse par séquençage à haut débit après cartographie chimique pour augmenter la sensibilité de la détection.

J'ai donc développé une technique de cartopgraphie chimique *in viro*, que nous appelons hSHAPE-Seq, et l'ai appliquée à l'étude de la région 5' du génome du VIH-1. Cette région est en effet cruciale pour la régulation de nombreuses étapes du cycle viral. Cette approche m'a permis d'analyser plusieurs mutants imitant le clivage séquentiel de $Pr55^{Gag}$ et l'effet des inhibiteurs de protéase (lopinavir et sulfate d'atazanavir) sur la structure de l'ARNg. De plus, les particules virales ont été traitées avec de l'aldritiol-2 (AT-2) afin d'expulser les ions Zn^{2+} des doigts de zinc NC et d'identifier leurs sites de liaison.

De plus, j'ai étudié *in vitro* les sites de fixation de $Pr55^{Gag}$, GagΔp6, NCp15, NCp9 et NCp7, afin de valider d'avantage les résultats de hSHAPE-Seq et de comparer la propriété chaperonne du domaine NC en fonction de son état de maturation.

II. Résultats et discussion

II.1. Développement et validation de la technique de hSHAPE-Seq

L'approche hSHAPE-Seq a été conçue pour déterminer la conformation de l'ARNg du VIH-1 pendant le traitement protéolytique de $Pr55^{Gag}$. Comme le montre ce manuscrit, l'approche hSHAPE-Seq a fourni des informations structurales détaillées sur les premiers 550 nucléotides du génome du VIH-1 à l'intérieur des particules virales. Le protocole hSHAPE-Seq a été largement optimisé pour générer les résultats présentés dans ce manuscrit. Le principal défi était la quantité très limitée de

matériel disponible pour les virus mutants et donc non réplicatifs. Par conséquent, plusieurs étapes clés du protocole ont nécessité une importante optimisation : i) production à grande échelle des particules virales PR- et intermédiaires de maturation, ii) traitement AT-2, qui a un effet dramatique sur le rendement de récupération de l'ARNg, iii) RTion et iv) préparation de la bibliothèque Illumina, comprenant la ligation des adaptateurs Illumina et la normalisation des échantillons.

Le profil de réactivité du génome issu de particules virales matures, que j'ai obtenu avec hSHAPE-seq, est globalement très similaire avec les résultats précédemment publiés par Wilkinson et al. (426). Dans cette étude, l'ARNg a également été modifié *in viro*, mais les modifications ont été analysées par la technique de hSHAPE, qui utilise l'électrophorèse capillaire plutôt que les nouvelles technologies de séquençage à haut débit, pour identifier leurs emplacements. Sur la base du profil de réactivité SHAPE, nous avons clairement montré que la région régulatrice 5' est plus structurée que la région codante de *gag* (Figure 94 A).

Cependant, contrairement aux travaux publiés précédemment, notre étude fournit une vue complète des réarrangements structuraux de la région 5' de l'ARNg lors de la maturation des particules virales. En effet, notre connaissance de la conformation immature de l'ARNg est très limitée, avec une seule analyse existante et se focalisant sur le domaine PBS seulement (375). En ce qui concerne la fixation de $Pr55^{Gag}$, des études *in vitro* ont été réalisées (33, 48, 67, 127, 225, 354, 388). Cependant, aucune étude similaire n'a été effectuée sur des particules immatures contenant $Pr55^{Gag}$. De plus, la conformation de l'ARNg des mutants bloqués à différentes étapes du processus de maturation (313) n'a pas été déterminée précédemment. Ainsi, nos résultats fournissent le premier aperçu de la maturation structurale de l'ARNg qui accompagne la transformation des particules virales immatures en particules infectieuses.

Deux et trois réplicas biologiques ont respectivement été obtenus pour les particules immatures et matures traitées par DMSO et un réplica non traité a été produit pour chaque étape de maturation. En comparant les réplicas, des variations significatives sont observées dans certaines régions même si les profils de réactivité sont globalement cohérents. L'un des trois réplicas traités avec DMSO pour les échantillons matures diffère sensiblement des deux autres et semble également incompatible avec l'échantillon non traité ; Ainsi, cet échantillon devrait probablement être retiré de l'analyse. Cette variabilité résulte probablement du fait que ces échantillons d'ARNg ont été recueillis à partir de différentes productions virales et suggère que certaines régions de la structure de l'ARNg pourraient adopter d'autres conformations. Néanmoins, le hSHAPE-Seq est hautement reproductible, comme en témoigne la comparaison des profils de réactivité obtenus pour les quatre concentrations d'inhibiteurs de protéase (annexes 16 à 23).

