



Relations structure-fonction du premier transfert de brin chez le vih-1

Ouerdia Maskri

► To cite this version:

Ouerdia Maskri. Relations structure-fonction du premier transfert de brin chez le vih-1. Microbiologie et Parasitologie. Université Paris Saclay (COmUE), 2016. Français. NNT : 2016SACLN062 . tel-01450381

HAL Id: tel-01450381

<https://theses.hal.science/tel-01450381>

Submitted on 6 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLN062

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À
“L'ECOLE NORMALE SUPERIEURE PARIS-SACLAY”

ECOLE DOCTORALE N° 577
SDSV: structure et dynamique des systèmes vivants
Spécialité de doctorat : Sciences de la vie et de la santé

Par

Mme Ouerdia Maskri

Relations structure-fonction du premier transfert de brin chez le VIH-1

Thèse présentée et soutenue à Cachan, le 09 décembre 2016

Composition du Jury :

Dr Jean-christophe PAILLART
Dr Bruno SARGUEIL
Dr Claudine MAYER
Dr Satoko YOSHIZAWA
Dr Philippe FOSSE

Directeur de Recherche, Université de Strasbourg
Directeur de Recherche, Université Paris Descartes
Professeure, Université Paris Diderot
Directeur de Recherche, Université Paris-Saclay
Directeur de Recherche, ENS Paris-Saclay

Rapporteur
Rapporteur
Présidente
Examinatrice
Directeur de thèse

Remerciements

Ce travail réalisé au niveau de l'ENS de Cachan, n'aurait pas été possible sans le soutien indéfectible de toute l'équipe « Structures et interaction des acides nucléiques » du Laboratoire de Biologie et Pharmacologie Appliquée –LBPA.

Je revendique avoir fait, au travers de ce travail de thèse et dans l'esprit du philosophe et essayiste Pierre Sansot « l'éloge de la lenteur » En effet, il m'a fallu du temps pour dessiner les contours de mon sujet, comprendre ce que j'observais, avoir un certain recul et mettre en forme une thèse. Il a fallu, aussi, que beaucoup de personnes s'investissent pour me permettre d'atteindre cet objectif. Ce travail est, pour moi, l'occasion de témoigner ma profonde gratitude à tous ceux et celles qui m'ont encadré et soutenu pour réaliser et finaliser cette thèse.

Je tiens, tout d'abord, à remercier Le Dr Malcolm Buckle pour m'avoir accepté au sein de son unité et permis, surtout, de travailler dans un environnement épanouissant. Les conditions de travail qui m'ont été offertes m'ont permis de me consacrer sereinement à l'élaboration de ma thèse

Je remercie mon directeur de thèse, le Dr Philippe Fossé, pour m'avoir encadré et accueilli dans son équipe. Je lui suis très reconnaissante pour le temps conséquent qu'il m'a consacré. Ses conseils avisés et sa capacité d'écoute ont été prépondérants pour la bonne réussite de cette thèse. Ses remarques, son ouverture d'esprit, sa franchise, sa gentillesse sont autant d'éléments qui m'ont permis d'atteindre cet objectif. Je dis, aussi, en paraphrasant Albert Einstein, qu'en travaillant avec vous, j'ai compris que dans la vie il faut agir pour être une personne de valeur et non un être de succès.

Je remercie également tous les membres du jury qui ont accepté de sacrifier une partie de leur temps pour évaluer mes travaux de thèse. Je mettrai à profit cette occasion pour profiter de votre culture scientifique et vos compétences professionnelles incontestables. Je dis tout simplement, que je suis très honoré que mon travail soit évalué par des scientifiques de très haut niveau. Vos remarques et vos conseils me donneront plus d'assurance pour être, davantage, persévérante et rigoureuse. Veuillez trouver ici le témoignage de mon profond respect.

Je souhaite remercier particulièrement Françoise Chaminade, pour sa gentillesse, ses encouragements, ses conseils avisés et pour m'avoir donné accès à sa grande expérience.

Je tiens à remercier, également, tous les membres de l'équipe : Olivier Mauffret, Brigitte Hartmann, Elizabeth Bugnard, Xu-guang Xi et Brigitte René, pour leur aide et leurs conseils avisés et leur bonne humeur.

Je souhaite remercier spécialement Loussiné Zargarian, Anissa Belfetmi, Akli Ben imeddouren, Ahmad elbahnsi et Cel Areny-Naves pour leur soutien et leur patience tout au long des quatre années passées ensemble, nous avons partagé de très bons moments dans un environnement amical et fraternel. Je remercie également tous les membres du laboratoire pour ces quatre belles années passées en leur compagnie.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille et mes amis et en particulier à ma mère, mon père et mes deux petites sœurs. Malgré mon éloignement, depuis maintenant, cinq longues années, votre amour, votre tendresse et vos encouragements, dans des moments de fatigue et de doute, m'ont toujours donné plus d'énergie pour me surpasser. Merci pour avoir fait de moi ce que je suis aujourd'hui. Est-ce un bon endroit pour dire ce genre de chose ? Je n'en connais, en tous cas, pas meilleur qu'un lieu du savoir.

Abréviations

ADAR1	Adenosine deaminase acting on RNA 1
AIDS	Acquired immunodeficiency syndrome
ALIX	ALG-2-interacting protein X
ALV	Avian leukosis virus
AMV	RT Avian myeloblastosis virus reverse transcriptase
AN	Acide nucléique
APOBEC3G	Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G
BLV	Bovine leukemia virus
CA	Capsid
CCD	catalytic core domain
CCR5	C-C-chemokine receptor 5
cDNA	Complementary DNA
CM	Cynomolgus monkey
cPPT	Central polypurine tract
CPSF	Cleavage polyadenylation specificity factor
CRF	Circulating recombinant form
CRM1	Chromosome region maintenance 1
CRS	<i>Cis</i> -acting repressive sequences
CTD	C-terminal domain / carboxy-terminal domain
CXCR4	C-X-C chemokine receptor type 4
CypA	Cyclophilin A
DC	Dendritic cell
DIS	Dimerization initiation site
dNTP	Deoxyribonucleoside triphosphate
dsDNA	Double-stranded DNA
ERT	Endogenous reverse transcription
ESCRT	Endosomal sorting complex required for transport
FRET	Fluorescence resonance energy transfer
gRNA	Genomic RNA
HERV	Human endogenous retrovirus
HFV	Human foamy virus
HIV	Human immunodeficiency virus

HR	Heptat repeat
HTLV	Human T-cell leukemia virus
IBD	IN binding domain
ICTV	International Committee of Taxonomy of Viruses
IN	Integrase
INS	Inhibitory/instability RNA sequences
LAV	Lymphadenopathy associated virus
LDF	Long DNA fragment
LEDGF	Lens epithelium-derived growth factor
LTR	Long terminal repeat
LysRS	Lysyl-tRNA synthetase
MA	Matrix
MLV	Murine leukemia virus
MMTV	Mouse mammary tumor virus
NA	Nucleic acid
NC	Nucleocapsid protein
NF- κ B	Nuclear factor κ -light-chain-enhancer of activated B cells
NTD	N-terminal domain
NWM	New world owl monkey
ORF	Open reading frame
OWM	Old world monkey
PAS	Primer activation signal
PBMC	Peripheral blood mononuclear cell
PBS	Primer binding site
PE	Psi element
PIC	Pre-integration complex
PI(4,5)P2	Phosphoinositol (4,5) bisphosphate
PPT	Polypurine tract
PR	Protease
Psi	Packaging signal
P-TEFb	Positive transcription elongation factor b
Rh	Rhesus monkey
RNase H	Ribonuclease H
RRE	Rev responsive element

RSV	Rous sarcoma virus
RT	Reverse transcriptase
RTC	Reverse transcription complex
SDF	Short DNA fragment
SHAPE	Selective 2'-hydroxyl acylation analysed by primer extension
SIV	Simian immunodeficiency virus
SIVage	Simian immunodeficiency virus isolated from an African green monkey
SIVcpz	Simian immunodeficiency virus isolated from chimpanzee
SIVmac	Simian immunodeficiency virus isolated from a macaque
SIVsm	Simian immunodeficiency virus isolated from sooty mangabey monkeys
SL	Stem-loop
SLIP	TTTTTT slippery site
SP	Self-priming
ssDNA	Minus-strand "strong-stop" DNA
STLV-III	Simian T-cell leukemia virus-III
TAR	Transactivator response element
TR2	Transportin-SR2
TRIM5 α	Tripartite motif 5 α
USE	Upstream enhancer element
UTR	Untranslated region
WDSV	Walleye dermal sarcoma virus
ZF	Zinc finger

Introduction.....	1
1. Les rétrovirus.....	1
1.1 Caractéristiques des rétrovirus.....	1
1.2 Classification des rétrovirus.....	2
2. Connaissances générales sur le VIH.....	4
2.1 Le VIH, l'agent causal du SIDA.....	4
2.2 Origine du VIH.....	5
2.2.1 VIH-1.....	7
2.2.2 VIH-2.....	7
2.3 Pathogenèse du VIH.....	8
2.4 Caractéristiques du VIH-1.....	9
2.4.1 Structure de la particule virale.....	9
2.4.2 Organisation génétique.....	10
2.5 Cycle réplcatif du VIH-1.....	11
2.5.1 Entrée du VIH-1.....	12
2.5.2 Transcription inverse de l'ARN viral.....	14
2.5.3 Intégration du génome viral dans l'ADN de la cellule hôte.....	14
2.5.4 Synthèse et maturation de l'ARN viral.....	15
2.5.5 Encapsidation de l'ARN génomique dans la particule virale.....	18
2.5.6 Assemblage et bourgeonnement de la particule virale.....	20
2.5.7 Maturation de la particule virale.....	21
3. Protéines intervenant dans le cycle réplcatif du VIH-1.....	24
3.1 Protéines cellulaires.....	24
3.1.1 TRIM 5.....	25
3.1.2 APOBEC3.....	26
3.1.3 Autres facteurs cellulaires intervenant dans le cycle réplcatif.....	28
3.2 Les protéines du VIH-1.....	29
3.2.1 La glycoprotéine d'enveloppe.....	30
3.2.2 Les protéines de structure.....	30
3.2.2.1 La protéine de la matrice (MA).....	30
3.2.2.2 La protéine de la capside (CA).....	30
3.2.2.3 La protéine de la nucléocapside NC.....	32
3.2.2.3.1 Structure de la NC.....	32
3.2.2.3.2 Propriétés de la NC.....	33
3.2.2.3.2.1 Activité déstabilisatrice des acides nucléiques.....	34
3.2.2.3.2.2 Activité agrégative des acides nucléiques.....	35
3.2.2.3.2.3 Cinétique rapide de liaison.....	35
3.2.2.3.2.4 Mode de fixation de la NC sur acides nucléiques.....	35
3.2.2.3.3 Role de la NC au cours du cycle réplcatif.....	38
3.2.2.3.3.1 Transcription inverse.....	38
3.2.2.3.3.2 Intégration de l'ADN viral.....	38

3.2.2.3.3	Assemblage et incorporation de l'ARNg viral.....	39
3.2.2.3.4	Bourgeonnement et maturation de la particule virale...	40
3.2.3	Les protéines non structurales.....	40
3.2.3.1	Les protéines de régulation et les protéines accessoires.....	40
3.2.3.1.1	Tat.....	40
3.2.3.1.2	Rev.....	40
3.2.3.1.3	Nef.....	41
3.2.3.1.4	Vpr.....	41
3.2.3.1.5	Vpu.....	41
3.2.3.1.6	Vif.....	42
3.2.3.2	La protéase (PR).....	42
3.2.3.3	L'intégrase.....	42
3.2.3.4	La transcriptase inverse (RT).....	43
3.2.3.4.1	Structure de la RT.....	43
3.2.3.4.2	Propriétés de la RT.....	44
3.2.3.4.2.1	Activité polymérase.....	44
3.2.3.4.2.2	Activité RNase H.....	46
4.	Transcription inverse de l'ARN génomique du VIH.....	49
4.1	La région de fixation de l'amorce ARNt ^(Lys,3)	50
4.2	Formation de l'hybride ARN viral-ARNt ^(Lys,3)	51
4.2.1	Hybridation de l'ARNt ^(Lys,3) à la séquence PBS.....	51
4.2.2	Interactions supplémentaires entre l'ARNt ^(Lys,3) et la région 5'UTR...	53
4.3	Initiation de la transcription inverse.....	54
4.4	Synthèse de l'ADN strong-stop.....	55
4.5	Le premier transfert de brin.....	55
4.6	Synthèse complète du brin (-) de l'ADN viral.....	56
4.7	Synthèse du brin (+) d'ADN à partir des PPT.....	56
4.8	Le second transfert de brin.....	59
4.9	Synthèse de l'ADN viral double-brin possédant deux LTR.....	59
4.10	Le transfert de brin et les recombinaisons génétiques.....	60
4.10.1	Modèles de transfert de brin.....	61
4.10.2	Mécanismes de transfert de brin.....	62
4.10.2.1	Le transfert de brin par invasion.....	63
4.10.2.2	Le transfert pause-dépendant.....	64
4.10.2.3	Le transfert pause-indépendant.....	64
5.	Le premier transfert de brin.....	66
5.1	Rôles de l'activité RNase H et de la NC au cours du premier transfert de brin	66
5.2	Inhibition du premier transfert de brin par les agents pharmacologiques.....	67
5.2.1	Inhibiteurs ciblant la RT.....	67
5.2.2	Inhibiteurs ciblant la NC.....	68
5.2.3	Inhibiteurs ciblant les acides nucléiques.....	68

5.3 Structure des acides nucléiques intervenant dans le premier transfert de brin.....	69
5.3.1 Structures ARN impliquées dans le premier transfert de brin.....	70
5.3.1.1 Repliement de la séquence R.....	70
5.3.1.2 Circularisation du génome par l'interaction TAR-TAR.....	71
5.3.1.3 Circularisation du génome par l'interaction U3- ARNt ^(Lys,3) ...	72
5.3.1.4 Circularisation du génome par l'interaction gag-U3, poly (A)..	73
5.3.2 Structure de l'ADN <i>strong-stop</i> (ADNss).....	73
5.4 Mécanismes intervenant dans le premier transfert de brin.....	74
5.4.1 La circularisation du génome du VIH-1 faciliterait le premier transfert De brin.....	74
5.4.2 Rôles des tige-boucles poly(A) et cpoly(A).....	75
5.4.3 Rôles des tige-boucles TAR et cTAR.....	76
Objectifs.....	79
Matériel et méthodes.....	80
1. Oligodésoxyribonucléotides	81
1.1 Purification des désoxyoligoribonucléotides	81
1.2 Marquage et purification des désoxyoligoribonucléotides.....	82
2. Analyse par retard sur gel de l'interaction NC-oligonucléotide	82
3. Analyse par retard sur gel de l'interaction ADNss-ARN en absence de NC.....	83
4. Analyse par les sondes structurales de l'interaction ADNss-ARN en absence de NC....	83
5. Analyse des cinétiques d'hybridation	84
5.1 Renaturation en absence de NC	84
5.2 Renaturation en présence de NC	85
Résultats et discussion.....	86
1. Analyse structurale de l'ADNss.....	87
2. Analyse par retard sur gel des sites de fixation de la NC.....	88
3. Analyse de la formation de l'hétéroduplexe r-R.....	95
3.1 Analyse en absence de NC.....	95
3.1.1 Essais d'hétérodimerisation de l'ADNss avec l'extrémité 3' du génome..	95
3.1.2 Analyse structurale de l'ADNss en présence de l'extrémité 3' du génome..	96
3.2 Analyse en présence de NC.....	98
3.2.1 Effet du protocole de renaturation dans la cinétique d'hybridation.....	98
3.2.2 Cinétiques d'hybridation d'ADNss mutés.....	100
Conclusions générales et perspectives.....	103
Articles publiés.....	108
Références bibliographiques.....	145

1. Les rétrovirus

1.1 Caractéristiques des rétrovirus

Les rétrovirus sont une famille de virus à ARN monocaténaire, diploïde et de polarité positive. Ces virus causent de nombreuses maladies chez les vertébrés : des désordres neurologiques, des tumeurs, le syndrome d'immunodéficience acquise (SIDA)...

À l'exception du spumavirus qui contient un génome viral à ADN (1), les rétrovirus présentent deux copies d'ARN simple-brin, ce qui facilite les événements de recombinaison génétique.

Ces virus ont la particularité de posséder une enzyme : la transcriptase inverse (RT). Cette dernière rétro-transcrit une molécule d'ARN simple-brin en molécule d'ADN bicaténaire qui est, ensuite, intégrée dans le génome de la cellule hôte par l'intégrase virale (IN). L'ADN bicaténaire viral intégré est désigné par le terme « provirus » : il est assimilé à un gène cellulaire et peut être transcrit en ARN messager (ARNm) ou en ARN génomique (ARNg). Environ 8% du génome humain présente des similarités de séquence avec des rétrovirus endogènes qui joueraient un rôle physiologique ou pathologique. Les rétrovirus endogènes sont des virus intégrés dans le génome humain et transmis à la descendance, dont certains sont des rétrovirus infectieux. Ces derniers possèdent trois gènes caractéristiques : *gag* (code pour les protéines de structure), *pol* (code pour les enzymes virales), *env* (code pour les protéines de surface) et les extrémités du génome viral sont caractérisées par les *Long Terminal Repeats* (LTR) (**Figure 1**)(2).

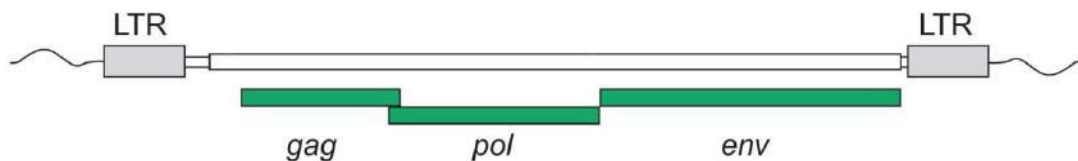


Figure 1. Structure des rétrovirus infectieux (2).

Introduction

1.2 Classification des rétrovirus

La classification a été tirée du comité international de taxonomie des virus (ICTV, <http://www.ictvonline.org/virusTaxonomy.asp>). Une famille est définie par des dénominateurs taxonomiques communs : l'information génétique et les propriétés de réplication. Des éléments de structure comme la présence d'enveloppe et la pathogénicité sont considérés dans certains cas de manière équivalente. Les rétrovirus ne sont pas assignés à un ordre, ils appartiennent à la famille des Retroviridae (Tableau 1). Les rétrovirus infectant les primates sont classés en sept groupes, définis par leur parenté évolutive, chacun avec le rang taxonomique du genre (**Figure 2**). Deux de ces groupes sont les lentivirus et les spumavirus, et cinq de ces groupes représentent des rétrovirus avec un potentiel oncogénique (oncovirus): beta, delta, gamma, epsilon et epsilon-like virus. En se basant sur l'organisation de leur génome, les rétrovirus peuvent être divisés en deux catégories, simple et complexe (3, 4). Tous les oncovirus, à l'exception du *human T-cell leukemia virus* (HTLV) et du *bovine leukemia virus* (BLV), sont des rétrovirus simples. HTLV, BLV, les lentivirus et les spumavirus sont complexes.

Tableau 1. Classification des rétrovirus (<http://www.ictvonline.org/virustaxonomy.asp>)

Famille	Sous-famille	Genre	Espèce
Retroviridae	Orthoretrovirinae	Alpharétrovirus	9 espèces
		Betarétrovirus	5 espèces
		Gammarétrovirus	4 espèces
		Deltarétrovirus	3 espèces
		Epsilon-rétrovirus	17 espèces
	Spumaretrovirinae	Lentivirus	9 espèces : <i>Human immunodeficiency virus 1</i> , <i>Human immunodeficiency virus 2</i> , <i>Bovine immunodeficiency virus</i> , <i>Caprine arthritis encephalitis virus</i> , <i>Equine infectious anemia virus</i> , <i>Feline immunodeficiency virus</i> , <i>Puma lentivirus</i> , <i>Simian immunodeficiency virus</i> et <i>Visna/maedi virus</i>
		Spumavirus	6 espèces

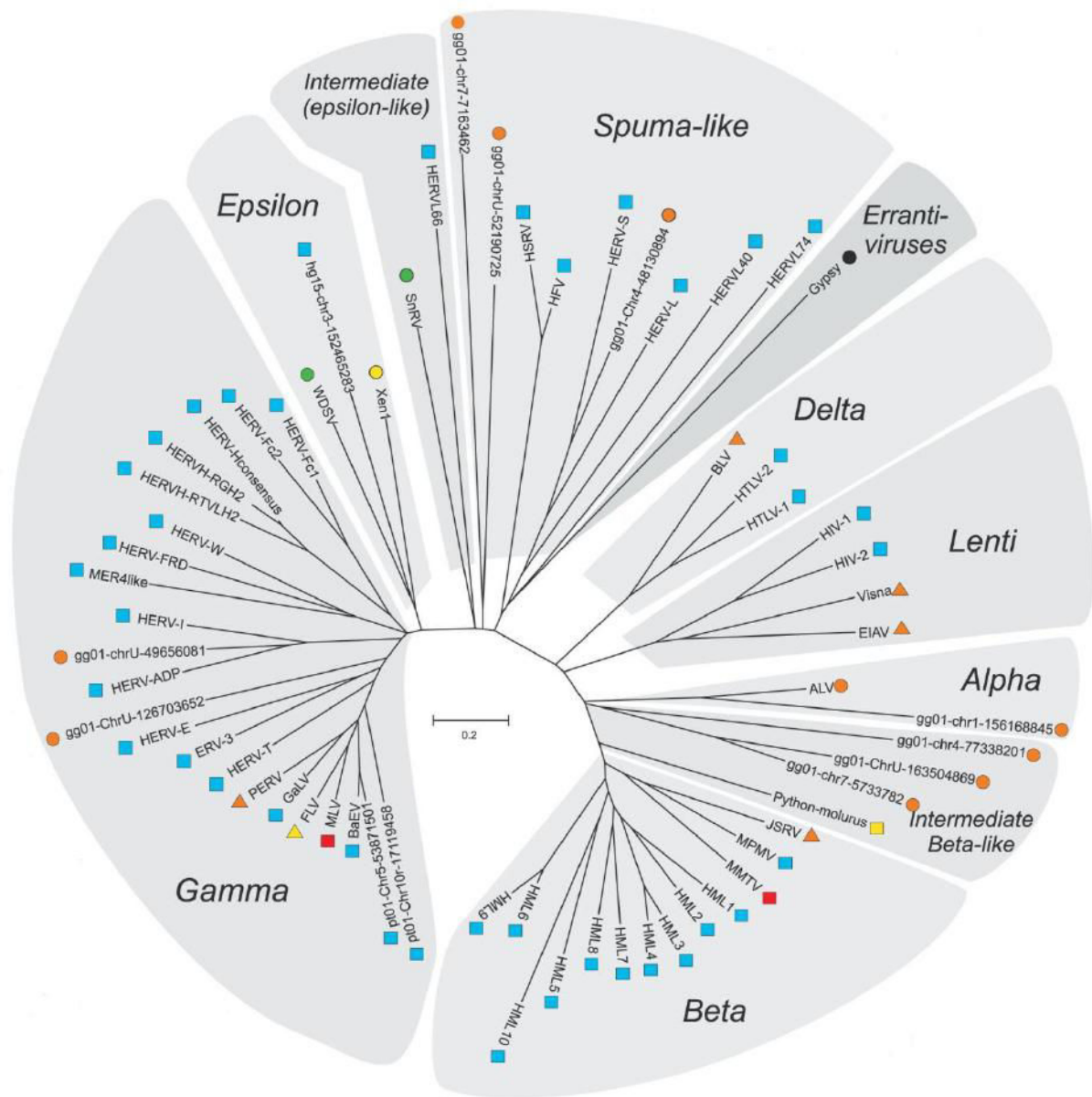


Figure 2. Classification des rétrovirus (4). ALV: Avian Leukosis Virus ; BLV: Bovine Leukemia Virus ; HERV: Human Endogenous Retrovirus ; HFV: Human Foamy Virus ; HIV: Human Immunodeficiency Virus ; HTLV: Human T-Lymphotropic Virus ; MLV: Murine Leukemia Virus ; MMTV: Mouse Mammary Tumor Virus ; WDSV: Walleye Dermal Sarcoma Virus.

2. Connaissances générales sur le VIH

2.1 Le VIH, agent causal du syndrome d'immunodéficience acquise (SIDA)

Le virus de l'immunodéficience humaine (VIH) est un rétrovirus qui a pour principale cible les lymphocytes T CD4+. En 1981, aux Etats Unis, une importante augmentation du taux de maladies infectieuses comme la pneumonie causée par *pneumocystis carinii*, a été constatée dans la communauté homosexuelle. A ces maladies était associée une forte diminution du taux de lymphocytes CD4+ dans le sang périphérique. Le rétrovirus a été isolé pour la première fois, en 1983 par Françoise Barré-Sinoussi et Luc Montagnier, à l'institut Pasteur (5). Ces derniers avaient observé une activité transcriptase inverse à partir de l'isolat d'un patient atteint de lymphadenopathie et le virus a ainsi été appelé *lymphadenopathy associated-virus* (LAV). En 1984, l'équipe de Robert Gallo a isolé un rétrovirus nommé *Human T-Lymphotropic Virus* (HTLV-III) décrit comme étant responsable du SIDA. L'équipe de Jay A. Levy a aussi identifié des rétrovirus nommés *AIDS-associated retrovirus* (ARV) (6). En 1985, la communauté scientifique a conclu que les trois rétrovirus étaient identiques. En 1986, un nouveau virus a été isolé de deux malades d'Afrique de l'Ouest, il a aussi pour cible les lymphocytes CD4+ et il induit un effet cytopathogène similaire au LAV-1, il a donc été appelé LAV-2. Cependant, il se distingue du LAV-1 par la séquence de son génome et les protéines qu'il code. Le virus LAV-2 présente une parenté antigénique au niveau de la protéine externe d'enveloppe avec *Simian T-Lymphotropic Virus* (STLV-III) qui est un rétrovirus isolé du macaque rhésus atteint d'immunodépression (7). En 1986, LAV-1 et LAV-2, ont été respectivement nommés VIH-1 et VIH-2.

En 2015, il a été estimé que 36.7 millions de personnes vivaient avec le VIH, dont 2,1 millions étaient des personnes nouvellement infectées (**Figure 3**).

Introduction

Global summary of the AIDS epidemic | 2015

Number of people living with HIV in 2015	Total	36.7 million	[34.0 million – 39.8 million]
	Adults	34.9 million	[32.4 million – 37.9 million]
	Women (15+)	17.8 million	[16.4 million – 19.4 million]
	Children (<15 years)	1.8 million	[1.5 million – 2.0 million]

People newly infected with HIV in 2015	Total	2.1 million	[1.8 million – 2.4 million]
	Adults	1.9 million	[1.7 million – 2.2 million]
	Children (<15 years)	150 000	[110 000 – 190 000]

AIDS deaths in 2015	Total	1.1 million	[940 000 – 1.3 million]
	Adults	1.0 million	[840 000 – 1.2 million]
	Children (<15 years)	110 000	[84 000 – 130 000]

WHO – HIV department | June 15, 2016



World Health Organization



UNAIDS



unicef

Figure 3: Rapport général du World Health Organization sur l'épidémie du VIH/SIDA en Juillet 2016 (<http://www.who.int/hiv/data/en/>).

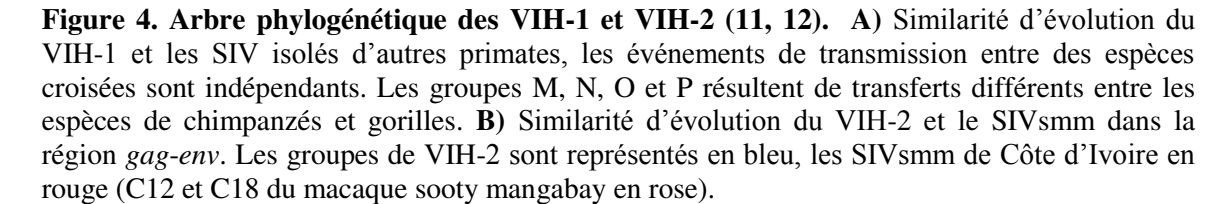
2.2 Origine du VIH

Chez les primates, le *simian immunodeficiency virus* (SIV) est un lentivirus qui attaque le système immunitaire. Des recherches menées, en 1999, sur une souche de SIV provenant du chimpanzé (SIVcpz) ont montré une forte similarité avec le VIH-1 (8).

En se basant sur la classification phylogénétique des lentivirus infectant les primates, il a été montré que les VIH-1 et VIH-2 proviennent de SIVs différents, ils ont donc des origines et une évolution différente (**Figure 4 et Tableau 2**). En effet, le VIH-1 descendrait du SIVcpz qu'on retrouve chez certaines sous-espèces de chimpanzé, principalement en Afrique centrale et la région équatoriale occidentale (9). En revanche, le VIH-2 descendrait du SIVsmm infectant le sooty mangabey qui vit en Afrique de l'ouest (10). Le passage des SIVs à l'homme semble s'être fait par consommation de la viande de brousse, la chasse ou encore la domestication de singes.

Tableau 2. Comparaison entre les espèces de VIH

Especies	Virulence	Infectivité	Prévalence	Origine
VIH-1	Forte	Forte	Global	Chimpanzé
VIH-2	Faible	Faible	Afrique de l'ouest	Sooty mangabey



2.2.1 VIH-1

Le VIH de type 1 est divisé en quatre groupes M, N, O et P, dont l'épidémiologie varie considérablement. En effet, le groupe M est le plus pandémique (environ 33 million d'individus infectés dans le monde). Les groupes M et N ont pour origine le SIVcpz retrouvé chez le chimpanzé commun (*pan troglodytes*) en Afrique occidentale (8). Le Groupe P, nouvellement découvert, et le groupe O présentent des similarités avec le SIVgor, identifié chez le gorille sauvage vivant au Cameroun (13). Il est possible que ces deux groupes dérivent de souches du SIVgor par transmission indépendante, du gorille à l'homme. Le groupe O, infecte quelques dizaines de milliers de personnes au Cameroun, Gabon et Guinée-équatoriale. Le groupe P a été identifié chez deux individus originaires de Yaounde au Cameroun (14). L'analyse d'un prélèvement sanguin effectué en 1959 à Kinshasa (République Démocratique du Congo) a démontré que l'infection à VIH existait déjà à cette époque. Il est probable que le VIH-1 (groupe M) s'est propagé à partir de Kinshasa des années 1920 (15). Le groupe M du VIH-1 qui est le plus étudié en laboratoire, regroupe 11 sous-types (A1, A2, B, C, D, F1, F2, G, H, J et K) et 48 formes circulaires recombinantes (CRFs) (**Figure 5**). Les isolats les plus étudiés en laboratoire sont HXB2, NL4-3, LAI (sous-type B) et MAL (recombinant des sous-types A1, D et K) (<http://www.hiv.lanl.gov/>).

2.2.2 VIH-2

Le VIH de type 2 est assez répandu en Afrique occidentale car il représente 10 à 20% des personnes infectées par le VIH. Il est décrit comme descendant du SIVsmm infectant le sooty mangabey (16). Par comparaison avec le VIH-1, les personnes séropositives au VIH-2 présentent une période de latence plus longue, une progression de la pathologie plus lente et une plus faible charge virale au cours du stade asymptomatique (17). Le VIH-2 est divisé en 8 groupes nommés de A à H (18), les groupes A et B étant les plus représentés. De nos jours, on a noté une diminution de la prévalence du VIH-2 dans la majorité des pays.

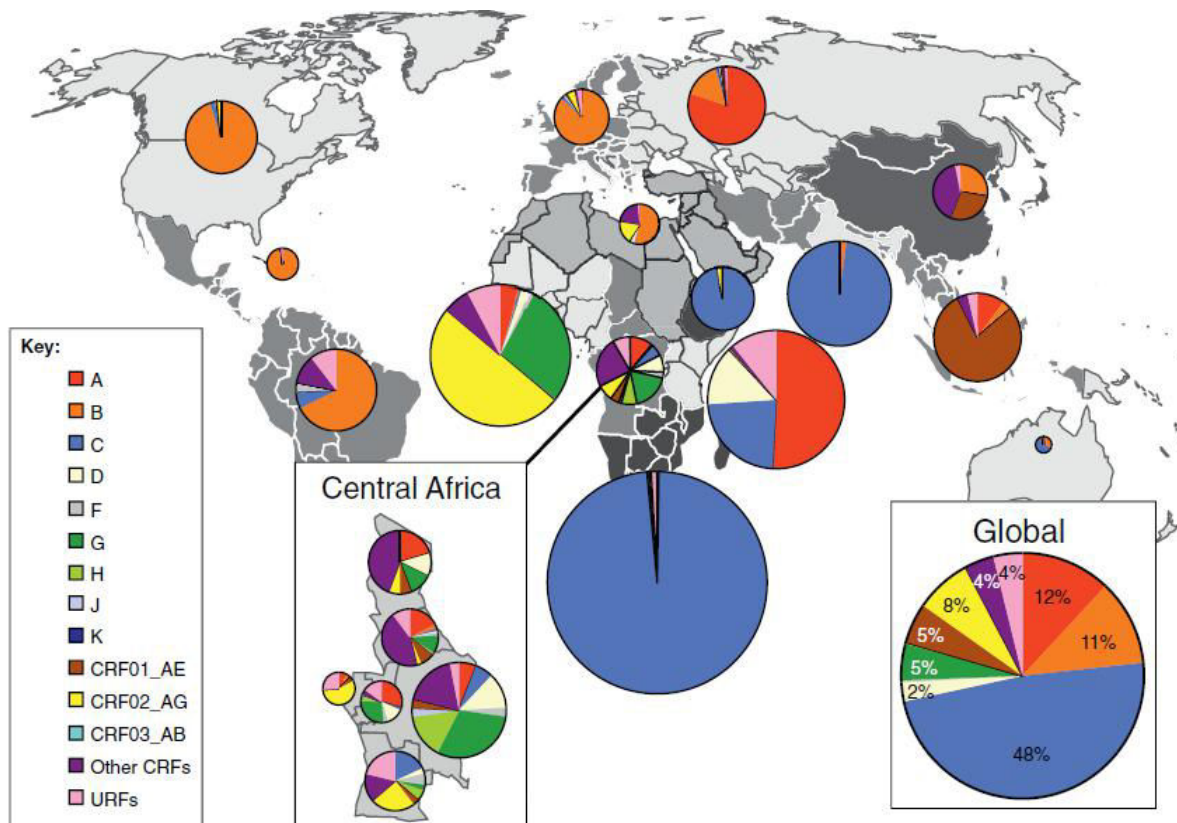


Figure 5. Distribution globale des sous-types du groupe M du VIH-1 et les formes recombinantes circulaires (CRFs) (19).

2.3 Pathogénèse du VIH

Les principales cibles du VIH sont les cellules du système immunitaire comme les lymphocytes T CD4+, les macrophages et les cellules dendritiques (DC). L'infection induite par le VIH conduit à une chute du taux des cellules T CD4+, principalement via trois mécanismes majeurs : i) le rétrovirus détruit des cellules infectées ; ii) un fort taux de cellules infectées entre en apoptose ; iii) les lymphocytes cytotoxiques CD8 reconnaissent les CD4+ infectée et les détruisent. Par conséquent, il se produit un déclin du nombre de lymphocytes T CD4+ en dessous du seuil critique, ce qui entraîne une perte d'immunité. Progressivement, l'organisme devient une cible pour les infections opportunistes. Ces différents stades d'infection par le VIH sont résumés dans la **Figure 6**.

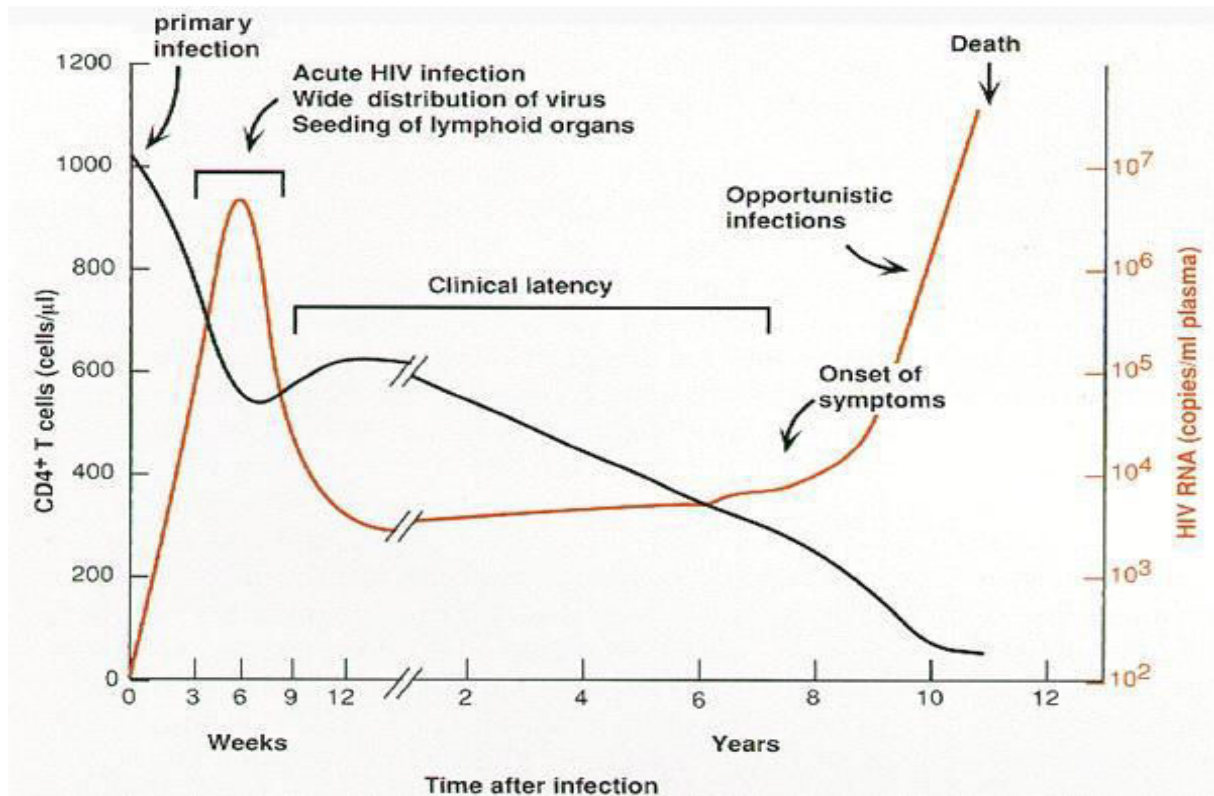


Figure 6. Evolution typique de l'infection par VIH-1 (adapté de 20).
 Courbe noire : évolution du taux des lymphocytes T CD4+ de la personne infectée.
 Courbe orange : évolution de la charge virale (copie d'ARN viraux) circulant dans le sang.

2.4 Caractéristiques du VIH-1

2.4.1 Structure de la particule virale

La particule virale du VIH-1 est de forme sphérique avec un diamètre d'environ 145 nm, le cœur de la particule virale mature présente une forme conique, contenant l'ARNg viral et les protéines nécessaires à la réplication (21).

Le VIH-1 est un virus enveloppé (**Figure 7**) dont l'enveloppe externe est composée d'une bicouche lipidique qui est enrichie en protéines d'enveloppe (Env). Celle-ci est constituée de trois glycoprotéines (gp120) formant une coiffe et de trois protéines transmembranaires (gp41) ancrées dans la membrane virale. La partie interne du virus forme une coque nommée capsid (CA). La capsid contient le génome du virus qui est constitué de deux molécules d'ARN en étroite association avec les protéines de nucléocapsid (NC) et les enzymes virales (protéase, transcriptase inverse et intégrase). La matrice est principalement constituée de protéines de matrice (MA) et est située entre l'enveloppe et la capsid (**Figure 7**).

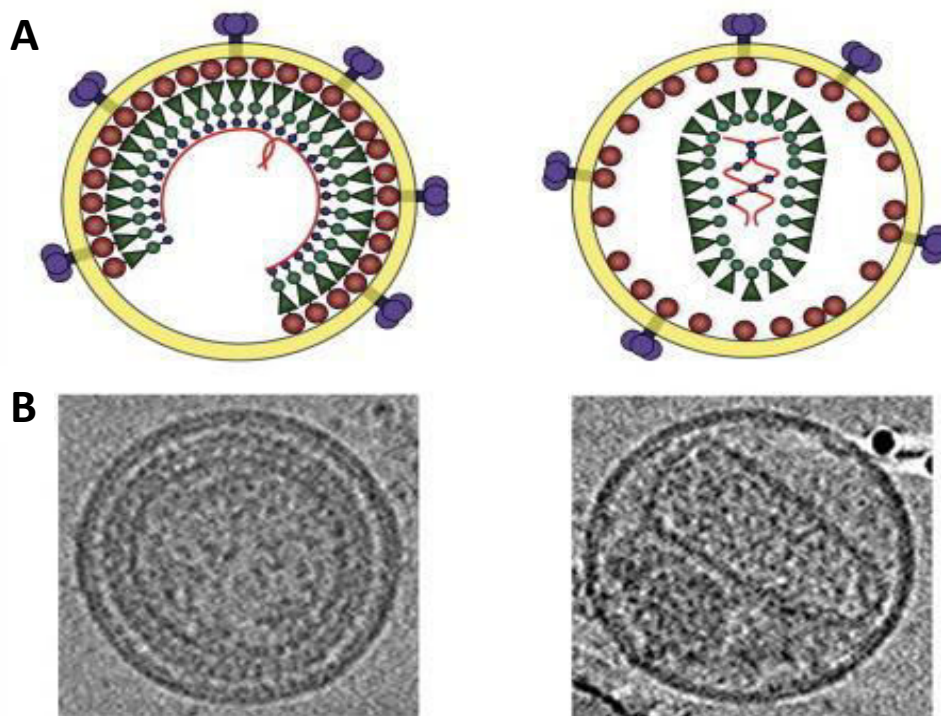


Figure 7. Structure des particules virales immature et mature du VIH-1 (22). A) organisation schématique des particules virales immature et mature. B) Observation des particules virales mature et immature par cryotomographie électronique. En mauve, gp120 ; en jaune, bicouche lipidique ; en rouge, MA ; en vert, CA ; en bleu, NC ; filament rouge, ARNg.

2.4.2 Organisation génétique du VIH-1

L'ARN viral est constitué de six marqueurs structuraux (TAR, RRE, Psi, SLIP, CRS, et INS) (<http://www.hiv.lanl.gov/>), et neuf gènes (*gag*, *pol*, *env*, *tat*, *rev*, *nef*, *vif*, *vpr*, *vpu*) codant pour 15 protéines différentes (**Figure 8**) (23). L'existence d'un dixième gène présent sur le brin antisens de l'ADN viral, au niveau du gène *env*, a récemment été montré, ce gène code pour la protéine ASP (*Anti-Sens Protein*) dont la fonction reste inconnue (24).