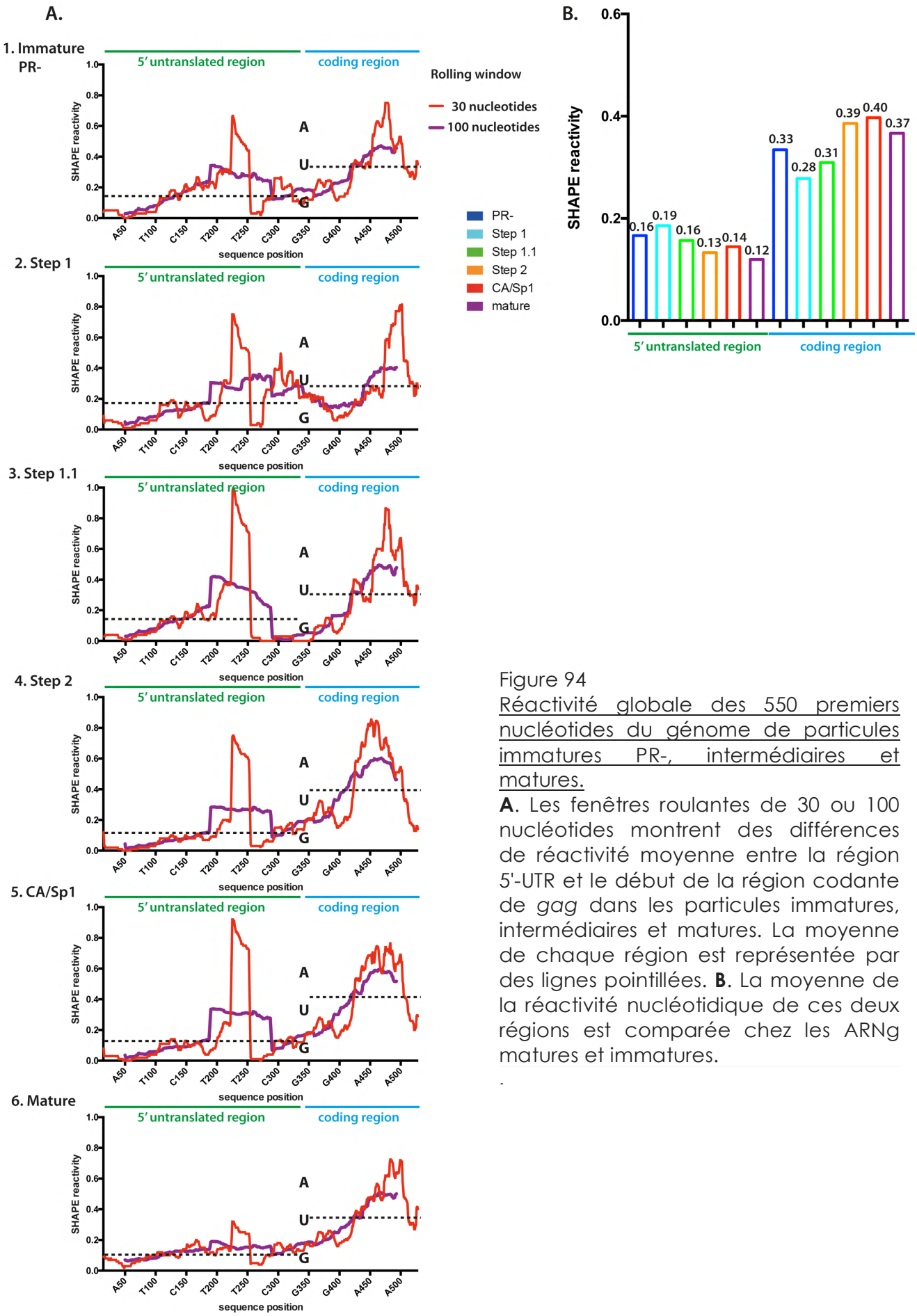


Figure 94

Réactivité globale des 550 premiers nucléotides du génome de particules immatures PR-, intermédiaires et matures.

A. Les fenêtres roulantes de 30 ou 100 nucléotides montrent des différences de réactivité moyenne entre la région 5'-UTR et le début de la région codante de *gag* dans les particules immatures, intermédiaires et matures. La moyenne de chaque région est représentée par des lignes pointillées. **B.** La moyenne de la réactivité nucléotidique de ces deux régions est comparée chez les ARNg matures et immatures.

.

II.2 La maturation génomique du VIH-1

II.2.1 Observations générales de la conformation structurale du génome du VIH-1 extrait de particules virales PR- et matures

Les profils de réactivité hSHAPE-Seq sont assez similaires, mais avec certaines différences principalement situées dans les domaines PBS et SL1. Cependant, le traitement AT-2 révèle des différences majeures entre les ARNs issus de particules PR- et matures. Ceci montre que la structure de l'ARNg est en effet différente entre ces deux états et que $Pr55^{Gag}$ et NCp7 se lient et agissent différemment. En effet, ce traitement est particulièrement intéressant et nous a permis de détecter les interactions de protéines contenant des domaines doigt de zinc avec l'ARNg. La rupture des interactions ARN-protéine entraîne à la fois des augmentations et des diminutions de la réactivité. La différence des réactivités obtenues avec DMSO et AT-2 est présentée à la figure 95 et met en évidence les sites de protection (réactivité obtenue dans la condition DMSO < réactivité obtenue dans la condition AT-2) et de déstabilisation de la structure du génome induite par la protéine (réactivité obtenue dans la condition DMSO > réactivité obtenue dans la condition AT-2).

Après traitement AT-2, la réactivité de plusieurs régions évolue de manière similaire dans les particules virales immatures et matures (figure 95, régions indiquées en rouge). Ces régions sont principalement situées dans le signal d'encapsidation de l'ARNg et protégées par $Pr55^{Gag}$ et NCp7. En revanche, $Pr55^{Gag}$ et NCp7 influent différemment la conformation de l'ARNg dans le domaine PBS et dans la région codante de *gag* (Figure 95, régions indiquées en vert). Le domaine PBS devient moins accessible lors du traitement AT-2 des particules virales immatures PR-, ce qui implique un effet déstabilisateur de $Pr55^{Gag}$, alors que certaines régions sont protégées par NCp7 dans les particules virales matures. La région codante de *gag* est fortement protégée par NCp7, tandis que $Pr55^{Gag}$ a un effet limité sur cette région.

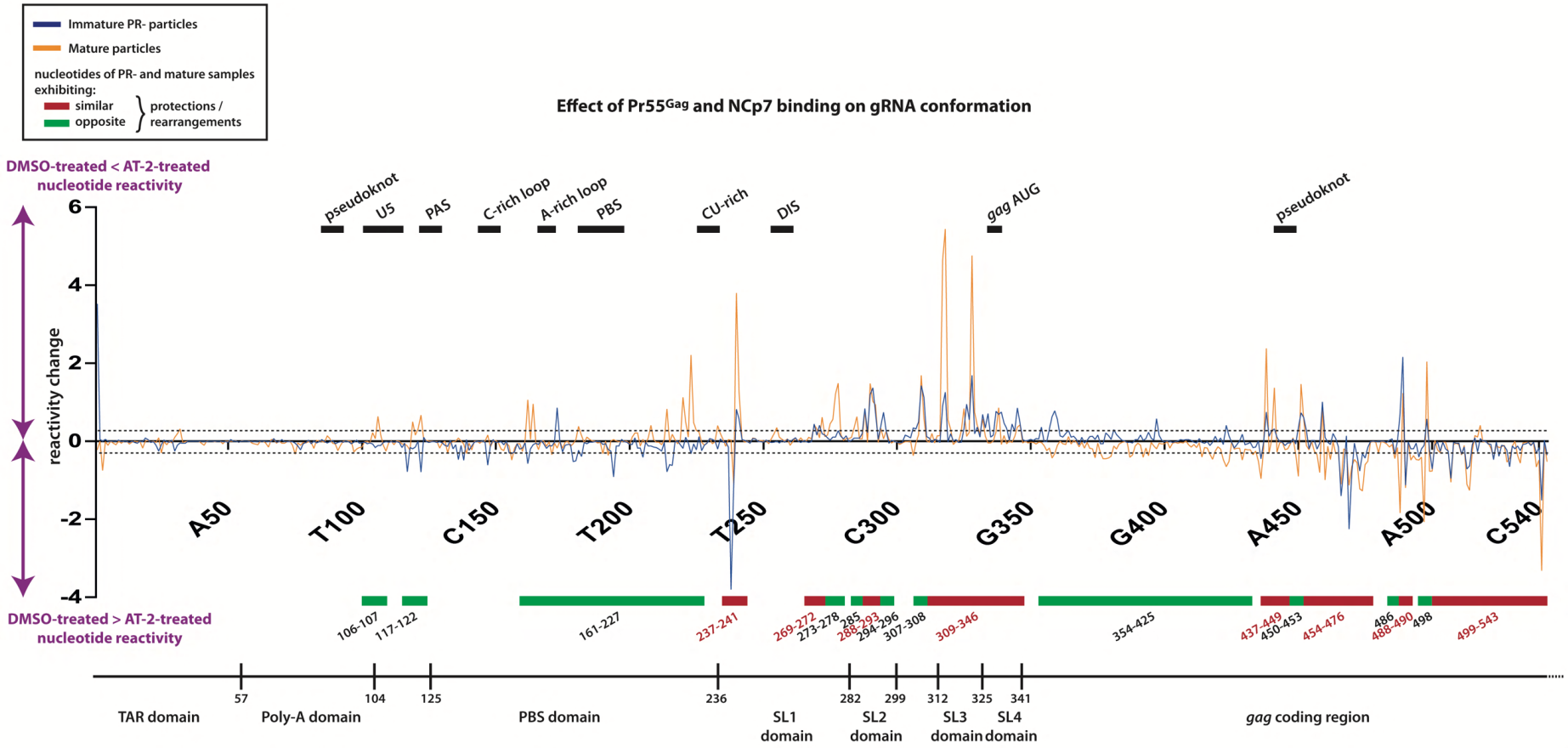


Figure 95

Effet de Pr55^{Gag} et NCp7 sur la conformation de l'ARNg.

Ce diagramme de différence illustre les changements de réactivité SHAPE lors du traitement AT-2, des particules PR- (bleu) et mature (orange), respectivement. Le changement de réactivité a été calculé en soustrayant la moyenne de réactivité des échantillons traités avec DMSO et AT-2. Une augmentation réactivité indique les sites de protection protéiques, alors qu'une réactivité réduite reflète la déstabilisation de la structure de l'ARNg par la protéine. Les sites ayant un changement de réactivité similaire dans les particules virales immatures et matures sont mis en évidence en rouge et les sites montrant des changements discordants sont représentés en vert. Un seuil à 0,2 et -0,2 est représenté par une ligne pointillée.

II.2.2 Pr55^{Gag} et NCp7 ont-ils un motif consensus de fixation à l'ARNg?

La composition nucléotidique des sites de protection de l'ARNg par Pr55^{Gag} et NCp7 a été comparée grâce au traitement AT-2 de particules PR- et matures (figure 96). Dix sites de fixation de Pr55^{Gag} sont présents dans les premiers 550 nucléotides du génome du VIH-1. Ils correspondent à un motif comprenant environ trois nucléotides avec une forte prédominance de nucléotides G et A à la troisième position. Il n'y a pas de consensus aux première et deuxième positions avec une fréquence relative similaire des résidus A, U et G. Les sites de fixation de NCp7 se composent d'un (12 sites), deux (5 sites) ou trois (8 sites) nucléotides. La prédominance des nucléotides A et G à la troisième position du site de 3 nucléotides est similaire à celle observée pour Pr55^{Gag}. La prédominance des résidus G à la fin du motif se retrouve également pour les sites de NCp7 à 1 et 2 nucléotides. La première et la deuxième position des sites à 3 nucléotides sont faiblement conservées, mais les résidus G sont relativement plus fréquents que dans les sites de fixation de Pr55^{Gag}. Ainsi, Pr55^{Gag} se fixe aux sites riches en A et G, alors qu'une proportion plus élevée de résidus G est nécessaire pour la liaison NCp7.

Les motifs du génome du VIH-1 qui lient Pr55^{Gag} et NCp7 pendant les étapes d'encapsidation, d'assemblage et de maturation ont été analysés par PAR-CLIP (246). Conformément à nos données, Kutluay et al. a montré que Pr55^{Gag} se lie préférentiellement à des motifs riches en A composés de cinq nucléotides avec au moins un nucléotide G. Les sites génomiques protégés par Pr55^{Gag} déterminés par hSHAPE-Seq sont plus petits (trois nucléotides) que les sites de liaison observés par PAR-CLIP, mais ces différences s'expliquent probablement par l'approche expérimentale utilisée, car le hSHAPE-Seq détecte uniquement les nucléotides protégés par Pr55^{Gag}, tandis que les sites de fixation peuvent impliquer une région plus vaste incluant éventuellement des nucléotides appariés. La prédominance de nucléotides G dans les sites de liaison NCp7 est en accord avec les structures RMN obtenues avec NCp7 (16, 112, 413) ainsi que les données de cartographie chimique *in viro* publiées par l'équipe de K. Weeks (426). Une préférence de NCp7 pour les motifs riches en nucléotides G et U est proposée par Kutluay et al. mais en regardant les motifs, une prédominance de résidus G est observée avec un nucléotide U pour deux nucléotides G (246).

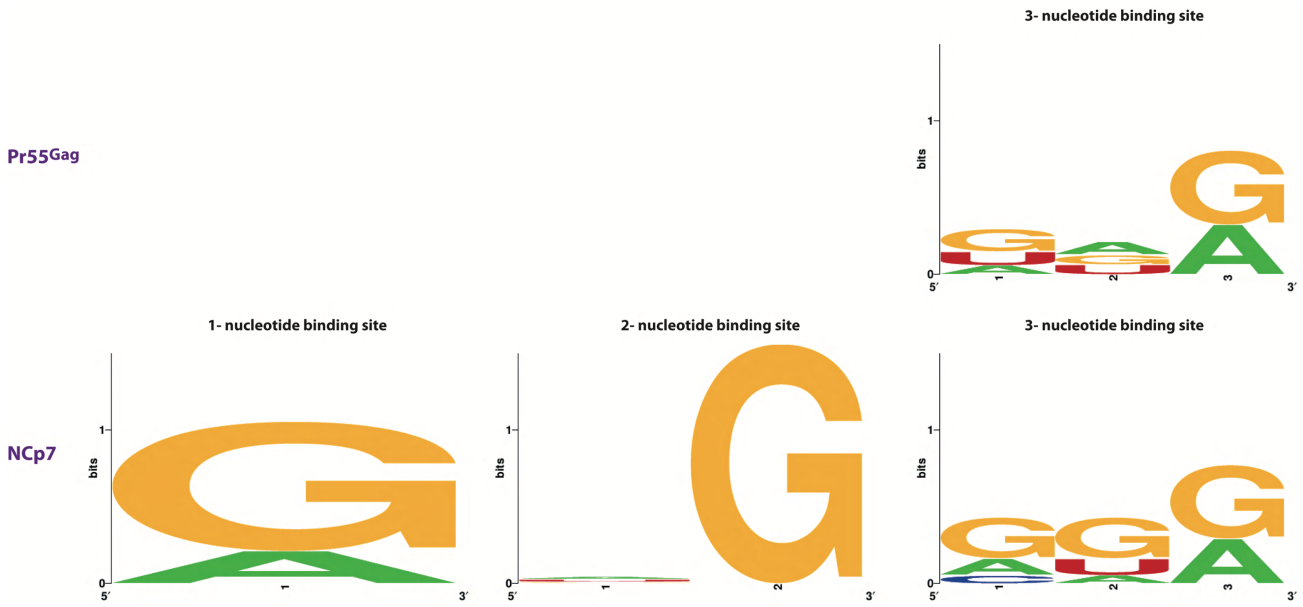


Figure 96
Présentation du site de fixation de Pr55^{Gag} et NCp7 au génome.
La composition nucléotidique des sites présentant une augmentation de réactivité après traitement AT-2 est présentée, dans le cas de particules immatures PR- et matures (http://weblogo.berkeley.edu/logo.cgi). La hauteur du motif indique la conservation de la séquence, tandis que la hauteur des symboles indique la fréquence relative des nucléotides à cette position.

II.2.3 Comment la structure du génome évolue-t'elle lors du clivage protéolytique de Pr55^{Gag}? Quel est l'impact de Pr55^{Gag} et des intermédiaires de maturation?

Les quatre virus mutants mimant la cascade protéolytique de Pr55^{Gag} affichent une évolution structurale cohérente du domaine PBS, du signal d'encapsidation et de la région codante de *gag*. En outre, cette évolution est en corrélation avec la conformation de l'ARNg dans les particules virales immatures PR- et matures. Les mutants Step 1 et Step 1.1 présentent les réarrangements les plus importants de l'ARNg. À partir de l'étape Step 2, cette dynamique est perdue et la conformation de l'ARNg ne semble pas être affectée au stade CA/Sp1, car aucun changement clair n'est observé dans nos données hSHAPE-Seq. La conformation finale est obtenue au stade mature, avec les derniers réarrangements des domaines PBS et SL4 (Figure 94 A).