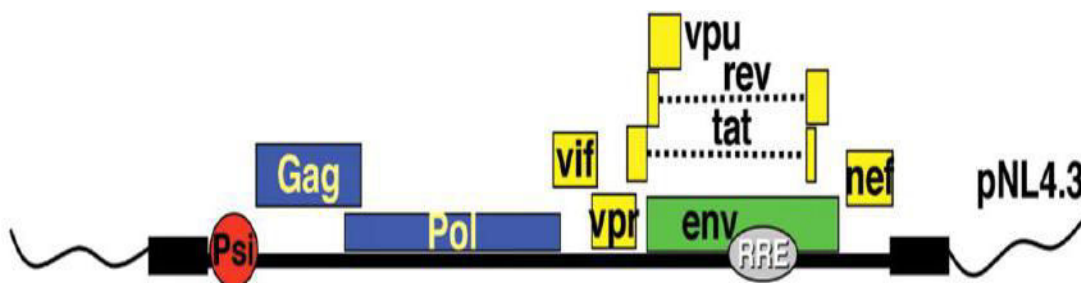


Figure 8. Organisation génétique du VIH-1 (25). ADN pré-intégré du VIH-1 (pNL4-3). Les rectangles noirs correspondent aux extrémités 5' et 3' LTR, les rectangles en couleur correspondent aux cadres de lecture ouverts. Le signal d'encapsidation (Psi) en rouge et le *rev-responsive element* (RRE) en gris y sont également représentés.

Introduction

TAR (*Transactivator response element*) est le site de liaison de la protéine Tat et de protéines cellulaires, il est constitué par les 57 premiers nucléotides de l'ARN viral du VIH-1. L'ARN TAR forme une tige-boucle contenant une boucle interne qui est le site de fixation de la protéine Tat (26).

RRE (*Rev responsive element*) est un domaine ARN contenu dans la région env du VIH-1. Il correspond à 203 nucléotides (position 7327 à 7530) englobant les extrémités des séquences gp120 et gp41). Le RRE est nécessaire à l'export des ARNm épissés et non épissés par fixation de la protéine Rev. Il contient environ 7 sites de fixation pour Rev (27).

Psi (*Packaging signal*) ou region psi, est constitué de quatre structures en tige-boucle (SL1-SL4). SL1 est la tige-boucle décrite comme site d'initiation de la dimérisation de l'ARN viral, SL2 contient le site donneur d'épissage. La tige-boucle SL3 a longtemps été considérée comme le principal élément d'encapsidation, en revanche il a été montré récemment que SL1 serait plus importante lors de la fixation de Gag et la sélection du génome viral (25, 28–31). Quant à la tige-boucle SL4, elle contient le codon d'initiation AUG, en revanche sa structure est instable.

SLIP (*Slippery site*) est un domaine contenant le motif TTTTTT suivi d'une structure en tige-boucle, il régule le décalage du cadre de lecture du ribosome en -1, générant ainsi le précurseur polyprotéique Gag-Pol qui permet la synthèse des enzymes suivantes : protéase, transcriptase inverse et intégrase (32, 33).

CRS (*cis-acting responsive sequences*) sont des séquences décrites comme inhibitrices de l'expression des protéines de structure en absence de la protéine Rev. Un seul site a été cartographié dans la région *pol* du VIH-1. A ce jour, la fonction exacte de CRS reste méconnue. Par ailleurs, certains sites d'épissage ont été décrits agissant comme des séquences CRS (32).

INS (*inhibitory/instability RNA sequences*) sont retrouvées dans les gènes de structure du VIH-1 ou d'autres rétrovirus. Les éléments INS multiples existent dans le génome et peuvent agir indépendamment. Des essais fonctionnels ont défini les INS comme éléments inhibant l'expression génique de manière post-transcriptionnelle (32, 34, 35).

2.5 Cycle réplcatif du VIH-1

Le cycle réplcatif du VIH-1 (**Figure 9**) s'effectue en deux phases majeures :

- Phase précoce : 1) attachement à la cellule hôte et désassemblage du virion ; 2) transcription inverse de l'ARNg ; 3) intégration de l'ADN viral bicaténaire.

Introduction

- Phase tardive : 1) transcription et épissage de l'ARNg ; 2) export des ARN viraux ; 3) synthèse des protéines virales, empaquetage de l'ARNg et assemblage de la particule virale ; 4) bourgeonnement et maturation de la particule virale.

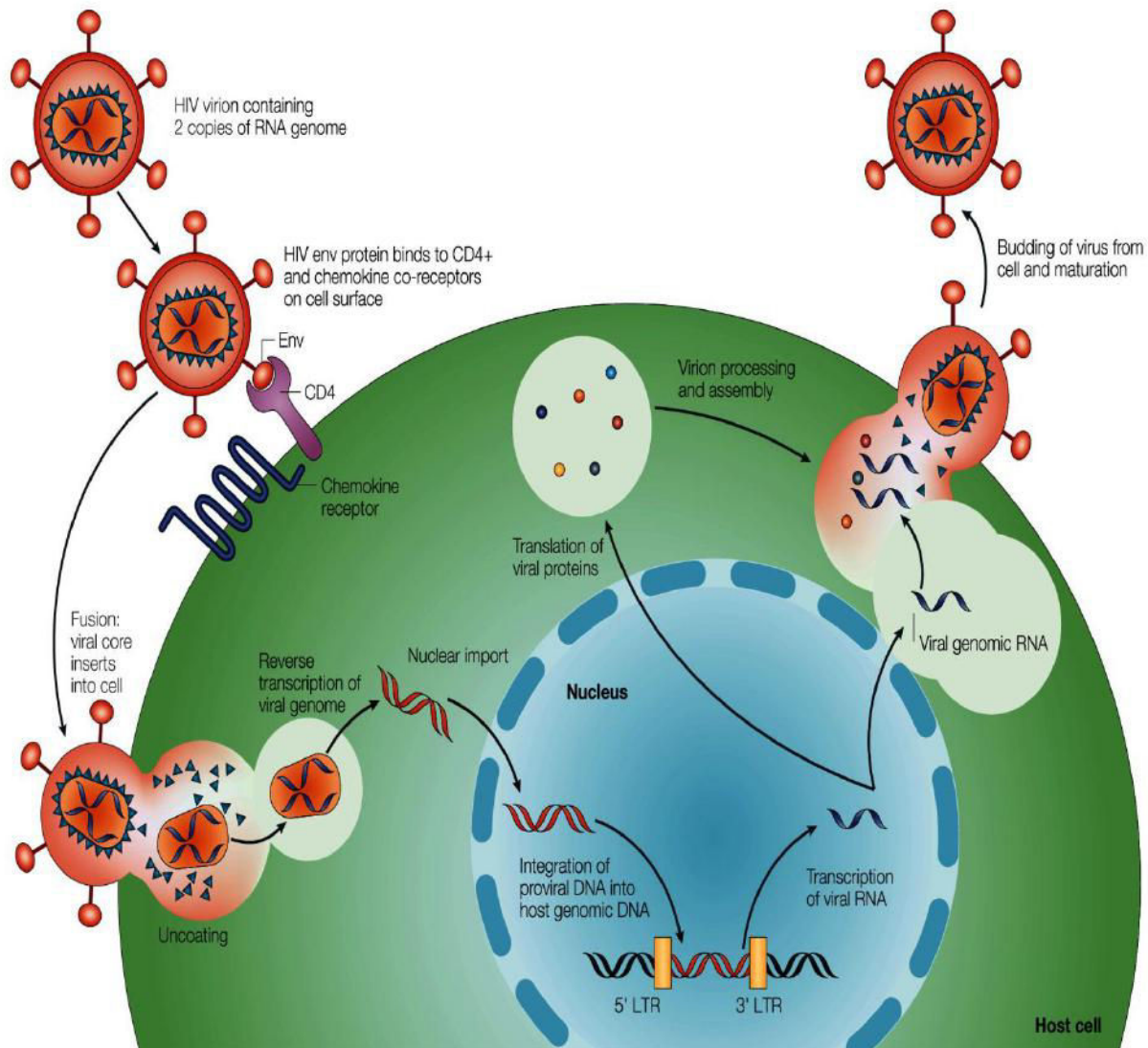


Figure 9. Cycle réplcatif du VIH-1 (9).

2.5.1 Entrée du VIH-1

Les cellules dendritiques (DC) sont les premières cellules que le virus rencontre au niveau des muqueuses lors de la transmission par voie sexuelle. Le VIH-1 peut infecter les DC qui expriment le récepteur CD4 et un co-récepteur de la famille des récepteurs à la chimiokine soit CCR5 ou CXCR4 (36). Elles sont soupçonnées de jouer un rôle important en transmettant le virus aux cellules T. Le VIH-1 pénètre dans la cellule par fusion membranaire de manière dépendante ou indépendante du pH, ce qui entraîne le relargage de la particule virale dans le cytoplasme. Par ailleurs, des études suggèrent que l'entrée et la fusion du VIH-1

Introduction

à la membrane de la cellule hôte est suivie d'une endocytose (36, 37). La principale protéine virale intervenant lors de l'entrée du virus est la protéine d'enveloppe Env qui fait 160 kDa lorsqu'elle est glycosylée (une protéine de surface gp120 et une protéine transmembranaire gp41 (38). La liaison à la cellule hôte est initiée par des interactions du complexe trimérique de l'enveloppe (la protéine gp120) et les récepteurs CD4-CCR5 ou CD4-CXCR4 à la surface cellulaire (**Figure 10A**) (37, 39). La première étape de fusion se fait par liaison du domaine gp120 au récepteur CD4 par une liaison de très haute affinité, ce qui entraîne un changement structural du complexe néoformé (**Figure 10B**). Ce réarrangement structural permet l'exposition de la boucle V3 de la gp120, ce qui conduit à la liaison au récepteur à la chimiokine (**Figure 10B et 10C**) (37, 39). Ces réseaux de liaisons formées sont très stables, ce qui induit la pénétration du domaine N-terminal de la protéine gp41 dans la membrane cellulaire (**Figure 10D**) (37, 39). Les séquences *heptad repeat* (HR) de la gp41 en interagissant, forment une structure en boucle (**Figure 10E**). Ainsi, les membranes cellulaire et virale sont à une très faible distance, ce qui conduit à leur fusion et à l'entrée de la capside virale (**Figure 10F**) (37, 39). Après la fusion, la capside contenant l'ARNg viral et les enzymes virales (RT, IN et PR) est injectée à l'intérieur de la cellule (37, 39).

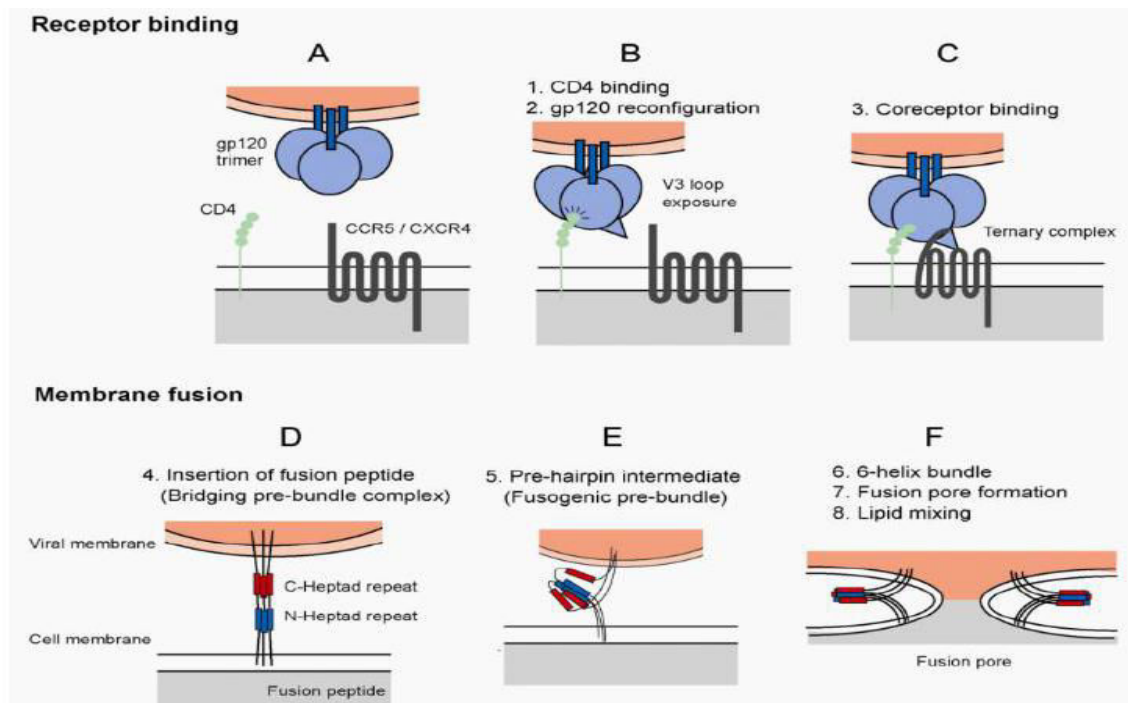


Figure 10. Entrée du VIH-1 dans la cellule hôte (36).

2.5.2 Transcription inverse de l'ARN viral

La transcription inverse débute principalement après l'entrée dans la cellule de la capside lorsque cette dernière est partiellement désassemblée (décapsidation). La transcription inverse est décrite en détail dans le chapitre 4.

2.5.3 Intégration du génome viral dans l'ADN de la cellule hôte

En fin du processus de transcription inverse, l'ADN viral double-brin est transloqué dans le noyau. L'ADN bicaténaire néo-synthétisé fait partie d'un large complexe nucléoprotéique nommé complexe de pré-intégration (PIC). Ce dernier inclut la présence de l'intégrase virale fixée aux deux extrémités 5' et 3' de l'ADN du VIH-1 (**Figure 11**) (38). Une fois intégré l'ADN peut rester en état de dormance lors de la phase de latence de l'infection par le VIH-1 (38). Une production optimale du virus nécessite la présence de certains facteurs de transcription comme le NF- κ B, qui est régulé positivement lorsque les cellules T sont activées (40, 41). Ce qui implique que les cellules détruites par le VIH-1 sont celles qui combattent l'infection.

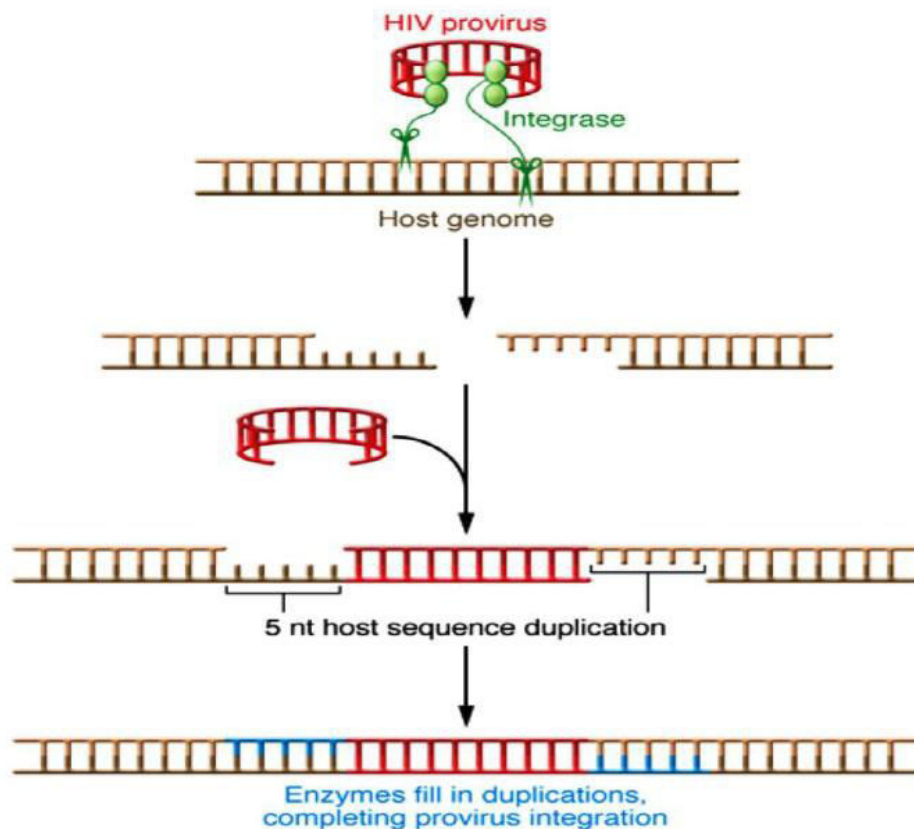


Figure 11. Intégration de l'ADN viral dans le chromosome de la cellule hôte (42).

Introduction

2.5.4 Synthèse et maturation de l'ARN viral

Au cours de la phase tardive du cycle de réplication du VIH-1, l'ADN intégré est transcrit en ARNm. La majorité des souches présente cinq différents sites donneurs d'épissage et huit différents sites accepteurs pour produire au moins 109 ARN épissés dans les cellules infectées (**Figure 12**) (43).

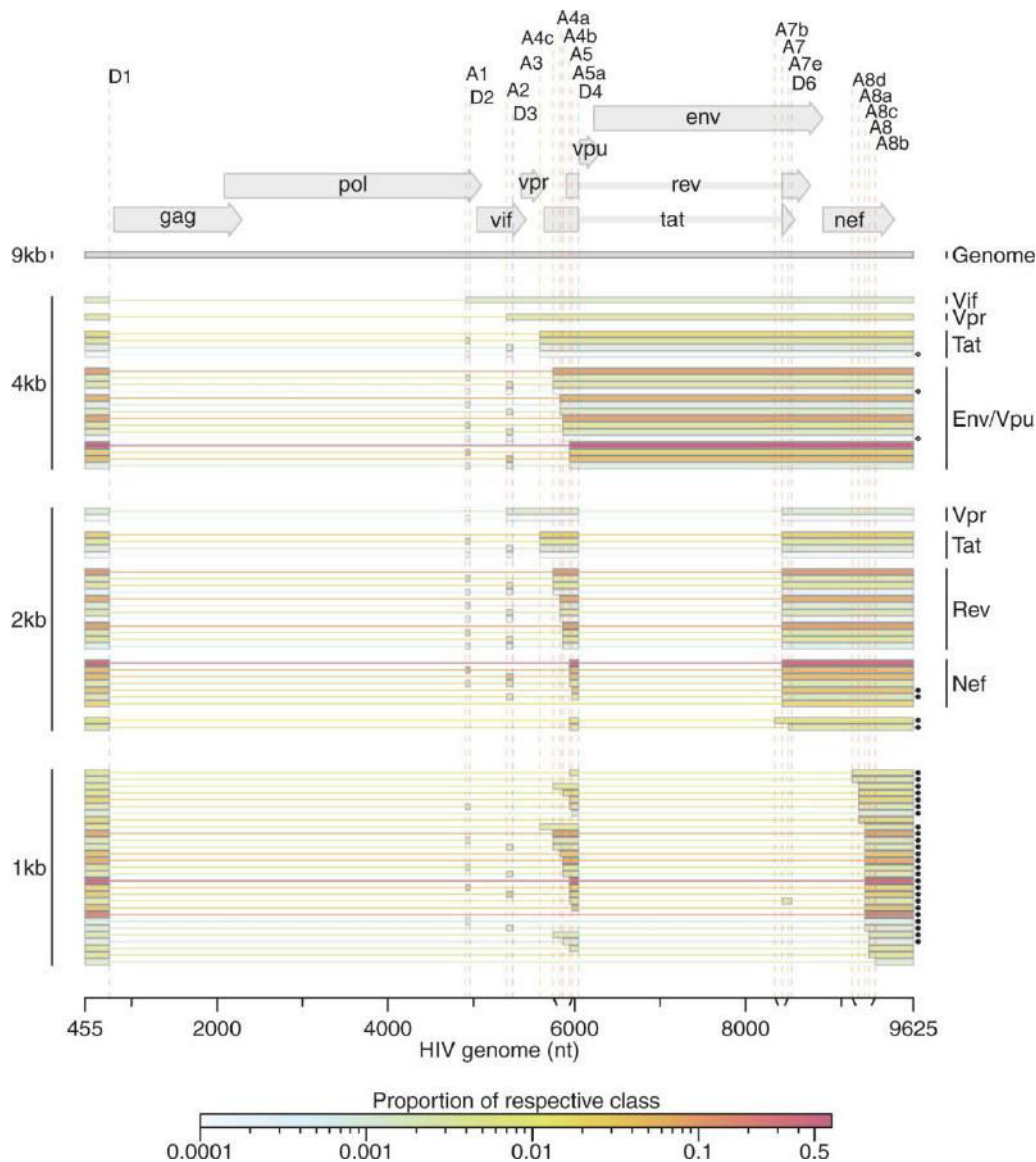


Figure 12. Localisation des sites d'épissage chez le VIH-1 (43). Organisation du génome du VIH-1 et les différents ARNm produits. Les sites donneur (D1-D4, D6) et accepteur (A1-A8) d'épissage sont indiqués. Les transcrits sont classés par taille (le génome non épissé, partiellement épissé ou 4 Kb, complètement épissé ou 2 Kb et une dernière classe d'ARN complètement épissés ou 1 Kb). Les exons sont représentés par des traits épais et les introns par des traits fins. Les protéines obtenues à partir de transcrits hautement conservés sont représentées. L'abondance relative à chaque transcrit au sein de sa classe constitutive est indiquée par une couleur suivant échelle affichée.

Introduction

La maturation de l'extrémité 3' et la polyadénylation des pré-ARNm du VIH-1 impliquent la reconnaissance du motif AAUAAA et des motifs riches en GU qui sont présents dans la séquence R, celle-ci étant dupliquée au niveau des extrémités 5' et 3' de l'ARN viral (**Figure 13**). Le VIH-1 utilise de multiples éléments de régulation pour générer une extrémité 3' polyadénylée. En effet, la séquence U3 du VIH-1 qui se situe en amont du site de polyadénylation en 3', contient des séquences activatrices (*upstream enhancer elements*) (USE) qui facilitent la liaison au facteur CPSF (*cleavage polyadenylation specificity factor*). Ce facteur optimise la polyadénylation à l'extrémité 3' des transcrits du VIH-1(44). D'autre part, il a été montré que l'USE situé à proximité de l'extrémité 5' près du gène Nef, se fixe aux protéines SR 9G8 cellulaires, qui recrutent les facteurs d'épissage CFIm et CPSF (45). De plus, la polyadénylation préférentielle de l'extrémité 3' par rapport à l'extrémité 5' de l'ARN viral s'effectue grâce au facteur d'épissage U1 snRNP. Ce dernier inhibe la reconnaissance du motif AAUAAA en 5' en se fixant sur le site d'épissage D1 (**Figure 13**). Une mutation du site D1 qui affaiblit la liaison de U1 snRNP conduit à l'utilisation du signal de polyadénylation en 5'(46).

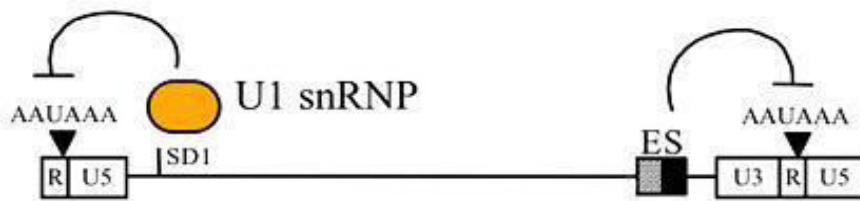


Figure 13. Régulation de la polyadénylation (47). La reconnaissance du signal de polyadénylation en 5' est empêchée par l'interaction de U1 snRNP avec D1. Le site de polyadénylation en 3' est régulé par l'action inhibitrice compétitive de ESS' (S) et l'action de ESE (E).

L'activité transcriptionnelle du VIH-1 est très faible en l'absence de la protéine virale Tat. En revanche, en présence de Tat et des facteurs cellulaires la transcription de l'ARN viral est fortement augmentée. Tat agit en se fixant sur l'élément TAR qui est situé dans la région R (48, 49). L'ARN pré-messager du VIH-1 subit de nombreuses combinaisons d'épissage pour produire les ARNm des protéines accessoires (Vpr, Vpu, Vif et Nef), de régulation (Tat et Rev) et de structure (MA, CA et NC) (50). Les ARNm faisant entre 1,8 à 2 kb en taille sont principalement polycistroniques et produisent préférentiellement Tat, Rev et Nef (**Figure 14**), selon le site accepteur d'épissage utilisé (51). Ces ARNm sont constitutivement exportés vers le cytoplasme et traduits. Au cours d'une phase plus tardive qui nécessite la présence de la protéine Rev, les ARN non épissés ou partiellement épissés

Introduction

(4kb) sont exportés vers le cytoplasme pour produire les protéines Gag, Pol, Env, Vif, Vpr et Vpu (**Figure 14**). L'export de ces ARN vers le cytoplasme nécessite l'interaction de la protéine Rev avec la séquence RRE et des facteurs cellulaires (46). Certains aspects de l'export des ARN viraux restent méconnus, notamment le choix du type d'export. En effet, deux voies d'export des ARN viraux ont été décrites : i) une voie constitutive à la cellule dépendante des facteurs cellulaires TAP/NXF1 via le complexe TREX (*transcription/export*) qui se fixe aux ARN épissés (2kb) ; ii) une voie qui fait intervenir la protéine cellulaire Crm1 (*Chromosome maintenance-1*) qui en reconnaissant la séquence NES (*nuclear export signal*) interagit avec Rev/RRE, ce qui crée un large complexe ribonucléoprotéique capable d'effectuer le transport des ARN partiellement ou non épissés (respectivement, 4 kb et 9 kb) vers le cytoplasme (27, 52). La voie constitutive à la cellule serait constamment inhibée par la voie Rev/RRE/Crm1 par un mécanisme qui n'est pas élucidé (52).

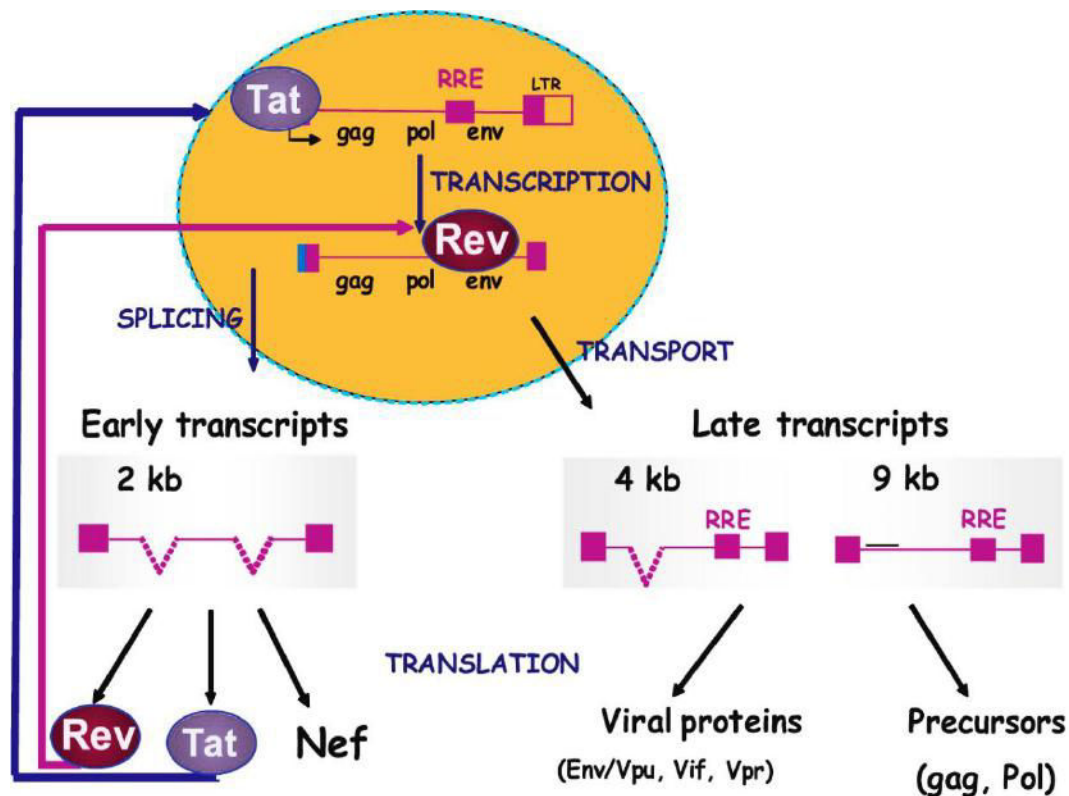


Figure 14. Expression des ARNm du VIH-1 (53). Rev et Tat sont produits à partir de transcrits qui sont les premiers à être exportés vers le cytoplasme et à être traduits. Rev et Tat entrent dans le noyau pour agir respectivement au niveau de la transcription, et l'export des ARNm non épissés ou partiellement épissés. Les transcrits non épissés ou partiellement épissés présentent un site de liaison (RRE) pour Rev conduisant à l'export à partir du noyau.

2.5.5 Encapsidation de l'ARNg dans la particule virale

Au cours de la phase tardive du cycle réplcatif, le VIH-1 sélectionne efficacement deux copies d'ARNg, non épissés, coiffées en 5' et polyadénylées en 3' par le mécanisme d'encapsidation qui a été très étudié (54–56). Ce mécanisme permet une incorporation préférentielle de l'ARNg dimérique dans le virion par rapport aux ARNm viraux épissés et aux ARNm cellulaires (57). Le précurseur Gag est l'acteur principal du processus d'encapsidation. En cas d'absence de l'ARNg viral, l'assemblage de la particule virale s'effectue en incorporant un taux équivalent d'ARN cellulaires (58). La sélection du génome viral s'effectue par l'interaction directe du domaine NC du précurseur Gag avec la région psi (ψ) qui est constituée des tiges-boucles SL1-SL3 (**Figure 15**).

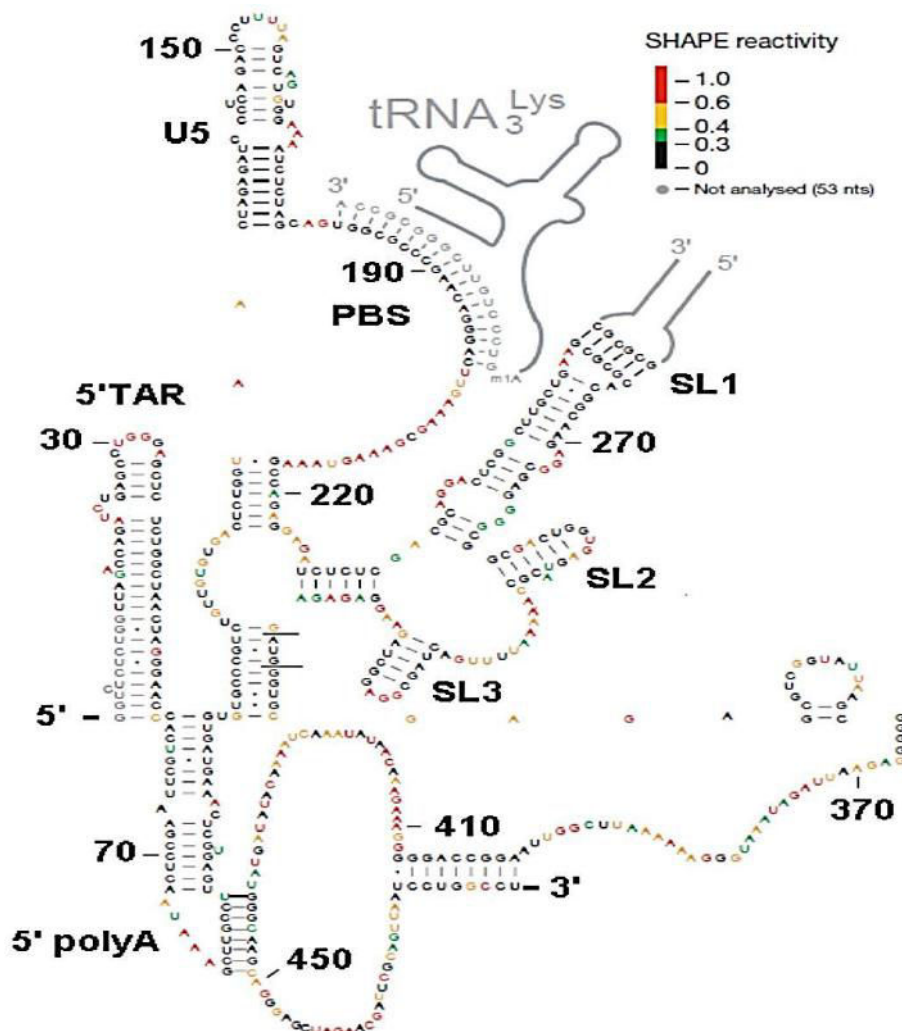


Figure 15. Modèle de structure secondaire de l'extrémité 5'UTR de l'ARNg du VIH-1 (59). Le code couleur indique le niveau de réactivité des nucléotides vis-à-vis d'un agent acylant (NMIA ou 1M7) qui modifie les nucléotides flexibles.

Introduction

Similairement aux autres rétrovirus, les nucléotides qui participent à la sélection du génome se situent dans la région 5'UTR du génome viral (25, 60). Par ailleurs, de courts éléments situés dans la région 5' UTR effectuent, de manière indépendante, la sélection d'ARN hétérologues dans des pseudo-particules virales en cours d'assemblage (*Virus-like particles* VLPs). Ces dernières ont été identifiées chez certains rétrovirus. Cependant, le VIH-1 semble nécessiter plus que la région 5' UTR (61). En effet, une séquence nucléotidique en aval de l'extrémité 5'UTR et se situant dans la région codante gag est nécessaire pour une encapsidation optimale (62–64). L'extrémité 5' UTR est la région la plus conservée du génome viral (65). De plus, cette région intervient dans la régulation et l'activation de la transcription, l'épissage, la fixation de l'amorce ARNt lors de la transcription inverse, la dimérisation de l'ARN viral. Par ailleurs, le site d'initiation de la dimérisation (SL1) semble être aussi déterminant pour l'encapsidation de l'ARN viral. L'interaction U5-AUG a été identifiée comme jouant un rôle important dans la régulation de l'encapsidation de l'ARN viral (56, 66, 67). Il a été proposé que la séquence du site donneur d'épissage (SL2) contribue à cette régulation *in vivo* (68). De plus, d'autres recherches ont montré que la tige-boucle poly(A) de l'extrémité 5' est un déterminant important dans l'encapsidation d'ARN viraux épissés et non épissés (61). Deux modèles ont été proposés pour l'encapsidation d'ARN épissés et non épissés (**Figure 16**).

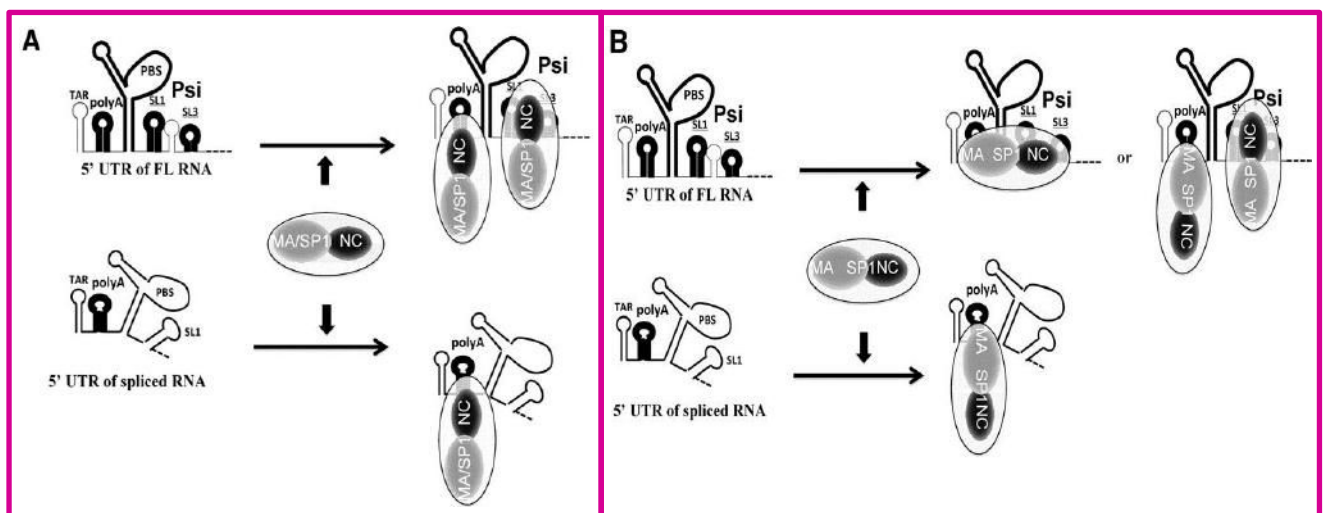


Figure 16. Modèles proposés pour l'encapsidation selective d'ARN épissés et non épissés (60). (A) L'ARN g non épissé (FL RNA) incluant la région psi formé par les tiges-boucles SL1 et SL3 est reconnu par le domaine NC de Gag. Ainsi, l'ARNg peut fixer plus de précurseurs Gag que les ARN épissés. (B) Modèle alternatif, le domaine MA/SP1 de gag se fixe, avec une interaction faible, sur la tige-boucle poly(A), tandis que, le domaine NC se fixe fortement sur la région psi. Ces deux types d'interactions Gag -ARN expliquent comment l'ARNg non épissé est plus compétitif que les ARN épissés, lors de l'assemblage de la particule viral. Ce domaine interagit aussi avec la tige-boucle poly(A) présente dans les ARN épissés et non épissés.

Une étude récente propose un troisième modèle pour l'encapsidation sélective (**Figure 17**) (30).

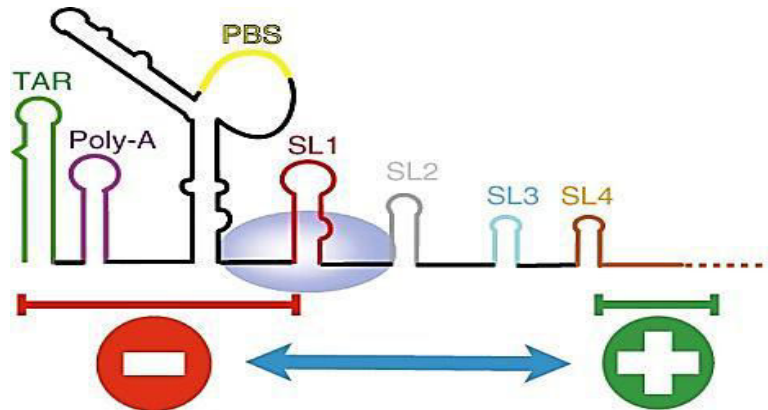


Figure 17. Troisième modèle d'encapsidation sélective (30). La fixation du précurseur Gag sur la partie inférieure de la tige-boucle SL1 est inhibée par la séquence en amont des ARN épissés. Cette régulation négative est contrebalancée par une courte séquence située en 3' de la tige-boucle SL4 dans l'ARNg.

2.5.6 Assemblage et bourgeonnement de la particule virale

Le génome joue un rôle structural important lors de l'assemblage de la particule virale (56, 61), cependant ce processus peut aussi se produire avec des ARNs cellulaires (69). L'assemblage de nouveaux virions débute au niveau de la membrane plasmique de la cellule hôte (**Figure 18**). La protéine Env (gp160) passe par le réticulum endoplasmique (RE) puis elle est transportée vers l'appareil de Golgi où elle est clivée par une protéine cellulaire en deux glycoprotéines, gp120 et gp41. Par la suite, ces dernières sont transportées vers la membrane plasmique et gp41 ancre gp120 à la membrane des cellules infectées. Les précurseurs protéiques Gag (Pr55) et Gag-Pol (Pr160) sont exprimés à partir de l'ARNg non épissé. Ils sont relargués puis se fixent à la membrane plasmique par l'extrémité N-terminale myristoylée. L'accumulation adéquate de précurseurs associés à l'ARNg induit l'assemblage du virion sous la membrane de la cellule hôte (22). De nombreux aspects de l'assemblage du virion comme le bourgeonnement et le relargage ne sont que partiellement compris. De nombreux facteurs cellulaires semblent aussi contribuer à ce processus (70). La formation de la NCp9 est concomitante de celle du peptide p6. Ce peptide est nécessaire au bourgeonnement de la particule virale, au moment du relargage de la membrane cellulaire (71). Le peptide p6 contient une région riche en prolines et un domaine di-leucine qui sont les cibles, respectivement, des protéines cellulaires Tsg101 et ALIX. Ces dernières sont impliquées dans le processus d'export cellulaire par les protéines de classe E et la machinerie nécessaire pour le bourgeonnement du VIH-1 (71–74).

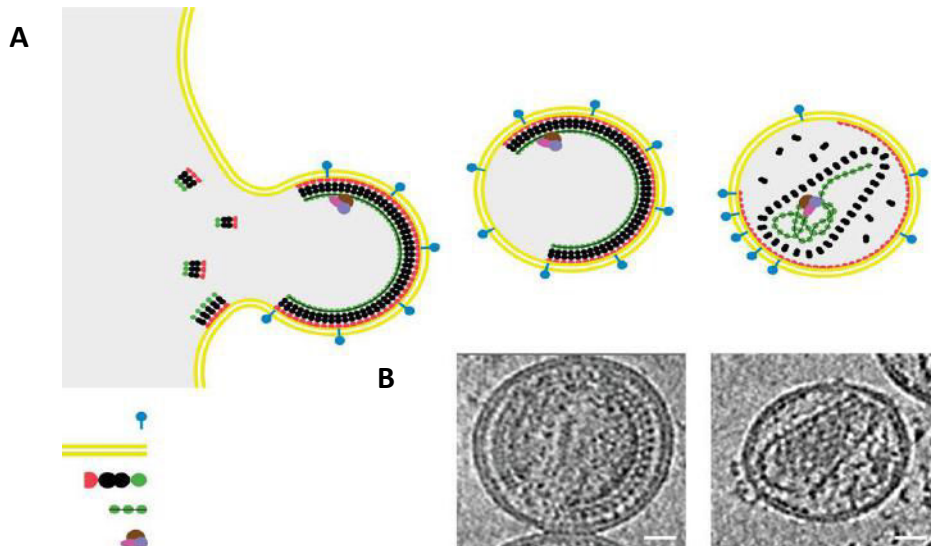


Figure 18. Assemblage de la particule virale (75). **A)** illustration schématique du processus d'assemblage et de bourgeonnement de la particule virale. MA (rouge), CA (noir), NC (vert), glycoprotéines (bleu clair), membrane (jaune) et protéines accessoires (violet, marron et rose). **B)** reconstitution tomographique d'une particule virale immature et mature.