Au stade Step 1, l'ARNg présente une réactivité accrue par rapport au stade immature PR- (figure 94 A2). En outre, le profil de réactivité de l'étape Step 1 est très similaire au profil immature traité avec AT-2 et suggère donc que NCp15 est incapable de protéger les premiers 550 nucléotides du génome du VIH-1. En effet, le signal d'encapsidation est hautement réactif, y compris les nucléotides du domaine SL4. Ces résultats sont en accord avec le fait que les intermédiaires contenant le domaine NC et NCp7 possèdent différentes propriétés de liaison à l'ARN *in vitro*.

NCp9 et NCp7 favorisent la formation de grands agrégats d'ARN, alors que ces agrégats n'ont pas été détectés avec NCp15. Ceci a été démontré par microscopie électronique de complexes NC-ADN (292, 293), analyse biophysique et biochimique des interactions NC-ARN (420, 432) et dimérisation de l'ARNg (201, 211, 313). De plus, nos données hSHAPE-Seq expliquent potentiellement la déficience observée de la dimérisation du génome (201, 211, 313), la RT et l'intégration (96) dans les particules virales mutantes dans lesquelles la maturation de $Pr55^{Gag}$ est arrêtée au stade NCp15 (313).

De manière surprenante, alors que NCp15 semble incapable de protéger l'ARNg de la même manière que $Pr55^{Gag}$, NCp15 déstabilise le domaine PBS aussi efficacement que $Pr55^{Gag}$, au niveau de la tige supérieure du domaine PBS, la boucle PBS et la séquence complémentaire au site PAS. L'accessibilité de ces sites est identique dans les particules PR- et Step 1. La déstabilisation du domaine PBS est perdue avec le traitement AT-2, ce qui indique que le réarrangement conformationnel n'est pas permanent au stade PR- immature. Comme NCp15 interagit faiblement avec l'ARNg, nos résultats suggèrent que les réarrangements du domaine PBS pourraient devenir permanents au stade Step 1, après clivage de $Pr55^{Gag}$ au site Sp1 / NC. En outre, cette réactivité accrue est conservée à l'étape 1.1, qui possède NCp9 et est progressivement perdue au fil de la maturation de $Pr55^{Gag}$, comme en témoigne la protection progressive de ces sites.

De manière intéressante, les nucléotides G257-G363 deviennent totalement non réactifs au stade Step 1.1 (Figure 97). Cette région comprend le signal d'encapsidation à partir du DIS et les 20 premiers nucléotides de la région codante de *gag*. En plus de cette région, le domaine PBS suit la même tendance observée au stade Step 1 et la réactivité des nucléotides C238-A242, le site présentant la plus forte réactivité dans les 550 premiers nucléotides, est même augmentée (figure 98).

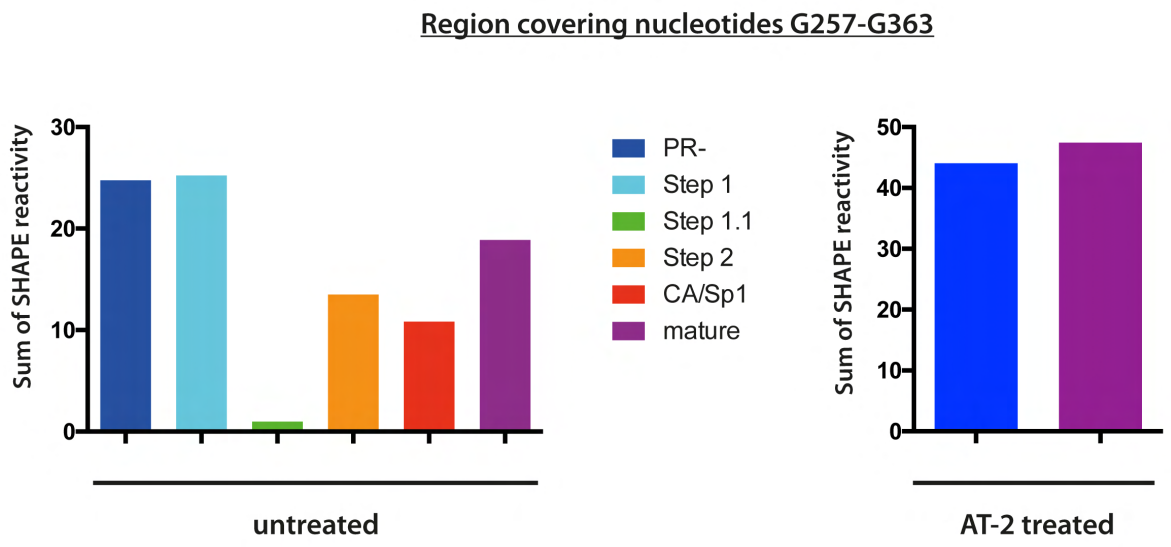


Figure 97
Evolution de la réactivité des nucléotides G257-G363 du génome du VIH-1 durant la maturation de $Pr55^{Gag}$.
La somme des réactivités SHAPE des nucléotides G257-G263 a été calculée pour les particules PR-, matures, les intermédiaires de maturation ainsi que les particules PR- et matures traitées avec AT-2.

Il serait intéressant de tester le traitement AT-2 à l'étape Step 1.1 afin de déterminer si les nucléotides G257-G363 sont protégés par NCp9 ou si l'ARNg est réarrangé. Des changements dans le profil de réactivité de cette région sont observés aux stades Step 1.1 et Step 2, même si la maturation de Pr55^{Gag} est bloquée à l'état NCp9 dans ces deux intermédiaires de maturation (Figure 97). La seule différence entre ces deux intermédiaires est le clivage MA / CA à l'étape Step 2, qui libère l'intermédiaire CA/Sp1. Ce clivage en amont de CA/Sp1 est nécessaire pour initier la déstructuration de la capside immature dans le modèle d'assemblage et de désassemblage responsable de la formation du core capsidique (109, 222). L'importance du désassemblage du core immature pour le remodelage conformationnel de cette région génomique reste inconnue.