2.5.7 Maturation de la particule virale

Le VIH-1, comme tous les autres rétrovirus possède deux copies de son matériel génétique (76). Plus précisément, le génome viral se présente sous la forme d'un dimère dont les sous-unités sont liées de manière non covalente dans les particules virales immatures et matures. La stabilité du dimère d'ARN dans la particule virale dépend de la maturation de celle-ci (77, 78). Après bourgeonnement, la particule virale n'est pas immédiatement infectieuse. Elle acquiert sa capacité d'infection lors de la maturation, un processus initié par la protéase virale (PR) qui clive les précurseurs Gag (Pr55) et Gag-Pol (Pr160) (79–81). Le mécanisme reste encore mal compris en ce qui concerne l'activation de la PR. La protéase virale va cliver les précurseurs Gag et Gag-Pol en différentes protéines et enzymes fonctionnelles. La maturation du virion semble être initiée pendant ou immédiatement après le relargage de la particule virale (79, 80, 82). L'étape de maturation permet le passage de la forme non infectieuse du virus à sa forme infectieuse. Environ 1000 à 1500 copies du précurseur Gag sont nécessaires à la formation d'un virion infectieux (23, 83). Gag est constitué de 6 protéines et peptides selon l'ordre suivant de l'extrémité N-terminale à l'extrémité C-terminale : MA-CA-sp1-NC-sp2-p6. Gag présente cinq sites de clivages protéolytiques ciblés par la PR virale (**Figure 19**). Le clivage de Pr55 est séquentiel car les taux de clivage des cinq sites ne sont pas équivalents (84). Le clivage entre SP1 et NC (SP1-NC) est le premier à se produire, tandis que les clivages entre SP2-p6 et MA-CA se

Introduction

produisent en second. Les deux sites restants CA-SP1 et NC-SP2 sont les derniers à être clivés (**Figure 19**).

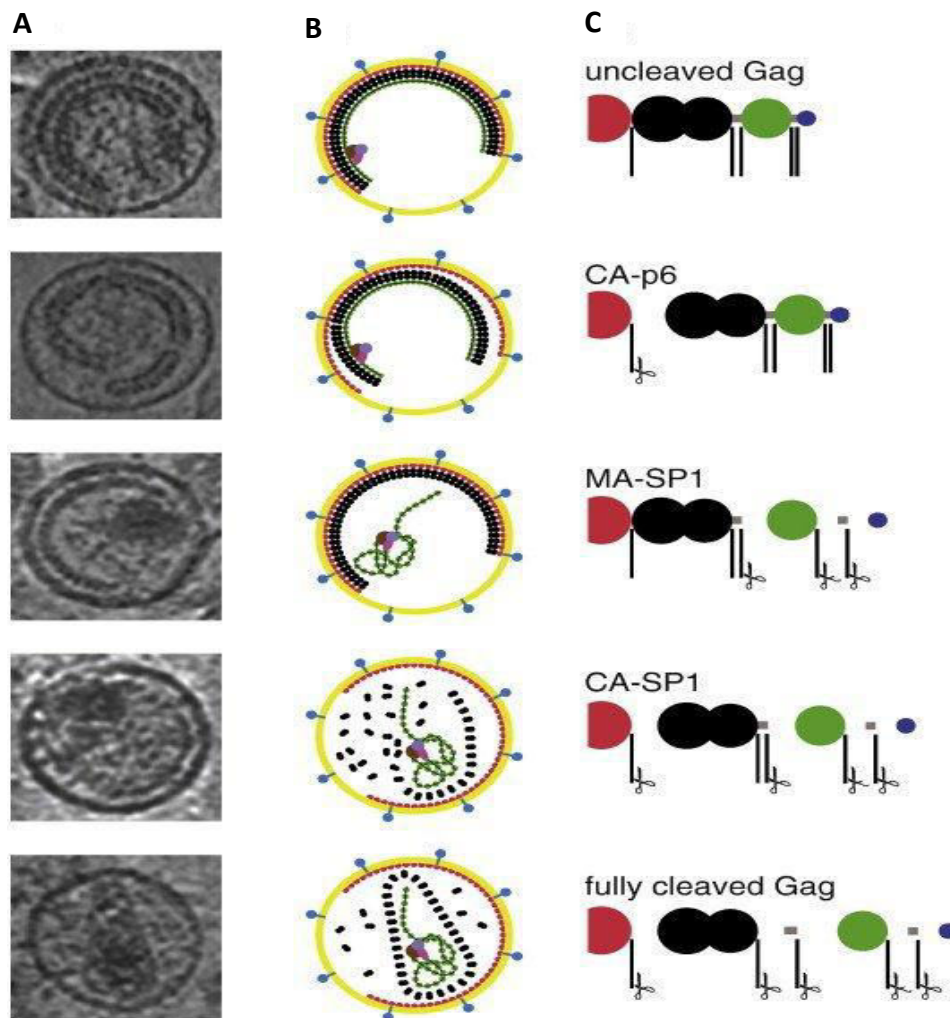


Figure 19. Les changements structuraux de la particule virale du VIH-1 au cours de la maturation (85). Les intermédiaires structuraux ont été générés par mutation individuelle des sites de clivage, chaque intermédiaire est illustré par : **A**) reconstitution tomographique de la particule virale. **B**) illustration schématique de la morphologie de la particule virale. MA (rouge), CA (noir), NC (vert), p6 (bleu), SP (gris), glycoprotéines (bleu clair), membrane (jaune) et protéines accessoires (violet, marron et rose). **C**) Schéma montrant l'état de clivage de la polyprotéine Gag en fonction des sites de clivage mutés ou non.

Le clivage protéolytique en SP1-NC conduit à séparer le domaine MA-CA du domaine NC associé à l'ARNg. Lors de la protéolyse, la NC existe sous deux formes intermédiaires, NCp15 (produit de clivage partiel comprenant les domaines NC-SP2-p6) et la NCp9 (produit de clivage partiel contenant les domaines NC-SP2). La forme finale NCp7 résulte du clivage total. Ces trois protéines présentent une activité chaperonne des acides nucléiques (86). Une

Introduction

étude récente suggère fortement que l'intégrase virale joue un rôle dans la maturation. Il a été montré qu'une mutation pouvant altérer la fixation de IN à l'ARNg conduit à la formation de particules virales non infectieuses et immatures (87).

3. Les protéines intervenant dans le cycle réplcatif du VIH-1

3.1 Les protéines cellulaires

De nombreuses protéines cellulaires interviennent dans le cycle réplcatif du VIH-1 (**Figure 20**). Le processus débute lorsque les glycoprotéines d'enveloppe se fixent au complexe de récepteurs CD4-CCR5 (étape 1), ce qui conduit à la fusion des membranes virale et cellulaire et entrée de la capside virale dans la cellule (étape 2). Le désassemblage partiel de la capside virale (étape 3), est associé à la transcription inverse (étape 4), ce qui donne le complexe de pré-intégration (PIC). A la suite de l'importation dans le noyau cellulaire (étape 5), l'intégrase associée au PIC coordonne l'intégration de l'ADN nouvellement synthétisé, à l'aide du *lens epithelium-derived growth factor* (LEDGF), une protéine cellulaire fixant la chromatine (étape 6). La transcription de l'ADN viral intégré (étape 7) est médiée par l'ARN polymérase II et le facteur *positive transcription elongation factor b* (P-TEFb). L'ARN non épissé qui est l'ARNg, a besoin de la protéine cellulaire CRM1 pour être exporté du noyau (étape 8). Le bourgeonnement et le relargage des particules virales sont médiés par les complexes ALIX et ESCRT (étapes 11 et 12). Chaque étape du cycle de réplication du VIH-1 est une cible potentielle pour les thérapies anti-rétrovirales (88).

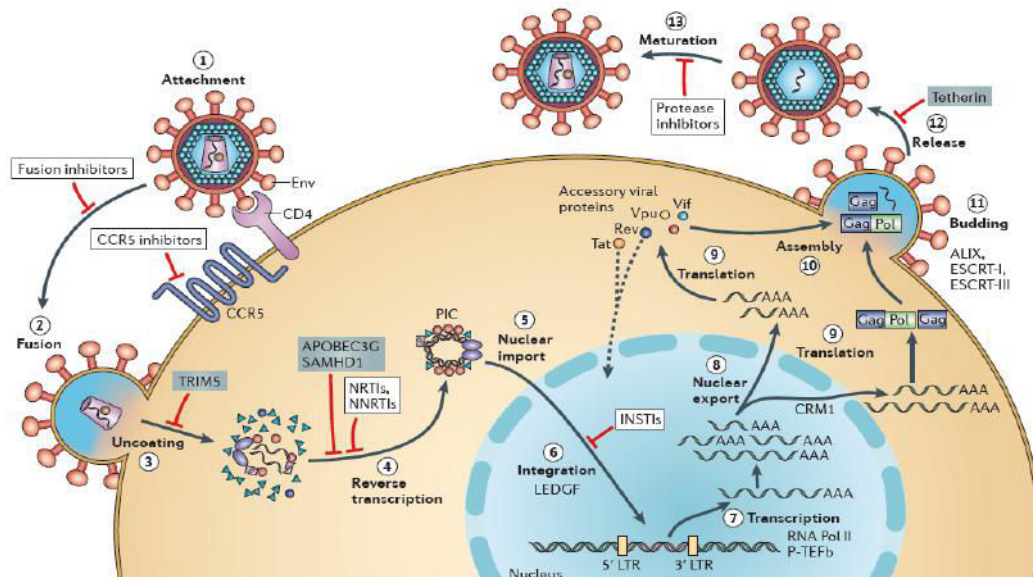


Figure 20. Rôle des protéines cellulaires dans le cycle réplcatif du VIH-1 (88). Les sites d'action des inhibiteurs cliniques et les facteurs cellulaires de restriction sont indiqués respectivement par des rectangles blancs et bleus.

Introduction

3.1.1 TRIM5

Le VIH-1 infecte l'homme et les chimpanzés, mais pas les *Old World Monkeys* (OWMs) tels que le singe rhésus (Rh) et le singe cynomolgus (CM) (89). Le VIH-1 entre efficacement dans les cellules d'OWMs mais il est cependant bloqué avant la transcription inverse. Il a été montré que la résistance à l'infection par le VIH-1 est dominante dans les hétérokaryons entre les cellules de l'homme et des OWM, ce qui suggère la présence d'un ou plusieurs facteurs inhibant l'infection des cellules OWM par le VIH-1 mais pas par le SIV (90). En 2004, le criblage d'une banque d'ADNc d'un Rh a permis d'identifier le *tripartite motif 5α* (TRIM5α) en tant que facteur antiviral cellulaire spécifique à certaines espèces (90) (**Tableau 3**). TRIM5α, présent dans le cytoplasme, reconnaît des motifs de la capsid et interfère avec l'étape de désassemblage de la capsid (décapsidation) (**Figure 21**), empêchant ainsi la transcription inverse et le transport du génome viral vers le noyau (89, 91).

Tableau 3. Restriction spécifique par TRIM5α selon les espèces (92).

TRIM5α	HIV-1	SIVmac	SIVagm	N-MLV	B-MLV	
Human	No	No	No	Yes	No	« Yes » désigne la restriction. « Weak » désigne une faible restriction. « No » désigne l'absence de restriction. « N.D. » désigne l'absence de résultats publiés. SIVmac, <i>simian immunodeficiency virus</i> isolé du macaque ; SIVagm, <i>simian immunodeficiency virus</i> isolé du <i>African green monkey</i> ; NMLV, <i>N-tropic murine leukemia virus</i> ; B-MLV, <i>B-tropic murine leukemia virus</i> .
Rhesus monkey	Yes	No	Yes	Weak	No	
Cynomolgus monkey	Yes	No	N.D.	N.D.	N.D.	
AGM (tantalus)	Yes	Yes	No	Yes	No	
AGM (pygerythrus)	Yes	No	No	Yes	No	
Squirrel monkey	No	Yes	Weak	No	No	
Owl monkey (TRIMCyp)	Yes	No	N.D.	No	No	

TRIM5α est l'un des variants d'épissage produit par le gène trim5. Les protéines TRIM5 font partie de la famille TRIM contenant les domaines *RING B-box 2* et *coiled-coil*. Les domaines RING possèdent une activité E3 ubiquitine ligase (93). Il a été démontré que TRIM5α forme un dimère (94, 95). Chez Le singe Rh, le domaine *B-box 2* peut s'auto-oligomériser formant un large réseau de TRIM5α oligomériques (96, 97). Le domaine *coiled-coil* de TRIM5α est important pour la formation d'homo-oligomères et est essentiel pour l'activité antivirale du TRIM5α (98). Parmi les variants d'épissage TRIM5, TRIM5α est le

Introduction

seul à posséder un domaine supplémentaire PRYSPRY C-terminal. TRIM5 α reconnaît la protéine de la capside (CA) multimérique d'un virus entrant, par son domaine PRYSPRY (96).

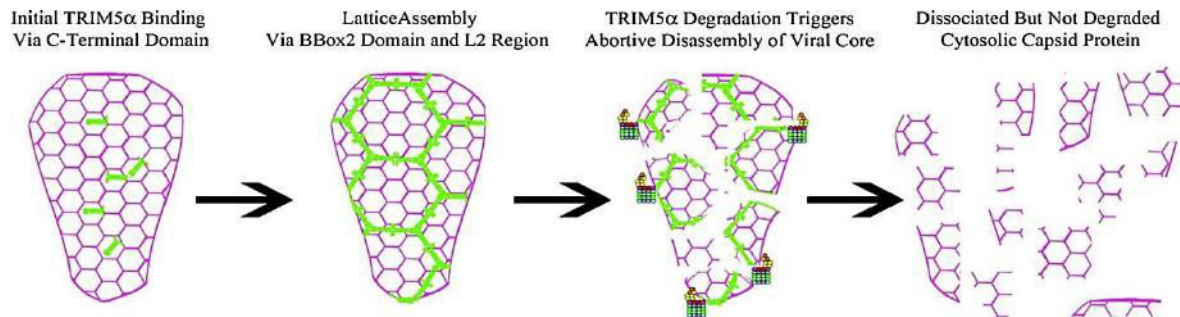


Figure 21. Modèle proposé pour la restriction rétrovirale par TRIM5 α (99). Suivant l'entrée de la capside (rose) dans le cytoplasme de la cellule hôte, les protéines TRIM5 α (vert) se fixent aux protéines de capside (CA) (rose). L'assemblage des TRIM5 α dimériques conduit à la formation d'une maille hexamérique entourant la capside. Ensuite, les protéines TRIM5 α sont dégradées par le protéasome ce qui déclenche un désassemblage abortif de la particule virale. Ainsi, il y a dissociation et non dégradation des protéines de capside.

3.1.2 APOBEC3

La famille de protéines *Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3* (APOBEC3) de mammifères a émergé comme un médiateur clé de la restriction intrinsèque des rétrovirus comprenant le VIH et le virus de l'hépatite B (100–102). Les protéines APOBEC3 font partie de la classe d'enzymes cytidine déaminase qui sont incorporées dans des virions. L'enzyme APOBEC3 contenant deux domaines en doigt de zinc édite l'ADN viral nouvellement synthétisé par déamination des cytidines (dC) en uraciles (dU), entraînant une hyper-mutation de G à A qui est létale. L'activité antivirale de APOBEC3 peut également être deaminase indépendante par interférence avec la transcription ou l'intégration de l'ADN viral (103, 104). En revanche, le VIH-1 détourne l'activité des enzymes cellulaires APOBEC3G et 3F sous l'action de son facteur d'infectivité virale (Vif). Ce dernier déclenche la dégradation d'APOBEC3G et 3F via la voie d'ubiquitination faisant intervenir des protéines cellulaires telles que : Cullin5, Elongins B et C, et Rbx1 (105). Les sept gènes de la famille APOBEC3 (A, B, C, DE, F, G et H) sont situés en tandem sur le chromosome 22q12-q13.2 qui couvre 150 kb. Leur activité anti-VIH et la sensibilité à la dégradation médiée par Vif est variable (**Tableau 4**).

Introduction

Tableau 4. Le profil anti-VIH-1 des protéines APOBEC3 (106).

Gene	Anti-HIV-1	mRNA expression in CD4+ T cells	Resistance to HIV-vif
APOBEC3A	No	Low, IFN- α inducible	No
APOBEC3B	Strong	Low	Complete
APOBEC3C	Weak	No	No
APOBEC3D/E	Weak	No	No
APOBEC3F	Strong	High	Partial
APOBEC3G	Strong	High	No
APOBEC3H	Strong/weak according to variant	High	Yes/no according to variant

Parmi toutes les protéines APOBEC3, APOBEC3G présente l'effet antiviral le plus élevé. L'impact des niveaux d'expression d'APOBEC3G sur la sévérité de l'infection au VIH-1 a été très étudié. La plupart des études ont rapporté une corrélation positive entre les niveaux d'expression et l'inhibition du virus (107). Une forte expression de APOBEC3G dans les cellules mononuclées du sang périphérique (PBMC) et les tissus du col utérin a été observée chez les femmes séronégatives exposés au VIH ; les PBMC sont également résistantes à l'infection par la souche macrophagotropic (R5) du VIH-1 (108). Les niveaux d'ARNm d'APOBEC3G sont inversement corrélés à la charge virale et présentent une corrélation positive avec le taux de cellules T CD4+ (109). Les observations indiquent que la variation génétique de APOBEC3 et son taux d'expression modulent le degré d'infection VIH-1 et la progression de la pathologie.

Dans les cellules infectées par le Δ Vif VIH-1, APOBEC3G n'est pas dégradé, il est incorporé dans la particule virale et transféré à la cellule cible suivante, où il exerce ses effets antiviraux à de multiples niveaux (110). APOBEC3G peut inhiber la polymérisation par la RT d'une manière déaminase indépendante (**Figure 22-1**). En effet, APOBEC3G se lie directement à l'ARN viral empêchant la RT de se déplacer le long de l'acide nucléique. De plus, APOBEC3G peut déclencher une déamination massive de dC à dU lors de la synthèse du brin (-) d'ADN viral (**Figure 22-2**). Ainsi, lors de la synthèse du brin (+) d'ADN, les adenosines sont incorporés à la place des guanines, entraînant une hypermutation de G à A. APOBEC3G inhibe également l'intégration de l'ADN viral, probablement, en induisant des défauts de clivage au niveau de l'ARNt, au cours du deuxième transfert de brin. Ces défauts de clivage conduisent à la formation d'un ADN viral aux extrémités modifiées par rapport à celles obtenues en présence du Vif sauvage (**Figure 22-3**). APOBEC3G interagit également avec IN, ce qui pourrait nuire à l'intégrité du complexe d'intégration, entraînant une diminution du taux d'intégration.

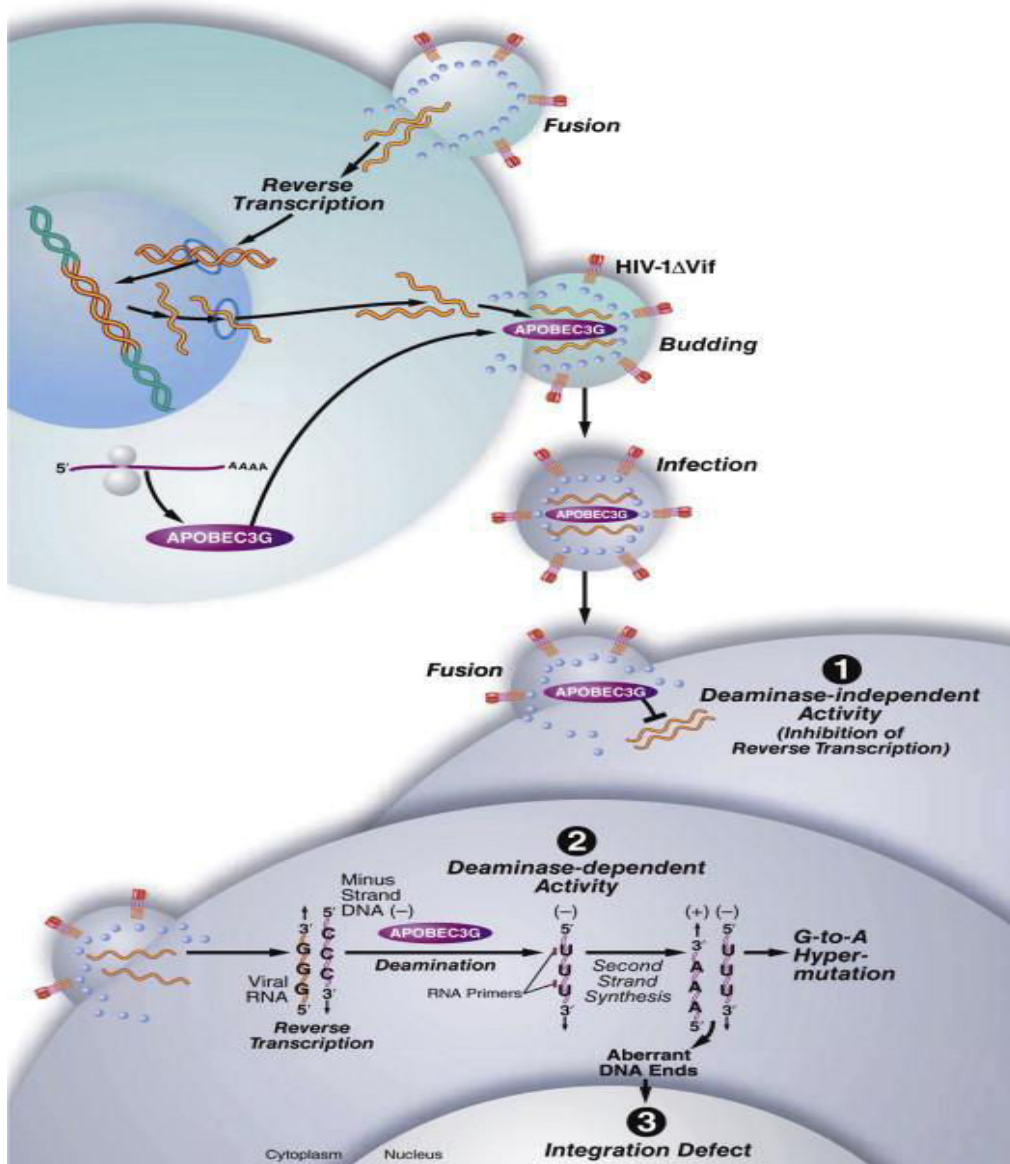


Figure 22. Action de APOBEC3G sur le cycle réplcatif du Δ Vif VIH-1 (111).

3.1.3 Autres facteurs cellulaires intervenant dans le cycle réplcatif

De nombreuses autres protéines cellulaires ont été décrites comme ayant un potentiel anti-rétroviral. La *Sterile alpha motif histidine-aspartate domain containing protein 1* (SAMHD1) est une phosphohydrolase qui catalyse la conversion des désoxyribonucleosides triphosphates (dNTP) en désoxyribonucleoside et triphosphates inorganiques (PPi). SAMHD1 est considérée comme un facteur antiviral du VIH-1 qui contrôle le taux de dNTP dans les cellules immunitaires en hydrolysant les dNTP (112, 113). Il a été suggéré que SAMHD1 inhibe plusieurs étapes du processus de réplication virale (114). SAMHD1 serait la cible du facteur Vpx présent chez certains rétrovirus comme le SIVmac ou encore le VIH-2 (113). Une étude effectuée *in cellulo* a montré que dans les virions du SIVmac dépourvus de Vpx,

Introduction

l'activité antivirale de SAMHD1 est traduite par une forte réduction du taux cellulaire de dNTP. Cette activité est inhibée en présence de Vpx qui déclenche une dégradation par le protéasome (113).

Endosomal sorting complexes required for transport (ESCRT) est un complexe de protéines cellulaires impliquées dans des processus d'exocytose, de réparation de la membrane plasmique et de bourgeonnement des particules virales (115). Parmi les protéines du complexe ESCRT, ALIX et TSG101 sont des facteurs cellulaires qui sont recrutés par le domaine p6 de Gag pour favoriser le bourgeonnement (116, 117). D'autre part, en se liant au précurseur Gag par son domaine p6 (118), une étude récente suggère que l'activité antivirale d'ESCRT au cours du bourgeonnement réside dans l'inactivation de la protéase virale (119).

3.2 Les protéines du VIH-1

Les protéines virales (**Tableau 5**) sont synthétisées à partir de plus d'une centaine d'ARNm différents dérivant tous d'un transcrit (**Figure 12**).

Tableau 5. Les protéines du VIH-1 (source : <http://www.hiv.lanl.gov/>)

Name	Size	Function	Localization
Gag			
MA	p17	membrane anchoring, <i>env</i> interaction, nuclear transport of viral core (myristylated protein)	virion
CA	p24	core capsid	virion
NC	p7	nucleocapsid, binds RNA	virion
	p6	binds Vpr	virion
Pol			
Protease (PR)	p15	Gag/Pol cleavage and maturation	virion
Reverse Transcriptase (RT)	p66, p51	reverse transcription, RNase H activity	virion
RNase H	p15		virion
Integrase (IN)	p31	DNA provirus integration	virion
Env	Gp120/gp41	external viral glycoprotein binds to CD4 and secondary receptors	plasma membrane, virion envelope
Tat	p16/p14	viral transcriptional activator	primarily in nucleolus-nucleus
Rev	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nucleolus-nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity	cytoplasm (cytosol, membranes)
Vpr	p10-15	promotes nuclear localization of preintegration complex, inhibits cell division, arrests infected cells at G2/M	virion nucleus (nuclear membrane?)
Vpu	p16	promotes extracellular release of viral particles; degrades CD4 in the ER; (phosphoprotein only in HIV-1 and SIVcpz)	integral membrane
Nef	p27-p25	CD4 and class I downregulation (myristylated protein)	plasma membrane, cytoplasm, (virion?)
Tev	p28	tripartite tat-env-rev protein (also named Tnv)	primarily in nucleolus-nucleus

3.2.1 La glycoprotéine d'enveloppe (Env)

La glycoprotéine d'enveloppe (Env) recouvre la face externe de la particule virale du VIH-1, elle intervient dans le processus de fusion membranaire et constitue la première cible de la réponse humorale. La protéine Env est constituée de trois sous-unités de gp120 (formant la tête) et trois sous-unités de gp41 (formant la tige). Les sous-unités gp120-gp41 de la protéine Env mature sont liées par des interactions non-covalentes et sont associées sous forme de trimère au niveau de la surface membranaire. La sous-unité gp120 présente cinq régions très variables appelées de V1 à V5 (120). La boucle V3 n'intervient pas dans la liaison au récepteur CD4, en revanche elle est importante pour le tropisme préférentiel du VIH-1 soit pour les lymphocytes T ou les macrophages (121, 122). La séquence représentée par la boucle V3 interagit avec le corécepteur CXCR4 et CCR5, ce qui détermine partiellement la susceptibilité des différents types cellulaires à générer des souches virales (123). La boucle V3 est également la cible principale d'anticorps qui bloquent l'activité virale du VIH-1.

3.2.2 Les protéines de structure

3.2.2.1 La protéine de la matrice (MA)

La protéine de la matrice représente le domaine N-terminal (NTD) du précurseur Pr55^{Gag}, elle est impliquée dans l'interaction de Gag avec la membrane plasmique, l'incorporation de la glycoprotéine Env dans l'enveloppe du virion et dans des événements de la phase précoce suivant l'entrée du virion. La structure tridimensionnelle de MA a été résolue par RMN et cristallographie aux rayons X (124, 125) (**Figure 23**). La protéine se replie en un domaine central comprenant principalement des hélices α et trois feuillets β . Environ 30 résidus du domaine C-terminal forment une hélice α servant à relier MA au domaine CA du précurseur Gag.

Dans la particule virale mature du VIH-1, MA est localisée sous l'enveloppe virale dérivant de la membrane de la cellule infectée. MA est myristoylée et contient des acides aminés basiques dans sa région N-terminale. La liaison à la membrane plasmique se fait par ancrage du groupement myristoyl dans la bicouche lipidique et par le domaine N-terminal basique qui lie les phospholipides acides, particulièrement le phosphoinositide PI(4,5)P2 qui est fortement présent dans la membrane plasmique (126).

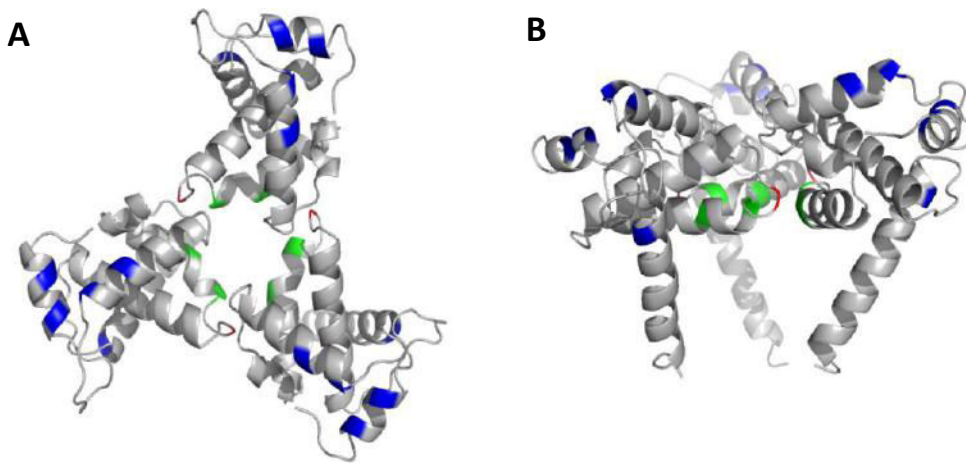


Figure 23. Organisation trimérique des protéines de la matrice (MA) (127). Vue longitudinale (A) et transversale (B) de la structure trimérique de MA. Le modèle a été généré en utilisant le logiciel pymol (www.pymol.org), le code de ce modèle sur la *Protein data bank* est 1HIW.

3.2.2.2 La protéine de la capsid (CA)

Dans la particule virale mature 216 hexamères de CA et 12 pentamères de CA forment une coque nommée capsid qui contient l'ARNg viral et les protéines associées (**Figure 24**). Le domaine CA du précurseur Gag joue un rôle important dans l'assemblage et la maturation du virion. La CA semble, également, être impliquée dans des événements de la phase précoce suivant l'entrée du virion (128). La protéine CA est composée de deux domaines : une région N-terminale dite « le domaine central » contenant les résidus 1-145, cette région intervient dans la maturation du virion et l'incorporation de la Cyclophiline A (CypA) (128); une région C-terminale dite « domaine de dimérisation » contenant les résidus 151-231, cette dernière est impliquée dans les interactions Gag-Gag. Le domaine central est composé de sept hélices α et deux feuillets β et une boucle fortement exposée servant de site de liaison canonique pour la CypA (129). Des études récentes montrent des interactions non canoniques avec la CypA ouvrant de nouvelles perspectives dans le développement thérapeutique qui ciblerait ces interactions (130).

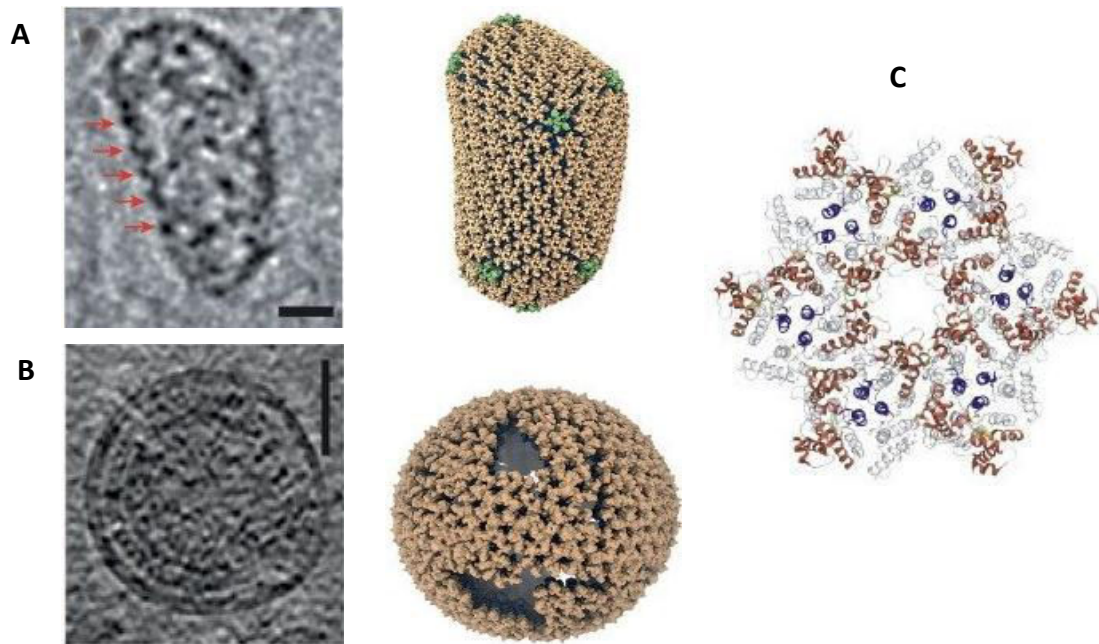


Figure 24. Modèle d'organisation des CA dans les particules virales matures et immatures (128). **A)** CA dans une particule de VIH-1 mature observée par cryo-microscopie électronique (cryo-ME), les flèches rouges indiquent la présence d'hexamères et pentamères à la périphérie. Sur le modèle computationnel obtenu à partir des données de cryo-ME les hexamères (marron) et pentamères (vert) sont observables. **B)** CA dans une particule de VIH-1 immature. **C)** Structure d'un hexamère de CA dans une particule virale immature (code d'accès PDB 4USN).

3.2.2.3 La protéine de la nucléocapside (NC)

Une particule virale infectieuse de VIH-1 contient environ 1000-1500 molécules de NC qui recouvrent l'ARN dimérique viral (23, 83). Le rapport NC:ARNg est de 1 NC pour 12-18 nucléotides.

3.2.2.3.1 *Structure de la NC*

La protéine de nucléocapside du VIH-1 est une petite protéine basique se fixant aux acides nucléiques. Elle est générée par clivage protéolytique du précurseur Gag, conduisant dans un premier temps à la formation de la NCp15, puis NCp9 et finalement NCp7 (communément appelée NC) (131). La NC contient 55 acides aminés comprenant deux doigts de zinc (ZF) avec le motif très conservé CCHC qui fixe le zinc avec une très forte affinité (**Figure 25**).

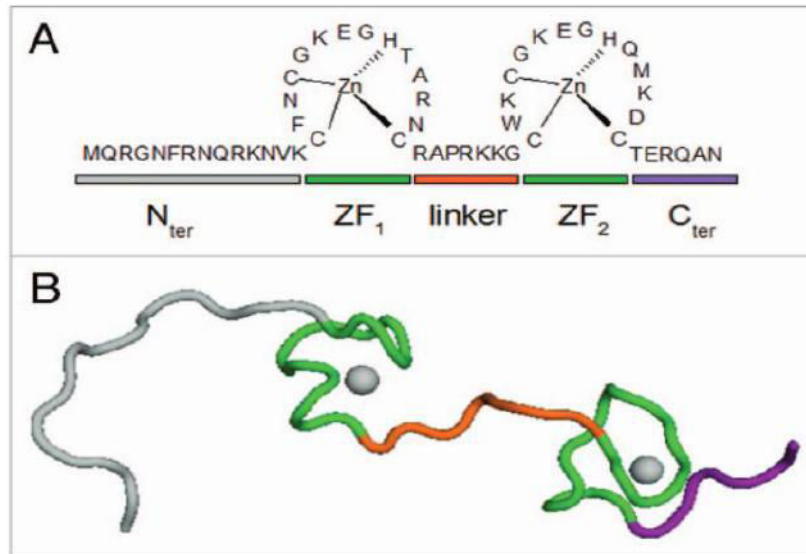


Figure 25. Structure primaire (A) et structure 3D (B) de la NC (132, 133). Les deux domaines en doigts de zinc (ZF₁ et ZF₂ en vert), sont séparés par une région charnière (orange) et les extrémités N (gris) et C-terminales (violet) sont faiblement structurées. La structure 3D a été obtenue à partir de données RMN (code PDB : 1AAF).

3.2.2.3.2 *Propriétés de la NC*

La NC présente une activité chaperonne des acides nucléiques (AN) (131, 134) grâce à laquelle elle remodèle la structure secondaire d'un acide nucléique de manière à ce qu'il adopte la conformation qui est thermodynamiquement la plus stable (135–137), contenant généralement le maximum de paires de bases (**Figure 26**). L'activité chaperonne est due à trois composantes principales : 1) la déstabilisation transitoire des structures secondaires par liaison des molécules de NC ; 2) l'hybridation des séquences complémentaires d'AN ; 3) une cinétique rapide de liaison (132, 133).

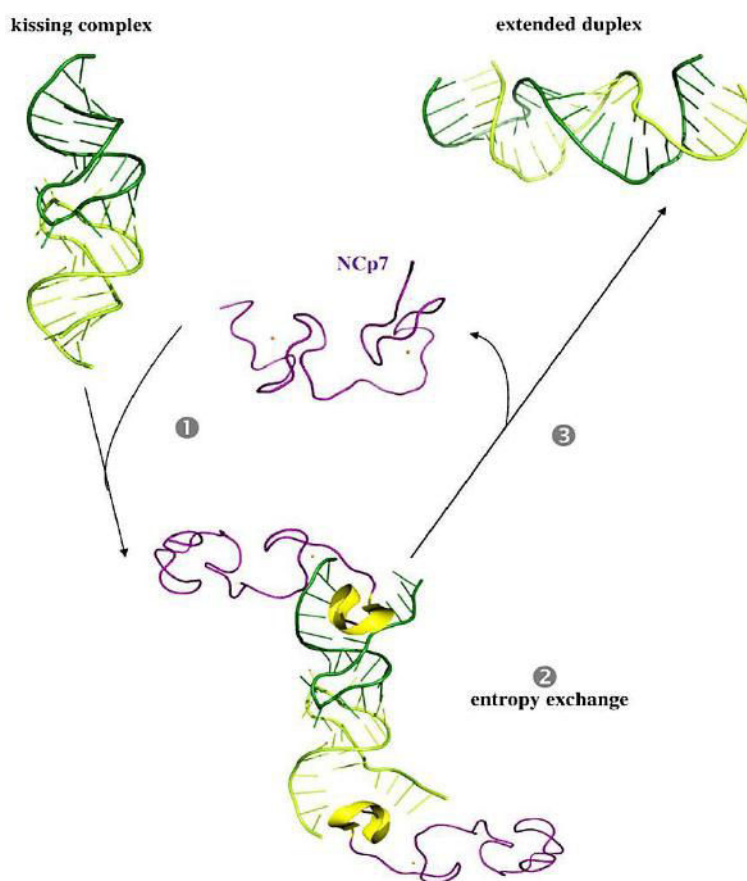


Figure 26. Remodelage de la structure d'un AN viral par la NC (138). Modèle schématique du changement d'entropie entre la NC et la séquence de dimérisation de l'ARNg du VIH-1. (1) La NC peu structurée se fixe au dimère d'ARN. (2) La fixation induit un réarrangement structural mutuel conduisant la partie N-terminale de la NC à former une hélice α (jaune) et à l'ouverture partielle des structures en tige-boucle de l'ARN. (3) La NC est libérée, permettant à l'ARN de rechercher une conformation stable, des cycles successifs de fixation puis dissociation (« on » et « off ») conduisent à la formation d'un long duplexe à partir d'un complexe boucle-boucle.

3.2.2.3.2.1 Activité déstabilisatrice des acides nucléiques

La capacité de la NC du VIH-1 à déstabiliser les AN structurés est due principalement à ses deux doigts de zinc. Cette activité déstabilisatrice dépend fortement de la stabilité des séquences d'AN. Les motifs comme les *bulges*, les boucles apicales, les mis-appariements, les paires de bases terminales jouent un rôle important dans la déstabilisation sélective par la NC (132, 139). La déstabilisation induite par la NC est initiée préférentiellement dans les régions double-brin qui sont à proximité de nucléotides non appariés. Par la suite, la NC dissocie les paires de bases afin que la déstabilisation se propage, ce qui permet le glissement de la NC et/ou la fixation de molécules de NC supplémentaires (132). L'activité déstabilisatrice de la NC induit une fluctuation des structures secondaires ce qui conduit à l'apparition d'espèces réactives nécessaires aux processus d'hybridation inter ou intramoléculaires au cours du cycle viral.

3.2.2.3.2.2 *Activité agrégative des acides nucléiques*

La transcription inverse nécessite de multiples réactions d'hybridation impliquant de nombreuses séquences d'AN. La capacité de la NC à agréger les AN de manière non spécifique (137) facilite l'étape de nucléation au cours du processus d'hybridation. La nucléation est une association de type diffusion limitée qui est ralentie par la répulsion électrostatique entre les séquences à hybrider et la faible probabilité du positionnement correcte des nucléotides à hybrider (140, 141). Des études effectuées avec des ligands cationiques multivalents présentant une forte mobilité lors de leur liaison aux AN, montrent que ces derniers constituent des agents d'agrégation ou de condensation efficaces (142, 143). Plusieurs recherches ont suggéré que la NC est fortement mobile lorsqu'elle se fixe aux AN (138, 144, 145), la région N-terminale de la NC est fortement cationique ce qui lui confère une propriété agrégative des AN. Le NTD est non structuré lorsque la NC est sous forme libre et forme une hélice α -3₁₀ lors de la liaison à un AN (146). Le NTD et les résidus basiques situés dans la région *linker* localisée entre les deux ZF contribuent à des réactions d'association avec les AN, assurant leur agrégation (86, 138, 147). Par ailleurs, la NC se lie préférentiellement aux séquences simple-brin riches en guanines par des interactions d'empilement via les résidus hydrophobes (64, 148, 149). La capacité de fixation et d'agrégation de la NC constituent une partie essentielle de son activité chaperonne.

3.2.2.3.2.3 *Cinétique rapide de liaison*

L'une des caractéristiques de la NC est qu'elle présente une cinétique de liaison rapide (86). En d'autres termes, la destabilisation et l'association des structures ADN/ARN ne nécessite pas que la NC réside longtemps sur les acides nucléiques cibles.

3.2.2.3.2.4 *Mode de fixation de la NC sur les acides nucléiques*

La constante d'association de la NC à un AN dépend fortement de la nature, la séquence et le repliement de la molécule cible et elle peut donc fortement varier. Cette forte variation de la constante d'association confère à cette protéine la capacité d'exercer de nombreuses fonctions, dépendant de la nature des acides nucléiques et de la concentration respective des protéines et des séquences d'AN (138, 150). Des études par RMN et par calorimétrie de titration isotherme (ITC) montrent que les boucles apicales des ARN présentant la séquence GNG constituent des sites de liaison de forte affinité pour la NC (146, 151). Des expériences de résonance plasmonique de surface (SPR) montrent que dans 0,15 M NaCl, la NC se fixe préférentiellement aux ADN simple-brin (152). Une étude basée sur la nano-manipulation de

Introduction

molécule unique (*single molecule stretching assay*) a montré que la NC présente une fixation préférentielle aux ADN simple-brin (153). Par ailleurs, dans 10 mM de phosphate de sodium (pH 7,0), l'affinité de la NC pour les AN double-brin est légèrement plus faible que l'affinité pour les AN simple-brin (138). La NC présente également des propriétés de fixation à des séquences spécifiques bien définies. Ces propriétés de forte liaison spécifique jouent un rôle critique dans la reconnaissance du signal d'encapsidation de l'ARNg viral, par le domaine NC du précurseur Gag, impliquant la reconnaissance et la sélection spécifique parmi de nombreux ARN cellulaires durant l'assemblage de la particule virale (54, 154–157). La NC se lie également avec une forte affinité aux motifs TG, GNG et TNG (64, 148, 152, 158, 159) (Tableau 6). Ces motifs sont spécifiquement reconnus par la NC dans trois tiges-boucles formées par l'ARN TAR et son complémentaire ADN (158–160) et l'ADN PBS (161).