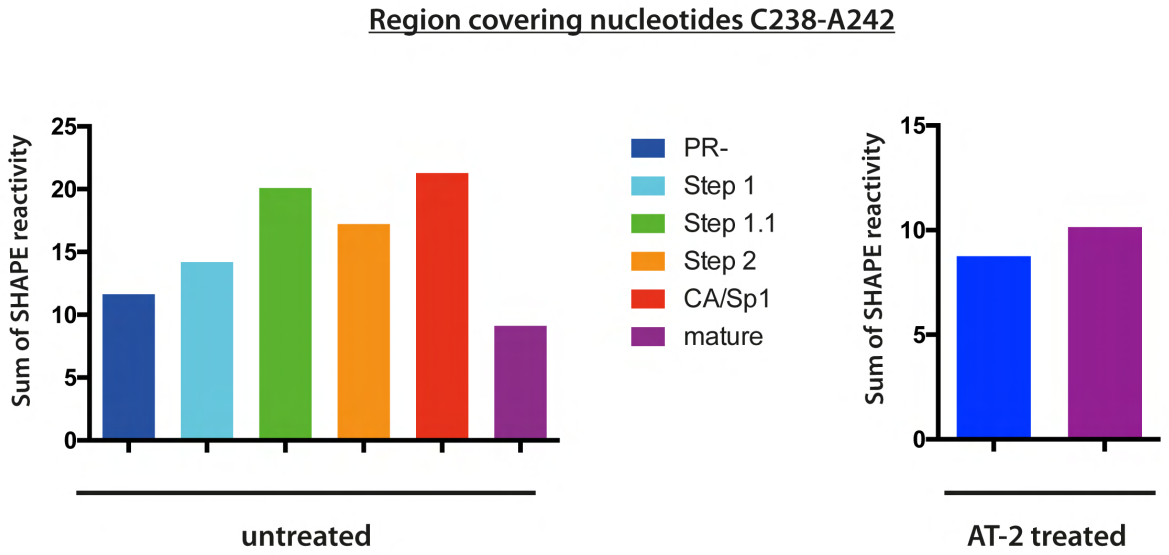


Figure 98
Evolution de la réactivité des nucléotides C238-A242 du génome du VIH-1 durant la maturation de Pr55^{Gag}.
La somme des réactivités SHAPE des nucléotides C238-A242 a été calculée pour les particules PR-, matures, les intermédiaires de maturation ainsi que les particules PR- et matures traitées avec AT-2.

III.1 Evolution du domaine PBS

Le domaine PBS, qui est déstabilisé par Pr55^{Gag} et se trouve dans une conformation "ouverte", est encore très réactif à l'étape Step 1 contenant NCp15 et aux étapes Step 1.1 et Step 2 contenant NCp9. La réactivité diminue à partir de l'intermédiaire Step 2, et le domaine PBS est presque totalement non réactif dans les particules virales matures. L'ARNt^{Lys,3} est fixé aux 18 nucléotides de la boucle PBS au stade PR- immature. Le traitement AT-2 des particules virales PR- immatures suggère que Pr55^{Gag} déstabilise la structure secondaire de la tige supérieure du domaine PBS (positions A136, A147, A138 et G162) et la séquence PAS (nucléotides A220, G221 et G223) proposé comme interagissant avec l'anti-PAS, de sorte que les interactions supplémentaires entre le tRNA^{Lys,3} et le domaine PBS peuvent déjà se produire, au moins de façon transitoire. Suite au traitement AT-2, la réactivité de ces régions

diminue, ce qui reflète une stabilisation de la structure intramoléculaire et indique que les interactions entre le domaine PBS et la boucle variable, la boucle anti-codon et le site TΨC ne subsistent pas en l'absence de $Pr55^{Gag}$. Cependant, le site accepteur du $tRNA^{Lys,3}$ reste fixé, bien qu'imparfaitement, aux 18 nucléotides de la boucle du domaine PBS en l'absence de $Pr55^{Gag}$, probablement en raison de la grande interface d'interaction. Cette interaction est améliorée au stade mature.

Puisque l'intermédiaire NCp9 est présent à la fois aux étapes Step 1.1 et Step 2, la diminution de réactivité du domaine PBS au dernier stade pourrait être dûe au clivage CA/Sp1 de l'intermédiaire p41 (MA-CA-Sp1) et à l'initiation subséquente de la formation du cone capsidique. Cette hypothèse est soutenue par le profil de réactivité inchangé du domaine PBS du stade Step 2 au stade CA/Sp1, où CA/Sp1 est encore présent, mais NCp9 est transformé en NCp7 et Sp2. La structuration finale du domaine PBS est finalement obtenue par le traitement de l'intermédiaire CA/Sp1 au stade mature et l'activité chaperonne de NCp7. À ce stade, les interactions entre le site accepteur, la boucle variable, la boucle anti-codon et le site TΨC du $tRNA^{Lys,3}$ sont stables lorsque NCp7 est dissocié par un traitement AT-2. Ces observations sont en accord avec la notion selon laquelle le recrutement du $tRNA^{Lys,3}$ est un processus effectué en plusieurs étapes, promu en premier lieu par $Pr55^{Gag}$, puis par NCp7 après achèvement du traitement protéolytique et éventuellement par les intermédiaires NCp9 également (71, 138, 168, 186, 375, 433). En outre, NCp7 peut affiner la fixation du $tRNA^{Lys,3}$ aux domaine PBS pour produire un complexe d'initiation entièrement fonctionnel pour la rétro-transcription, ce qui n'est probablement pas le cas avec NCp9. Cette hypothèse est en corrélation avec l'infectivité des particules virales mutantes CA / Sp1, contrairement aux mutants bloqués au stade Step 2, qui ne sont pas infectieux (313).

III.2 Evolution du signal d'encapsidation

Le domaine SL1 est hautement structuré de l'étape 1 au stade CA / Sp1, en accord avec le profil de réactivité des étapes immatures PR- et matures. La séquence palindromique de 6 nucléotides est non réactive lors de la cascade de maturation, indiquant que l'ARNg dimérise grâce au domaine SL1. La boucle interne inférieure est réactive à toutes les étapes de maturation (Figure 98) et constitue le site le plus réactif des 550 premiers nucléotides du génome (figure 94 A). Au stade immature, la boucle interne supérieure est protégée par $Pr55^{Gag}$. Ces données sont en accord avec les résultats précédemment publiés par notre équipe, montrant que cette boucle est le site de liaison principal de $Pr55^{Gag}$ (4, 48, 388). Lors du traitement AT-2 des particules virales matures, non seulement la réactivité de la boucle interne supérieure augmente, reflétant un site de liaison de NCp7, mais également la région entière comprenant les positions A269-G278. Rien de semblable n'a été observé dans les particules virales immatures. Nos données suggèrent fortement que la tige intermédiaire SL1, associant les deux boucles internes, n'existe pas au stade mature. Nous suggérons que cette conformation ouverte reflète la formation d'un duplexe

étendu impliquant la tige supérieure du domaine SL1 (nucléotides C248-G270). Ce duplexe étendu induit des contraintes structurales qui sont résolues grâce à l'ouverture de la tige intermédiaire de SL1. La potentielle existence de ce duplexe étendu remet également en question l'existence de la tige inférieure de SL1 dans le modèle SL1 étendu, du fait de la contrainte stérique. En effet, Mujeeb et al. a montré que l'ajout de deux paires de bases à la base de la tige intermédiaire SL1 créait un domaine SL1 stable et l'incapacité de former un duplexe étendu en présence de NCp7 (303).

Les boucles apicales des domaines SL2 et SL3 sont protégées par $Pr55^{Gag}$, NCp7 et vraisemblablement l'intermédiaire NCp9. Le profil de réactivité du domaine SL2 lors du traitement AT-2 révèle l'instabilité de la tige aux stades immature et mature et la stabilisation de la conformation du domaine SL2 médiée par $Pr55^{Gag}$ et NCp7.