La NC présente une polarité de fixation opposée selon que l'acide nucléique est un ADN ou un ARN (148). Une étude réalisée au sein de notre équipe a montré que la NC se fixe à l'ARN avec une orientation anti-parallèle, c'est-à-dire que l'extrémité N-terminale de la NC est orientée vers l'extrémité 3' de l'ARN (**Figure 27A**). Inversement, la NC se fixe parallèlement à l'ADN avec son extrémité N-terminale orientée vers l'extrémité 5' de l'ADN (**Figure 27B**) (139, 148).

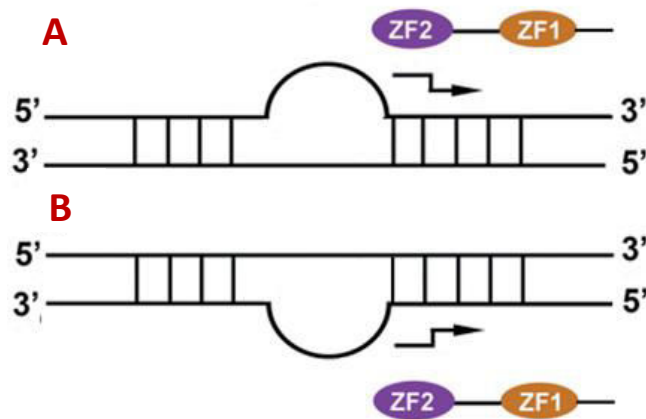


Figure 27. Polarité de fixation de la NC aux acides nucléiques (139). **(A)** La NC se fixe à l'ARN avec une orientation antiparallèle. La NC se fixe à proximité de la région simple-brin induisant une déstabilisation de la tige dans le sens indiqué par la flèche. **(B)** La NC se fixe à l'ADN avec une orientation parallèle.

Tableau 6. Affinité de la NC pour différentes séquences oligonucléotidiques cibles (138).

Name	Sequence	NA	NC	K_{obs} (nM)	Salt conditions	Method	References
SL3 (SL)	ggpCUAGCGGAGGCUAGuoc	RNA-14nt	NCp7	21-28	200 mM NaCl	TFQ	Paoletti <i>et al.</i> ⁹¹
		RNA-20nt					Shubsda <i>et al.</i> ⁹³
	GGACUAGCGGAGGCUAGUCC	RNA-20nt	NCp7	100	25 mM CH ₃ COONa 25 mM NaCl	Native GE	Athavale <i>et al.</i> ⁸²
		DNA-20nt	NC(1-72)	70	100 mM NaCl	TFQ	De Guzman <i>et al.</i> ⁵⁷
	GGACTAGCGGAGGCTAGTCC	DNA-20nt	NC(12-53)	1000	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		DNA-20nt	NC(12-53)	1450	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
	AGGGCCCUAGCGGAGGCUAGCGGCC	RNA-27nt	NCp9		100 mM NaCl	Filter binding	Berglund <i>et al.</i> ⁹⁸
		RNA 46nt	NCp7	~200	1 mM MgCl ₂ 50 mM KCl	Native GE	Clever <i>et al.</i> ¹⁶²
	751nt(-)CUAGCGGAGGCUAG(-)796nt	RNA-20nt	NCp7	85	100 mM NaCl	Filter binding	Xi <i>et al.</i> ¹⁶³
		RNA-20nt	NCp7	170	50 mM NaCl	ITC, NMR	Amarasinghe <i>et al.</i> ⁵⁴
SL2 (SL)	GGACUAGCGGAGGCUAGUCC	RNA-20nt	NCp7	178	150 mM CH ₃ COONH ₄	ESI-FTMS	Hagan and Fabris ¹⁶⁴
		RNA-20nt	NCp7	551	150 mM NaCl	FA	Takahashi <i>et al.</i> ¹⁶⁵
	GGACUAGCGGAGGCUAGUCCdCdA	RNA-23nt	NCp7	110	1 mM MgCl ₂ 25 mM CH ₃ COONa	ITC, NMR	Amarasinghe <i>et al.</i> ⁵⁴
		RNA-19nt	NCp7	23	25 mM NaCl	TFQ	Shubsda <i>et al.</i> ⁹³
	GGCGACUGGUGAGUACGCC	RNA-19nt	NCp7	62	150 mM CH ₃ COONH ₄	ESI-FTMS	Hagan and Fabris ¹⁶⁴
		RNA-19nt	NCp7	~400	50 mM KCl	Native GE	Clever <i>et al.</i> ¹⁶²
	GGCGCGCGACUGGUGAGUACGCCAAAAA	RNA-29nt	NCp7	80	30 mM NaCl	Filter binding	Bourbigot <i>et al.</i> ⁵⁶
		DNA-14nt	NC(11-55)	250	0.2 mM MgCl ₂ 30 mM NaCl	TFQ, ITC	Bazzi <i>et al.</i> ⁹⁷
	GTCCCTGTTCGGGC	DNA-26nt	NC(11-55)	260	0.2 mM MgCl ₂	ITC	Vuilleumier <i>et al.</i> ⁹⁴
		DNA-6nt	NC(1-72)	285	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
PBS (SL)	CCAGAGAGCTCCGGCTCGACCTGG	DNA-6nt	NCp7	200	100 mM NaCl	TFQ	Avilov <i>et al.</i> ⁹⁶
		RNA-6nt	NC(11-55)	400	30 mM NaCl	2FCS	Didier <i>et al.</i> ¹⁶⁶
	AATGCC	DNA-12nt	NCp7	1200	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
		RNA-12nt	NCp7	560	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	AATGACTGAAAC	DNA-6nt	NC(1-72)	1100	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		RNA-6nt	NCp7	710	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	ATGCC	DNA-12nt	NCp7	120	100 mM NaCl	ITC	Spriggs <i>et al.</i> ¹⁰¹
		RNA-11nt	NCp7	3-5	150 mM NaCl	SPR, FA, TFQ	Fisher <i>et al.</i> ¹⁰⁰
	TGACC GTGACCG	DNA-8nt	NCp7				
		RNA-8nt	NCp7				
cTAR (SL)	AATGCC	DNA-6nt	NC(1-72)	260	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		RNA-6nt	NC(1-72)	285	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
	AATGCC	DNA-6nt	NCp7	200	100 mM NaCl	TFQ	Avilov <i>et al.</i> ⁹⁶
		RNA-6nt	NC(11-55)	400	30 mM NaCl	2FCS	Didier <i>et al.</i> ¹⁶⁶
	AATGACTGAAAC	DNA-12nt	NCp7	1200	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
		RNA-12nt	NCp7	560	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	AATGACTGAAAC	DNA-6nt	NC(1-72)	1100	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		RNA-6nt	NCp7	710	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	ATGCC	DNA-12nt	NCp7	120	100 mM NaCl	ITC	Spriggs <i>et al.</i> ¹⁰¹
		RNA-11nt	NCp7	3-5	150 mM NaCl	SPR, FA, TFQ	Fisher <i>et al.</i> ¹⁰⁰
Linear single-stranded DNA or RNA	AATGCC	DNA-6nt	NC(1-72)	260	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		RNA-6nt	NC(1-72)	285	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
	AATGCC	DNA-6nt	NCp7	200	100 mM NaCl	TFQ	Avilov <i>et al.</i> ⁹⁶
		RNA-6nt	NC(11-55)	400	30 mM NaCl	2FCS	Didier <i>et al.</i> ¹⁶⁶
	AATGACTGAAAC	DNA-12nt	NCp7	1200	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
		RNA-12nt	NCp7	560	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	AATGACTGAAAC	DNA-6nt	NC(1-72)	1100	100 mM NaCl	TFQ	Vuilleumier <i>et al.</i> ⁹⁴
		RNA-6nt	NCp7	710	100 mM NaCl	2ApF	Avilov <i>et al.</i> ⁹⁵
	ATGCC	DNA-12nt	NCp7	120	100 mM NaCl	ITC	Spriggs <i>et al.</i> ¹⁰¹
		RNA-11nt	NCp7	3-5	150 mM NaCl	SPR, FA, TFQ	Fisher <i>et al.</i> ¹⁰⁰

Les nucléotides en gras et soulignés, correspondent aux sites de fixation identifiés de la NC. Les affinités des sites de fixation ont été obtenues à partir de complexes NC-AN 1:1. TQF, tryptophan fluorescence quenching ; GE, gel electrophoresis ; 2FCS, two-focus fluorescence correlation spectroscopy ; FA, fluorescence anisotropy ; 2ApF, 2-aminopurine fluorescence intensity ; SPR, surface plasmon resonance ; ITC, isothermal titration calorimetry ; ESI-FTMS, electrospray ionization-fourier transform mass spectrometry.

3.2.2.3.3 Rôle de la NC au cours du cycle réplcatif

La NC joue de multiples rôles dans le cycle réplcatif viral grâce à sa capacité de se lier de manière spécifique ou non à presque toutes les séquences d'acides nucléiques (138). La NC du VIH-1 a été fortement étudiée au cours des 25 dernières années, ces recherches ont montré l'implication de la NC, sous forme libre ou comme domaine du précurseur Gag, dans l'assemblage de la particule virale, la dimérisation et l'empaquetage de l'ARNg, la transcription inverse et l'intégration dans le génome de la cellule hôte (162) (**Figure 28**).

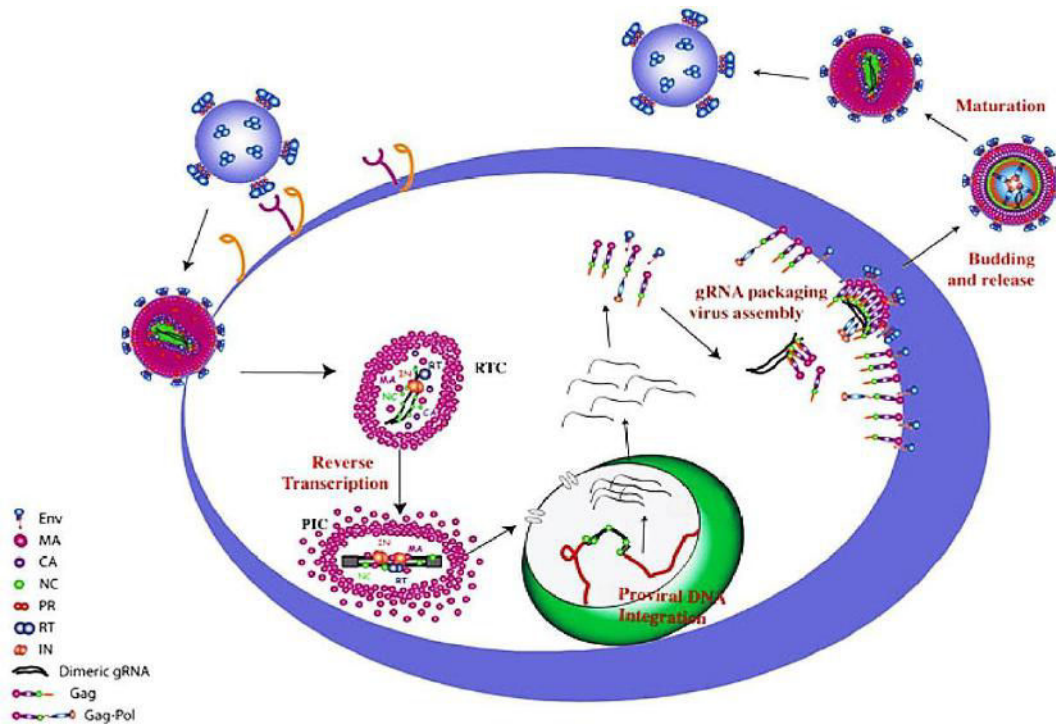


Figure 28. Etapes du cycle réplcatif du VIH-1 où la NC intervient (162). La NC mature (NCp7) assiste la RT lors de la conversion de l'ARNg simple-brin en ADN double-brin. La NC participerait à l'intégration de l'ADN viral dans le génome de la cellule hôte. Le domaine NC de la polyprotéine Gag reconnaît et encapside l'ARNg au cours du processus d'assemblage.

3.2.2.3.3.1 Transcription inverse

Le rôle de la NC dans l'étape de transcription inverse sera décrit dans le chapitre 4.

3.2.2.3.3.2 Intégration de l'ADN viral

Le virion du VIH-1 ayant pénétré dans le cytoplasme serait constitué d'un ilot de protéines CA entourant le complexe constitué d'ARNg dimérique, de l'amorce ARN_t^(lys,3) et d'environ 1000 à 1500 molécules de NC. De plus, des complexes purifiés ont montré la présence de RT, IN, PR, Vpr et un faible taux de Pr55^{Gag} et P41^{MA-CA} (163, 164). Cet assemblage nucléoprotéique est désigné comme le complexe de transcription inverse (RTC).

Introduction

Lorsque le RTC est dans le cytoplasme, la transcription inverse est amorcée et le complexe est transporté à travers le cytoplasme vers le noyau où il est converti en *preintegration complexe* (PIC) qui est activement transporté vers le noyau (165). La présence de NC dans le complexe de pré-intégration est attendue dans une certaine mesure en raison de sa présence dans les RTC, des recherches ont montré une localisation de la NC dans le noyau après infection (166). En revanche, la NC n'a pas été détectée dans des PIC isolés du VIH-1 (167). Cet échec à détecter la NC pourrait être dû à la faible interaction de la NC et les acides nucléiques dans les PIC. Il a été proposé deux modèles pour expliquer la non détection de NC dans les PIC isolés du VIH-1. Le premier modèle suggère que la NC se fixe non spécifiquement aux ADN double-brin, obligeant l'intégrase (IN) à se fixer aux extrémités LTR (168). Lorsque IN se fixe à la bonne position, il n'est pas nécessaire que la NC soit présente. Le second modèle propose que la NC se fixe, à très faible concentration, à des régions spécifiques de l'ADN viral. La visualisation par microscopie à transmission électronique a montré que l'ADN double-brin contient en son centre une zone nommée FLAP qui peut fixer la NC (169, 170). Cette zone est une structure triple brin contenant deux segments d'ADN(+) se chevauchant et un brin (-) complémentaire. Des études ont montré que le FLAP joue un rôle important dans le transport du PIC, il est également désigné comme étant un facteur déterminant influençant l'étape d'intégration de l'ADN viral (171). De plus, le clivage de la NCp15 par la protéase virale n'est pas complet, il est probable que la NCp9 soit présente en très faible quantité. Cette dernière se fixe sur les ADN simple-brin et double-brin (172). Des travaux ont montré que la NCp9 possède un mode de liaison coopérative, c'est-à-dire, l'accroissement des interactions Gag-Gag augmente la constante d'association NC/ADN (86), ce qui suggère qu'une faible concentration de NCp9 peut permettre à IN de former un complexe aux extrémités LTR. Il a également été observé *in vitro* que l'intégration stimulée par la NCp9 est plus efficace qu'en présence de NC (173).

3.2.2.3.3.3 Assemblage et incorporation de l'ARNg viral

La fixation de Gag via son domaine NC, à l'ARN génomique, est initiée dans la région Psi, cette dernière agit comme un point de nucléation pour permettre l'agrégation des unités de Gag et Gag-Pol par des interactions Gag-Gag et Gag-ARN. Des analyses génétiques ont montré que le domaine NC de Gag est nécessaire à la reconnaissance spécifique et l'encapsidation de l'ARNg, les mutations des doigts de zinc (174, 175) ou la présence de multiples résidus basiques dans le domaine NC de Gag (176, 177) peuvent significativement réduire l'empaquetage du génome. L'assemblage de la particule virale est médié par

Introduction

l'empaquetage de l'ARN et les interactions Gag-Gag, suivie de l'étape de bourgeonnement à partir de la cellule hôte sous la forme d'une particule immature non infectieuse. La délétion du domaine NC du précurseur Gag induit une accumulation de la poly-protéine Gag au niveau de la membrane plasmique *in cellulo* (175).

3.2.2.3.3.4 *Bourgeonnement et maturation de la particule virale*

Une étape nécessaire à la maturation de la particule virale est la formation d'un dimère stable entre les deux molécules d'ARNg qui sont présentes dans la particule virale. Des études ont montré que la NC est importante pour la formation de ce dimère (78, 178, 179). Deux études *in vitro* suggèrent que les sous-unités des dimères stables sont liées par un duplexe relativement long qui nécessite l'action de la NC pour être formé à température physiologique (180, 181). Des mutations au niveau des doigts de zinc ont un impact négatif sur le bourgeonnement et le clivage du précurseur Gag (178, 179).

3.2.3 Les protéines non structurales

3.2.3.1 *Les protéines de régulation et les protéines accessoires*

Le VIH-1 code deux protéines de régulation, le *Transactivator of transcription* (Tat) et le *Regulator of expression of virion proteins* (Rev). Le virus code également pour des protéines dites accessoires : *negative effector* (Nef), *viral infectivity factor* (Vif) et les *viral protein r* et *u* (respectivement, Vpr et Vpu).

3.2.3.1.1 *Tat*

(86 à 101 acides aminés) se fixe sur la structure en tige-boucle TAR (182). Elle est produite par des ARNm entièrement ou partiellement épissés (**Figure 12**). Tat est l'un des deux facteurs de régulation viraux essentiel pour l'expression génique. Il recrute de nombreuses protéines cellulaires permettant de rendre la structure de la chromatine accessible à la machinerie de transcription (183, 184). Un faible taux de Tat est détecté dans les cellules infectées. Tat a été localisé principalement au niveau du nucléole/noyau par immunofluorescence. La fixation de Tat à TAR se produit conjointement avec les protéines cellulaires qui contribuent à l'activité de Tat. La fixation de Tat à TAR active d'un facteur 1000 la transcription à partir du LTR 5' (184, 185).

3.2.3.1.2 *Rev*

(116 acides aminés) est une protéine qui se fixe spécifiquement sur la région RRE qui est situé dans le gène *env*. Elle est produite à partir d'ARNm entièrement épissés (**Figure 12**).

Introduction

La fixation de Rev à la structure RRE facilite le transport des ARN viraux non épissés ou partiellement épissés, du noyau vers le cytoplasme (186). Des travaux ont montré que la fixation de Rev est inhibée par l'enzyme *Adenosine Deaminase Acting on RNA 1* (ADAR1) (187, 188).

3.2.3.1.3 *Nef*

(206 acides aminés) est produite à partir d'ARNm entièrement épissés (**Figure 12**). Elle présente une région centrale structurée (acides aminés 62-147 et 179-200) ; les régions N- et C-terminales flexibles (acides aminés 2-61 et 201-206) ; une boucle interne flexible (acides aminés 148-178) (numérotation standard NL4-3) (189). Nef est considérée comme étant un facteur pathogène, son rôle dans le développement du SIDA reste peu compris. Bien que ses effets soient dus aux interactions protéine-protéine, une seule protéine cellulaire, la protéine tyrosine kinase Hck, a été caractérisée comme se fixant directement à Nef (Kd environ 200 nM) (190–192). Il a été montré que Nef présente de multiples activités dont la régulation négative de l'expression des CD4 à la surface cellulaire, la perturbation de l'activation des lymphocytes T et la stimulation de la pathogénicité du VIH-1 (193).

3.2.3.1.4 *Vpr*

(96 acides aminés) joue un rôle dans la capacité du VIH-1 à infecter des cellules ne se divisant pas en facilitant la localisation nucléaire du PIC (194). Cette protéine est produite à partir d'ARNm partiellement épissés (**Figure 12**), elle intervient dans : la coactivation transcriptionnelle des gènes viraux et cellulaires, l'induction de l'arrêt du cycle cellulaire, la contribution directe et indirecte au dysfonctionnement des lymphocytes T et la régulation de l'activité de Nf-κB (194).

3.2.3.1.5 *Vpu*

(81 acides aminés) est propre aux VIH-1 et SIVcpz (le virus SIV le plus apparenté au VIH-1). Chez le VIH-2 et les autres SIV, il n'y a pas de protéine similaire. Vpu est produite à partir d'ARNm complètement épissés (**Figure 12**). Elle présente deux fonctions biologiques bien distinctes : (a) régulation négative de l'expression des récepteurs CD4 en modulant la dégradation par le protéasome des CD4 nouvellement synthétisés dans le réticulum endoplasmique (ER), (b) stimulation du relargage des virions à partir de la membrane plasmique des cellules infectées par le VIH-1 (195, 196). Des études ont montré que ces deux fonctions pourraient avoir influencé la transmission inter-espèces du virus de

l'immunodéficience chez les primates. De plus, cette protéine accessoire permet au VIH-1 d'échapper au système immunitaire de l'hôte (197).

3.2.3.1.6 *Vif*

(220 acides aminés) est produite à partir d'ARNm complètement épissés (**Figure 12**). Vif est essentielle pour la réplication du VIH-1 dans les lymphocytes du sang périphérique, les macrophages et certaines lignées cellulaires (198). Vif inhibe l'activité antivirale de la protéine cellulaire APOBEC3G en empêchant son incorporation dans les virions en phase de bourgeonnement. Ainsi, APOBEC3G est ubiquitinylée ce qui déclenche sa dégradation par le protéasome (199, 200).

3.2.3.2 *La protéase (PR)*

L'activité de la protéase du VIH-1 n'est pas nécessaire pour la production et le relargage du virion *per se*, en revanche elle est essentielle pour la maturation conduisant à des particules virales infectieuses. Elle génère des particules virales infectieuses par clivage des précurseurs polyprotéiques Gag et Gag-Pol. La PR du VIH-1 est classée dans la famille des protéases à acide aspartique (cet acide aminé se trouve dans le site actif en position 25) et est un homodimère formé de deux sous-unités identiques de 99 acides aminés (201, 202). Le centre de l'enzyme contient le site de fixation au substrat qui interagit avec les différents sites de clivage des précurseurs Gag et Gag-Pol. La dimérisation de la polyprotéine Gag-Pol est nécessaire à l'activation de la protéase virale qui est présente dans le domaine Pol du précurseur Gag-Pol (203).

3.2.3.3 *L'intégrase (IN)*

L'intégrase du VIH-1 est une enzyme essentielle pour l'intégration de l'ADN viral dans le génome de la cellule hôte. Le processus d'intégration de l'ADN se produit suivant deux étapes bien distinctes appelées maturation en 3' et transfert de brin.

Fonctionnellement et structuralement, l'intégrase appartient à la famille des polynucléotidyl transférase (204). Chaque particule virale contient environ 50 à 100 copies de l'enzyme (205). IN est une protéine de 32 kD comprenant trois domaines structuralement indépendants : le domaine amino-terminal (NTD), le domaine du cœur catalytique (CCD) et le domaine carboxy-terminal (CTD).

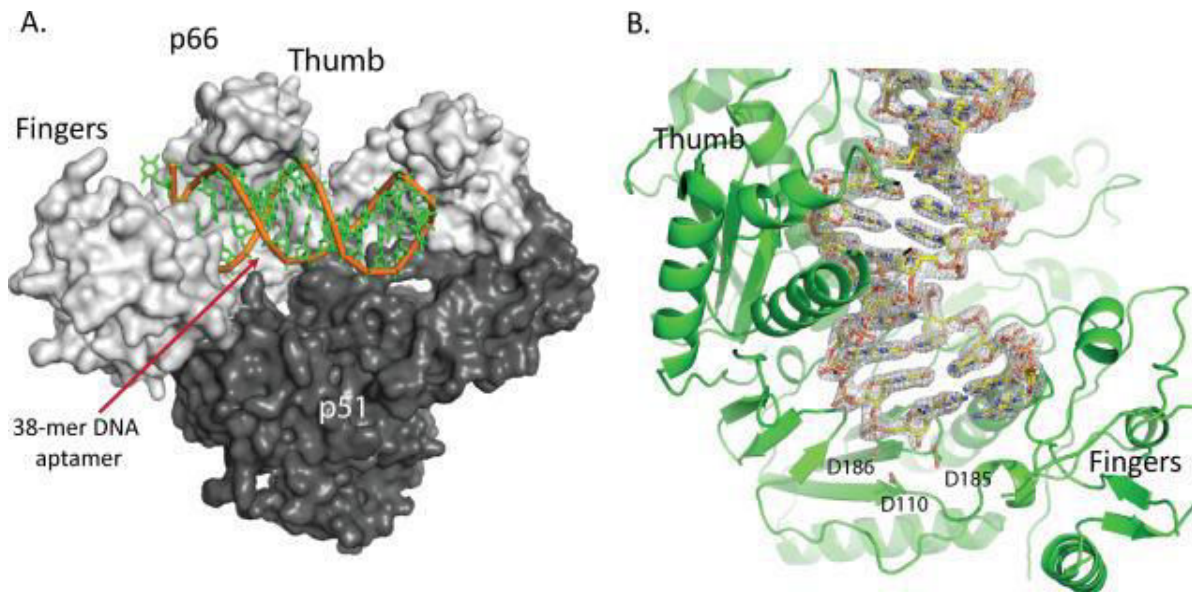
IN semble être en équilibre entre les états monomérique, dimérique, tétramérique et d'autres états oligomériques. L'intégrase se lie à l'ADN uniquement sous sa forme dimérique, ce dernier représente la forme active de l'enzyme (206, 207). Deux protéines cellulaires ont été caractérisées comme intervenant dans l'import du PIC. Le complexe IN-ADN est

transporté vers le noyau par la transporine-SR2 (TR2) (208). LEDGF/p75 est un autre cofacteur cellulaire de l'intégrase du VIH-1 qui interagit avec IN par son *IN binding domain* (IBD) et fixe le PIC à la chromatine de la cellule hôte (209).

3.2.3.4 La transcriptase inverse (RT)

3.2.3.4.1 Structure de la RT

La RT est un hétérodimère qui est généré par le clivage du précurseur Gag-Pol et se présente comme deux sous-unités, p66 (560 acides aminés) et p51 (les 440 premiers acides aminés de la sous-unité p66) (210). La sous-unité p66 possède deux domaines : le domaine N-terminal contenant une activité ADN polymérase qui est responsable de l'étape critique de conversion de l'ARNg viral simple brin en ADN double-brin ; le domaine C-terminal possédant une activité RNase H qui dégrade le brin d'ARN dans un hybride ARN-ADN. Les deux sous-unités sont codées par la séquence génomique virale. Le domaine RNase H constitue les 120 derniers acides aminés de la sous-unité p66, qui correspond au fragment p15 généré par le clivage de la sous-unité p66 par la protéase virale. Ce clivage génère aussi la sous-unité p51 (211, 212) qui ne présente pas d'activité catalytique. Similairement aux autres polynucléotides polymérases, la structure tridimensionnelle globale de la sous-unité p66 est souvent comparée à une main droite (**Figure 29**), avec un sous-domaine dit *Fingers* (acides aminés 1-85 et 118-155), un sous-domaine *Palm* (acides aminés 86-117 et 156-237) et un sous-domaine *Thumb* (acides aminés 238-318) (213). Le sous-domaine *Palm* possède l'activité polymérase avec les trois acides aspartiques formant le site actif (D110, D185, D186). Les dNTP se fixent entre les sous-domaines *Palm* et *Fingers* et d'importants contacts se produisent entre les riboses et les bases azotées avec les résidus L74, Y115, M184 et Q151. La poche catalytique est formée par le sous-domaine *Fingers* qui se rabat sur le sous-domaine *Palm* (214). A proximité du domaine catalytique, la sous-unité p66 contient le domaine RNase-H (acides aminés 427-560) lié par le domaine de connexion (acides aminés 319-426) au sous-domaine *Thumb*. Le domaine de connexion est aussi impliqué dans des interactions avec les AN et la sous-unité p51. Malgré leur homologie de séquence, la sous-unité p66 présente une structure ouverte et flexible, tandis que la sous-unité p51 est plus globulaire et elle aurait un rôle structural dépourvue de l'activité catalytique avec les trois acides aspartiques enfouis à l'intérieur.



3.2.3.4.2 *Propriétés de la RT*

La RT possède deux activités enzymatiques : une activité ADN polymérase qui peut copier les matrices ADN ou ARN et une activité RNase-H qui clive le brin d'ARN uniquement si ce dernier est hybridé à un ADN. Les deux fonctions enzymatiques de la RT coopèrent pour convertir l'ARN en ADN linéaire double-brin. La RT ne possède pas une activité de relecture exonucléolytique qui confirme que le transcrit est la copie exacte de l'ARN matrice. Par conséquent, la RT effectue des erreurs avec un taux de mutations d'environ $3,4 \times 10^{-5}$ mutations par paire de bases en un cycle de réplication (215). Étant donné que la taille du génome du VIH-1 est d'environ à 10^4 paires de base, et l'infection par le VIH est caractérisée par un fort taux de réplication avec production d'environ 10 milliards de nouvelles particules virales par jour chez les individus infectés non traités, des millions de variants viraux sont produits chaque jour chez une personne infectée (216, 217). Une fois que la synthèse de l'ADN est complétée, il en résulte un ADN viral linéaire et double-brin qui est transloqué vers le noyau pour être inséré dans le génome de la cellule hôte par l'intégrase.

3.2.3.4.2.1 *Activité ADN polymérase*

Le mécanisme de polymérisation de l'ADN par la RT du VIH-1 est relativement bien caractérisé et des données cristallographiques ont permis de définir les étapes de ce processus.

Introduction

La réaction débute par liaison de la RT à l'acide nucléique substrat ce qui induit un changement conformationnel du sous-domaine *Thumb* de la sous-unité p66, d'une conformation fermée à une conformation ouverte. Similairement aux autres ADN polymérases, la RT nécessite une amorce pour copier l'acide nucléique qui sert de matrice. L'enzyme se lie préférentiellement aux AN double-brin, ainsi, l'extrémité 3' de l'amorce se fixe sur le *Priming site* (site P), adjacent au site actif de l'activité polymérase (212, 214, 218, 219). L'étape initiale d'incorporation nucléotidique est la fixation des dNTP entrants au niveau du *Nucleotide-binding site* (site N) pour former un complexe tertiaire (214). La limitation de la cinétique de la réaction de polymérisation est due à un changement conformationnel du sous-domaine *Fingers* qui se rabat sur le dNTP entrant et le site actif de l'activité polymérase, permettant d'aligner le 3'-OH de l'amorce, le phosphate α et le site actif de la polymérase (214, 220, 221). Par la suite, une étape de réactions chimiques induit la formation d'un pont phospho-diester entre le nucléoside nouvellement incorporé et l'amorce, générant un pyrophosphate. Le sous domaine *Fingers* s'ouvre et conduit le pyrophosphate hors du site actif. Au cours du processus de synthèse d'ADN, l'acide nucléique substrat est transloqué de manière à libérer le site de liaison nucléotidique, ainsi la RT peut fixer le prochain dNTP entrant. Cette étape de limitation de la cinétique de polymérisation par fixation et appariement du dNTP adéquat constitue la base de la fidélité de polymérisation de la RT (222, 223). L'étape de réactions chimiques nécessite la présence de deux ions Mg^{2+} , ces derniers coordonnent les oxygènes des trois phosphates d'un dNTP entrant et la chaîne latérale des résidus aspartate catalytiques (D110, D185, D186). Cette étape de coordination au niveau du site polymérase permet au 3'OH d'attaquer le phosphate α du nucléotide entrant, activant ainsi le groupement hydroxyl (**Figure 30**). Les deux ions métalliques contribuent, également, à la stabilisation de la charge des intermédiaires de réaction (224).

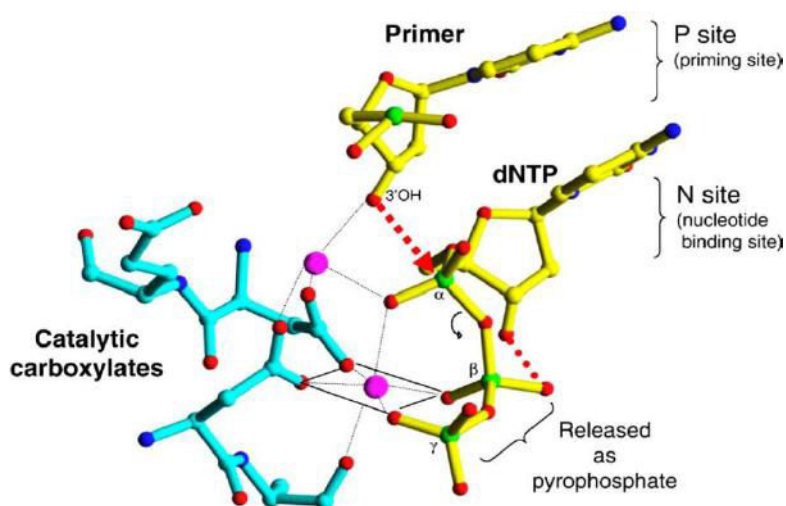


Figure 30. Chélation d'ions métalliques au niveau du site actif de l'activité polymérase de la RT (225). Les deux ions Mg^{2+} sont représentés par des boules mauves.

3.2.3.4.2.2 Activité RNase H

Le site actif de l'activité RNase H de la RT est localisé dans le CTD et est séparé du site de polymérisation par 18 paires de bases de l'hétéroduplexe substrat ADN-ARN (226, 227). Cette configuration permet à la RT de cliver l'ARN au cours de la synthèse de l'ADNc, ce processus est dit *polymerisation-dependant RNase H activity* (**Figure 31**) (228). Toutefois, le taux de polymérisation de la RT est supérieur au taux d'hydrolyse de l'ARN. Des coupures indépendantes de la polymérisation sont donc effectuées lorsque la RT se fixe sur les hybrides ARN-ADN non hydrolysés, cette action est nécessaire pour la dégradation complète du génome ayant servi de matrice (**Figure 31**) (228, 229). Des études ont montré que l'activité RNase H de la RT est activée en présence d'ions métalliques, un mécanisme similaire aux autres polymérases associées à des nucléases. La structure cristallographique du domaine RNase H montre deux Mn^{2+} liés au site actif. Des expériences de spectroscopie par RMN et de calorimétrie ont démontré la présence de deux ions Mn^{2+} et Mg^{2+} liés au site actif de la RNase H (230, 231). Des essais biochimiques ont mis en évidence une activité RNase H optimale obtenue en présence d'un ion Mg^{2+} et un ion Mn^{2+} (232). Toutefois, il est possible que les deux ions soient du Mg^{2+} *in vivo*. Il a été suggéré que la spécificité de l'activité RNase H aux hybrides ARN-ADN soit liée à la largeur du petit sillon (233, 234). Cette suggestion a été proposée car il a été observé que les hybrides ARN-ADN présentent une conformation intermédiaire entre les formes A et B des acides nucléiques double-brin dite « la forme H », possédant un petit sillon d'une largeur d'environ 9 à 10 Å (235). La structure cristallographique de la RT liée à un hybride ARN-ADN suggère que la largeur du petit sillon serait responsable du clivage spécifique des duplexes ARN-ADN (235).

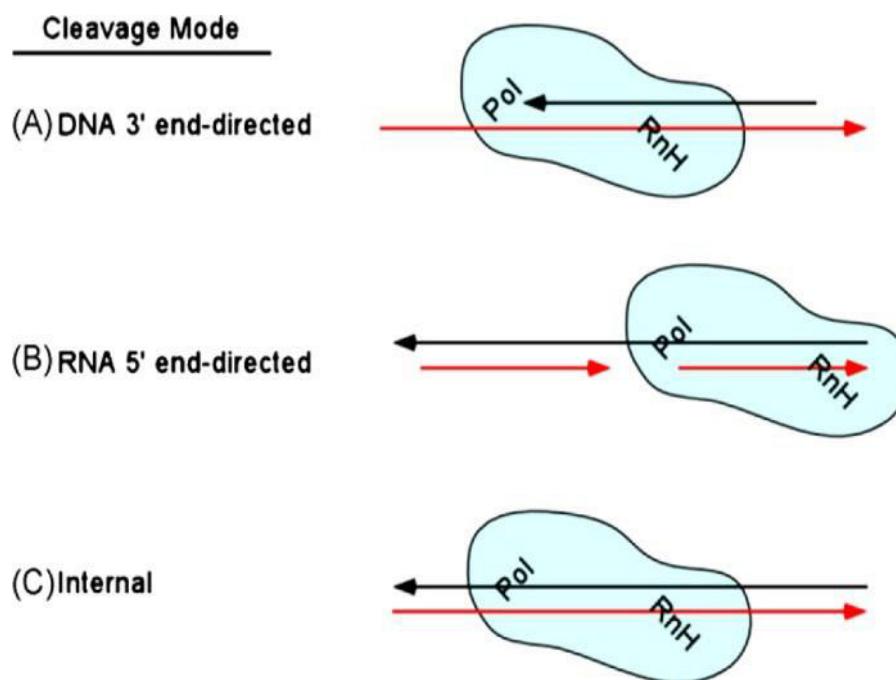


Figure 31. Trois modes de clivage par la RT du VIH-1 (236). La RT se fixe aux acides nucléiques substrats et peut effectuer trois modes de clivage : Le clivage dirigé par l'extrémité 3' l'ADN en cours de synthèse (A), Le clivage dirigé par l'extrémité 5' de l'ARN copié (B) et le clivage interne (C). La RT est schématisée en bleu, le brin d'ADN en noir, le brin d'ARN en rouge et les sites actifs des domaines polymérase (pol) et RNase H (RnH) sont indiqués.

L'activité RNase H est responsable de la dégradation du brin ARN dans l'hybride ARN-ADN formé pendant la synthèse du brin (-) d'ADN. Elle est également impliquée dans la dégradation de l'amorce $ARN_t^{(Lys,3)}$ (237) et le *Polypurine Tract* (PPT) (238). Les virus ayant une activité RNase H déficiente ne sont pas infectieux (239). En général, la dégradation de l'ARNg est indépendante de la séquence. La RNase H de la RT du VIH-1 effectue cependant certains clivages spécifiques.

Lors de la synthèse de l'ADN, la RT se fixe aux hybrides ARN-ADN substrats et utilise trois modes de clivage distincts pour la dégradation des séquences d'ARNg (**Figure 31**) : (A) clivage dirigé par l'extrémité 3' de l'ADN et situé à 15-20 nucléotides de celle-ci. Ce clivage est couplé à la synthèse d'ADN (clivage dépendant de l'activité polymérase) ; (B) Clivage dirigé par l'extrémité 5' de l'ARN et situé à 13-19 nucléotides de celle-ci, dans lequel l'extrémité 5' de l'ARN est en retrait sur un long fragment d'ADN ; (C) clivage interne. Les clivages de type (B) et (C) sont indépendants de l'activité polymérase de la RT (228). Des expériences de clivage par RNase H ont révélé que les substrats ARN-ADN sont en contact avec les sites polymérase et RNase H en même temps (240). Au cours des clivages polymérase-indépendant, la NC contribue à dissocier les courts duplexes résultants du clivage

Introduction

primaire de l'ARNg (241, 242). Il a également été suggéré que la NC coopère avec la RNase H par formation d'un complexe NC-RT (131, 242–244), par un mécanisme doigt de zinc-dépendant (230).

De nombreuses recherches effectuées par l'équipe de Zhuang *et al.* montrent clairement que l'orientation de la RT lorsqu'elle est complexée à des duplexes ADN-ARN détermine son action, soit la polymérisation et synthèse d'ADN ou hydrolyse du brin d'ARN (245–247).

4. Transcription inverse de l'ARN génomique du VIH

Bien que dans quelques particules virales la transcription inverse (**Figure 32**) est initiée avant l'entrée dans la cellule (248, 249), ce processus débute principalement lorsque le virion pénètre dans le cytoplasme de la cellule hôte.

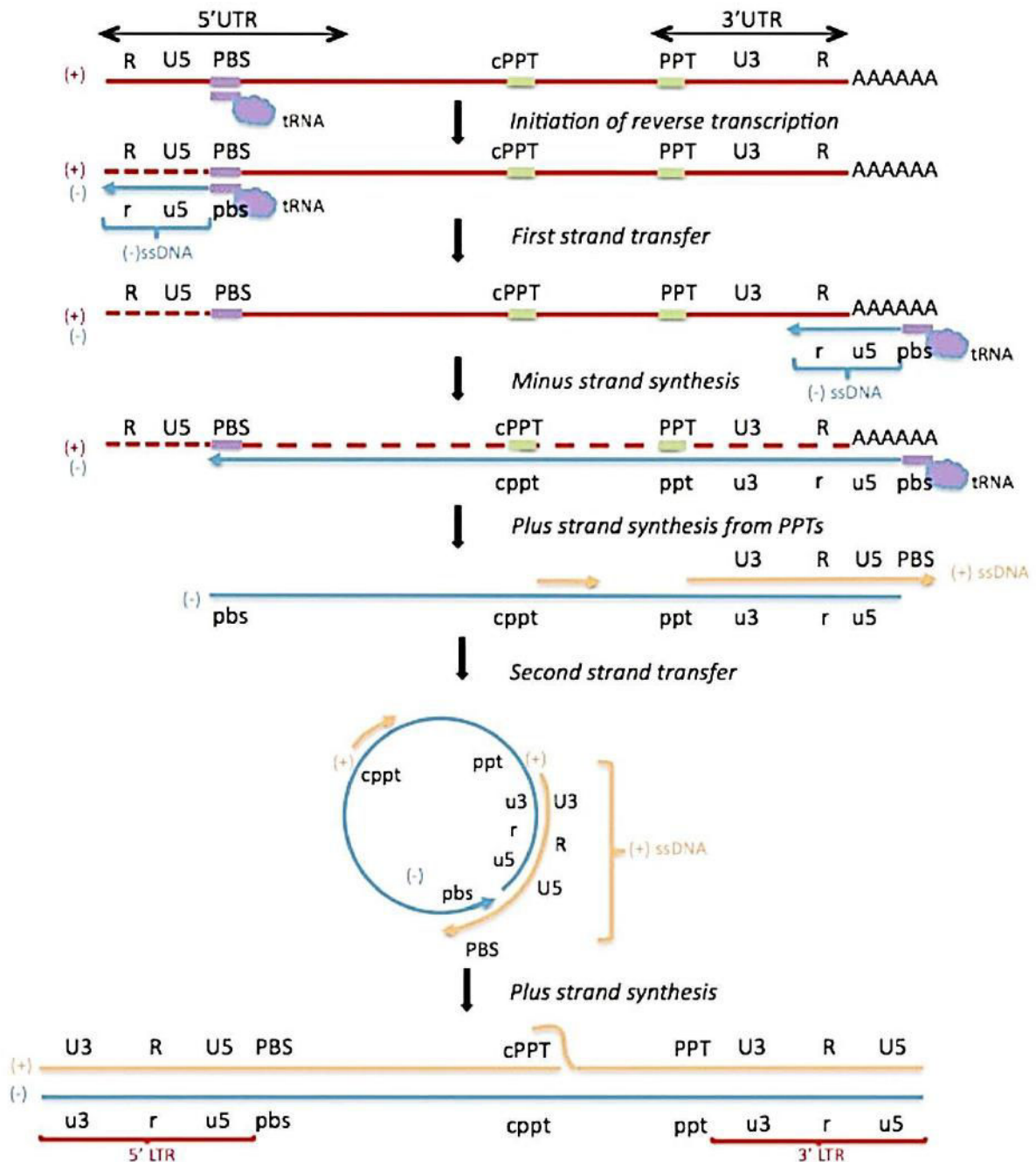


Figure 32. Représentation schématique du déroulement de la transcription inverse chez le VIH-1. L'ARN^{lys} est représenté par un trait et un nuage en violet ; l'ARN viral en rouge ; le brin (-) d'ADN en bleu et le brin (+) en orange.

Introduction

Récemment, il a été montré que la SAMHD1 bloque la réplication du VIH-1 en réduisant le taux intracellulaire de désoxyribonucléoside triphosphate (114). Des études ont montré que des mutations dans le domaine NC affectent le déroulement de la transcription inverse (250–253). Plus précisément, la transcription inverse se produit avant le relargage des particules virales. Ces études suggèrent que la NC bloque la synthèse de l'ADNc viral lors de la formation et la maturation du virion. Plusieurs laboratoires ont développé des systèmes acellulaires afin d'identifier les mécanismes impliqués dans les étapes clés de la transcription inverse. Ainsi, les intermédiaires de la transcription inverse et les produits identiques à ceux produits par les cellules infectées peuvent être synthétisés dans des virions purifiés (appelé transcription inverse endogène naturelle ou NERT) (249, 254–256). La transcription inverse dans les virions purifiés peut être amplifiée en la présence de désoxyribonucléotides et de détergents doux (appelé transcription inverse endogène ou ERT) (257, 258). En revanche, le rendement du produit complet obtenu dans les essais ERT et NERT est extrêmement bas (254).

La transcription inverse de l'ARNg simple-brin est caractérisée par une suite d'événements complexes qui aboutissent à la synthèse d'une copie ADN double-brin linéaire qui est entièrement intégrée dans l'ADN de la cellule l'hôte (**Figure 32**) (225, 259). Ce processus catalysé par la RT est un événement précoce post-entrée se produisant dans le cytoplasme des cellules infectées. L'activité chaperonne de la NC est nécessaire au processus de transcription inverse (64, 134, 167).

4.1 La région de fixation de l'amorce ARNt^(Lys, 3)

Localisée au niveau de l'extrémité 5'UTR, la séquence *Primer Binding Site* (PBS) est complémentaire des derniers 18 nucléotides de l'ARNt^(Lys,3), elle est nécessaire pour l'initiation de la transcription inverse (59, 259). Des expériences de *Selective 2'-Hydroxyl acylation Analysed by Primer Extension* (SHAPE) effectuées avec l'ARNg extrait à partir de virions et cellules, ont permis au groupe de Weeks de proposer un modèle de structure secondaire de la région 5'UTR (260) (**Figure 33**). Dans ce modèle, la séquence PBS est totalement appariée avec les 18 nucléotides de l'extrémité 3' de l'ARNt^(Lys,3). Les trois premiers nucléotides à être copiés sont, en revanche, non appariés et dans une conformation accessible à la RT (261). Des études *in vitro* réalisées avec des sondes de structure (ribonucléases et agents chimiques) suggèrent qu'en absence d'ARNt^(Lys,3), la région PBS de l'isolat LAI est quasiment non appariée (262, 263) tandis qu'elle est majoritairement appariée chez l'isolat MAL (155).

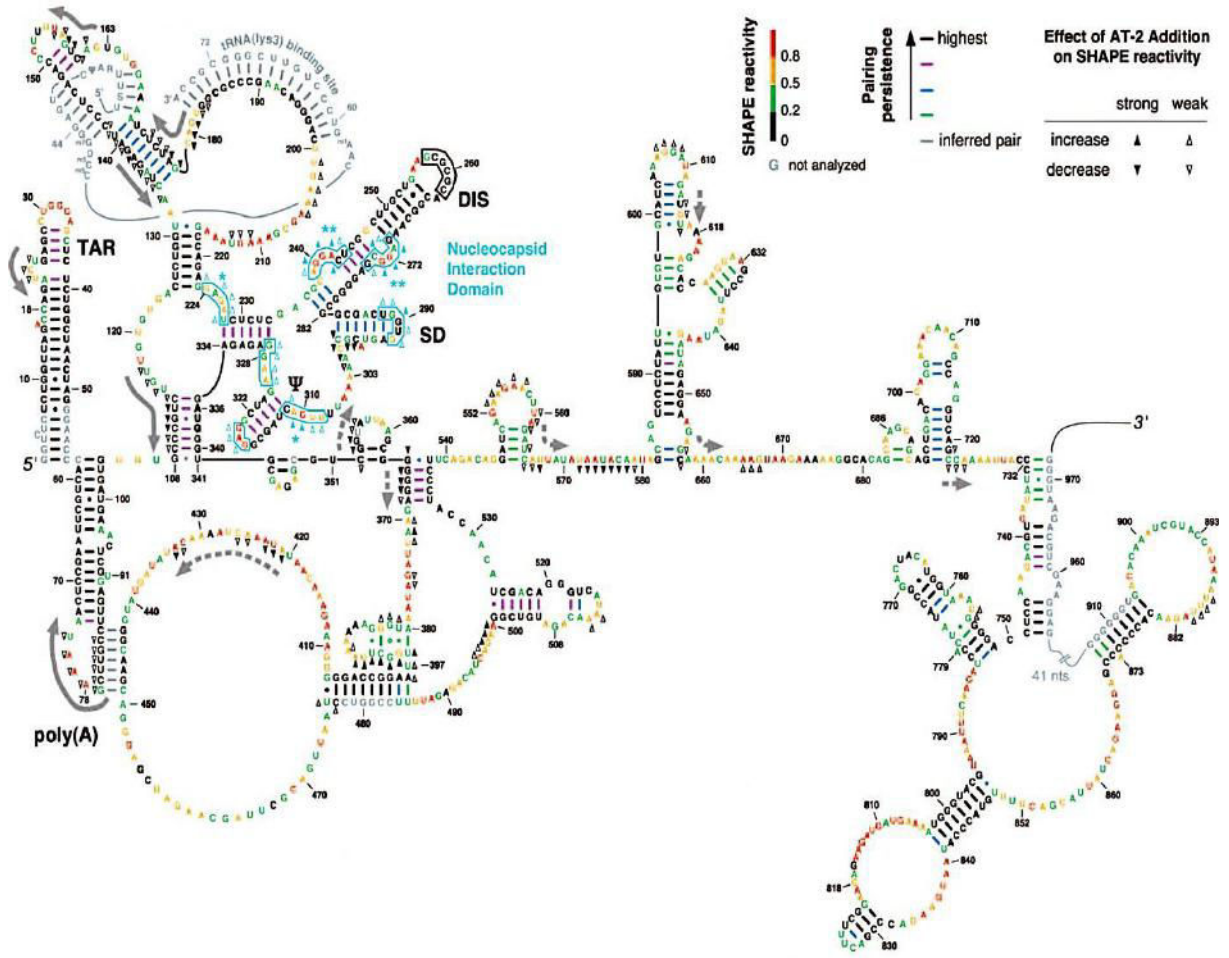


Figure 33. Modèle de structure secondaire de la région 5'UTR du VIH-1 (260). Les régions interagissant avec l'ARNt^(Lys,3) (gris) et les sites préférentiels de fixation de la NC (cyan) sont indiqués, DIS (SL1), SD (SL2) et ψ (SL3) sont également indiqués. Les nucléotides sont colorés selon leur réactivité vis-à-vis d'un agent SHAPE (voir le diagramme de réactivités), les séquences nucléotidiques entourées en bleu correspondent à des sites de fixation de la NC. AT-2 (*Aldrithiol-2*) est un agent chimique qui rompt l'interaction entre la NC et l'ARN a permis d'identifier les nucléotides liés à la NC, les triangles orientés vers le haut ou vers le bas indiquent respectivement, l'augmentation ou la diminution de la réactivité des nucléotides après traitement par AT-2. Les sites primaires et secondaires de liaison à la NC sont représentés par un ou deux astérisques, respectivement.

4.2 Formation de l'hybride ARN viral-ARNt^(Lys, 3)

4.2.1 Hybridation de l'ARNt^(Lys,3) à la séquence PBS

Les 18 nt de l'extrémité 3' de l'amorce ARNt^(Lys,3) sont parfaitement complémentaires des 18 nt du PBS situés dans la région 5'UTR de l'ARNg viral (**Figure 33**). Au cours du processus d'assemblage de la particule virale, il y a incorporation sélective de trois ARNt^{Lys} cellulaires iso-accepteurs (264–266) avec une abondance équivalente entre l'ARNt^(Lys,3) et les ARNt^(Lys-1,2) (267), en revanche seul l'ARNt^(Lys,3) sert d'amorce lors de la transcription inverse

Introduction

de l'ARNg (268, 269). Des études ont montré qu'une mutation ponctuelle de la séquence PBS (T au lieu de C en position 8 dans la séquence PBS) due aux éventuelles erreurs de la RT est réversible (270). En effet, 50% des molécules d'ARNg ayant la séquence PBS mutée peuvent fixer l'ARNt^(Lys-3), ce qui restaure la séquence PBS sauvage au niveau de l'ADN pré-intégré. En d'autres termes, par sa fréquence de fixation à la séquence PBS sauvage ou mutée, l'ARNt^(Lys) amorce détermine la séquence PBS dans les virions filles (270). Si l'hybridation de l'ARNt^(Lys,3) à l'ARNg a lieu avant ou après la sélection du génome est peu claire, en revanche il est probable que l'hybridation *in vivo* est facilitée par l'activité chaperonne de la NC (64, 167, 271). Les acides aminés basiques de la NC sont principalement impliqués dans l'appariement des séquences complémentaires ARNt et séquence PBS, tandis que les deux domaines à doigts de zinc sont responsables de la déstabilisation des régions structurées de l'ARNg (64, 167, 271). Des études ont montré que le placement de l'ARNt sur le PBS se produit, également, dans les virions dépourvus d'activité PR, ce qui suggère que le domaine NC de Gag facilite l'hybridation *in vivo* (272). Des études plus récentes ont montré que le complexe ARNt-PBS formé par Gag n'est pas aussi stable que celui formé par la NC (273), ce qui suggère que le processus d'hybridation s'effectue en deux étapes : initiation de l'hybridation de l'ARNt par le domaine NC de Gag, suivie d'un remodelage du complexe ARNt-PBS par la NC mature. Par ailleurs, Roldan *et al.* (274) ont constaté que l'hybridation médiée par la NC est plus efficace lorsque la NC est combinée au domaine C-terminal de CA. Ces observations suggèrent que d'autres domaines protéiques des précurseurs Gag ou Gag-Pol peuvent influencer les interactions ARNg-ARNt.

Le remodelage de l'ARNg par la NC facilite probablement le processus de transcription inverse. Par exemple, une étude *in vitro* suggère que la NC déstabilise les tiges-boucles et d'autres structures secondaires dans la région 5'UTR pour empêcher les pauses ou le décrochage de la RT (64). La méthode SHAPE a été utilisée pour localiser les régions d'ARNg affectées par l'activité chaperonne de la NC (260). De manière surprenante, les tiges-boucles déstabilisées par la NC ne sont pas distribuées de façon uniforme dans l'ARNg. En effet, six sites double-brin situés entre l'extrémité 5' de l'ARNg et la séquence PBS (les 185 premiers nt) sont préférentiellement déstabilisés par la NC (260) (**Figure 33**). Ces sites sont dans la région contenant le site d'hybridation de l'ARNt, ainsi que la séquence rétro-transcrite par la RT lors de la synthèse de l'ADN *strong-stop* (ADNss). Par conséquent, cette activité de déstabilisation localisée est compatible avec le rôle de la NC consistant à faciliter, à la fois, l'hybridation de l'ARNt et le processus de transcription inverse.

Introduction

Le moment où l'ARNt^(Lys,3) est déplacé de la séquence PBS reste encore inconnu, mais le déplacement est nécessaire pour permettre à la RT de copier le PBS (+) et les 18 nt de l'extrémité 3' de l'ARNt (Figure 32). Le processus de transcription inverse qui est largement accepté (Figure 32), est compatible avec le déplacement de l'amorce ARNt^(Lys,3) avant ou après le premier transfert de brin.

4.2.2 Interactions supplémentaires entre l'ARNt^(Lys,3) et la région 5'UTR

Des interactions supplémentaires entre le 5' UTR et l'ARNt^(Lys,3) ont été identifiées *in vitro* (Figure 34). La première interaction qui a été identifiée se fait entre la boucle de l'anticodon de l'ARNt^(Lys,3) et une boucle conservée riche en A (nucléotides 168-171 dans la Figure 33) située en amont du PBS (275, 276). Cette interaction boucle-boucle entre la matrice et l'amorce a été observée avec un ARNt^(Lys,3) naturel, mais pas avec un ARNt^(Lys,3) synthétique dépourvu des nucléotides naturels modifiés (276, 277). La seconde interaction se produit entre la séquence 48-55 de l'ARNt^(Lys,3) et une séquence conservée de 8 nt, en amont du PBS (nt 123-130), nommée *primer activation signal* (PAS) (Figure 33) (278–280).

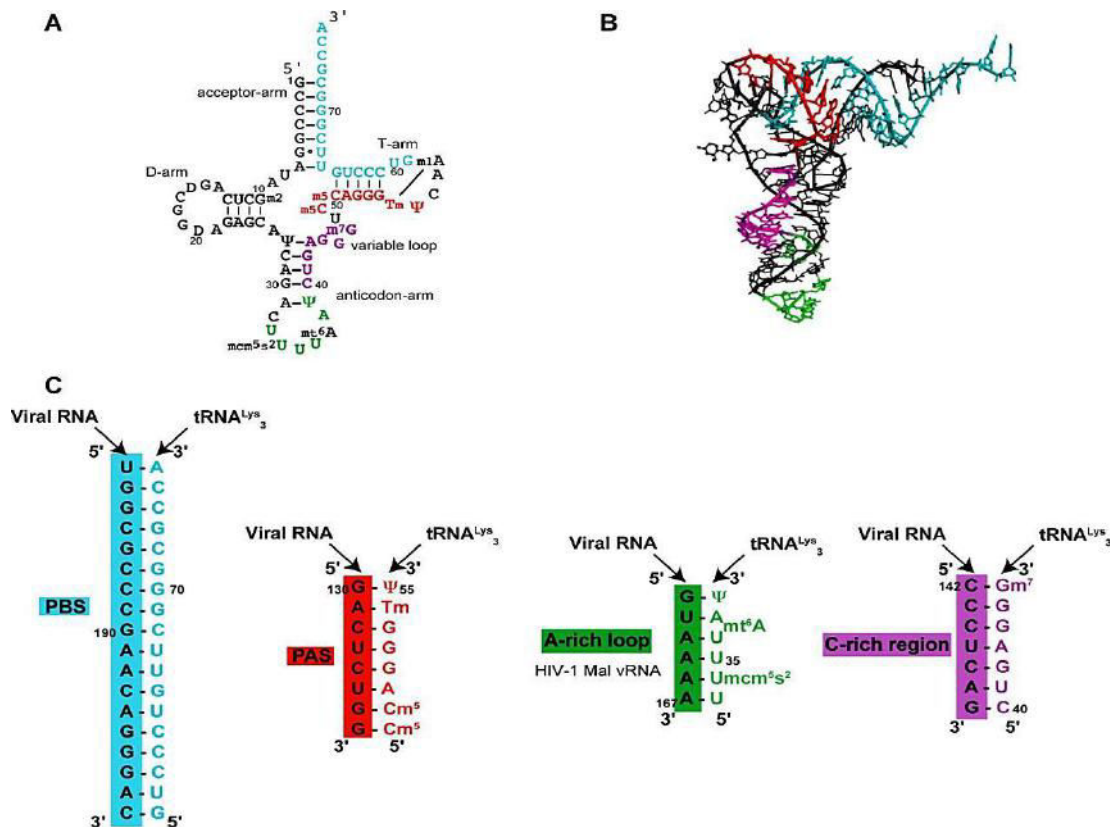


Figure 34. Interactions entre le 5'UTR et l'ARNt^(Lys,3) (adapté de (265). **A)** Structure secondaire de l'ARNt^(Lys,3). **B)** Structure 3D de l'ARNt^(Lys,3). Les motifs complémentaires à l'ARN viral sont indiqués ainsi : séquence complémentaire au PBS en bleu, PAS en rouge, boucle riche en A en vert et la région riche en C en violet. **C)** Interactions par appariement de bases entre l'ARNg et l'ARNt^(Lys,3).

Des expériences de mutagenèse dirigée réalisées sur l'ARNg suggèrent l'idée que la NC favorise une interaction entre l'ARNt^(Lys,3) (en particulier, le bras 3' de la tige de l'anticodon et une partie de la boucle variable) et la séquence 142-148 de l'ARN NL4-3 (281).

4.3 Initiation de la transcription inverse

Des études au moyen de nucléases réalisées en l'absence de NC ont montré que l'interaction de l'ARNt^(Lys,3) avec la boucle riche en A est plus importante dans la souche MAL que dans la souche HXB2 pour une initiation efficace de la transcription inverse *in vitro* (276, 282–284). En outre, l'interaction entre la séquence 48-55 de l'ARNt^(Lys,3) et le PAS favorise l'initiation de la transcription inverse *in vitro* dans les souches HXB2 et LAI (262, 278, 280). Il a été montré que les interactions supplémentaires entre l'ARNt amorce et l'ARNg facilitent la liaison de la RT au substrat en évitant des encombrements stériques entre la RT et le duplexe d'AN (261, 285). En l'absence de NC, une synthèse d'ADNss efficace *in vitro* nécessite la présence d'au moins 24 nt en aval de la séquence PBS de l'ARN matrice, lorsque l'amorce est l'ARNt^(Lys,3) (281). En revanche, les 24 nt ne sont pas nécessaires en présence de NC. Des expériences de gel retard ont montré qu'en l'absence de dNTP, la NC n'affecte pas la liaison de la RT aux complexes formés soit avec l'ARN sauvage ou des matrices mutantes ayant des modifications dans la séquence 143-149 (281). Un essai similaire a été réalisé avec l'ARNt amorce allongé d'un nucléotide, les résultats ont révélé que la NC stimule l'incorporation avec le type sauvage, mais pas avec les mutants (64). Par conséquent, l'action de la NC dépendrait de la liaison de la RT au substrat et à l'extension de l'amorce d'au moins 1 nt. D'autre part, l'équipe de Zhuang a montré qu'au cours du processus d'initiation de la transcription inverse, la RT adopte deux orientations : une orientation favorisant l'activité polymérase, et une orientation inversée dans laquelle le site actif de la polymérase se place à quelques nucléotides de l'extrémité 3' de l'ARNg (246). L'équilibre entre ces deux orientations jouerait un rôle dans la régulation de l'activité de la RT au cours de l'initiation de la transcription (**Figure 35**). En effet, la présence de structures secondaires orienterait la RT vers la position non favorable à la polymérisation conduisant à la pause de la RT. Ainsi, l'ajout des premiers nucléotides à l'amorce ARNt conduit à un changement d'orientation de la RT vers la position favorisant la polymérisation (**Figure 35**).

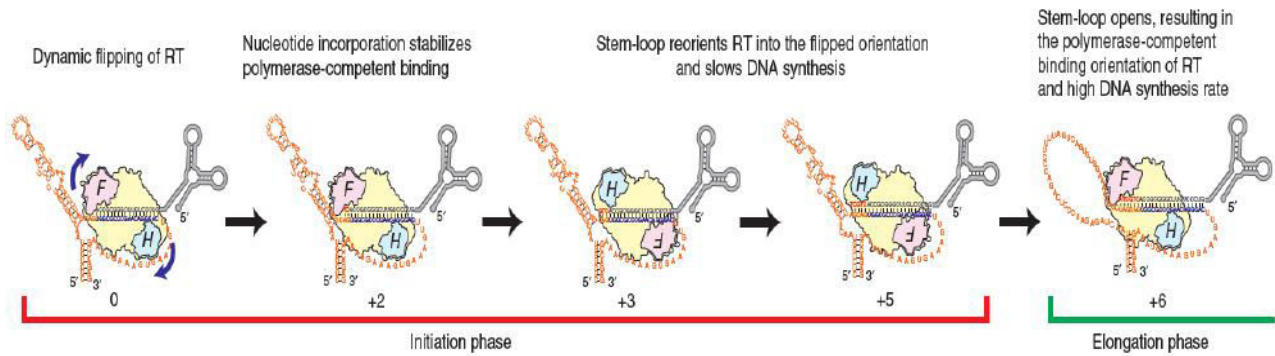


Figure 35. Orientation de la RT au cours de l'initiation de la transcription inverse (246). La RT se lie au complexe d'initiation de la transcription suivant deux orientations : une orientation favorable à l'activité polymérase et une orientation inversée dépourvue d'activité polymérase. La majeure partie du temps, la RT se lie au complexe ARNt-ARNg suivant l'orientation inversée. L'ajout de deux nt à la l'ARNt amorce induit un changement d'orientation de la RT vers la conformation active ce qui augmente le taux de transcription inverse. La synthèse d'ADN se ralentit lorsque la RT rencontre une structure secondaire au niveau de l'ARNg en position +3, ce qui force la RT à se repositionner en conformation inversée augmentant ainsi la probabilité de pause. La synthèse par déplacement de brin jusqu'en position +6 pourrait conduire à la destabilisation de la structure secondaire. Ainsi, la RT change de nouveau de conformation vers l'orientation active et la synthèse d'ADN s'accélère au cours de la phase d'élongation.

4.4 Synthèse de l'ADN strong-stop

L'extension de l'ARNt amorce par la RT conduit à la synthèse de l'ADNss, un court ADN qui est la copie de la région R-U5 de l'ARNg (**Figure 32**). L'ADNss a été découvert lors de la synthèse *in vitro* de l'ADN du virus du sarcome de Rous (RSV) par l'ADN polymérase endogène du virus (286). Le produit majeur de cette synthèse a été un fragment d'ADN de 101 nucléotides de long mais pas l'ADN viral de pleine longueur, il a ainsi été nommé ADN *strong-stop* (ADNss). Il y a très peu d'ADN viral synthétisé avant l'entrée du virion dans la cellule hôte (248, 249, 287). En effet, il y a environ une molécule d'ADNss pour mille molécules d'ARNg (249). Par ailleurs, dans les particules virales du VIH-1, le nombre de molécules d'ADNss est 10 à 100 fois plus élevé que le nombre de molécules d'ADN synthétisées après le premier transfert de brin (248, 249). Lorsque les particules virales sont mises en contact avec des dNTP, environ une copie d'ADNss est synthétisée pour une molécule d'ARNg (249).

Le clivage par la RNase H se déroule pendant la synthèse de l'ADNss, et génère de courts hybrides dont la stabilité thermodynamique est réduite. L'activité destabilisatrice de NC permet la dissociation des courts duplexes et la libération de l'ADNss qui a la possibilité de se replier (288).

4.5 Le premier transfert de brin

La synthèse complète du brin (-) d'ADN viral nécessite que l'ADNss soit transféré vers l'extrémité 3' de la même molécule d'ARNg ou de la seconde molécule d'ARNg incorporée dans la même particule virale (289). Ce processus est désigné comme « le premier transfert de brin » et est très efficace *in vivo*, car l'ADNss ne s'accumule pas dans les cellules infectées (290) (**Figure 32**). Cette partie sera détaillée dans le chapitre suivant.

4.6 Synthèse complète du brin (-) de l'ADN viral

Une fois l'étape de transfert de brin achevée, la RT catalyse la polymérisation de l'ADNss, qui est accompagnée d'une dégradation supplémentaire de l'ARNg par l'activité RNase H de la RT (**Figure 32**). De très faibles quantités du brin (-) d'ADN de pleine longueur sont détectée dans les essais ERT (258). Par ailleurs, la synthèse du brin (-) d'ADN peut également être réalisée *in vitro* dans des réactions totalement reconstituées comprenant uniquement une amorce, une matrice d'ARN, la RT et la NC. En utilisant un ARNg matrice de 874 nt *in vitro*, Drummond et al. (291) ont montré qu'en présence de concentrations élevées de NC, presque tous les produits d'ADN synthétisés sont de pleine longueur avec, en revanche, un faible rendement. Toutefois, une copie de l'ADN de pleine longueur d'un ARNg matrice de 7 kb ne peut être détecté dans ces essais reconstitués (292). En présence de NC, un brin (-) d'ADN d'approximativement 4 kb peut être synthétisé *in vitro*, cependant, le rendement est faible (292). Les positions relatives des domaines à doigts de zinc de la NC ne sont pas importantes au cours du processus d'élongation. Etant donné que l'ARNg matrice est très structuré (59), la polymérisation peut être ralentie lorsque la RT rencontre une structure secondaire. L'activité déstabilisatrice de la NC réduit la pause, permettant ainsi à la RT de poursuivre la polymérisation sans obstacle (292, 293). En outre, il a été montré que d'importants agrégats sont formés en présence de NC, de RT et des acides nucléiques (131, 172, 244, 294). Ainsi, l'activité d'agrégation de la NC est probablement plus importante que son activité de déstabilisation des duplexes lors de la synthèse de longs ADN (292).

4.7 Synthèse du brin (+) d'ADN à partir des PTT

Bien que presque l'ARNg entier est dégradé par la RNase H au cours de la synthèse du brin (-) d'ADN, il subsiste deux régions de 15 nt contenant uniquement des bases puriques. Ce sont les *Poly-purine tract* (PPT) qui fonctionnent comme des amorces pour la synthèse du brin (+) d'ADN : (1) le 3' PPT se situe à l'extrémité 5' de U3 ; (2) le cPPT est au centre du gène IN (**Figure 32**). Les séquences des deux PPT (5'-AAAAGAAAAGGGGGG-3') sont

Introduction

identiques (236, 295, 296). Les duplexes PPT ne sont pas comme les autres hybrides ARN-ADN, car ils présentent des motifs structuraux qui confèrent une résistance au clivage par la RNase H. L'enchaînement des A au niveau de l'extrémité 5' contribuerait à la reconnaissance des PPT et au positionnement du site actif de l'activité polymérase (295) (**Figure 36A**). Des études utilisant différentes techniques biochimiques, biophysiques (incluant la cristallographie aux rayons X), les sondes chimiques seules ou associées à la spectrométrie de masse et la RMN ont révélé que la structure adoptée par le PPT est cruciale (295–298).

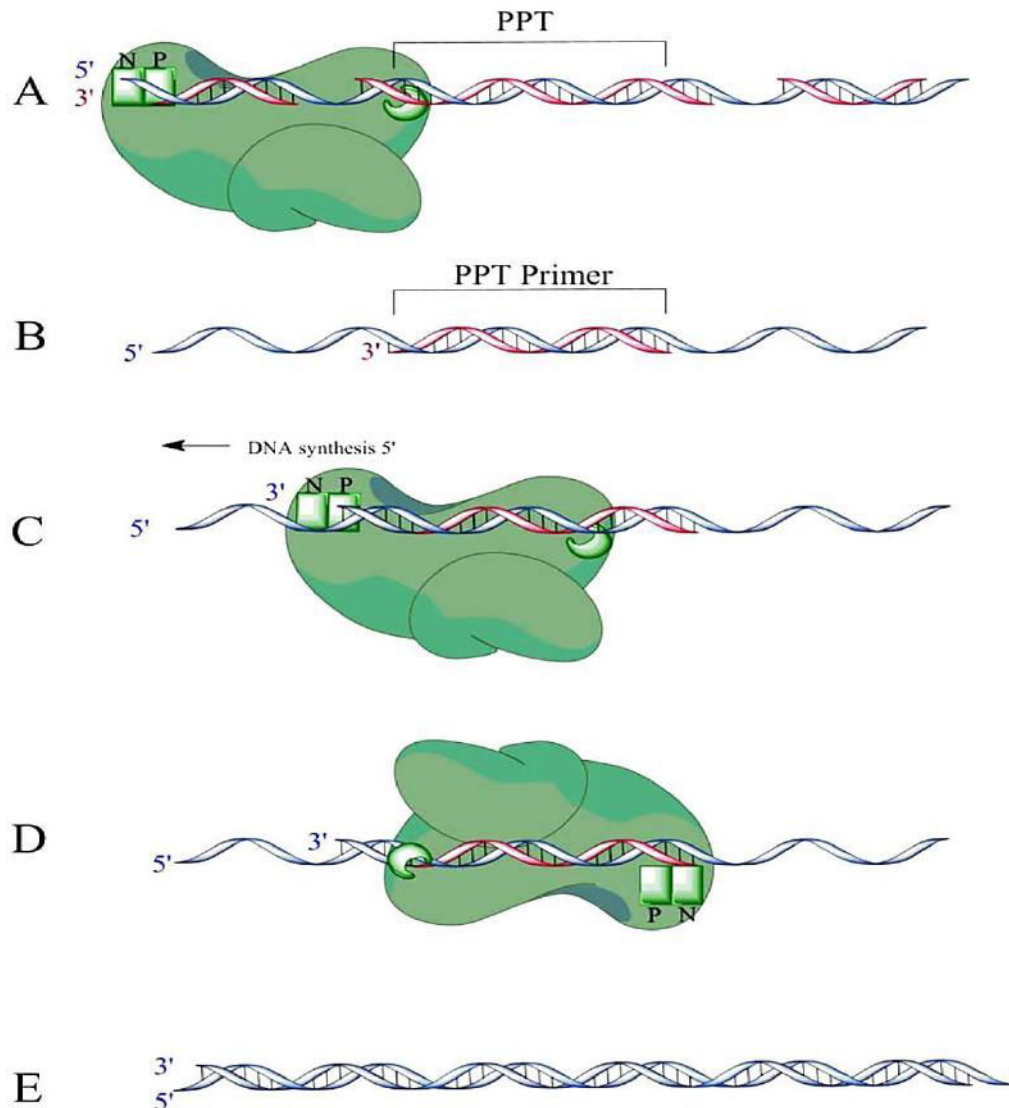


Figure 36. Synthèse du brin (+) d'ADN à partir d'un PPT (adapté de (299)). **A**) Lors de la synthèse du brin (-) d'ADN, la région PPT est résistante au clivage par la RNase H tandis que le reste de l'ARNg est dégradé. **B**) La séquence PPT ainsi formée sert d'amorce à la synthèse du brin (+) d'ADN. **C**) L'amorce PPT est allongée de 12 nt. **D**) La RT effectue une pause et change d'orientation ce qui lui confère une activité RNaseH polymérase indépendante lui permettant d'effectuer des clivages au niveau des jonctions ARN-ADN et de dégrader le PPT. **E**) la séquence de 12 nucléotides générée est allongée jusqu'à l'extrémité 5' du brin (-) d'ADN. A la fin du second transfert de brin, le brin (+) d'ADN est entièrement synthétisé formant ainsi un ADN double-brin.

Introduction

L'extrémité 5' du duplexe PPT-ADN comprend un empilement classique de paires de bases de type Watson Crick et adopte une conformation hélicoïdale de forme A, bien que certaines anomalies structurales aient été détectées par RMN (298). Les résultats de l'étude par RMN (298) et ceux d'une étude ancienne (237) sont compatibles avec une flexibilité inhérente du duplexe PPT-ADN. Il a été suggéré que lors de la liaison de la RT, des perturbations structurales se produisent le long de la séquence PPT permettant l'accès aux résidus présents dans différents sous domaines de la RT, réalisant une liaison optimale et un clivage précis à la jonction PPT-U3 (**Figure 36 A-B**) (296). L'amorce 3'PPT est générée par un clivage précis de la RNase H à la jonction PPT-U3. L'extrémité 3' de l'amorce ARN nouvellement formée est prolongée par la RT jusqu'à la position +12 en aval de l'ARN PPT (**Figure 36C**). La RT fait une pause puis, se réoriente et fait un clivage spécifique à la jonction PPT-U3 (**Figure 36D**). Ce clivage est important car il définit l'extrémité 3' LTR servant de substrat à l'IN virale. Les amorces PPT présentant des clivages aberrants ou qui initient mal la synthèse d'ADN ne sont pas utilisées de manière efficace, empêchant, ainsi, l'avancement de la transcription inverse (209, 300). Etant donné que les sites actifs de la RT sont positionnés simultanément sur chaque brin d'AN, la RT doit reconnaître sa position sur le génome et changer dynamiquement son orientation selon la tâche à accomplir (245). Ainsi, afin d'initier efficacement la synthèse du brin (+) d'ADN, la RT doit se lier comme une polymérase pour étendre l'amorce et comme une RNase H pour cliver à la jonction ARN-ADN. De ce fait, la réaction d'élimination de l'amorce nécessite une activité RNase H polymérase indépendante. En revanche, le clivage de l'extrémité 5' servant à supprimer le PPT est moins précis (236).

La synthèse du brin (+) d'ADN nécessite que la RT allonge l'amorce 3' PPT en utilisant comme matrice le brin (-) d'ADN, encore lié à ARNt^(Lys,3) (**Figure 32**). Le produit de cette extension d'ADN est constitué des régions U3-R-U5-PBS et est nommé ADN *strong-stop* (+) (ADNss(+)). La RT ne copie pas l'intégralité de l'ARNt^(Lys,3) car la polymérisation est stoppée par la présence d'une adénine méthylée en position 58 (**Figure 34A**). L'élimination de l'amorce nécessite un clivage primaire par la RNase H qui ne se produit pas à la jonction ARNt-ADN mais plutôt entre le résidu A (rA) de l'extrémité 3' de l'ARNt et l'avant-dernier résidu C (rC). Cela conduit à la présence d'une adénosine et non d'un désoxyribonucléotide à l'extrémité 5' de l'ADN(-) et formation d'un duplexe de 17 pb entre l'extrémité 3' de l'ARNt et la séquence PBS de l'ADNss(+) (64).

L'utilisation de systèmes reconstitués *in vitro* modélisant les étapes du second transfert de brin a montré que le clivage primaire par RNase H n'est pas suffisant pour l'élimination

complète de l'ARNt, probablement en raison de la stabilité de l'hybride de 17 pb (301, 302). Ainsi, des clivages secondaires par la RNase H sont également nécessaires. D'autre part, l'activité chaperonne de la NC pourrait participer aussi à l'élimination de l'extrémité 3' de l'amorce ARNt par dissociation des duplexes ARN-ADN. En effet, l'élimination de l'amorce est plus efficace lorsque la RNase H et la NC sont présentes. Des études ont démontré que les doigts de zinc sont nécessaires à l'élimination de l'amorce, ce qui indique que l'activité déstabilisatrice de la NC est critique au cours de cette étape (174, 303).

4.8 Second transfert de brin

Le second transfert de brin est un processus intramoléculaire qui est nécessaire à la synthèse intégrale du brin (+) d'ADN viral (304). Au cours de ce transfert l'ADNss(+) est transféré à l'extrémité 3' du brin d'ADN(-). Ce processus repose sur l'appariement des 18 nt de la séquence PBS, située à l'extrémité 3' de l'ADNss(+), appelé (+)PBS et la séquence complémentaire située à l'extrémité 3' de l'ADN(-) (**Figure 32**).

Des études *in vitro* ont montré que la NC facilite l'hybridation responsable du second transfert de brin (302). Une étude par RMN suggère que les changements conformationnels induits par la NC dans la région pbs(-) facilitent l'hybridation avec le PBS(+) (305). Une autre étude RMN, réalisée avec la NC ou une forme tronquée (12-55) et le pbs(-), montre que la fixation de la protéine dépend des résidus hydrophobes des doigts de zinc (161). Des études de fluorescence ont montré que la NC déstabilise faiblement les tiges-boucles PBS et elle joue un rôle dans le processus d'hybridation du PBS par son activité agrégative des AN (306).

4.9 Synthèse de l'ADN viral double brin possédant deux LTR

L'utilisation des deux PPT implique la synthèse d'un ADN(+) discontinu, qui génère un fragment d'ADN simple-brin d'environ 99 nt au centre de l'ADN(+) double-brin non intégré (**Figure 32**) (307). Cela conduit à la formation d'une structure triple brin nommée *Flap* qui contient deux segments d'ADN(+) se chevauchant et un brin (-) complémentaire. Le segment non apparié a été initié à partir du cPPT, tandis que le segment en amont a été initié à partir du 3'PPT. L'élimination du *Flap* et la réparation des coupures dans l'ADN double-brin n'a pas été démontrée *in vivo*, on suppose que des enzymes cellulaires seraient impliquées (295). Deux études indépendantes ont présenté des résultats indiquant que le cPPT protège le VIH-1 de l'activité des facteurs de restriction APOBEC3 humains (308, 309).

La formation d'un intermédiaire circulaire au cours du second transfert de brin permet à la RT d'allonger les brins (+) et (-) d'ADN viral (**Figure 32**). Suite à la circularisation, la RT

synthétise 636 nt pour générer la séquence LTR (U3-R-U5) aux extrémités des deux brins d'ADN. L'achèvement de la transcription inverse nécessite la synthèse par déplacement de brin, qui correspond à la synthèse d'un brin d'ADN à travers une région d'ADN double-brin. Ce processus est facilité *in vitro* par la NC (229, 310–312). Un échec dans la finalisation de la synthèse par déplacement de brin entraînerait la formation d'extrémités LTR incomplètes, ce qui créerait un ADN viral double-brin ne pouvant être intégré dans le génome de la cellule hôte.

4.10 Le transfert de brin et les recombinaisons génétiques

Les deux transferts de brin sont des étapes essentielles de la transcription inverse car ils sont requis pour la synthèse de l'ADN double-brin viral et la création des séquences LTR nécessaires à l'intégration. Puisque les particules rétrovirales contiennent deux copies identiques d'ARNg, les transferts de brin peuvent se produire de manière intra- ou intermoléculaires (313–316). Dans les cas où les particules virales contiennent deux génomes d'ARN non identiques, les transferts de brin intermoléculaires conduisent à la recombinaison, qui génère à son tour la diversité virale qui peut conduire à un mécanisme de résistance aux facteurs antirétroviraux de l'hôte et aux agents pharmacologiques (317–320).

En plus des deux transferts de brins terminaux, il a été montré qu'au cours de la transcription inverse, la RT peut changer fréquemment de matrice au niveau des régions internes. De nombreuses caractéristiques des premier et second transferts de brin sont également conservées au cours de ces transferts de brin internes. Lorsque les deux ARNs encapsidés sont génétiquement distincts, le transfert interne peut conduire à la production d'un virus recombinant (**Figure 37**). La faible fidélité de la RT crée des mutations qui peuvent être rapidement disséminées par les transferts de brins internes. Cela crée un environnement favorable pour l'évolution virale rapide par des recombinaisons.

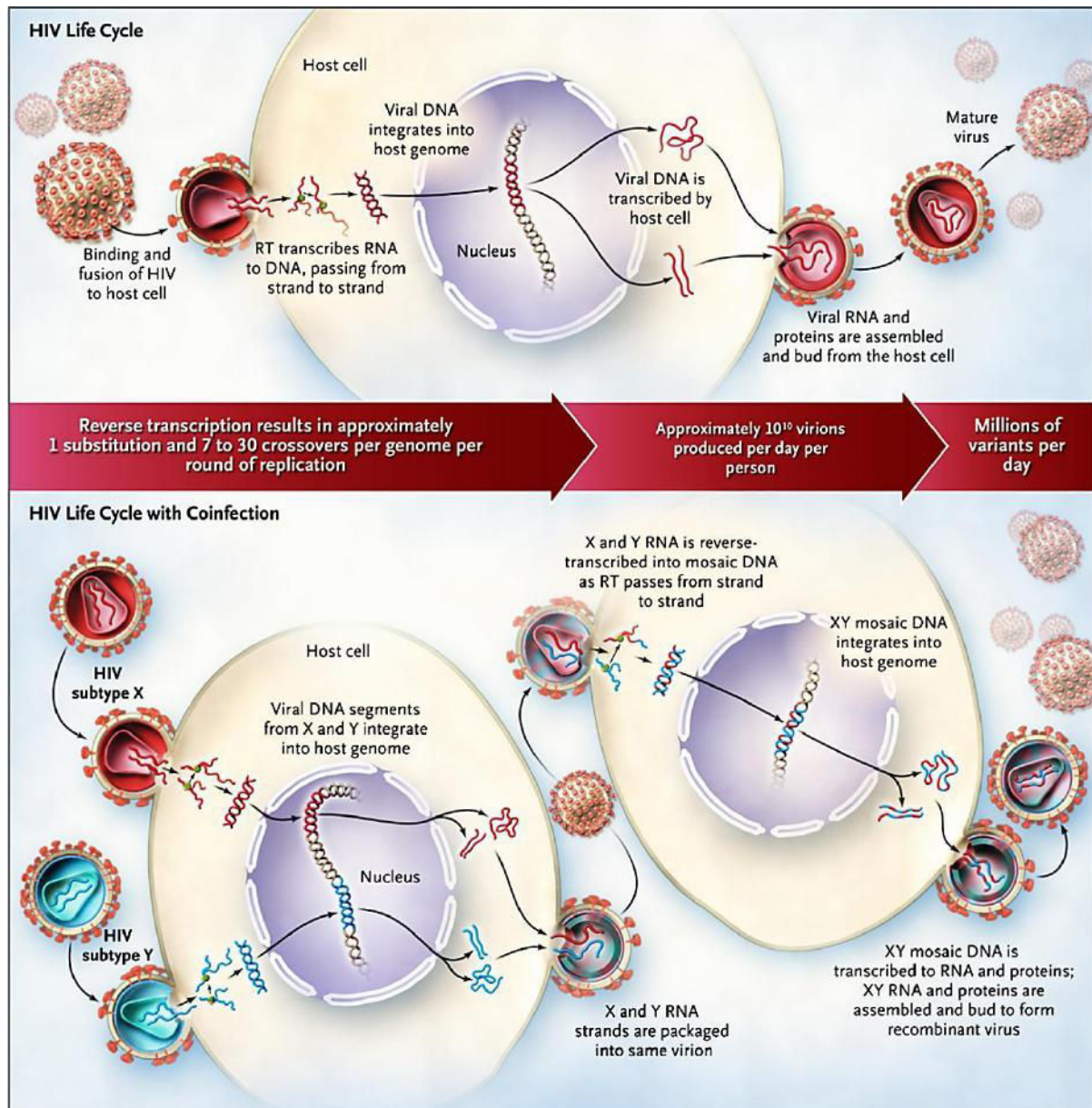


Figure 37. Evolution de la diversité génétique chez le VIH-1 au cours du cycle réplcatif viral et création de formes recombinantes uniques dans le cas d'une co-infection par deux sous-types (321).

4.10.1 Modèles de transfert de brin

Les modèles de transfert de brin pendant la synthèse du brin (-) comprennent le *forced copy choice*, le *copy choice* et le *dynamic copy choice* (322–324) (**Figure 38**). Les modèles *forced copy choice* et *copy choice* décrivent les configurations qui favorisent le transfert, tandis que le *dynamic copy choice* prend en compte les activités de la RT dans le transfert par *forced copy choice* ou *copy choice*. Dans le modèle *forced copy choice* proposé par Coffin (322), les coupures dans l'ARNg induisent le transfert de brin pendant la synthèse du brin (-). Dans ce modèle, l'extension d'amorce est initiée sur un ARNg et se poursuit jusqu'à la

Introduction

rencontre d'une coupure au niveau du brin matrice. A ce stade, sur la base de l'homologie de séquences, le brin d'ADN(-) est transféré vers le second ARNg co-encapsidé et la synthèse se poursuit. Le modèle de transfert de brin *copy choice* augmente la possibilité de changement de matrice pendant la synthèse du brin (-) incluant des régions du génome intactes (324). Similairement au modèle *forced copy choice*, l'homologie de séquence est une condition nécessaire au transfert de type *copy choice*. Dans le cas de recombinaisons non homologues, le taux de recombinaison moyen est d'environ 0,1 à 1,0 % du taux de la recombinaison homologue (249).