Le domaine SL4 contient le codon d'initiation de *gag* et est généralement décrit comme étant associé avec les nucléotides U105-G116 dans l'élément U5 et la région riche en CU (positions U228-C233). Cette région présente un intérêt particulier car il a été proposé comme régulant l'encapsidation de l'ARNg grâce à un réarrangement structural de la 5'-UTR (1, 220, 269). Nos données obtenues par hSHAPE-Seq révèlent un fort effet de $Pr55^{Gag}$ et des intermédiaires contenant le domaine NC sur la réactivité du domaine SL4 plutôt que l'appariement nucléotidique de ce domaine. En effet, $Pr55^{Gag}$ protège fortement le domaine SL4 et cette protection est perdue lors du premier clivage de $Pr55^{Gag}$ permettant la libération de NCp15. Le domaine SL4 n'est pas réactif du stade Step 1.1 au stade CA/Sp1, vraisemblablement protégé par NCp9 et dans une moindre mesure par NCp7 au stade mature, comme le montre le traitement AT-2 des particules immatures PR- et matures. Ainsi, alors que l'intermédiaire NCp15 est incapable de lier efficacement le domaine SL4, cette capacité est restaurée avec NCp9 et NCp7. La déprotection du domaine SL4 observée lors du traitement AT-2 est en défaveur de l'interaction de cette région avec les nucléotides U105-G116 dans U5 et la région riche en CU au stade immature et au moins avec la région riche en CU au stade mature. Ainsi, nos résultats ne sont pas compatibles avec l'interaction U5-AUG proposée pour l'encapsidation de l'ARNg et les modèles proposés à partir de résultats obtenus *in viro* (426), *ex viro* (383, 422) et *in vitro* (191, 220). Il est à noter que la comparaison de la structure de l'ARNg avec et sans traitement AT-2 a largement contribué à ces conclusions.

III.3 Evolution des premiers 200 nucléotides de la region codante de gag

La région codante de *gag* est fortement modifiée par rapport à la 5'UTR (Figure 94 A). Deux régions riches en GC, non réactives à toutes les étapes de maturation, sont présentes en absence de traitement AT-2. Plusieurs domaines sont très réactifs et la fixation de $Pr55^{Gag}$ n'influence pas fortement ces domaines, ni en les protégeant ni en les déstabilisant. Cependant, la réactivité des positions G369-U397 et U486-A533

est perdue lors du traitement AT-2 de particules matures. Ainsi, NCp7 et éventuellement NCp9 déstabilisent fortement ces sites. En ce qui concerne les nucléotides G408-A475, leur réactivité augmente au cours du traitement protéolytique de $Pr55^{Gag}$, à l'exception du stade Step 1. De plus, nos résultats hSHAPE-Seq démontrent un effet déstabilisant assez important de NCp7 au début de la région codante de *gag*, alors que l'équipe de K. Weeks a suggéré que NCp7 n'ait peu d'activité de déstabilisation dans cette région (426).

II.2.4 Comparaison des données *in vitro* et *in viro* de cartographie chimique

Des expériences d'empreinte protéique ont été effectuées *in vitro* avec $Pr55^{Gag}$, GagΔp6, NCp15, NCp9 et NCp7 afin de valider les résultats de hSHAPE-Seq et de comparer les propriétés de chaque protéine. Ces résultats ont été comparés avec les mutants viraux mimant la cascade protéolytique de $Pr55^{Gag}$ et les résultats *in vitro* et *in viro* obtenus sont globalement similaires. Certaines divergences concernant, par exemple, les nucléotides situés entre les domaines SL2 et SL3 ont été trouvées, mais il est donc important de garder à l'esprit que les conditions *in vitro* représentent un système simplifié par rapport à l'environnement présent *in viro*. En outre, les données *in vitro* sont compatibles avec l'existence de l'interaction U5-AUG ainsi que l'interaction entre les régions riches en AG et riches en CU (résultats non présentés). Il est à noter que les expériences d'empreinte *in vitro* ont été effectuées en absence du $tRNA^{Lys,3}$. Nous avons montré *in viro* que la structuration finale du domaine PBS est obtenue par le traitement de l'intermédiaire CA / Sp1 et l'activité chaperonne de NCp7, stabilisant la fixation du $tRNA^{Lys,3}$. Ainsi, l'influence de la fixation du $tRNA^{Lys,3}$, sur la structuration finale du domaine PBS, explique potentiellement l'écart entre les résultats obtenus *in viro* et *in vitro*.

Nos résultats sont en corrélation avec une amélioration de l'activité chaperonne des protéines contenant le domaine NC pendant la cascade protéolytique, ce qui est cohérent avec les résultats *in viro* et l'hypothèse que la libération de NCp7 augmente sa flexibilité et influence ses propriétés de liaison / dissociation des acides nucléiques. En ce qui concerne l'intermédiaire NCp15, aucune forte différence de comportement n'est observée *in vitro* et *in viro*.

Dans cette étude, nous avons également testé la protéine GagΔp6, un intermédiaire qui n'existe pas pendant le processus de maturation naturel, car cette protéine est souvent utilisée comme substitut de $Pr55^{Gag}$ dans des études *in vitro* (110, 111, 206, 225, 306, 423 , 431). De manière surprenante, Pr55Gag et GagΔp6 ont des caractéristiques très différentes, comme clairement visualisé au niveau des boucles apicales des domaines SL1 et SL3 et à la boucle interne inférieure de SL1. Fait intéressant, les profils de réactivité obtenus en présence de GagΔp6 et NCp9, qui possèdent tous deux le domaine p6, sont très similaires. Contrairement à $Pr55^{Gag}$, NCp15 et NCp7, ces deux protéines ne se sont pas complètement dissociées de la matrice d'ARNg en l'absence de SDS. Ces résultats mettent en évidence les propriétés d'agrégation de NCp9 et GagΔp6 et suggèrent que Sp2 et / ou p6

affectent le mode de liaison de ces protéines. Fait intéressant, il a été proposé que le domaine NC interagisse de manière transitoire avec p6 (101, 292, 293, 420). GagΔp6 a été souvent utilisé comme un substitut de $Pr55^{Gag}$ dans les études *in vitro* parce que $Pr55^{Gag}$ est sensible au clivage protéolytique lors de son expression et purification (65). Notre laboratoire a récemment testé l'impact du domaine p6 sur la spécificité de liaison de $Pr55^{Gag}$ aux ARN viraux (résultats non publiés de Dubois et al.). De manière surprenante, cette comparaison révèle que GagΔp6 lie l'ARNg, les ARN viraux épissés et les ARN cellulaires avec une affinité très similaire. En outre, le domaine SL1 ne constitue pas le principal élément de reconnaissance de GagΔp6, contrairement à $Pr55^{Gag}$. Ces résultats démontrent que le domaine p6 joue un rôle crucial dans la liaison spécifique de $Pr55^{Gag}$ à l'ARNg du VIH-1. Ainsi, $Pr55^{Gag}$ et GagΔp6 ne peuvent être considérés comme équivalents en ce qui concerne leurs propriétés de liaison à l'ARN.

II.2.5 Quel est l'impact des inhibiteurs de protéase sur la conformation structurale de l'ARNg?

Les faibles concentrations d'IP utilisées dans des milieux cliniques inhibent efficacement la réplication virale tout en produisant uniquement des défauts subtils ou non détectables sur le clivage des précurseurs $Pr55^{Gag}$ et $Pr160^{GagPol}$. En outre, les IP affectent le processus de rétro-transcription, mais pas l'activité enzymatique de l'enzyme elle-même (304), et une accumulation des intermédiaires NCp9 a été signalée (304). Sur la base de ces faits, nous avons émis l'hypothèse que l'activité des IP pourrait être liée au clivage du domaine NC et aux défauts de maturation structurale de l'ARNg. La comparaison de la conformation de l'ARNg avec quatre concentrations différentes de LPV et ATV n'a révélé aucune évolution en corrélation avec l'augmentation de la concentration d'IP. En outre, le profil structural de l'ARNg obtenu à 1,1 nM de LPV est différent des conformations immatures PR- et matures. Ainsi, nos résultats indiquent que la maturation structurale de l'ARNg est bloquée à un stade intermédiaire en présence d'IP.