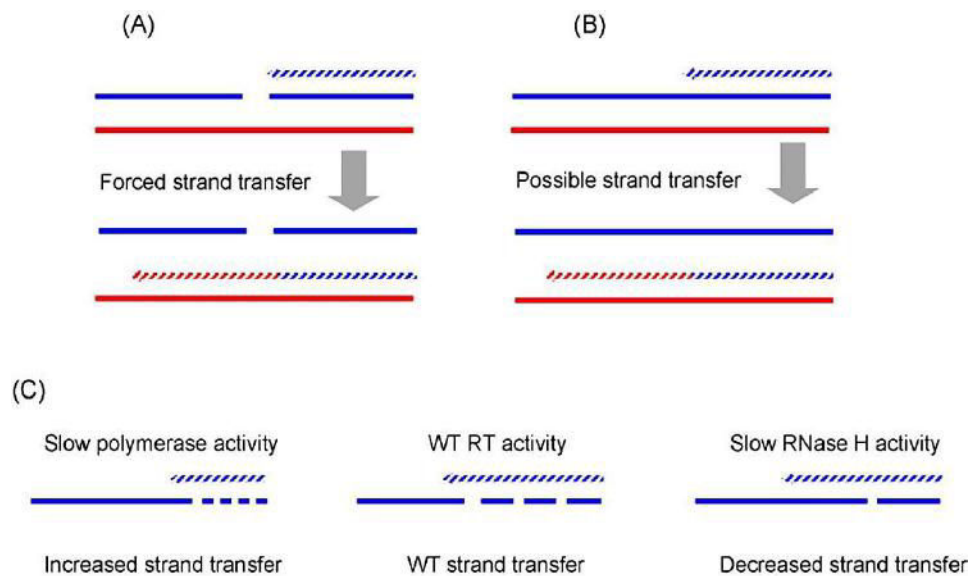


Figure 38. Modèles de transfert de brin au cours de la synthèse de l'ADN(-) (289). **A)** Le transfert de brin *forced copy choice*. L'ADNc (trait bleu discontinu) est polymérisé le long d'un ARN donneur (trait bleu continu) jusqu'à ce qu'une coupure dans l'ARN donneur est atteinte. Afin de poursuivre la synthèse d'ADNc, ce dernier doit être transféré vers l'ARN accepteur (trait rouge continu) pour être allongé (trait rouge à discontinu). **B)** Le transfert de brin *copy choice*. L'ADNc est polymérisé le long d'un ARN donneur (trait bleu simple) puis il est transféré vers l'ARN donneur. Il n'y a pas de configuration d'un ARN donneur matrice, comme la présence d'une coupure, forçant le transfert de brin. **C)** Le transfert de brin *dynamic copy choice*. Les activités enzymatiques de la RT (polymérase et RNase H) influencent l'efficacité du transfert de brin. Une activité polymérase lente induit des coupures RNase H plus fréquentes et une augmentation du transfert de brin. Par ailleurs, une activité RNase H plus lente réduit le transfert de brin.

Le modèle *dynamic copy choice* a été proposé par Pathak *et al*, (325, 326). Ce modèle est descriptif de certains facteurs qui influencent le transfert de brin de type *copy choice*. Le modèle propose que le rapport entre les activités polymérase et RNase H de la RT dicte l'efficacité du transfert de brin. En effet, dans ce modèle, un ralentissement de la vitesse de polymérisation favorise le clivage de la matrice par la RNase H et donc le transfert.

4.10.2 Mécanismes du transfert de brin

Les mécanismes du transfert de brin ont été étudiés *in vitro* en utilisant des systèmes d'ARN donneur-accepteur en tant que matrices. Plus précisément, la synthèse est initiée à partir d'une amorce hybridée à l'extrémité 3' de l'ARN donneur. Ainsi, l'extension de l'amorce est effectuée le long de l'ARN donneur qui est homologue à l'extrémité 3' de l'ARN accepteur. Ensuite, l'ADN nouvellement synthétisé est transféré vers l'ARN accepteur qui sert de matrice pour la poursuite de la synthèse d'ADN (327).

4.10.2.1 Le transfert de brin par invasion

Le transfert de brin par invasion est un mécanisme dans lequel la première interaction entre l'ADNc et l'ARN accepteur se produit en amont de l'extrémité 3' de l'ADNc, le processus est appelé « invasion de brin » et le site est appelé « site d'invasion » (**Figure 39**) (328). Ce dernier, est une région d'ADNc dégagée par l'activité RNase H de la RT et permet ainsi l'interaction entre l'ARN accepteur et l'ADNc. Un premier modèle proposé par Negroni et Buc (328) a proposé qu'au cours de la synthèse, l'activité chaperonne de la NC facilite l'hybridation de l'ARN accepteur à l'ADNc via des séquences d'AN complémentaires. La synthèse de l'ADNc se poursuivant, il y a extension de l'hybride ADNc-ARN accepteur nécessitant la déstabilisation des tiges-boucles préexistantes, tandis que des structures secondaires sont formées de manière transitoire. Le transfert est achevé par déplacement des duplexes ARN donneur-ADNc et copie de la région 5' de l'ARN accepteur.

Pour déterminer la position du site d'invasion, une série d'oligodésoxyribonucléotides ont été conçus pour interférer avec le transfert de brin (329). Chaque oligodésoxyribonucléotide est homologue à une région de l'ARN accepteur et peut, par conséquent, s'apparier avec une partie de l'ADNc. Lorsque les deux ARN et l'oligodésoxyribonucléotide sont ajoutés dans la réaction de transcription inverse, ce dernier entre en compétition avec l'ARN accepteur pour la liaison à l'ADNc complémentaire au site d'invasion. La formation de l'hybride ADNc-oligodésoxyribonucléotide diminue l'efficacité du transfert de brin. L'interaction de l'ARN accepteur avec l'ADNc a été aussi examinée en mesurant, en fonction du temps, la sensibilité de l'accepteur au clivage par la RNase H. Les premiers clivages définissent le point d'invasion. Dans le but de déterminer comment et où l'extrémité de l'ADNc en cours de synthèse change de matrice, des mutations ont été introduites tout le long de l'accepteur. Le point du transfert de brin peut, ainsi, être déterminé par séquençage du produit de transfert et identification des mutations ayant été incorporées. En utilisant ces méthodes, les études de plusieurs systèmes avec différentes matrices ont mis

en évidence que l'invasion se produit en amont de l'extrémité 3' de l'ADNc, soutenant ainsi le mécanisme de transfert par invasion (330–332).

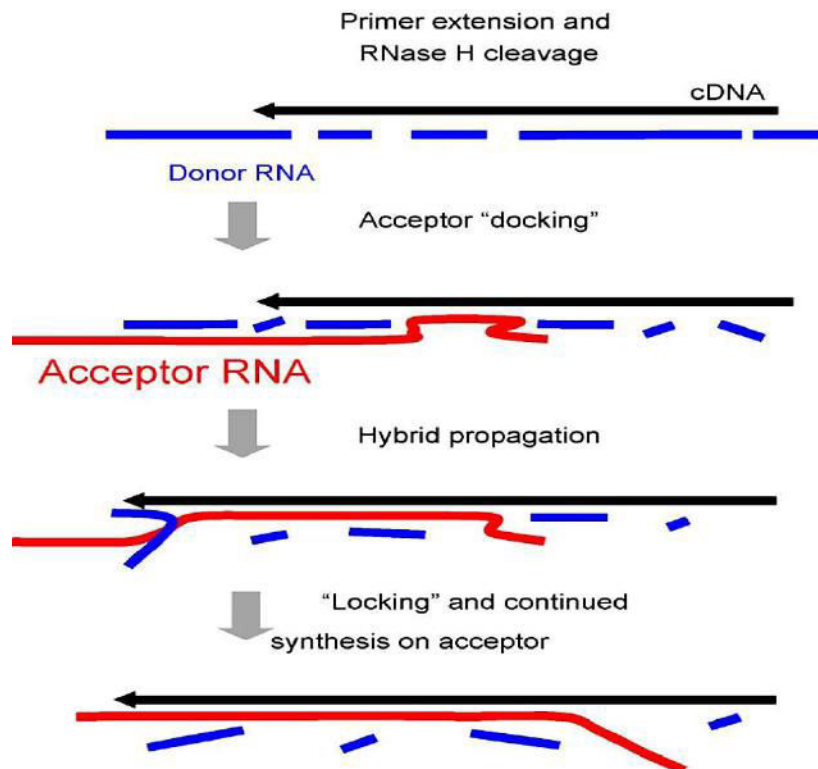


Figure 39. Transfert de brin par invasion (289). Au cours de la synthèse d'ADN (noir), la RT dégrade l'ARN donneur (bleu). La RNase H génère une série de coupures rapprochées créant une région d'ADNc simple-brin, ce qui permet à l'ARN accepteur (rouge) de s'hybrider à l'ADNc. Cette étape est désignée par le terme «docking» ou «invasion». L'hybride ARN accepteur-ADNc se propage par migration de branche, déplaçant les fragments de l'ARN donneur clivé. Finalement, l'extrémité 3' de l'ADNc se dissocie de l'ARN donneur et s'hybride avec l'ARN accepteur, dont la partie 5' est ensuite copiée par la RT.

4.10.2.2 Le transfert pause-dépendant

Les structures secondaires dans l'ARNg, tels que les tiges-boucles et tétrades de G peuvent ralentir la RT pendant la synthèse d'ADN et favoriser ainsi le clivage de l'ARN donneur par l'activité la RNase H (246, 333, 334), ce qui facilite l'interaction entre l'ADNc et l'ARN accepteur (271, 327, 335–337). Dans le mécanisme pause-dépendant, il existe une corrélation positive entre la pause de la RT et le transfert de brin (271, 327, 335–337).

4.10.2.3 Le transfert pause-indépendant

Un transfert efficace peut également être facilité par une interaction favorable entre l'accepteur et l'ADNc, en absence d'une forte pause de la RT au niveau de l'ARN donneur (338, 339). Des études réalisées par le groupe de DeStefano ont montré qu'en absence de NC,

Introduction

l'efficacité de transfert était significativement plus élevée, avec des séquences faiblement structurées de la région pol-vif qu'avec des séquences de la région gag-pol hautement structurée (339). Un autre mécanisme de transfert a également été proposé en étudiant des ARN contenant la séquence SL1 (DIS) qui est responsable de la dimérisation du génome viral (340–342). Des mutations dans SL1 ou la délétion de cette tige-boucle peuvent conduire à une diminution d'environ quatre fois le taux de transfert de brin *in vitro* (343–345).

Ces données *in vitro* suggèrent que la structure dimérique de l'ARNg, en rapprochant les deux brins d'ARNg, facilite les transferts de brin internes et par conséquent la recombinaison entre ces deux ARN encapsidés dans la même particule virale. En accord avec cette hypothèse, les études de l'équipe de Hu montrent que le taux de recombinaison augmente lorsque les deux ARN encapsidés présentent des SL1 complémentaires (346, 347). Il est intéressant de souligner que la délétion de la tige-boucle SL1 altère la synthèse d'ADN viral au niveau du second transfert de brin, bien que la structure dimérique de l'ARNg ne semble pas directement impliquée dans ce transfert (348).

5. Le premier transfert de brin

Le premier transfert de brin est médié par l'appariement des séquences complémentaires R et r de l'extrémité 3' de l'ARNg et l'ADNss (**Figure 32**). Cette étape est obligatoire pour la transcription inverse et la réplication virale. Etant donné que les particules virales du VIH-1 contiennent deux copies d'ARNg, le premier transfert de brin peut se produire de manière intra- ou intermoléculaires (313–315, 349). Bien que le premier transfert de brin puisse se produire avec un ADNss tronqué, la grande majorité des premiers transferts de brin se produisent après la synthèse de l'ADNss complet (350, 351). Le premier transfert de brin est inefficace dans les virions intacts (248, 257, 304). En revanche, la majeure partie de l'ADN viral détecté dans les cellules infectées est d'une longueur très supérieure à l'ADNss, ce qui suggère que le premier transfert de brin se produit rapidement *in vivo* (257). La majorité des études portant sur le mécanisme du premier transfert de brin ont été réalisées sur des systèmes *in vitro*.

5.1 Rôles de l'activité RNase H et de la NC au cours du premier transfert de brin

Les premières études sur le premier transfert de brin ont été réalisées *in vitro* avec des systèmes contenant des matrices virales, relativement non structurées, des ARN donneur et accepteur (5' et 3' du génome, respectivement) et en absence de NC (352–354). Ces travaux ont montré que, lors de la synthèse de l'ADNss, l'extrémité 5' de la matrice doit être dégradée par l'activité RNase H de la RT pour permettre le transfert de brin (352–354). Cette activité RNase H libère l'ADNss qui peut s'hybrider avec la séquence R située dans l'extrémité 3' de l'ARNg et permettre, ainsi, la poursuite de la synthèse du brin (-) d'ADN. Les molécules de RT qui ne sont pas impliquées dans la polymérisation sont importantes dans la réalisation de clivages par l'activité RNase H polymérase-indépendante, ce qui permet la libération du brin (-) d'ADN et ainsi un transfert de brin efficace (289, 355). En tenant compte de l'activité RNase H, l'équipe de Bambara a suggéré que le premier transfert de brin est un transfert par invasion (**Figure 39**) (329).

Le premier transfert de brin est grandement facilité par la NC. Ceci est dû principalement à la capacité de la NC à faciliter l'appariement de la région 5' r de l'ADNss avec la région complémentaire R située à l'extrémité 3' de l'ARNg (174, 356). Etant donné que les régions complémentaires sont engagées dans des structures secondaires, l'appariement dépend de la capacité de la NC à déstabiliser ces structures (174, 357). Par ailleurs, il a été suggéré que la déstabilisation des structures secondaires constituerait l'étape limitante de l'appariement R (358). De façon inattendue, Guo et al. (303) ont démontré *in vitro* qu'en

absence d'ARN accepteur, les ADN synthétisés sont plus longs que l'ADNss, ces derniers ont été désignés comme « produits d'auto-amorçage » *Self priming* (SP). Etant donné que l'ADNss contient la séquence cTAR (séquence complémentaire de l'ARN TAR) pouvant former une tige-boucle, l'ADNss peut s'auto-hybrider à son extrémité 3', permettant ainsi à l'ADNss d'agir à la fois comme amorce et matrice en absence d'ARN accepteur. Il est clair que si l'autoamorçage se produisait *in vivo*, la réplication du virus serait affectée. Cependant, aucun ADN SP n'a été détecté dans les réactions endogènes de virions traités avec des détergents, ce qui démontre que le VIH-1 a contourné ce problème potentiel. La NC semble être la principale protéine qui inhibe la synthèse d'ADN SP (303). Une étude suggère que l'inhibition de l'autoamorçage par la NC ne nécessite pas la présence d'ARN accepteur (359). Ce résultat est probablement dû à un excès de NC qui inhibe la transcription inverse. En effet, plusieurs études ont montré que la NC inhibe faiblement la synthèse d'ADNss SP en l'absence d'ARN accepteur (360–363). Cependant, lorsque l'ARN accepteur et la NC sont présents, le taux d'ADN SP est considérablement réduit et il y a une augmentation concomitante du transfert de brin (360–363).

5.2 Inhibition du transfert de brin par les agents pharmacologiques

5.2.1 Inhibiteurs ciblant la RT

Puisque le premier transfert de brin est un processus essentiel de la transcription inverse, il serait intéressant de disposer d'agents pharmacologiques qui inhibent ce processus. La RT est l'une des enzymes du VIH-1 les plus étudiées, principalement dans le développement des inhibiteurs de l'activité polymérase (364, 365). Ces inhibiteurs peuvent être des analogues nucléotidiques/nucléosidiques (NtRTIs) ou non nucléosidiques (NNRTIs) différant par leur structure et leur mode d'action (**Figure 40A**) (366). L'émergence de résistances à ces inhibiteurs a placé l'activité RNase H de la RT comme cible supplémentaire très prometteuse. Dans le but de caractériser des composés présentant un mode de fixation allostérique, agissant simultanément sur les deux fonctions polymérase et RNase H, des composés ont été conçus pour se fixer simultanément sur les sites polymérase et RNase H. Ces composés inhibent la transcriptase inverse *in vitro*, chez certains mutants résistants aux inhibiteurs de la polymérase du VIH-1 (227, 367). D'autres études ont montré que des composés pyrido-pyrimidinone (**Figure 40B**), dérivant d'inhibiteurs de l'intégrase virale, inhibent l'activité RNase H de la RT et la réplication virale (368, 369). La RT peut effectuer différents types de clivages,

Introduction

dependants ou non de l'activité polymérase, au cours de la transcription inverse. Ainsi, il a été suggéré que la caractérisation du type de clivage le plus limitant au cours de la transcription inverse est nécessaire pour cibler l'inhibition de la RNase H (370).

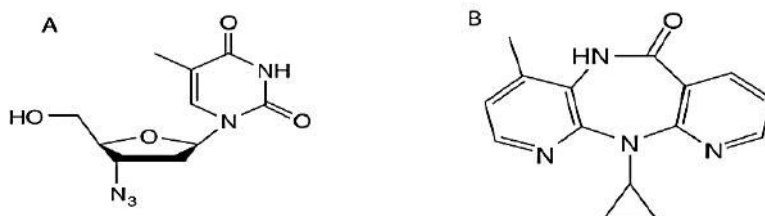


Figure 40. Structure chimique des inhibiteurs de la RT (366). A) Structure d'un *nucleoside RT inhibitor* (NRTI) Zidovudine (AZT). B) Structure d'un *nonnucleoside RT inhibitor* (NNRTIs) qui est la Nevirapine (NVP).

5.2.2 Inhibiteurs ciblant la NC

La protéine de nucléocapside intervient au niveau de plusieurs étapes du cycle réplcatif du VIH-1, notamment dans le mécanisme de transfert de brin au cours de la transcription inverse (162). Des études ont montré que des mutations ponctuelles au niveau de la NC induisent une forte réduction de la charge virale et du potentiel infectieux du virus (371–373). Ainsi, l'inhibition de la NC devrait aussi empêcher le premier transfert de brin de se produire (134, 138, 144, 241, 243, 265). De nombreuses études se sont intéressées au ciblage des doigts de zinc de la NC, en développant des composés dits éjecteurs de zinc (**Figure 41A**) (374–378). Cependant, ces composés présentent une faible spécificité et une forte toxicité (379, 380). D'autres approches pharmacologiques se sont focalisées sur des composés inhibant l'interaction de la NC avec les acides nucléiques. Une de ces approches a été réalisée par criblage de 4800 composés de faible masse moléculaire, en présence d'une forme tronquée de la NC. Cette étude a montré que cinq des composés testés empêchaient la fixation de la NC tronquée sur l'ADN cTAR en se fixant sur les doigts de zinc mais sans éjecter le zinc (381). Un autre criblage réalisé avec la NC, l'ADN SL2 et 14400 composés a permis d'identifier deux composés qui ont une affinité de l'ordre du nano-molaire pour la NC (382). Ces deux composés inhibent la fixation de la NC sur l'ADN SL2, en revanche, ils présentent une faible activité anti-VIH-1 *ex vivo* (382).

5.2.3 Inhibiteurs ciblant les acides nucléiques

Deux études ont montré que l'actinomycine D inhibe *in vitro* le premier transfert de brin en se fixant sur l'ADNss (383, 384). Il faut cependant noter que l'actinomycine D est toxique

Introduction

pour les cellules, elle ne présente pas d'activité antirétrovirale spécifique car elle est principalement un inhibiteur de la transcription et elle active la réplication du VIH-1 dans certaines lignées cellulaires (385). L'inhibition *in vitro* du premier transfert de brin par l'actinomycine D suggère, cependant, qu'il existe une possibilité d'inhiber spécifiquement ce processus, si on dispose d'une connaissance précise des structures des acides nucléiques impliqués dans ce transfert. Une investigation de petites molécules pouvant se fixer spécifiquement aux acides nucléiques substrats constitue, aussi, une autre stratégie de recherche. L'équipe de Gatto a caractérisé *in vitro* des anthraquinones (**Figure 41B**) ayant une forte affinité pour les tiges-boucles ARN TAR et ADN cTAR intervenant dans le premier transfert de brin (386). Ces composés empêchent l'hybridation des tiges-boucles complémentaires en inhibant l'activité chaperonne de la NC (386). Cependant, les essais *ex vivo* n'ont montré aucune inhibition rétrovirale, de plus aucune détection de ces composés n'a été observée dans les cellules traitées, suggérant que les anthraquinones testées ne traversent pas la membrane cellulaire (386). La même équipe a récemment étudié une nouvelle anthraquinone qui inhibe l'hybridation TAR-cTAR en utilisant trois mécanismes différents (387) : 1) stabilisation des tiges-boucles TAR et cTAR ; 2) compétition avec la NC au niveau des boucles du TAR ; 3) fixation sur la NC.

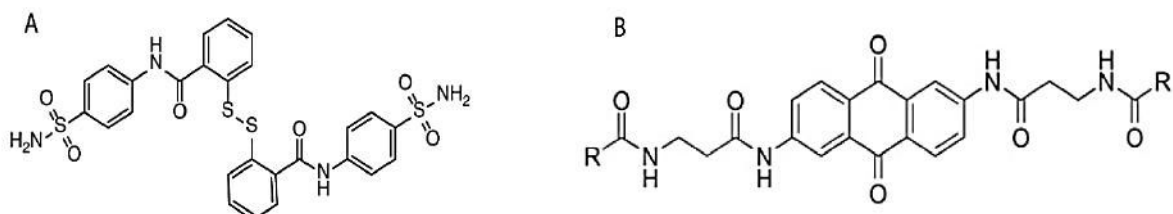


Figure 41. Structure chimique des inhibiteurs ciblant la NC ou les acides nucléiques (162). **A)** Structure d'un composé éjecteur de zinc le disulfide benzamide (DIBA). **B)** Structure d'un composé anthraquinone ciblant les éléments TAR et cTAR.

5.3 Structure des AN intervenant dans le premier transfert de brin

L'isolat NL4-3 du VIH-1 est largement utilisé par plusieurs équipes de recherche dans le monde. Notre équipe et d'autres équipes en France utilisent l'isolat MAL du VIH-1 qui est un recombinant complexe des sous-types A, D et I du VIH-1. Il contient une duplication de 23 nucléotides, incluant des parties du PBS et des régions en aval (155). Cependant, l'alignement des séquences des extrémités 5' et 3'UTR des deux isolats montre qu'il n'y a pas de grandes différences dans les séquences R, les motifs SL1 à SL3, le segment 1 et le motif '9nt' qui sont complémentaires de parties de l'ARNt^(Lys,3) (**Figure 45**), tandis que les séquences PBS et PPT sont très conservées.

5.3.1 Structures ARN impliquées dans le premier transfert de brin

5.3.1.1 Repliement de la séquence R

Le premier transfert de brin nécessite la présence de la séquence R (96 nt dans l'isolat MAL) aux deux extrémités du génome (**Figure 42**). Il est intéressant de noter que la taille de la séquence R varie considérablement chez les rétrovirus, 15 nt chez le MMTV (*Mouse mammary tumor virus*) et 247 bases chez le HLT-2 (388–390). Une étude *ex vivo* réalisée par mutagenèse dirigée a montré que le premier transfert de brin se produit dans la majorité des cas après la copie complète de la séquence R située en 5' de l'ARN génomique du VIH-1 (351). La réplication du VIH-1 est cependant possible avec des séquences R tronquées (391). Des études *in vitro* suggèrent fortement que la structure adoptée par la séquence R joue un rôle important dans le premier transfert de brin chez le VIH-1 (392–394).

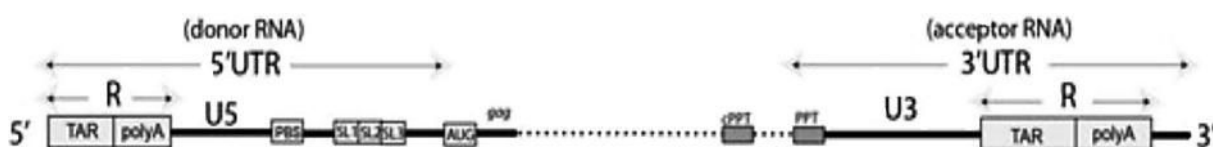


Figure 42. Les extrémités 5' et 3' UTR de l'ARNg du VIH-1 (isolat MAL, GenBank : X004415.1)

La séquence 5' R du VIH-1 est repliée en structures secondaires correspondant aux tiges-boucles TAR et Poly(A) (**Figure 33**) (59, 159). L'étude *in vitro* d'un ARN de 104 nucléotides au moyen de la diffusion des rayons X aux petits angles (SAX) suggère qu'au niveau tridimensionnel les tiges-boucles TAR et poly(A) adoptent une structure coaxiale (**Figure 43**) (395). Au niveau du 3'UTR, la séquence 3' R est aussi repliée en deux tiges-boucles, la partie supérieure de la tige-boucle poly(A) et la tige-boucle TAR (159). Cette dernière est importante pour une transcription inverse efficace (396). L'équipe de Harrich a montré que les mutations dans TAR altérant l'orientation de la tige-boucle apicale ou la séquence de la boucle apicale réduisent fortement la synthèse de l'ADNss (396, 397). Une analyse phylogénétique (398) et une étude *in vitro* (159) suggèrent que la séquence 3' poly(A) forme une tige-boucle dans l'ARNg, il n'existe cependant pas de résultat montrant sa présence *in vivo*. Des travaux *in vitro* suggèrent que la tige-boucle poly(A) n'est pas critique pour le transfert de brin (159, 394).

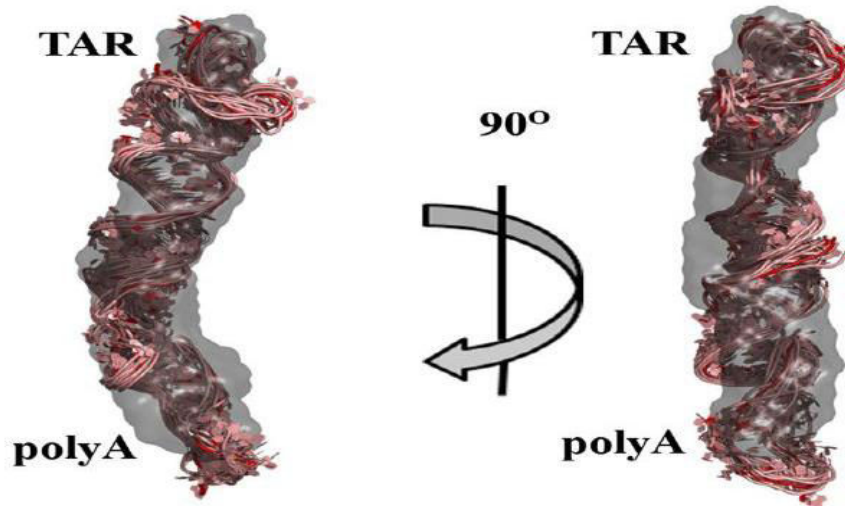


Figure 43. Modèle de structure des éléments TAR et poly(A) (395). Modèle obtenu par la diffusion des rayons X aux petits angles (SAX) qui montre les séquences TAR et poly(A) adoptant une structure coaxiale.

5.3.1.2 Circularisation du génome par l'interaction TAR-TAR

Chez certains virus de plantes, des séquences particulières présentes aux deux extrémités du génome participent au processus de circularisation, en formant des complexes boucle-boucle ou duplexe étendu (399). En plus du site de dimérisation du génome viral (DIS), la tige-boucle apicale du TAR pourrait être impliquée dans le processus de dimérisation grâce à la présence d'une séquence palindromique de 10 nucléotides (400).

D'autre part, TAR est présent aux deux extrémités du génome, ainsi les deux séquences palindromiques pourraient interagir conduisant à la circularisation du génome (**Figure 44**). Une étude *in vitro* n'est cependant pas en faveur de cette hypothèse (401). En effet, elle montre que les mutations qui inhibent l'interaction TAR-TAR n'empêchent pas la formation d'hétérodimères entre les ARN représentant les extrémités 5' et 3' de l'ARNg.

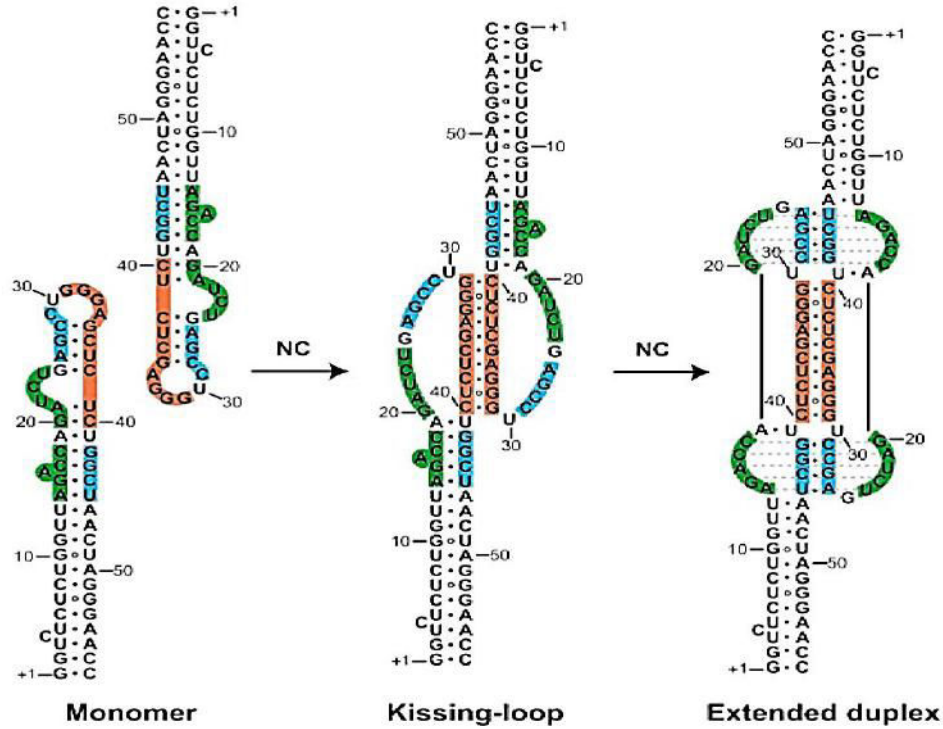


Figure 44. Modèle de dimérisation via la tige-boucle TAR (400). La dimérisation induite par la NC conduit à la formation d'un complexe par interaction boucle-boucle (*kissing complexe*) et un duplexe étendu (*extended duplex*), les séquences impliquées dans l'interaction intermoléculaire sont représentées par les couleurs rose, bleu et vert.

5.3.1.3 Circularisation du génome par l'interaction U3-ARN^{t(Lys-3)}

La séquence U3 (457 nt dans l'isolat MAL) est située à l'extrémité 3' de l'ARNg. Une première étude a identifié une séquence de 9 nucléotides hautement conservée dans la région U3. Cette séquence est complémentaire de la tige de l'anticodon de l'ARN^{t(Lys-3)} (402). Il a donc été suggéré que l'ARN^{t(Lys-3)} qui sert d'amorce lors de l'initiation de la transcription inverse, sert aussi de pont pour relier les extrémités du génome viral. L'ARN^{t(Lys-3)} pourrait interagir avec l'extrémité 3' de l'ARNg en formant un duplexe de 9 nucléotides (402). Une étude plus récente a montré que la séquence de 9 nucléotides fait partie d'une région présentant une forte complémentarité de séquence avec l'ARN^{t(Lys-3)} entier (**Figure 45**) (403).

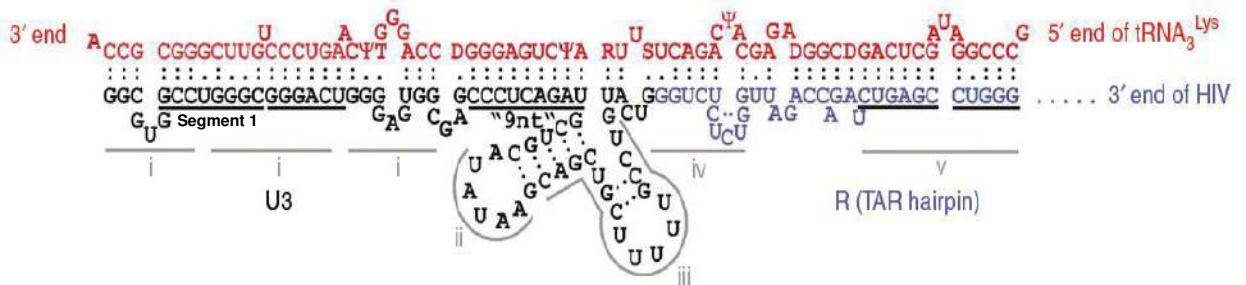


Figure 45. Complémentarité de séquence entre U3 R et l'ARN^{t(Lys-3)} (403).

5.3.1.4 Circularisation du génome par l'interaction gag-U3, poly(A)

Une analyse des hétérodimères formés *in vitro* entre les extrémités 5' et 3' de l'ARNg a montré que des séquences situées dans les régions poly(A) et U3 peuvent s'hybrider avec des séquences situées dans le gène gag. L'analyse phylogénétique des sous-types du groupe M suggère que l'interaction gag-U3, poly(A), a été conservée au cours de l'évolution du VIH-1 (404). A titre d'exemple dans le sous-type B, cette interaction fait intervenir des nucléotides de l'ORF de gag, 16 nucléotides de la région U3 et 28 nucléotides de la séquence poly(A) située dans la séquence 3'(401, 404). Il a été proposé que cette interaction contribue à la circularisation du génome viral *in vivo* (**Figure 46**) (404).

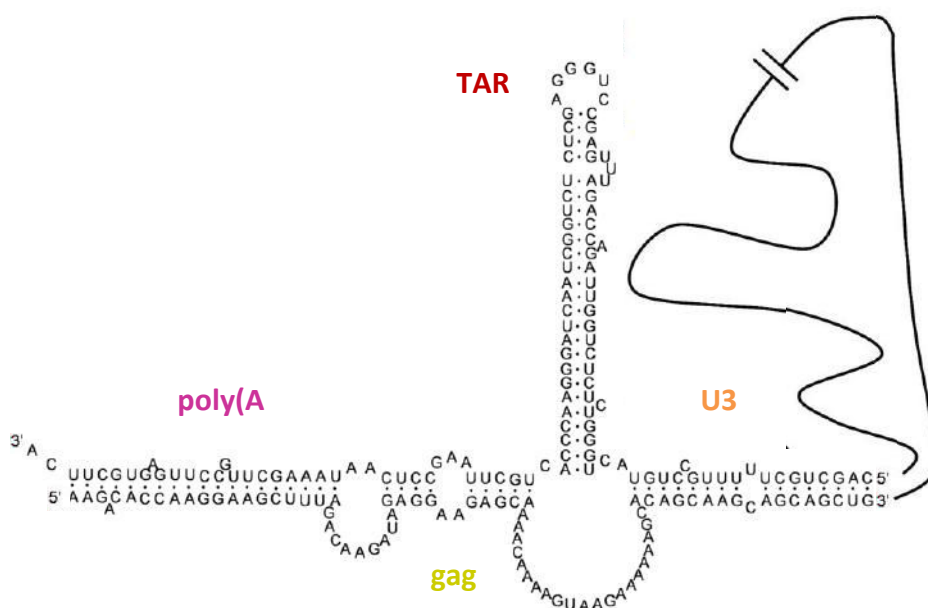


Figure 46. Modèle de structure secondaire proposé pour l'interaction gag-U3poly(A)(404).

5.3.2 Structure de l'ADNss

La région r de l'ADNss (**Figure 32**) est entièrement complémentaire de la région R de l'ARNg. Dans r, les séquences complémentaires des éléments TAR et poly(A) sont respectivement nommées cTAR et cpoly(A). Les séquences TAR, poly(A), cTAR et cpoly(A) peuvent adopter chacune une structure en tige-boucle (158, 159, 393, 405). Plusieurs études réalisées avec la séquence cTAR isolée ont montré que cette séquence forme une tige-boucle dynamique en équilibre entre une conformation fermée et une conformation en « Y » (132, 362, 405, 406). Récemment, notre équipe a montré que la NC déplace l'équilibre vers la conformation « Y » (158). Elle a aussi montré que les boucles apicale et interne de la tige-boucle cTAR sont, respectivement, des sites faible et fort pour la NC (158).

Jusqu'à notre étude, une seule étude avait examiné la structure secondaire de l'ADNss utilisant des ADNss tronqués (50, 63 et 123 nucléotides) (363). Lorsque j'ai débuté ma thèse,

la structure secondaire de l'ADNss entier (178 nt chez l'isolat MAL) n'avait donc pas été déterminée.

5.4 Mécanismes intervenant dans le premier transfert de brin

5.4.1 La circularisation du génome du VIH-1 faciliterait le premier transfert de brin

La circularisation du génome, de manière à ce que les séquences R situées aux extrémités soient côte à côte, faciliterait le premier transfert de brin. Comme décrit précédemment, différentes interactions ARN-ARN (**Figure 47**) contribuent à la circularisation du génome *in vitro* (407). Il a été montré que la séquence de 9 nucléotides (motif 9nt) (5'-CCCUCAGAU, **Figure 45**) favorise le premier transfert de brin *in vitro* (402). Une étude plus récente a ensuite montré que le motif 9nt et une séquence voisine nommée *segment 1* (5'-GCCUGGGCGGGACU-3', **Figure 45**) augmentent le transfert *in vitro*, principalement lorsqu'ils sont présents ensemble (403). De plus, des mutations dans la région U3 au niveau du motif 9nt et du segment 1, réduisent de 9% à 26% le transfert dans des cellules infectées (407). Les mutations dans le segment 1 ont plus d'effet sur le transfert, tandis que l'altération du motif 9nt ou des deux séquences diminue plus facilement le transfert en comparaison avec les mutations ciblant uniquement le segment 1 seul (407). L'interaction de l'extrémité 3' de l'ARNt^(Lys-3) avec le segment 1 pourrait faciliter le déplacement de l'ARNt^(Lys-3) de la séquence PBS et faciliter ainsi le second transfert de brin. Des expériences de mutagenèse dirigée effectuées avec des hétérodimères d'ARN ont montré que l'interaction gag-U3,poly(A) stimule le premier transfert de brin *in vitro* (401). Cette circularisation du génome serait exclusive aux ARNg non épissés car les ARN viraux épissés ne possèdent pas la séquence gag. A ce jour, la circularisation de l'ARNg du VIH-1 dans les particules virales ou dans les cellules infectées n'a pas été démontrée.

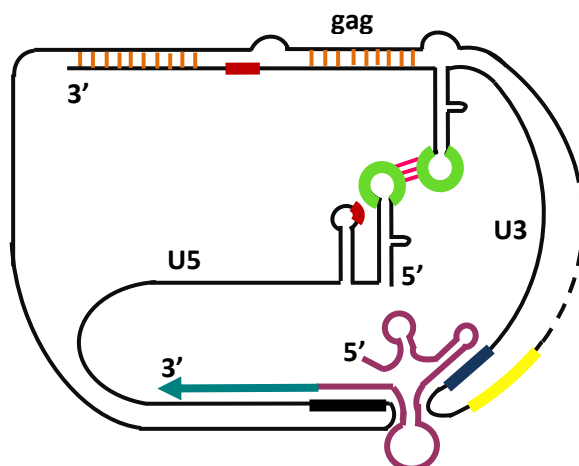


Figure 47. Modèle d'interaction des extrémités 5' et 3' de l'ARNg du VIH-1 (407). Le schéma montre l'interaction entre les séquences *gag* et 3'U3poly(A), ainsi qu'entre l'ARNt^(Lys-3) (mauve) fixé à la région PBS (trait noir épais) et son interaction avec le motif 9nt (bleu foncé) qui est situé dans la région U3. Les boucles apicales des éléments TAR et poly(A) sont respectivement indiquées en vert et granat. Le segment 1 (jaune) et l'ADNss en cours de synthèse (bleu clair) sont également indiqués.

5.4.2 Rôle des tiges-boucles poly(A) et cpoly(A)

Dans l'ARNg extrait du virion, la séquence R située dans le 5' UTR forme les tiges-boucles TAR et poly(A) (59), la tige inférieure de cette dernière est appariée à une partie de la région U5 (59). Au niveau du 3'UTR, la tige-boucle TAR est formée et caractérisée dans l'ARNg extrait du virus (59), tandis que la tige-boucle poly(A) a été prédite et caractérisée *in vitro* (159, 398) mais non démontrée *in vivo* (59). L'équipe de Berkhout a été la première à proposer un modèle dans lequel le transfert serait facilité par des interactions boucle-boucle, impliquant d'une part les tiges-boucles TAR et poly(A) de l'ARN accepteur et d'autre part les tiges-boucles cTAR et cpoly(A) de l'ADNss (393). L'étude *in vitro* du transfert de brin réalisée à partir d'ARN artificiels contenant la séquence R, suggère que cette séquence a été sélectionnée au cours de l'évolution pour favoriser le transfert (394). Cette étude suggère aussi que dans R c'est la tige-boucle TAR qui joue un rôle important dans le transfert tandis que la tige-boucle poly(A) n'est pas impliquée car elle n'est pas formée dans certains des ARN artificiels. Les travaux de notre équipe ont en effet montré l'absence de la tige-boucle poly(A) dans des ARN artificiels (159). En revanche, le modèle de transfert de brin par invasion de l'ARN accepteur, qui a été proposé à partir d'études *in vitro*, suggère un rôle important de la séquence poly(A) dans le transfert (331, 408). Dans ce modèle, au cours de la synthèse de l'ADNss, la RT fait une pause au niveau de la tige-boucle TAR 5' (ARN donneur), ceci initie une dégradation de la séquence poly(A) 5' par l'activité RNase H,

permettant, ainsi, l'appariement de la séquence poly(A) 3' (ARN accepteur) avec la séquence cpoly(A) de l'ADNss.

5.4.3 Rôle des tiges-boucles TAR et cTAR

Une étude *ex vivo* de mutagenèse dirigée de la séquence R suggère que la majorité des premiers transferts de brin se produisent après l'achèvement de la synthèse de l'ADNss (351), autrement dit, la séquence ADN cTAR jouerait un rôle important dans le premier transfert de brin (351). Des mutations dans la boucle apicale du TAR réduisent fortement, le transfert *in vitro* (393). Ce résultat est en faveur de l'appariement TAR-cTAR initié par l'interaction boucle-boucle impliquant la boucle apicale de chaque AN (**Figure 48**).

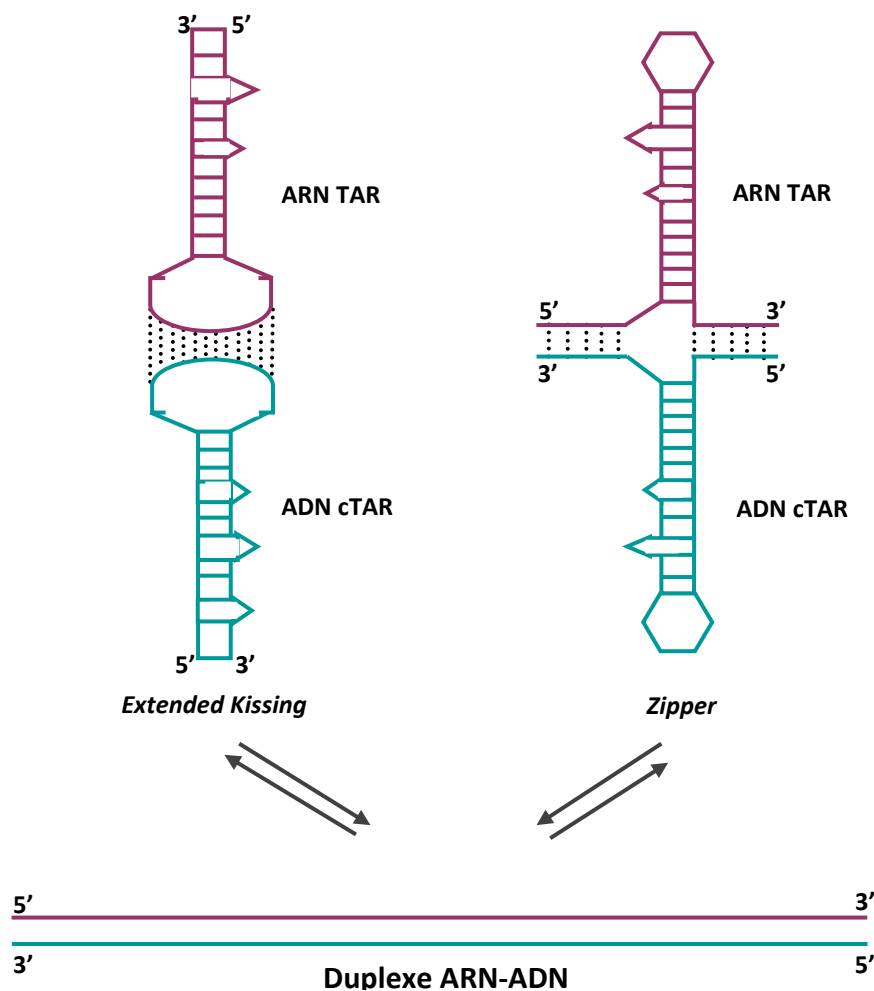


Figure 48. Deux voies d'hybridation de l'ARN TAR-ADN cTAR (409). Modèle de gauche : initiation de l'appariement par l'interaction boucle-boucle (*kissing complex*) qui est convertie en un *extended kissing*. Modèle de droite : initiation de l'appariement au niveau des extrémités 5' et 3' (*zipper*).

Introduction

Une étude *in vitro* de l'équipe a montré que les mutations empêchant l'éventuelle interaction boucle-boucle, réduisent fortement l'hybridation entre la tige-boucle cTAR et l'ARN 3'UTR (159). La même étude a montré que la tige apicale de la tige-boucle TAR (nucléotides 17-44, **Figure 44**) est conservée au sein de différents ARN accepteurs qui contiennent la séquence R et favorisent le transfert de brin *in vitro* (394). Le transfert de brin *in vitro* n'est pas réduit lorsque les ARN donneur et accepteur contiennent la partie supérieure de la tige-boucle TAR et pas sa partie inférieure (393). En revanche, le transfert *in vitro* est fortement réduit lorsque la séquence R de l'ARN 3'UTR est limitée aux 29 premiers nucléotides et ne peut donc pas former la partie supérieure de la tige-boucle TAR car (410). Ces résultats suggèrent que la partie supérieure de la tige-boucle TAR, dans l'ARN accepteur, joue un rôle important dans le premier transfert de brin. La plupart des travaux *in vitro*, dont ceux de mon équipe, portant sur la formation de l'hybride R-r n'ont été réalisées qu'avec de courts acides nucléiques ayant une taille inférieure ou égale à 59 nucléotides (ADN cTAR, ARN TAR et des versions tronquées de ces deux molécules). Une partie importante de ces travaux a eu pour but de comprendre comment est initié l'appariement TAR-cTAR. En utilisant l'ADN mini-cTAR et l'ARN mini-TAR en absence de NC, notre équipe a montré la formation d'un complexe boucle-boucle impliquant la boucle apicale de chaque tige-boucle (158). Vo *et al.* (409), ont aussi conclu que l'hybridation TAR-cTAR en absence de NC est initiée à partir des boucles apicales, tandis qu'elle est initiée à partir des extrémités 3'/5', en présence de fortes concentrations de NC, conduisant à la formation d'un intermédiaire comparable à une fermeture éclair (*zipper*) (**Figure 48**). En outre, les séquences TAR et cTAR peuvent s'hybrider via les deux voies dans des conditions de sous-saturation en NC (409). Godet *et al.* (411), ont montré que dans des conditions de faibles concentrations en sels, les tiges-boucles TAR et cTAR s'hybrident quasi-exclusivement selon le modèle *zipper* en présence de la NC (12-55) qui est une forme tronquée dépourvue du domaine N-terminal (411). Il faut souligner que, les deux voies d'appariement n'ont pas été démontrées pour la réaction d'hybridation entre l'ADNss entier et l'extrémité 3' de l'ARNg.

Plusieurs études ont eu aussi pour but de comprendre comment la NC déstabilise les deux tiges-boucles TAR et cTAR pour pouvoir les appairer complètement. Une première étude réalisée avec une NC tronquée suggère que la tige-boucle TAR est très stable et faiblement déstabilisée par la NC (une paire de bases par molécule) (405). Notre groupe a montré que les doigts de zinc de la NC interagissent fortement avec deux guanines situées dans la boucle apicale de l'ARN TAR (159). Des expériences de FRET effectuées avec la séquence cTAR ont révélé une structure très dynamique de cette dernière, impliquant un

Introduction

équilibre entre la conformation fermée et la conformation partiellement ouverte « Y » (362, 405, 406, 412). Dans la conformation Y, la tige inférieure est ouverte et la tige supérieure est fermée. Il a été montré que la NC déstabilise fortement les extrémités de cTAR déplaçant l'équilibre vers les conformations les plus ouvertes (158, 362, 405, 413). Récemment, une étude de notre équipe à laquelle j'ai participé a montré que la NC, à faible concentration, déstabilise uniquement la tige apicale de la tige-boucle ARN TAR (139). La déstabilisation est initiée à partir d'un site secondaire de la NC, c'est-à-dire une paire de base contenant une guanine et adjacente à un bulge. Il est intéressant de souligner qu'une étude réalisée à l'échelle de la molécule unique est aussi en faveur de la déstabilisation de la tige apicale initiée par la guanine adjacente à un bulge (414). Ces résultats récents permettent d'expliquer comment l'interaction boucle-boucle peut conduire à la formation d'un long duplexe.

Objectifs

Comme cela a été mentionné dans l'introduction, le premier transfert de brin repose en grande partie sur l'appariement de la séquence r de l'ADNss avec la séquence R située à l'extrémité 3' de l'ARN 3'UTR. Les deux séquences forment des structures secondaires qui doivent s'ouvrir pour pouvoir s'apparier. La NC joue un rôle prépondérant dans l'ouverture de ces structures et la formation du duplexe r-R.

Dans la majorité des travaux *in vitro* publiés dont ceux de mon équipe d'accueil, la formation du duplexe r-R n'a été étudiée qu'avec de courts acides nucléiques longs d'environ 55-59 nucléotides (ADN cTAR et ARN TAR) qui ne reflètent que partiellement les séquences et structures impliquées dans le premier transfert de brin *ex vivo*. Une étude *ex vivo* (351) suggère fortement que dans 90 à 99 % des cas le premier transfert de brin a lieu après la synthèse complète de l'ADNss (178 nucléotides chez l'isolat MAL). Cependant, à ce jour, il n'existe pas d'études ayant caractérisé les mécanismes moléculaires gouvernant l'hybridation de l'ADNss avec l'ARN 3'UTR en présence de NC, au cours du premier transfert de brin.

L'objectif principal de ma thèse est de caractériser les mécanismes moléculaires qui régissent l'hybridation de l'ADNss avec l'ARN 3'UTR. En d'autres termes, caractériser par les techniques de la biologie moléculaire et les sondes de structure (agents chimiques et nucléases) : i) la structure secondaire de l'ADNss entier ; ii) les sites dans l'ADNss qui interagissent préférentiellement avec la NC; iii) le ou les sites d'initiation de l'hybridation au niveau de l'ADNss et de l'ARN 3'UTR ; iv) l'étendue de l'appariement des séquences complémentaires r et R.

Matériel et Méthodes

Dans cette partie sont décrits les matériel et méthodes qui ne sont pas mentionnés dans l'article publié.

1. Oligodésoxyribonucléotides

Les oligodésoxyribonucléotides représentent des fragments de l'ADNss. Ces oligodésoxyribonucléotides contiennent chacun un site préférentiel de fixation de la protéine de nucléocapside, ils sont présentés en couple de sauvage/mutant. Dans les séquences suivantes, les lettres colorées en rouge indiquent les mutations par rapport à la séquence sauvage d'ADNss du VIH-1 (isolat MAL). Quant à la taille et la position de la séquence au niveau de l'ADNss, elles sont entre parenthèses.

18-38C29 : 5'-ACCGTCTAGAGCGGTCTGAGG-3' (21 nt ; 18-38)

18-38C29m : 5'-ACCGTCTAGAGCGGTCTCAGG-3' (21 nt ; 18-38)

57-77wt : 5'-TCACACAACAGATGGGCACAC-3' (21 nt ; 57-77)

57-77m : 5'-TCACACAACAGATCCGCACAC-3' (21 nt ; 57-77)

91-111wt : 5'-CTCAAGGCAAGCTTTATTGAG (21 nt ; 91-111)

91-111m : 5'-CTCAAGGCAAACCTTTATTGAG (21 nt ; 91-111)

96-107wt : 5'-GGCAAGCTTTAT-3' (12 nt ; 96-107)

96-107m : 5'-GGCAACCTTTAT -3' (12 nt ; 96-107)

124-135wt : 5'-GGTTCCTTGCTA-3' (12 nt ; 124-135)

124-135m : 5'-GGTTCCTTCCTA -3' (12 nt ; 124-135)

1.1 Purification des oligodésoxyribonucléotides

Les oligodésoxyribonucléotides ont été achetés chez Eurogentec. 25 pmol d'oligonucléotides sont repris dans un volume équivalent du tampon de charge bleu urée (B/U) (7 M urée; 0,05% bleu bromophénol ; 0,05% xylène cyanol), à volume égal. Ils ont été purifiés par électrophorèse sur gel dénaturant 15 % polyacrylamide/7,2 M urée, 1X TBE (90 mM Tris-Borate, 2 mM d'EDTA). À la fin de la migration, les oligonucléotides ont été révélés par *UV-shadowing*, les bandes du gel les contenant ont été découpées, puis l'extraction a été effectuée par élution passive (processus de diffusion), pendant la nuit à température ambiante, dans une solution contenant 0,3 M d'acétate de sodium, 0,1 % de SDS et 1 mM d'EDTA. Au jour J+1, l'éluat est récupéré puis centrifugé sur colonne spinX à 10000 rpm pendant 2 min dans la mini-centrifugeuse de paillasse BECKMAN Microfuge^R Lite. S'en suit une étape de précipitation avec 2,5 volumes d'éthanol absolu et une centrifugation à 14000 rpm pendant 1 h à 4°C dans la mini-centrifugeuse de paillasse BECKMAN Microfuge^R

Lite. Le surnageant est ensuite retiré et un lavage du culot avec 200 µl d'éthanol 70% est effectué pendant 5 min. Les oligonucléotides purifiés ont ensuite été séchés 5 min au speed-vac et remis en solution dans 20 µl d'eau milli-QTM. La concentration de chaque oligonucléotide a été déterminée par spectrophotométrie d'absorption à 260 nm.

1.2 Marquage et purification des oligodésoxyribonucléotides

Les oligodésoxyribonucléotides ont été marqués en 5' au ³²P en utilisant la T4 polynucléotide kinase (New England Biolabs) et le [γ-³²P] ATP (Perkin Elmer). La réaction a été effectuée dans un volume final de 8 µl. Vingt-cinq pmol d'oligodésoxyribonucléotide dans 3,7 µl d'eau milli-QTM, 0,8 µl de tampon réactionnel (10X) (concentrations finales : 70 mM Tris-HCl pH 7,6 ; 10 mM MgCl₂ et 5 mM DTT), 2,5 µl de [γ-³²P] ATP (25 µCi, 16,7 pmol) et 1 µl de T4 polynucléotide kinase (10 U/µl) constituent le mélange réactionnel. Après une incubation à 37°C pendant 45 min, l'oligodésoxyribonucléotide marqué a été purifié sur gel de polyacrylamide (19:1)/7,2 M urée (respectivement, 12% et 15% pour les oligonucléotides d'une longueur de 21 nt et 12 nt). Pour vérifier la pureté de l'oligonucléotide marqué en 5', une aliquote correspondant à l'équivalent de 20000 cpm a été analysée par électrophorèse sur gel dénaturant 8% polyacrylamide (19:1)/7,2 M urée dans le tampon d'électrophorèse 1X TBE (90 mM Tris-Borate, 2 mM d'EDTA). Les oligonucléotides ont été révélés par autoradiographie, les bandes du gel les contenant ont été découpées, puis l'extraction a été effectuée par élution passive (processus de diffusion), pendant la nuit à température ambiante, dans une solution contenant 0,3 M d'acétate de sodium, 0,1 % de SDS et 1 mM d'EDTA. Au jour J+1, l'éluat est récupéré puis centrifugé sur colonne spinX à 10000 rpm pendant 2 min dans la mini-centrifugeuse de paillasse BECKMAN Microfuge^R Lite. S'en suit une étape de précipitation avec 2,5 volumes d'éthanol absolu et une centrifugation à 14000 rpm pendant 1 h à 4°C dans la mini-centrifugeuse de paillasse BECKMAN Microfuge^R Lite. Le surnageant est ensuite retiré et un lavage du culot avec 200 µl d'éthanol 70% est effectué pendant 5 min. Les oligonucléotides purifiés ont ensuite été séchés 5 min au speed-vac et remis en solution dans 20 µl d'eau milli-QTM.

2. Analyse par retard sur gel de l'interaction NC-oligonucléotide

Les expériences de retard sur gel ont été réalisées dans un volume final de 10 µl. Chaque oligonucléotide (5 pmol ; 5000 cpm/pmol) dans 6 µl d'H₂O, a été dénaturé à 90°C pendant 2 min puis plongé dans la glace pendant 2 min. Deux microlitres d'un tampon réactionnel (BYC 0,2 (4X)) ont été ajoutés à chaque essai (concentrations finales dans chaque

Matériel et Méthodes

essai : 50 mM Tris HCl pH 7,8 ; 75 mM KCl ; 0,2 mM MgCl₂ et 1 mM DTT). Ensuite, 2 µl de NC, à différentes concentrations, ont été ajoutés pour obtenir les rapports 1:12 ; 1:7 ; 1:3,5 ; 1:1,75 et 1:1, puis les essais ont été incubés à 37°C pendant 15 min. Ensuite, 4 µl du tampon de charge bleu de glycérol (BG) (50% glycérol; 0,05% bleu de bromophénol ; 0,05% xylène cyanol) ont été ajoutés et les essais ont été déposés sur gel, 16% (oligonucléotides 18-38C29, 57-77 et 91-111) ou 17% (oligonucléotides 124-135 et 96-107), de polyacrylamide non dénaturant (29:1). La migration a été effectuée à 4°C dans le tampon 0,5X TBE (45 mM Tris-Borate, 1mM EDTA) à 300 V et 25 mA. En fin de migration, le gel de polyacrylamide est plongé dans une solution de fixation (acide acétique 6% et éthanol 10%) pendant 15 min puis séché pendant 30 min à 80°C. Les oligonucléotides libres et liés à la NC ont été quantifiés en utilisant le TyphoonTM TRIO (GE healthcare) et le logiciel ImageQuant. Le pourcentage de complexe formé = intensité de la radioactivité du complexe/(intensité de la radioactivité du complexe + intensité de l'oligonucléotide libre) X 100.

3. Analyse par retard sur gel du mélange ADNss-ARN 3'UTR

Ces expériences de retard sur gel l'agarose ont été effectuées dans un volume final de 5 µl. L'ADNss (0,6 pmol à 5 X 10⁴ cpm/pmol) dans 2 µl d'eau milli-QTM a été dénaturé à 90°C pendant 2 min et plongé 2 min dans la glace, avant d'ajouter 0,5 µl de tampon réactionnel (concentrations finales : 75 mM KCl, 2 mM MgCl₂ et 50 mM Tris-HCl. pH 7,8). L'échantillon a ensuite été incubé à 37°C pendant 15 min avant une incubation de 15 min à 4°C. L'ARN non marqué a été utilisé à différentes concentrations (0,6 ; 1,2 ; 2,4 et 4,8 pmol) et a subi le même traitement de dénaturation-renaturation que l'ADNss avant d'être ajouté à ce dernier. Le mélange réactionnel a ensuite été incubé 30 min à 4°C. En fin d'incubation, 4 µl du tampon de charge bleu de glycérol (BG) (50% glycérol; 0,05% bleu de bromophénol ; 0,05% xylène cyanol) ont été ajoutés. Les échantillons ont été analysés par électrophorèse sur gel d'agarose 1,2% dans le tampon de migration 0,5X TBM (45 mM Tris-borate, 2 mM MgCl₂) à 4°C, 150 V et 25 mA. En fin de migration, le gel d'agarose est plongé dans une solution de fixation (acide trichloroacétique 10%) pendant 10 min, séché pendant 1h à température ambiante puis radiographié.

4. Analyse structurale de l'ADNss en présence de l'extrémité 3'UTR

Les analyses structurales ont été réalisées en absence de NC avec l'ADNss mis en présence de l'ARN 3'UTR, l'ARN Δ3' ou l'ARN Δ3'-1. L'ADNss (0,6 pmol à 5 X 10⁴ cpm/pmol) dans 2 µl d'eau milli-QTM a été dénaturé à 90°C pendant 2 min et plongé 2 min

dans la glace, avant d'ajouter 0,5 µl de tampon réactionnel (concentrations finales : 75 mM KCl, 2 mM MgCl₂ et 50 mM Tris-HCl. pH 7,8). L'échantillon a ensuite été incubé à 37°C pendant 15 min avant une incubation de 15 min à 4°C. L'ARN non marqué a été utilisé à différentes concentrations (0,6 ; 1,2 ; 2,4 et 4,8 pmol) et a subi le même traitement de dénaturation-renaturation que l'ADNss avant d'être ajouté à ce dernier. Après incubation des mélanges réactionnels 30 min à 4°C, ont été ajoutés aux essais le KMnO₄ (concentrations finales : 0,75 ou 1 mM), la DNaseI (0,2 ou 0,4 unités) ou la MB (2 ou 4 unités), les essais ont été incubés à 4°C pendant 1 min dans le cas du KMnO₄ et 20 min dans le cas des nucléases. Les essais avec les nucléases ont subi une extraction au phénol/chloroforme suivie d'une précipitation à l'éthanol puis séchage des culots au speed-vac. Les essais KMnO₄ ont été stoppés par ajout de 40 µl d'une solution d'arrêt (0,7 M βmercaptoéthanol ; 0,4 M NaOAc pH 7 ; 10 mM EDTA et 25 g/ml d'ARNt), avant de subir une extraction au phénol/chloroforme suivie d'une précipitation à l'éthanol puis séchage des culots au speed-vac. Les culots sont ensuite remis en solution dans 100 µl de pipéridine (1M) et incubés à 90°C pour cliver l'ADNss au niveau des thymines modifiées. À la fin de l'incubation, les essais sont lyophilisés et remis en solution dans 20 µl d'eau milli-QTM. Les essais sont lyophilisés une seconde fois avant de les remettre en solution dans 15 µl d'eau milli-QTM et d'effectuer une troisième lyophilisation. Les culots de tous les essais (DNaseI, KMnO₄ et MB) sont repris dans 7 µl de tampon de charge bleu urée (B/U) (7 M urée; 0,05% bleu bromophénol ; 0,05% xylène cyanol). Le séquençage de l'ADNss est effectué par la méthode Maxam-Gilbert (415). Les échantillons ont ensuite été analysés par électrophorèse sur gel dénaturant polyacrylamide (19:1)/7,2 M d'urée, en effectuant des migrations courte (8%) et longue (12%) afin d'identifier les sites de clivage au nucléotide près.

5. Analyse des cinétiques d'hybridation

Les expériences d'hybridation ont été effectuées dans un volume final de 12 µl suivant deux méthodes de renaturation des acides nucléiques.

5.1 Renaturation en absence de NC

L'ADNss wt (0,15 pmol à 4×10^5 cpm/pmol) dans 4 µl d'eau milli-QTM a été dénaturé à 90°C pendant 2 min et plongé 2 min dans la glace. Ensuite 1 µl de tampon réactionnel (concentrations finales : 75 mM KCl, 0,2 mM MgCl₂ et 50 mM Tris-HCl. pH 7,8) a été ajouté avant une incubation à 37°C pendant 15 min. L'ARN (0,45 pmol) non marqué, a subi le même traitement de renaturation et a été ajouté à l'ADNss. Deux microlitres de NC (11,67 pmol) a été ajouté et le mélange réactionnel a été incubé à 37°C pendant 1, 3, 9, 15, 30, 60 ou 120

min. En fin d'incubation, les échantillons ont subi une extraction au phénol/chloroforme puis 4 µl de tampon de charge (50% glycérol; 0,05% bleu de bromophénol ; 0,05% xylène cyanol) ont été ajoutés à la phase aqueuse. Les échantillons ont été analysés par électrophorèse (150 V, 25 mA) à 25°C sur gel 2% agarose (QA Agarose TM, Obiogene) - 0,5X TBE. En fin de migration, le gel d'agarose a été fixé dans une solution d'acide trichloroacétique (10%) pendant 10 min, séché pendant 1h à température ambiante puis radiographié.

5.2 Renaturation en présence de NC

L'ADNss wt (0,15 pmol à 4×10^5 cpm/pmol) dans 4 µl d'eau milli-QTM a été dénaturé à 90°C pendant 2 min et plongé 2 min dans la glace, avant d'ajouter 1 µl de tampon réactionnel (concentrations finales : 75 mM KCl, 0,2 mM MgCl₂ et 50 mM Tris-HCl. pH 7,8) et 1 µl de NC (2,67 pmol). L'essai a ensuite été incubé à 37°C pendant 15 min. L'ARN (0,45 pmol) non marqué, a subi le même traitement de renaturation en présence de NC (9 pmol) et a été ajouté à l'ADNss. Le mélange réactionnel a ensuite été incubé à 37°C pendant 1, 3, 9, 15, 30, 60 ou 120 min. En fin d'incubation, les échantillons ont subi une extraction au phénol/chloroforme puis 4 µl de tampon de charge (50% glycérol; 0,05% bleu de bromophénol ; 0,05% xylène cyanol) ont été ajoutés à la phase aqueuse. Les échantillons ont été analysés par électrophorèse (150 V, 25 mA) sur gel 2% agarose (QA Agarose TM, Obiogene) - 0,5X TBE. En fin de migration, le gel d'agarose a été fixé dans une solution d'acide trichloroacétique (10%) pendant 10 min, séché pendant 1h à température ambiante puis radiographié.

Résultats et Discussions

1. Analyse structurale de l'ADNss

La première étape de l'étude a été d'obtenir un ADNss entier. A cause de sa longueur et de sa séquence, plusieurs sociétés n'ont pas accepté de synthétiser par voie chimique l'ADNss (178 nucléotides chez l'isolat MAL) car elles pensaient qu'il n'était pas possible de le produire et le purifier en assez grande quantité. Les sociétés Eurogentec et IBA GmbH ont essayé mais elles ont échoué et n'ont pu nous fournir qu'un ADNss tronqué. La synthèse d'ADNss est réalisée en utilisant la RT du VIH-1 achetée chez Worthington Biochemicals, l'oligodésoxyribonucléotide PR qui sert d'amorce et les ARN sauvage ou mutés synthétisés *in vitro*. Ce type de synthèse qui produit juste assez d'ADNss pour les analyses structurales, a été mise au point par Yingying Chen, une ancienne doctorante du laboratoire. L'ADNss est marqué au ^{32}P en 5' en utilisant l'amorce PR marquée en 5'. L'ADNss est ensuite purifié à homogénéité par une longue migration sur gel de séquence, élution de l'ADNss de la bande de gel prélevée et précipitation à l'éthanol. Nous avons ensuite vérifié par électrophorèse sur gel d'agarose que l'ADNss purifié ne forme pas un homoduplexe car si cela avait été le cas, l'analyse des essais d'hybridation et la détermination de la structure secondaire de l'ADNss auraient été plus complexes.

Les résultats que j'ai obtenus avec l'ADNss purifié ont donné lieu à une publication dans *J.Biol.Chem.* qui est présentée à la fin de ce mémoire. Les résultats obtenus sont résumés ci-après.

En utilisant trois sondes de structure (permanganate de potassium, DNase I et *mung bean nuclease*) et le logiciel mfold, nous avons été les premiers à déterminer la structure secondaire de l'ADNss en l'absence et présence de NC. Nous avons montré que l'ADNss, en l'absence de NC, se replie principalement en une seule conformation qui contient trois tiges-boucles. Parmi ces tiges-boucles, deux sont formées par les séquences cTAR et cpoly(A), la troisième est formée par l'extrémité 5' de la région u5. En présence de NC, la structure secondaire globale de l'ADNss n'est pas modifiée. On observe cependant une ouverture de la partie inférieure de la tige cTAR. De plus, nos données suggèrent que la fixation de la NC induit une ouverture de la structure tridimensionnelle de l'ADNss et facilite ainsi une interaction de la séquence r de l'ADNss avec la séquence R de l'ARN 3'UTR. Dans l'ADNss replié, nous avons identifié quatre sites où la NC se fixe préférentiellement. Finalement, l'analyse phylogénétique qui a été réalisée avec Julien Godet (Université de Strasbourg) a

montré que le repliement de l'ADNss et les sites forts pour la NC sont conservés parmi les différents groupes et isolats du VIH-1.

2. Analyse par retard sur gel des sites de fixation de la NC

Au cours de l'étude structurale de l'ADNss, quatre sites préférentiels de fixation de la NC ont été identifiés dans l'ADNss (**Figure 8 de l'article**). Dans le but de réaliser ultérieurement des études RMN et de mieux comprendre la fixation de la NC au niveau de ces sites, dix oligodésoxyribonucléotides contenant ces sites ont été utilisés dans des expériences de retard sur gel. Pour chaque séquence sauvage, un mutant a été créé de sorte à abolir la fixation de la NC (**Figure 49**). Les bases mutées sont des guanines qui sont supposées jouer un rôle essentiel dans la fixation de la NC (152, 158, 416–418).

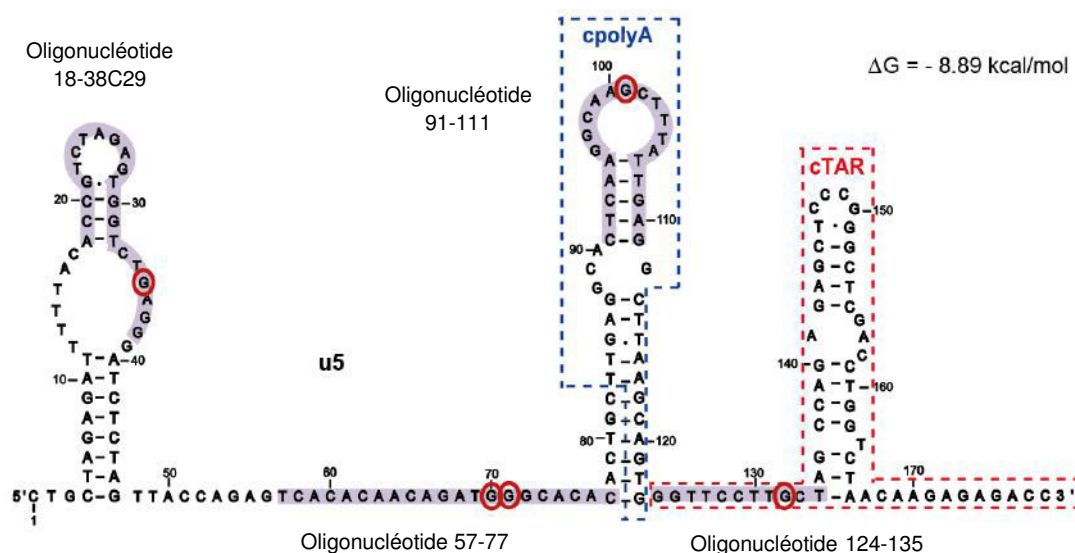


Figure 49. Structure secondaire de l'ADNss en présence de NC. Les séquences des oligodésoxyribonucléotides utilisés dans les essais de retard sur gel sont surlignées en mauve, les nucléotides remplacés par des cytosines sont entourés en rouge.

2.1 18-38C29 et mutant :

L'oligonucléotide 18-38C29 n'est pas totalement identique à la séquence sauvage. En effet, il contient une mutation ponctuelle (C29 au lieu de T29) afin qu'il puisse adopter une structure en tige-boucle (**Figure 50**). Les expériences réalisées avec les deux oligonucléotides montrent qu'il n'y a pas de différence significative entre la fixation de la NC sur les séquences 18-38C29 et 18-38C29m (**Figure 51**). En effet, les résultats révèlent qu'au rapport NC:nt égal à 1:7, le pourcentage de complexe NC-oligonucléotide formé est de 7,1% pour le

Résultats et discussion

18-38C29 et 5,9% pour le 18-38C29 m. De plus, nous remarquons une augmentation progressive du complexe formé avec l'augmentation de la concentration en NC, jusqu'à atteindre 59,6% avec le 18-38C29 et 62,5% avec le 18-38C29 m, au rapport NC:nt égal à 1:1. En présence de concentrations croissantes en NC nous observons une seule bande supplémentaire par rapport à l'oligonucléotide incubé sans NC (**Figure 51**). Cette observation suggère fortement que chaque oligonucléotide fixe spécifiquement une seule NC, c'est-à-dire il y aurait un seul site par oligonucléotide. Il faut cependant souligner qu'aux fortes concentrations de NC, il y a formation d'agrégats, minoritaires, qui restent dans les puits.



Figure 50. Structures secondaires des oligodésoxyribonucléotides 18-38C29 (A) et 18-38C29m (B). Les structures ont été obtenues à partir du programme mfold (419) avec les paramètres : 75 mM NaCl ; 0,2 mM MgCl₂ et 37°C. La mutation sensée supprimer ou diminuer fortement la fixation de la NC est entourée d'un cercle orange.

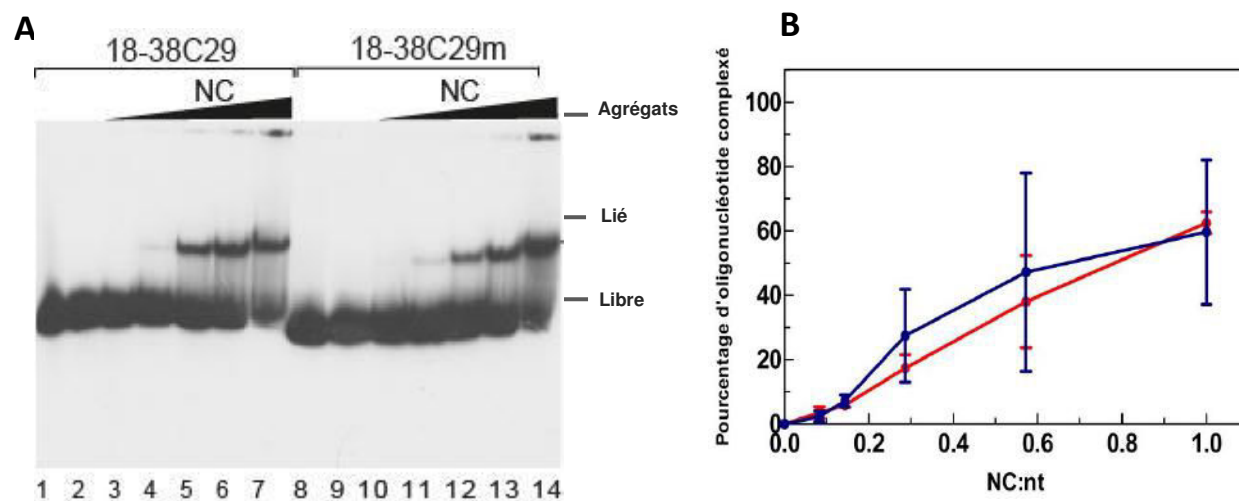


Figure 51. Etude de l'interaction entre la NC et les oligodésoxyribonucléotides 18-38C29 et 18-38C29m. **A)** Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les échantillons ont été analysés par électrophorèse sur gel de polyacrylamide 16% à 4°C dans le tampon 0,5X TBE. Les pistes 1 et 8 représentent les échantillons dénaturés dans l'eau. Les pistes 2 et 9 correspondent aux échantillons incubés à 37°C en absence de NC. Les pistes 3-7 et 10-14 représentent les oligonucléotides incubés à 37°C en présence de concentrations croissantes de NC pour avoir respectivement un rapport NC:nt de 1:12 ; 1:7 ; 1:3,5 ; 1:1,75 et 1:1. **B)** Graphe représentant le pourcentage de complexe NC-oligonucléotide en fonction du rapport NC:nt. Les courbes bleu et rouge indiquent respectivement le 18-38C29 et le 18-38C29m. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

Résultats et discussion

Nos analyses des interactions NC-ADNss qui sont présentées dans l'article, ont montré qu'en présence de la protéine, la coupure *mung bean* (MB) en T34 disparaît complètement et la coupure MB en A36 est fortement réduite en comparaison avec les essais réalisés en absence de NC. En revanche, nous remarquons que la mutation du G35 en cytidine n'a pas eu d'impact significatif sur la capacité de fixation de la NC sur l'oligonucléotide. Ceci pourrait être dû à la présence du motif GAG au niveau de la boucle apicale. En effet, il a été montré que la NC présente une bonne affinité pour le motif GNG au niveau de courtes séquences d'ADN (420). De plus, dans les deux oligonucléotides, le motif GAG a probablement une conformation différente de celle présente dans l'ADNss. En effet, la cytidine en position 29 stabilise la tige et permet la formation de la paire de bases 22T.G28. Dans le contexte de l'ADNss entier en l'absence de NC, les expériences de coupure à la MB montrent que la boucle apicale de la région u5 est difficilement coupée ce qui suggère qu'elle n'est pas accessible dans la structure 3D de l'ADNss. Nos expériences d'empreinte à la MB suggèrent fortement que la séquence T₃₂CTGA₃₆ est un site fort pour l'ADNss (421). Cette séquence contient le motif TG non apparié qui est préférentiellement reconnu par la NC (152, 158, 417, 418). Il est donc surprenant que l'oligonucléotide 18-38C29m, qui ne possède pas le motif TG, fixe la NC aussi bien que l'oligonucléotide contenant ce motif. Ce résultat suggère que la séquence T₃₂CTGA₃₆ n'est pas reconnue par la NC lorsqu'elle est présente dans l'oligonucléotide 18-38C29. Une hypothèse est donc que cette séquence n'est pas accessible dans l'oligonucléotide, il est cependant difficile d'expliquer cette inaccessibilité.

2.2 57-77wt et mutant :

Le logiciel mfold prédit que les oligonucléotides 57-77wt et 57-77m (séquences dans la partie Matériel et Méthodes) ne sont pas capables de former des structures secondaires stables. Les résultats obtenus montrent que l'oligonucléotide sauvage forme un complexe NC-oligonucléotide détectable à partir du rapport NC:nt égal à 1:12 (**Figure 52**). Au rapport NC:nt égal à 1:1, 72% de l'oligonucléotide 57-77wt est sous la forme d'un complexe tandis que l'oligonucléotide 57-77m ne dépasse pas 25% de complexe formé au même rapport. On observe la présence d'une seule bande supplémentaire en présence de NC, ce qui suggère la présence d'un seul site dans les oligonucléotides. Ces résultats sont en accord avec les expériences d'empreinte à la MB qui suggèrent fortement que la séquence G₆₇ATGG₇₁ est en interaction avec la NC dans l'ADNss (421). Ainsi, la substitution des G70 et G71 en cytidines a fortement altéré la capacité de la NC à se fixer sur cette séquence.

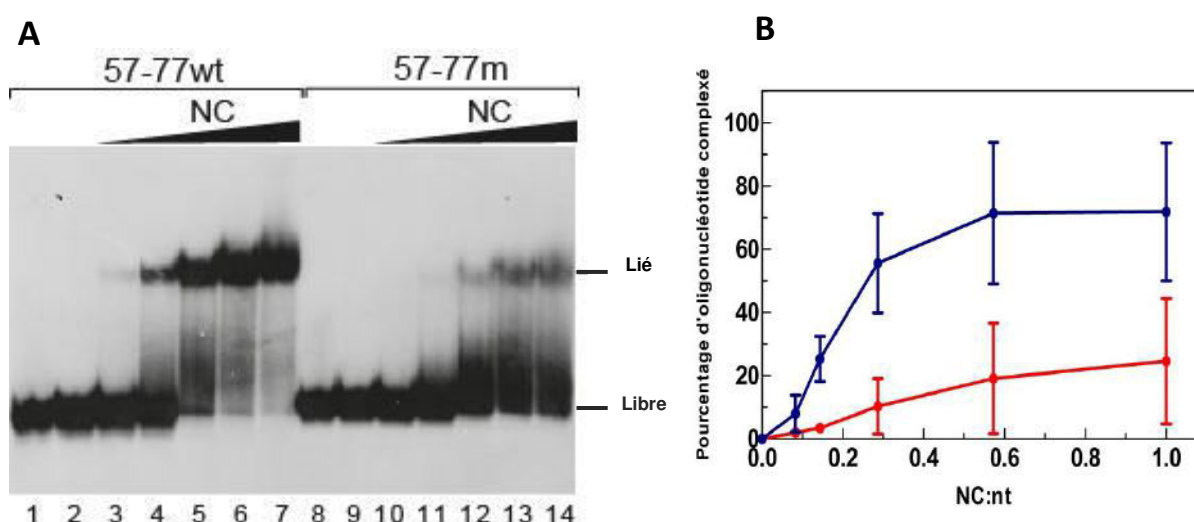


Figure 52. Étude de l'interaction entre la NC et les oligodésoxyribonucléotides 57-77wt et 57-77m. **A)** Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les échantillons ont été analysés par électrophorèse sur gel de polyacrylamide 16% à 4°C dans le tampon 0,5X TBE. Les pistes 1 et 8 représentent les échantillons dénaturés dans l'eau. Les pistes 2 et 9 correspondent aux échantillons incubés à 37°C en absence de NC. Les pistes 3-7 et 10-14 représentent les oligonucléotides incubés à 37°C en présence de concentrations croissantes de NC pour avoir respectivement un rapport NC:nt de 1:12 ; 1:7 ; 1:3,5 ; 1:1,75 et 1:1. **B)** Graphe représentant le pourcentage de complexe NC-oligonucléotide en fonction du rapport NC:nt. Les courbes bleu et rouge indiquent respectivement le 57-77wt et le 57-77m. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

2.3 91-111 wt et mutant :

L'oligonucléotide 91-111wt représente la région contenant la tige-boucle apicale cpoly(A) (**Figure 53**). Les expériences réalisées avec les deux oligonucléotides montrent qu'il n'y a pas de différence significative entre la fixation de la NC sur la séquence sauvage ou mutée à la position 101 (**Figure 54**). En effet, les résultats révèlent qu'au rapport NC:nt égal à 1:7, le pourcentage de complexe NC-oligonucléotide formé est de 8,4% pour le 91-111wt et 8,8% pour le 91-111m.

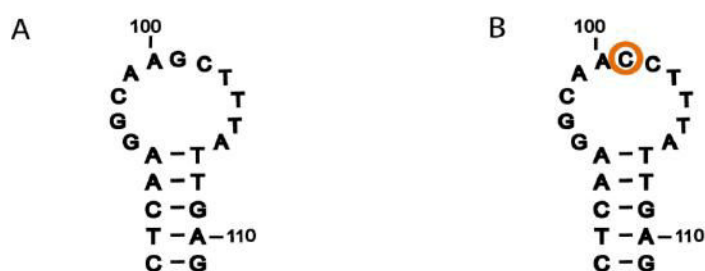


Figure 53. Structures secondaires des oligodésoxyribonucléotides 91-111wt (A) et 91-111m (B). La mutation sensée supprimer ou diminuer fortement la fixation de la NC est entourée d'un cercle orange.

Résultats et discussion

De plus, nous remarquons une augmentation progressive du complexe formé avec l'augmentation de la concentration en NC, jusqu'à atteindre 55,5% avec la séquence sauvage et 55,5% avec la séquence mutée, au rapport NC:nt égal à 1:1. En présence de concentrations croissantes en NC nous observons une seule bande supplémentaire par rapport à l'oligonucléotide incubé sans NC (**Figure 54**). Cette observation suggère fortement que chaque oligonucléotide fixe spécifiquement une seule NC, c'est-à-dire il y aurait un seul site par oligonucléotide. De plus, nos résultats suggèrent que G101 n'est pas la guanine interagissant avec la NC dans les oligonucléotides 91-111.

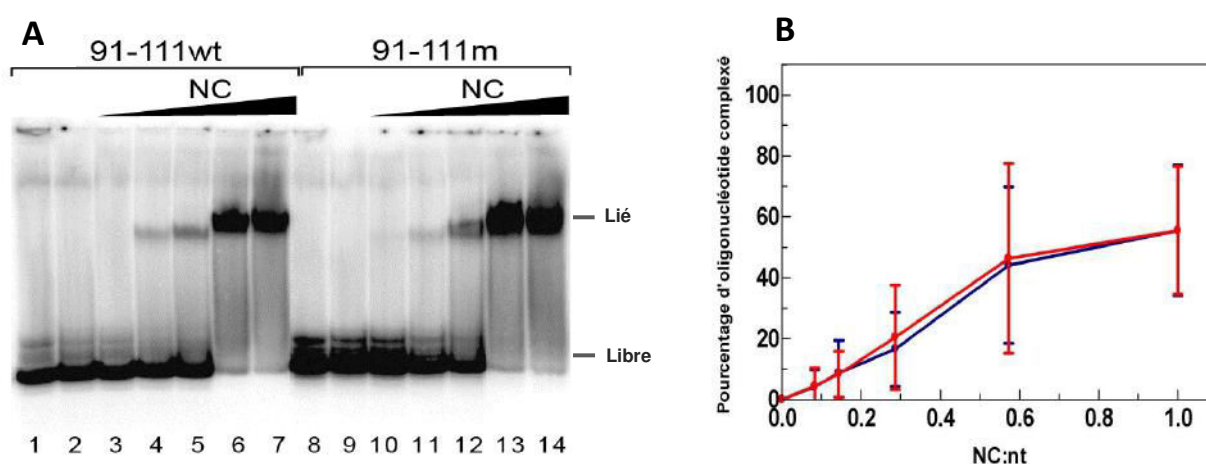


Figure 54. Étude de l'interaction entre la NC et les oligodésoxyribonucléotides 91-111wt et 91-111m. **A)** Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les échantillons ont été analysés par électrophorèse sur gel de polyacrylamide 16% à 4°C dans le tampon 0,5X TBE. Les pistes 1 et 8 représentent les échantillons dénaturés dans l'eau. Les pistes 2 et 9 correspondent aux échantillons incubés à 37°C en absence de NC. Les pistes 3-7 et 10-14 représentent les oligonucléotides incubés à 37°C en présence de concentrations croissantes de NC pour avoir respectivement un rapport NC:nt de 1:12 ; 1:7 ; 1:3,5 ; 1:1,75 et 1:1. **B)** Graphe représentant le pourcentage de complexe NC-oligonucléotide en fonction du rapport NC:nt. Les courbes bleu et rouge indiquent respectivement le 91-111wt et le 91-111m. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

Nos analyses des interactions NC-ADNss dont les résultats sont résumés dans la figure 5 l'article, ont montré qu'en présence de la NC, les coupures MB en A99 et A100 sont fortement réduites en comparaison avec les essais réalisés en absence de NC. En revanche, nous remarquons que la mutation du G101 en cytosine n'a pas eu d'impact significatif sur la capacité de fixation de la NC sur l'oligonucléotide. Ceci pourrait être dû à la présence de deux guanines aux positions 96 et 97 au niveau de la boucle apicale de cpoly(A). En effet, il a été montré que la NC nécessite la présence d'au moins une guanine pour se fixer aux acides nucléiques (158, 160, 417). Ainsi, ces deux guanines pourraient être un site de fixation de la NC au niveau des deux oligonucléotides. En revanche, elles seraient inaccessibles dans le

contexte de l'ADNss entier. En accord avec cette hypothèse, les expériences de coupure à la DNaseI montrent la présence d'une coupure en position 97 dont la réactivité augmente fortement en présence de NC (421). Ces observations suggèrent qu'au niveau des G96 et G97, il se produirait des appariements qui rendraient les deux guanines inaccessibles à la NC. Ces appariements, probablement transitoires, pourraient avoir lieu au niveau d'une zone simple brin de l'ADNss. Une autre hypothèse est que les oligonucléotides forment une tige instable et par conséquent la NC interagirait avec le motif TG.