En effet, la conformation de l'ARNg à 1,1 nM LPV ressemble aux stades Step 2 et CA/Sp1, aucune différence claire n'étant observée entre ces deux profils de réactivité. En outre, le LPV a été signalé comme affectant le clivage NC / Sp2 (304), reflétant éventuellement un défaut de maturation similaire à celui observé pour le mutant Step 2. L'intermédiaire NCp9 est déjà présent à l'étape Step 1.1, où CA/Sp1 est toujours attaché à MA et donc lié à la membrane virale. Cependant, le profil de réactivité des échantillons traités par IP est différent des particules bloquées à l'étape Step 1.1, ce qui pourrait indiquer une influence de la formation du cone capsidique sur le réarrangement structural du génome. Moore et al. a en outre étudié la stabilité de l'ARNg en présence d'IP à l'IC_{50} et l'IC_{90} (295). L'analyse par Northern blot n'a montré qu'un léger impact des IP sur la stabilité de l'ARNg dimérique, tandis qu'environ 40% des particules virales ont montré une morphologie aberrante avec un complexe RNP excentrique. Ainsi, les auteurs ont conclu que la stabilisation de l'ARNg dimérique est indépendante de la formation du core viral.

Des changements de morphologie des particules du VIH-1 sont remarqués en présence de faibles doses d'IP. Une augmentation graduelle des particules aberrantes avec des cores diffus ou vides, ainsi qu'une accumulation de complexes RNP excentriques a été observée (215, 295, 304). Selon Müller et al., la proportion de particules présentant ces défauts est trop faible pour expliquer la perte d'infectivité. Cependant, la proportion de particules morphologiquement matures à l'IC_{50} de LPV diminue d'environ 40%, avec 20% de particules présentant un défaut de maturation et 30% de particules immatures (304). Ceci est conforme avec l'étude de Moore et al. montrant environ 25% des particules avec une région dense et excentrée du cone et 15% de particules immatures en présence d'IC_{50} de LPV et ATVs (296). La proportion de particules virales avec une morphologie mature diminue avec des concentrations croissantes d'IP, avec 20-20% restant à l'IC_{90} de LPV. En outre, le degré d'altération de la maturation de $Pr55^{Gag}$ et la réduction de l'infectivité à faible concentration d'IP ne sont pas directement et quantitativement corrélés.

La relocalisation du complexe RNP, observée en présence d'IP est similaire au phénotype morphologique observé avec les NCINI (144, 210, 228). Dans le cas des NCINI, le phénotype s'expliquerait en raison de la multimérisation de l'IN, ce qui perturberait l'interaction IN-RNP et donc l'initiation de la morphogenèse du cone capsidique. Il est donc tentant de lier l'action des IP à la formation du cone capsidique.

II.2.5 Modèle proposé de la structure secondaire de la région 5' du génome du VIH-1 au stade mature

Sur la base de nos données hSHAPE-Seq obtenues lors du traitement AT-2 de particules matures, nous proposons un nouveau modèle de structure secondaire des 550 premiers nucléotides du génome du VIH-1 (figure 99). Les sites de liaison de NCp7 proposés et les sites déstabilisés par NCp7 sont indiqués sur le modèle.

Dans ce modèle, les nucléotides U105-G116 de U5 interagissent avec les nucléotides G361-G365 de la région codante de *gag*, au lieu de l'interaction U5-AUG (191, 220, 383, 422, 426) (figure 99 A). De plus, la région riche en CU est proposée comme interagissant avec le domaine SL4, mais de manière décalée, avec les nucléotides G331-A336 interagissant avec la région riche en CU au lieu des nucléotides C329-A334 communément proposés (Figure 99 B). La structuration globale du domaine PBS est similaire à la structure consensus trouvée dans les modèles publiés, à l'exception du PAS (Figure 99 C). En effet, nos résultats indiquent que le PAS est lié au $tRNA^{Lys,\ 3}$. L'anti-PAS est protégé par NCp7 et est donc proposé comme étant non apparié. La boucle PBS comprend un nombre plus petit de nucléotides car deux interactions supplémentaires sont proposées (positions U131-A133 avec positions U200-G202 et positions U115-U118 avec positions A205-G208). Le domaine SL1 est dessiné comme une tige-boucle intramoléculaire irrégulière. La boucle apicale et la tige forment probablement un duplex prolongé. La tige intermédiaire est bio-informatiquement prédite (logiciel RNA structure (346)), même lorsque la réactivité

SHAPE est appliquée. Les domaines SL2 et SL3 plus les nucléotides liant ces deux domaines adoptent la structure communément admise (figure 99 D). Les premiers 200 nucléotides de la région codante de *gag* sont fortement structurés, mais cette région est déstabilisée par NCp7, comme le montre la réactivité réduite observée lors du traitement AT-2 (figure 99 E). Nous avons également essayé de modéliser la région 5' du génome du VIH-1 au stade immature, en utilisant les résultats obtenus après traitement AT-2 par hSHAPE-Seq. Cependant, le fort effet déstabilisant de $Pr55^{Gag}$ sur le domaine PBS rend difficile la modélisation du domaine PBS, et donc de la totalité de la région 5' du génome du VIH-1. Il sera intéressant de construire un modèle 3D de la structure mature de la région 5' du génome du VIH-1. Ce sera un test important pour notre modèle de structure secondaire, car il indiquera s'il est topologiquement réalisable. La même stratégie pourrait être appliquée pour discriminer entre plusieurs modèles de structure secondaire de la conformation de l'ARNg immature.

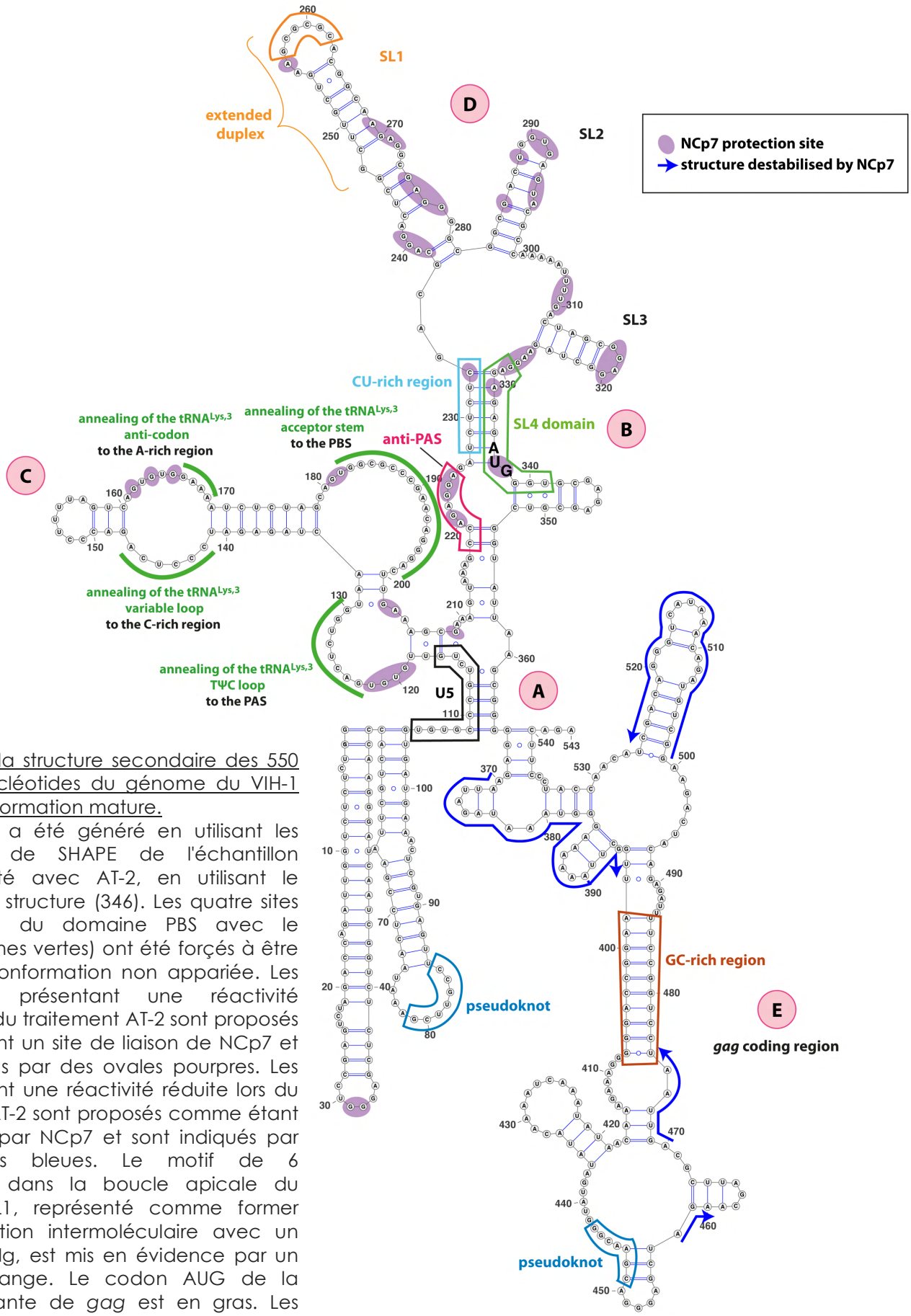


Figure 99

Modèle de la structure secondaire des 550 premiers nucléotides du génome du VIH-1 dans la conformation mature.

Ce modèle a été généré en utilisant les contraintes de SHAPE de l'échantillon mature traité avec AT-2, en utilisant le logiciel RNA structure (346). Les quatre sites d'interaction du domaine PBS avec le $tRNA^{Lys,3}$ (lignes vertes) ont été forcés à être dans une conformation non appariée. Les nucléotides présentant une réactivité accrue lors du traitement AT-2 sont proposés comme étant un site de liaison de NCp7 et sont indiqués par des ovales pourpres. Les sites montrant une réactivité réduite lors du traitement AT-2 sont proposés comme étant déstabilisés par NCp7 et sont indiqués par des flèches bleues. Le motif de 6 nucléotides dans la boucle apicale du domaine SL1, représenté comme former une interaction intermoléculaire avec un second ARNg, est mis en évidence par un encadré orange. Le codon AUG de la région codante de *gag* est en gras. Les deux régions impliquées dans le pseudoknot entre la boucle poly-A et la région codante de *gag* (324), non réactif dans nos données hSHAPE-Seq, sont indiquées par des encadrés bleus.

Structural rearrangements of the HIV-1 genomic RNA during maturation of the viral particle

During the replicative cycle, the HIV-1 particle buds from the infected cell as an immature particle and has to undergo a maturation process to become infectious. Proteolytic processing of $Pr55^{Gag}$ triggers morphological rearrangements of the immature $Pr55^{Gag}$ shell in order to form a mature particle with the characteristic cone-shaped core containing the genomic RNA (gRNA) dimer. Concomitantly, the gRNA dimer becomes more stable and more compact. Whereas $Pr55^{Gag}$ processing is well established, gRNA structural maturation remains poorly understood. Structural rearrangements of the genome are facilitated by the RNA chaperone activity of the NCp7 protein. Moreover, NC-containing intermediates, albeit transiently produced, are crucial for viral infectivity. Our goal was to determining **the different steps leading to the formation of the mature dimeric gRNA** and to better **understand the link between proteolytic processing and genomic maturation**.

To this end, the structure of the gRNA 5' region was assessed by **1.** *in vitro* and **2.** *in viro* chemical probing. **1.** *In vitro* footprinting assays of the 5' region of the HIV-1 gRNA were performed with $Pr55^{Gag}$, Gag∆p6, NC-containing intermediates and NCp7. These experiments allowed us to compare the RNA binding sites of these proteins and their effects on the gRNA structure. **2.** In order to analyse the steps leading to the mature gRNA conformation, we developed an *in viro* chemical probing technique, termed hSHAPE-Seq. Thanks to this powerful approach, we analysed the first 550 nucleotides of the HIV-1 genome from several wild type and mutant viruses mimicking the sequential processing of $Pr55^{Gag}$. The gRNA secondary structure was also studied within immature PR- and mature viral particles treated with the AT-2 zinc ejector, in order to identify the $Pr55^{Gag}$ and NCp7 binding sites as well of the gRNA destabilising activity of these proteins.

The hSHAPE-Seq methodology allowed us to detect multiple RNA-RNA and RNA-protein interactions during the sequential maturation process Differences in the PR- and mature gRNA conformations were mostly located at the PBS and SL4 domains. Major differences were observed upon AT-2 treatment. Indeed, $Pr55^{Gag}$ and NCp7 strongly influence the gRNA conformation by progressively stabilising the annealing of four different regions of $tRNA^{Lys,3}$ to the PBS domain. At the immature stage, the acceptor arm of $tRNA^{Lys,3}$ is annealed to the 18 nucleotides of the PBS, while the three additional interactions seem to transiently exist thanks to the chaperone activity of $Pr55^{Gag}$. Step 1 and Step 1.1 stages exhibit the most important rearrangements. In addition, Step 1 highlights limited interactions of NCp15 with the HIV-1 genome. From the Step 2 stage, this dynamic is lost and the gRNA conformation does not make apparent progress at the CA/Sp1 stage. The final conformation is obtained at the mature stage with multiple stable interactions between $tRNA^{Lys,3}$ and the PBS domain and important rearrangements of the SL4 domain. Interestingly, SL4 becomes reactive upon dissociation of $Pr55^{Gag}$ and NCp7, strongly suggesting that this region is unpaired. Thus, our results support neither the U5/AUG interaction nor the interaction between the CU-rich and the AG-rich regions.

The structure of the gRNA 5' region has also been investigated within viral particles treated with two different protease inhibitors (PIs) in order to better understand the mechanism of action of these antivirals and to assess their potential effect on gRNA maturation. Our hSHAPE-Seq results suggest that the structural maturation of the genome cannot proceed to completion in the presence of PIs. The gRNA reactivity profile in the presence of PIs correlate with the profile of the Step 2 stage, the last maturation stage where NC/Sp2 product is present. As PIs were previously shown to impair maturation of the NC/Sp2 intermediate, and considering the crucial role of the NCp7 RNA chaperone activity, we suggest that impairment of the NC/Sp2 cleavage blocks gRNA structural rearrangements and further inhibit viral replication.