2.4 96-107 wt et mutant :

Pour confirmer que dans l'oligonucléotide le G101 n'interagit pas préférentiellement avec la NC, nous avons utilisé deux oligonucléotides représentant la boucle apicale du cpoly(A) et ayant une longueur de 12 nucléotides. La NC est prédite ne pas se fixer sur G96 et G97 car ils sont à l'extrémité 5' des oligonucléotides tandis qu'elle devrait pouvoir interagir avec G101 qui occupe une position centrale dans l'oligonucléotide.

Le logiciel mfold prédit que les oligonucléotides 96-107wt et 96-107m ne sont pas capables de former des structures secondaires stables. Les expériences de gel retard effectuées avec les deux oligonucléotides n'ont montré aucun complexe formé (résultats non montrés). En revanche, les expériences d'empreinte à la MB suggèrent que la séquence C₉₈AAGC₁₀₂ est en interaction avec la NC dans l'ADNss (421). Bien que cette séquence présente une guanine non appariée qui est un critère majeur dans la reconnaissance par la NC, elle ne possède pas les motifs TG et GNG qui constituent des sites de haute affinité pour la NC. Les résultats obtenus avec la DNaseI suggèrent que la boucle apicale du cpoly(A) est en partie structurée (421). Il est donc possible que la guanine 101 dans l'ADNss est dans une conformation adéquate pour interagir avec la NC.

2.5 124-135 wt et mutant :

Le logiciel mfold prédit que les oligonucléotides 124-135wt et 124-135m ne sont pas capables de former des structures secondaires stables. Les expériences de retard sur gel réalisées avec les deux oligonucléotides montrent que la séquence sauvage fixe la NC tandis que la séquence mutée la fixe peu (**Figure 55**). Ces résultats sont en accord avec les expériences de coupure à la MB qui suggèrent fortement que la séquence C₁₂₉TTGC₁₃₃ est en interaction avec la NC (421). Ainsi, une mutation ponctuelle (C132 à la place de G132) est suffisante pour altérer fortement la reconnaissance par la NC. Nos résultats renforcent la notion que la NC reconnaît préférentiellement les motifs TG non appariés.

Résultats et discussion

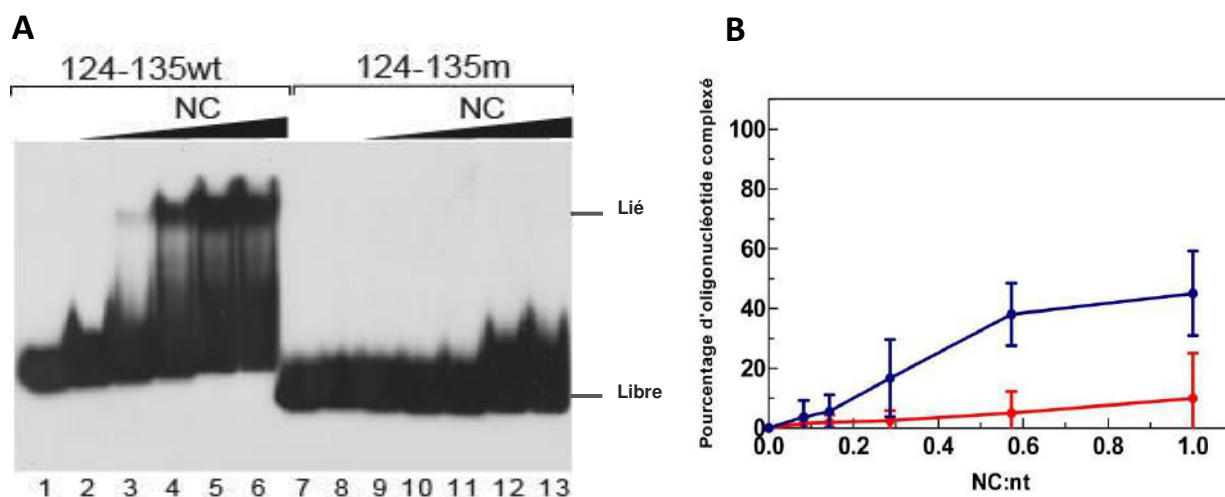


Figure 55. Etude de l'interaction entre la NC et les oligodésoxyribonucléotides 124-135wt et 124-135m. A) Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les échantillons ont été analysés par électrophorèse sur gel de polyacrylamide 17% à 4°C dans le tampon 0,5X TBE. Les pistes 1 et 8 représentent les échantillons dénaturés dans l'eau. Les pistes 2 et 9 correspondent aux échantillons incubés à 37°C en absence de NC. Les pistes 3-7 et 10-14 représentent les oligonucléotides incubés à 37°C en présence de concentrations croissantes de NC pour avoir respectivement un rapport NC:nt de 1:12 ; 1:7 ; 1:3,5 ; 1:1,75 et 1:1. B) Graphe représentant le pourcentage de complexe NC-oligonucléotide en fonction du rapport NC:nt. Les courbes bleu et rouge indiquent respectivement le 124-135wt et le 124-135m. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

En conclusion, les résultats obtenus avec les oligonucléotides 57-77wt et 124-135wt sont parfaitement en accord avec la notion que les séquences de l'ADNss qu'ils représentent interagissent préférentiellement avec la NC. Il est intéressant de noter que ces oligonucléotides ne sont probablement pas structurés. Par conséquent, ils ne sont pas de bons modèles pour comprendre comment la NC déstabilise les tiges présentes dans l'ADNss. Les oligonucléotides 18-38C29 et 91-111wt adoptent probablement des structures en tige-boucle et seraient de bons modèles pour analyser par RMN la déstabilisation des tiges par la NC. Malheureusement, ces oligonucléotides ne reflètent probablement pas les interactions qui s'établissent entre la NC et les séquences de l'ADNss qu'ils représentent.

3. Analyse de la formation de l'heteroduplexe R-r

L'un des objectifs de ce travail a été d'étudier les mécanismes moléculaires qui régissent l'hybridation de l'ADNss avec l'extrémité 3' du génome en présence et en absence de NC. A notre connaissance, cette étude est la première à s'intéresser à l'appariement des séquences entières des acides nucléiques (R et r) intervenant dans le premier transfert de brin.

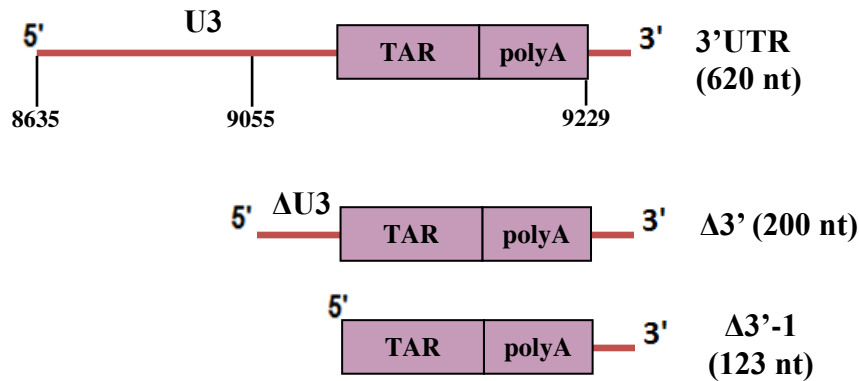


Figure 56. Structures primaires des ARN utilisés.

3.1 Analyse en absence de NC

3.1.1 Essais d'heterodimérisation de l'ADNss avec l'extrémité 3' du génome

Dans le but de déterminer s'il existe des interactions faibles (par exemple du type boucle-boucle) entre l'ADNss et la région 3' du génome, des expériences d'heterodimérisation ont été effectuées, en l'absence de NC, entre l'ADNss et les ARN Δ3'-1 (123 nt), Δ3' (200 nt) ou 3'UTR (620 nt) correspondants à la région 3'UTR du génome complète ou tronquée (**Figure 56**). Les analyses ont été effectuées par électrophorèse sur gel natif d'agarose dans le tampon de migration 0,5X TBM. En effet, des études ont montré que l'analyse par électrophorèse à 4°C sur gel d'agarose contenant du MgCl₂ permet d'observer les duplexes dont les sous-unités sont liées par des interactions faibles (158, 422–424). Nos résultats montrent qu'en absence de NC l'ADNss reste sous forme monomérique en présence de concentrations croissantes des ARN Δ3' et 3'UTR (**Figure 57**). Ces résultats suggèrent qu'il n'existe pas d'interactions faibles entre l'ADNss et l'extrémité 3' de l'ARNg du VIH-1. On ne peut cependant pas totalement exclure l'hypothèse qu'il existe des interactions faibles qui ne résistent pas à la migration électrophorétique bien que celle-ci soit réfrigérée à 4°C en présence de MgCl₂.

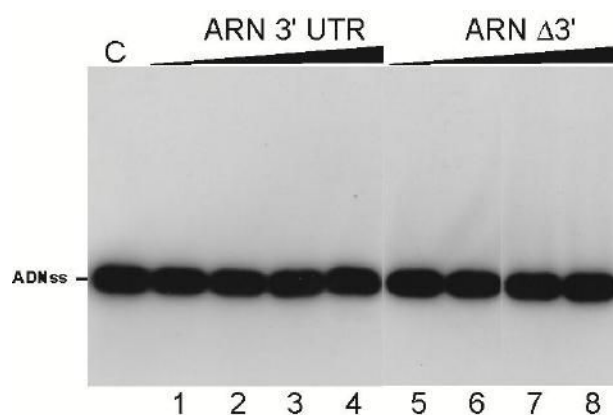


Figure 57. Essais d'heterodimérisation de l'ADNs avec l'extrémité 3' du génome. Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les échantillons ont été analysés par électrophorèse sur gel d'agarose 1,2 % à 4°C dans le tampon 0,5X TBM. La piste C correspond à l'ADNs seul. Les pistes 1-4 et 5-8 représentent l'ADNs en présence de concentrations croissantes en ARN $\Delta 3'$ ou 3'UTR, les rapports ARN:ADN utilisés sont 1:1 ; 2:1 ; 4:1 ; 8:1.

3.1.2 Analyse structurale de l'ADNs en présence de l'extrémité 3' du génome

Les essais d'hétérodimérisation entre l'ADNs et l'extrémité 3' du génome n'ont pas permis de mettre en évidence des interactions faibles ou fortes entre ces deux acides nucléiques en absence de NC. On ne peut cependant pas totalement exclure l'existence d'interactions faibles entre ces deux acides nucléiques qui se dissocient au cours de l'électrophorèse.

Pour identifier ces éventuelles interactions, des expériences de caractérisation structurale par les sondes de structures ont été effectuées dans les mêmes conditions que les expériences d'hétérodimérisation. C'est-à-dire qu'après l'incubation de l'ADNs avec l'ARN $\Delta 3'-1$, $\Delta 3'$ ou 3'UTR pendant 30 min à 4°C, les essais ont été traités avec la MB, la DNaseI ou le KMnO_4 afin d'identifier des protections induites par l'extrémité 3' du génome. Nos résultats ont révélé qu'il n'y a pas de changement de réactivité au KMnO_4 et à la DNaseI dans l'ADNs en présence des ARN $\Delta 3'-1$, $\Delta 3'$ ou 3'UTR (résultats non montrés). En revanche, en présence de l'ARN 3'UTR, la plupart des coupures MB dans la région u5 de l'ADNs sont légèrement réduites (**Figure 58**). D'autre part, quatre coupures MB ont fortement été réduites à savoir T15 et C17 (situés dans la boucle interne de u5), A₁₈ et T₂₂ (situé dans la boucle apicale de u5) (**Figure 59**). Ceci suggère qu'en absence de NC, l'ADNs est engagé dans des interactions faibles avec l'ARN 3'UTR mais pas avec les ARN $\Delta 3'-1$ et $\Delta 3'$. Ainsi, la séquence 8635-9055 de la région U3, qui est absente chez les ARN $\Delta 3'-1$ ou $\Delta 3'$ (**Figure 56**), possède probablement des points de contacts avec l'ADNs.

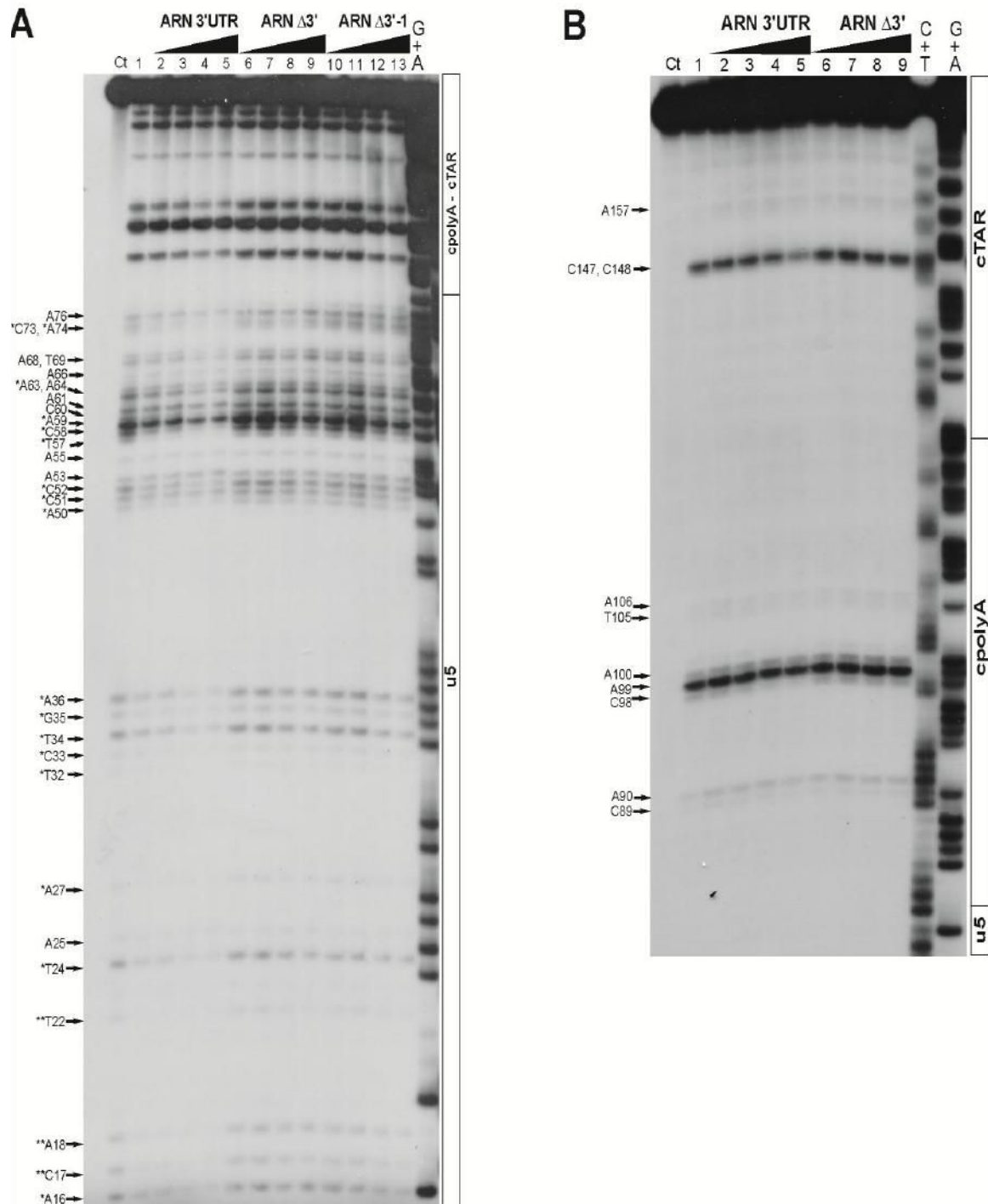


Figure 58. Analyse structurale de l'ADNss en présence de l'extrémité 3' du génome. Les expériences d'analyse structurale par les sondes de structure ont été effectués comme décrit dans la partie « Matériel et Méthodes ». La piste Ct correspond à l'ADNss non incubé avec la MB et en absence d'ARN. La piste 1 correspond à l'ADNss seul et incubé avec la MB (4 unités). Les pistes 2-5, 6-9 et 10-13 correspondent aux échantillons incubés avec la MB (4 unités) en présence des ARN 3'UTR, $\Delta 3'$ ou 3'-1. Les échantillons ont été analysés sur gel 12% (A) et 8% (B) de polyacrylamide/7,2 M d'urée dans le tampon 1X TBE. Une étoile indique un nucléotide qui est faiblement ou moyennement protégé en présence de l'ARN. Deux étoiles indiquent un nucléotide qui est fortement protégé en présence d'ARN.

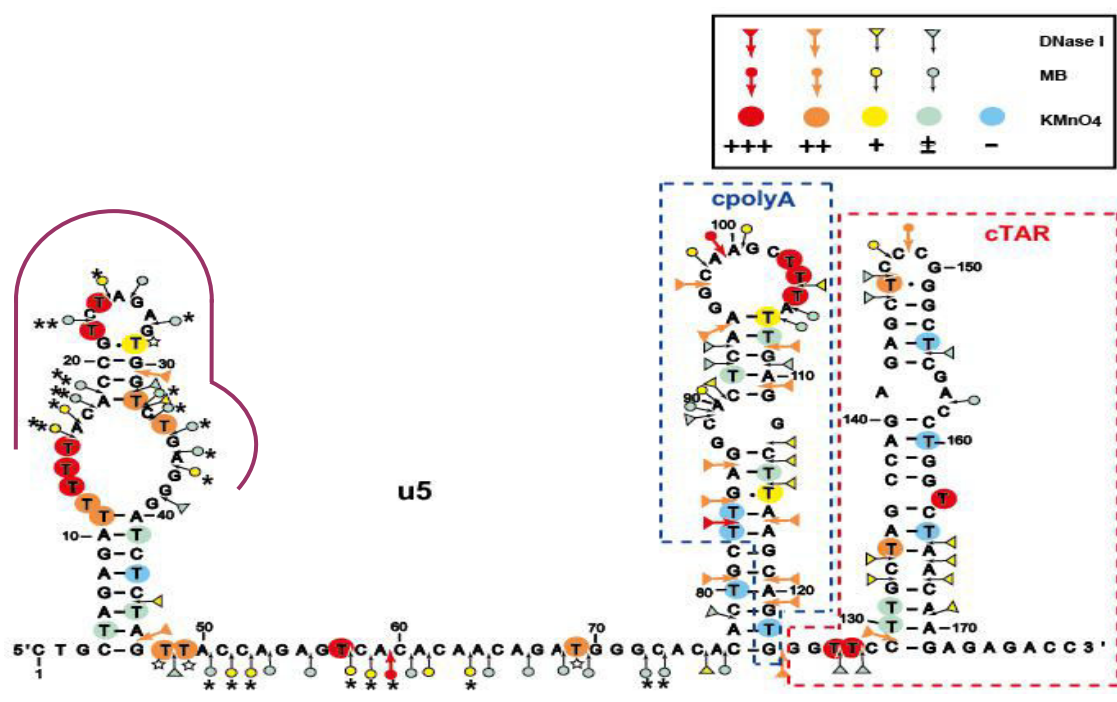


Figure 59. Modèle de structure secondaire de l'ADNs en présence de l'extrémité 3' du génome. Le code couleur utilisé pour désigner les réactivités est encadré. Deux étoiles indiquent un nucléotide qui est fortement protégé en présence d'ARN. Les régions effectuant des interactions faibles avec la séquence (8635-9055) sont indiquées par un trait violet.

3.2 Analyses en présence de NC

Jusqu'à notre étude, il n'avait pas été reconstitué *in vitro*, en présence de NC, le duplexe r-R avec l'ADNs entier et une partie de l'extrémité 3' de l'ARNg du VIH-1.

3.2.1 Effet du protocole de renaturation dans la cinétique d'hybridation

Dans des expériences préliminaires, nous avons montré qu'il est possible de former le duplexe r-R en utilisant deux méthodes différentes d'hybridation, nous avons donc comparé les cinétiques d'hybridation des deux méthodes. La cinétique d'hybridation de l'ADNs wt avec l'ARN $\Delta 3'$ a été d'abord étudiée, en effectuant la renaturation des acides nucléiques (après dénaturation à 90°C) séparément et en absence de NC pendant 15 min à 37°C (voir la partie Matériel et Méthodes), avant d'incuber ensemble et en présence de NC les deux acides nucléiques (**Figure 60 A**). Les résultats montrent qu'il y a 18,7 % d'heteroduplexe formé à 1 min, ce pourcentage augmente progressivement jusqu'à atteindre 72,7 % après 120 min d'incubation (**Figure 60 C**). Puisque *in vivo* les deux acides nucléiques sont toujours en présence de NC, une seconde stratégie expérimentale a été développée de sorte à ce que les

Résultats et discussion

deux acides nucléiques soient renaturés séparément en présence de NC (voir la partie Matériel et Méthodes) avant d'être incubés ensemble (**Figure 60 B**).

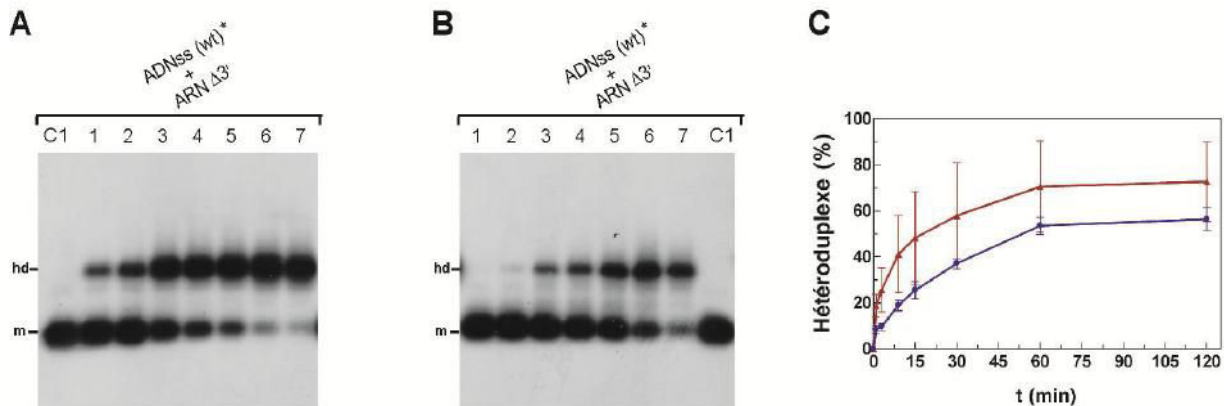


Figure 60. Effet du protocole de renaturation dans la cinétique d'hybridation. Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les essais ont été analysés par électrophorèse à 25°C sur gel d'agarose 2% dans le tampon 0,5X TBE. Les deux acides nucléiques ont été renaturés séparément en absence (A) ou en présence de NC (B). La piste C1 correspond à l'ADNss incubé à 37°C pendant 120 min en présence de l'ARN $\Delta 3'$ et en absence de NC. Les pistes 1-7 correspondent respectivement aux essais incubés à 37°C pendant 1, 3, 9, 15, 30, 60 et 120 min en présence de NC au rapport NC:nt égal à 1:10. Les formes monomérique et hétéroduplexe sont désignées respectivement par m et hd. C) Graphe représentant le pourcentage de complexe NC-oligonucléotide en fonction du temps. La courbe bleue indique le pourcentage d'hétéroduplexe obtenu lorsque les deux acides nucléiques sont renaturés en présence de NC. La courbe rouge indique le pourcentage d'hétéroduplexe obtenu lorsque les deux acides nucléiques sont renaturés en absence de NC. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

Les résultats montrent qu'à 1 min il y a 8,6 % de complexe formé, ce pourcentage augmente progressivement jusqu'à atteindre 56,5 % après 120 min d'incubation (**Figure 60 C**). Ainsi, la première méthode de renaturation permet de générer plus d'hétéroduplexe que la seconde. Toutefois, nous avons choisi d'utiliser la seconde méthode dans la suite de nos expériences car elle se rapproche plus des conditions *in vivo* et il y a peu de variabilité des résultats lorsque l'on utilise cette méthode (voir les barres d'erreur dans la **Figure 57 C**). Il faut souligner qu'il n'y a quasiment pas d'hétéroduplexe formé en l'absence de NC quelque soit la méthode de renaturation (contrôles C1 dans la **Figure 57**). Nos résultats montrent donc que la NC est essentielle pour la formation du duplexe r-R lorsque l'on utilise l'ADNss entier et une partie de l'extrémité 3' du génome.

3.2.2 Cinétiques d'hybridation des ADNss mutés

A partir d'expériences réalisées avec des acides nucléiques dont la taille était inférieure ou égale à 59 nucléotides (158, 411), deux voies possibles d'hybridation ont été décrites, le *kissing complex* initiant l'hybridation avec les boucles apicales ; le *zipper complex* qui initie l'hybridation par la base des tiges. Nous avons conçu deux mutants pour tester les deux voies possibles d'hybridation, SL1 est un ADNss qui est muté au niveau de la boucle apicale du cTAR (**Figure 61**) afin d'empêcher l'initiation de l'appariement via le *kissing complex*, INV est muté au niveau des extrémités 5' et 3' de la tige-boucle cTAR (**Figure 61**) afin d'empêcher l'initiation de l'appariement via le *zipper complex*.

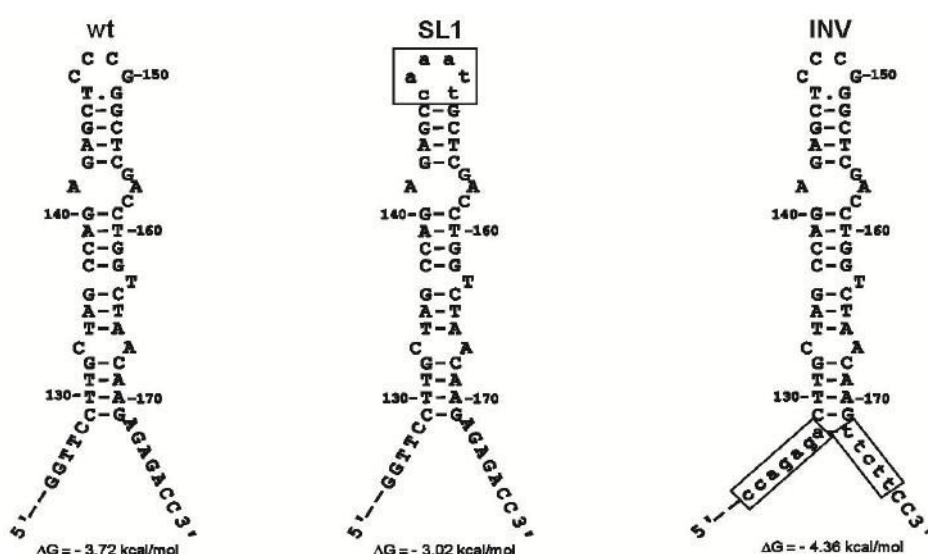


Figure 51. Structure secondaire prédite des séquences cTAR dans les ADNss wt et mutants. Le logiciel mfold de zucker (419) a été utilisé pour prédire la structure secondaire la plus stable pour chaque séquence cTAR. Le misappariement T.G a été observé par RMN de la tige supérieure de cTAR (418). Les mutations sont indiquées par les lettres encadrées.

Les expériences ont été réalisées en présence du rapport NC:nt égal à 1:18 (**Figure 62 A-C**) ou 1:7 (**Figure 62 D-F**). Au rapport 1:18, les ADNss mutés forment plus d'hétéroduplexes que l'ADNss wt (**Figure 62 C**). Par exemple après 120 min, 43% d'ADNss wt est sous la forme d'hétéroduplexes tandis que c'est au moins 65% dans le cas des mutants. Une hypothèse pour expliquer ces résultats est que les structures des mutants sont moins stables que celle du sauvage et donc plus facilement déstabilisables par la NC. Le programme mfold prédit cependant que la tige-boucle cTAR est plus stable dans le mutant INV que dans l'ADNss sauvage (**Figure 62**). Toutefois, nos résultats suggèrent que les mutations au niveau de la boucle apicale ou de la base de la tige cTAR, facilitent le processus d'hybridation en réduisant la stabilité des ADNss mutants. En accord avec cette hypothèse, les expériences

Résultats et discussion

d'hybridation effectuées en présence d'une forte concentration en NC montrent que les ADNss mutants et sauvage présentent les mêmes cinétiques d'hybridation (**Figure 62 F**). De plus, nous observons également la présence de complexes de haut poids moléculaire avec les ADNss mutants qui sont absents avec l'ADNss sauvage, suggérant la formation de multimères composés d'au moins trois molécules

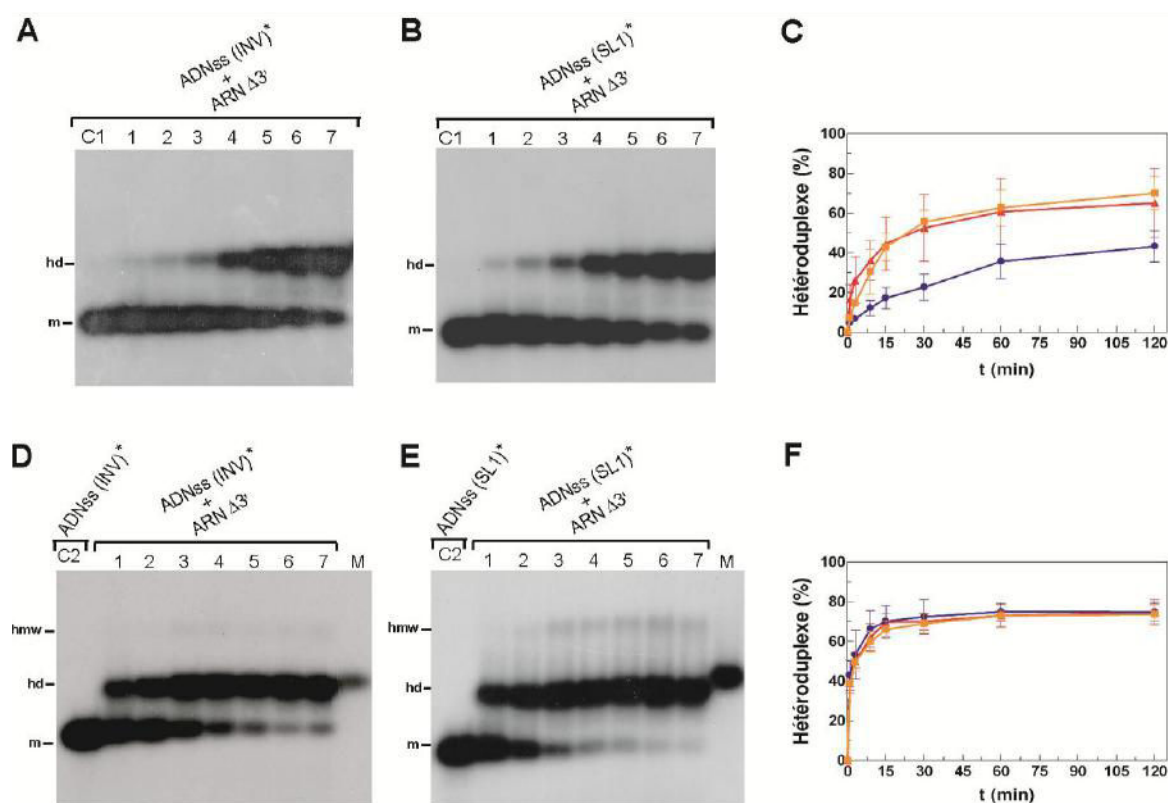


Figure 62. Cinétiques d'hybridation des ADNss mutés. Les essais de retard sur gel ont été effectués comme décrit dans la partie « Matériel et Méthodes ». Les essais ont été analysés par électrophorèse à 25°C sur gel d'agarose 2% dans le tampon 0,5X TBE. **A, B)** La piste C1 correspond à l'ADNss incubé en présence de l'ARN $\Delta 3'$ et en absence de NC pendant 120 min. Les pistes 1-7 correspondent respectivement aux essais incubés à 37°C pendant 1, 3, 9, 15, 30, 60 et 120 min en présence de NC au rapport NC:nt égal à 1:18. Les formes monomérique et hétéroduplexe sont désignées respectivement par m et hd. **D, E)** La piste C2 correspond à l'ADNss mutant seul incubé en présence de NC au rapport 1 NC pour 7 nucléotides pendant 120 min. Les pistes 1-7 correspondent respectivement aux essais incubés à 37°C pendant 1, 3, 9, 15, 30, 60 et 120 min en présence de NC au rapport NC:nt égal à 1:7. Les complexes de haut poids moléculaire sont désignées par hmw. La piste M représente le marqueur de taille ADN (356 nucléotides). **C, F)** Graphes représentant le pourcentage de complexe NC-oligonucléotide en fonction du temps. Les courbes bleue, orange et rouge indiquent respectivement l'ADNss wt, l'ADNss INV et l'ADNss SL1. Chaque point est la moyenne d'au moins trois expériences. Les barres d'erreur indiquent les déviations standards.

Résultats et discussion

En résumé, nos résultats suggèrent fortement que le processus d'hybridation médié par la NC ne repose pas sur une seule des deux voies décrites. Ainsi, la NC peut initier l'hybridation au niveau de séquences complémentaires qui ne font pas intervenir les boucles apicales et les extrémités 3'/5' des tiges boucles cTAR et TAR. De plus, nos résultats sont en accord avec le modèle faisant intervenir les tiges-boucles poly(A) et cpoly(A). En effet, nos résultats sont compatibles avec le modèle de transfert de brin par invasion proposé par Bambara *et al*, (331, 407). Dans ce modèle, la pause de la RT au niveau de la base de la tige-boucle TAR en 5' initie les coupures par la RNase H au niveau de la séquence poly(A) en 5' (ARN donneur) permettant ainsi l'interaction de la tige-boucle poly(A) en 3' (ARN accepteur) avec la tige-boucle cpoly(A) de l'ADNss naissant. Le transfert de brin est ainsi initié par un processus de propagation jusqu'au transfert total de l'ADNss (**Figure 36**).

Conclusions et perspectives

Conclusions et perspectives

Mon sujet de thèse a porté sur la compréhension des mécanismes moléculaires qui régissent le premier transfert de brin chez le VIH-1. Ce transfert joue un rôle crucial dans la transcription inverse du génome du VIH-1. Il repose en grande partie sur la formation d'un hybride ADN-ARN impliquant la séquence *r*, qui est une partie de l'ADNss, et la séquence *R* située à l'extrémité 3' de la région 3'UTR de l'ARNg viral. On sait que la formation de l'hybride ADN-ARN dépend de la structure des acides nucléiques impliqués et est facilitée par la NC, mais les structures, interactions et dynamiques qui gouvernent cette hybridation sont mal connues. Jusqu'à notre étude, la formation de l'hybride ADN-ARN n'avait été étudiée qu'avec des courts acides nucléiques qui ne reflètent que partiellement les structures formées par l'ADNss et l'ARN 3'UTR dans le virus en cours de réplication.

Nous avons été les premiers à déterminer, en l'absence et présence de NC, la structure secondaire de l'ADNss entier du VIH-1. A notre connaissance, les structures secondaires des ADNss d'autres rétrovirus n'ont pas été déterminées. Nos résultats obtenus avec trois sondes de structure (DNase I, *mung bean nuclease* et permanganate de potassium) sont en faveur d'une structure secondaire de l'ADNss du VIH-1 constituée de trois tiges-boucles (u5, cpoly(A) et cTAR) et trois régions simple-brin en l'absence ou présence de NC (Figure 5 de l'article). Il est intéressant de noter qu'une partie de la séquence u5, qui a priori n'interagit pas avec l'ARN 3'UTR, est la plus longue (29 nt) des trois régions simple-brin et qu'elle ne dépend pas de la présence de la NC. Il est aussi intéressant de souligner que la région *r*, qui est constituée des séquences cpoly(A) et cTAR, ne se replie pas sous la forme d'un domaine structural totalement indépendant. En effet, six paires de bases sont formées entre les séquences u5 et cpoly(A) (Figure 8 de l'article). Les courts ADN (55 ou 59 nt) correspondant à l'ADN cTAR forment, en absence de NC, deux types de tige-boucle qui sont en équilibre : une conformation fermée et une conformation en Y qui est partiellement ouverte (158, 362, 405). Nos données structurales indiquent que dans l'ADNss la séquence cTAR adopte la conformation en Y semi-ouverte (conformère 1 de la Figure 8 de l'article) mais pas la conformation fermée. Par conséquent, l'ADN cTAR, un modèle d'étude très utilisé, adopte la conformation fermée qui n'existe pas dans l'ADNss qui est le substrat naturel du premier transfert de brin.

Notre analyse phylogénétique réalisée avec les séquences des différents groupes du VIH-1 est aussi en faveur d'une structure secondaire de l'ADNss constituée de trois tiges-boucles (u5, cpoly(A) et cTAR) et trois régions simple-brin (Figure 7 de l'article). L'appariement impliquant les nucléotides 4-11 et 41-58 (numérotation relative au groupe M du VIH-1) pourrait jouer un rôle important dans la synthèse des brins (-) et (+) de l'ADN viral

Conclusions et perspectives

car il est très conservé. Cet appariement est aussi conservé au niveau de l'ARNg et pourrait faciliter l'initiation de la transcription inverse à partir de l'amorce ARNt^(Lys,3) (425). De même, les structures en tige-boucle des séquences TAR et poly(A) sont conservées au niveau de l'ADNss (cette étude) et de l'ARNg (398, 426). Par conséquent, nous ne pouvons pas exclure la possibilité que la conservation des éléments structuraux dans l'ADNss est due à une conservation au niveau de l'ARNg. Une hypothèse attrayante est que les éléments structuraux ont été conservés au niveau de l'ARNg et de l'ADNss pour permettre une réplication optimale du virus, c'est-à-dire faciliter la transcription inverse, la synthèse et polyadénylation de l'ARNg.

Nous avons été les premiers à identifier quatre sites préférentiels de fixation de la NC au niveau de l'ADNss. Ces sites sont constitués de nucléotides non appariés et trois d'entre eux contiennent le motif TG qui est connu pour interagir préférentiellement avec la NC (152, 418, 420). Dans la perspective d'études RMN ultérieures, nous avons étudié par retard sur gel l'interaction de la NC avec des oligonucléotides contenant chacun l'un des sites forts. Nos résultats suggèrent que seulement deux oligonucléotides miment probablement bien l'interaction qu'établit la NC avec les sites forts 67-71 et 129-133 dans l'ADNss (Figure 8 de l'article). Ces oligonucléotides ne sont pas structurés et ne sont donc pas intéressants pour déterminer par RMN comment la NC déstabilise des éléments structuraux de l'ADNss. Nos données structurales suggèrent que la structure secondaire de l'ADNss et la partie inférieure de la tige-boucle cTAR sont plus ouvertes en présence de NC qu'en son absence (Figure 5 de l'article). Il est intéressant de noter que l'effet de la NC sur l'ADN cTAR modèle (158) est proche mais pas identique à celui observé sur le domaine cTAR dans l'ADNss. A partir de nos résultats, nous proposons un modèle d'hybridation dans lequel la première étape est la fixation de la NC au niveau de ses sites forts pour déclencher l'ouverture de la structure tridimensionnelle de l'ADNss et favoriser ainsi l'accessibilité de la région r ; la seconde étape étant la déstabilisation par la NC des tiges-boucles ; la troisième étape étant l'appariement des séquences r et R par la NC.

Nos essais d'hétérodimérisation analysés par retard sur gel n'ont pas permis de révéler des interactions entre l'ADNss et l'ARN 3'UTR en l'absence de NC. En revanche, l'étude structurale réalisée avec trois sondes (DNase I, *mung bean nuclease* et permanganate de potassium) suggère qu'une partie de la région u5 établit des interactions faibles avec une partie de la région U3 de l'ARNg. Il est intéressant de noter que les séquences complémentaires r et R ne présentent aucune interaction en absence de NC. Nous n'avons donc pas mis en évidence d'interactions entre les boucles apicales des TAR et cTAR et celles

Conclusions et perspectives

des poly(A) et cpoly(A). En ce qui concerne le processus d'hybridation en présence de NC, nous avons été les premiers à l'étudier *in vitro* avec l'ADNss entier et une partie de l'ARN 3'UTR nommé ARN $\Delta 3'$. Nos résultats montrent clairement que la NC est requise pour former l'hybride ADNss-ARN $\Delta 3'$. Les cinétiques d'hybridation que nous avons réalisées avec l'ADNss sauvage et deux ADNss mutés montrent que l'initiation de l'hybridation ne repose pas sur une seule voie, à faible ou forte concentration de NC. L'hybridation pourrait être simultanément initiée au niveau des deux voies classiquement proposées (complexe boucle-boucle et interaction par la partie inférieure des tiges) ou d'une troisième impliquant l'appariement des tiges-boucles ARN poly(A) et ADN cpoly(A). En accord avec cette hypothèse, des études *in vitro* suggèrent que l'appariement des séquences poly(A) et cpoly(A) joue un rôle important dans le transfert (331, 408).

Les résultats que nous avons obtenus vont permettre d'initier de nouvelles études *in vitro* qui auront pour objectif principal de valider nos hypothèses. Des mutations ponctuelles empêchant la fixation de la NC au niveau de ses quatre sites forts devraient avoir un effet sur la déstabilisation de la structure tridimensionnelle de l'ADNss. D'autres mutations ponctuelles au niveau de certaines guanines devraient pouvoir avoir un effet sur la déstabilisation des tiges de l'ADNss. En effet, des travaux récents dont ceux de notre équipe ont montré que la déstabilisation d'une tige par la NC peut être initiée à partir d'un site secondaire, c'est-à-dire une paire de bases contenant une guanine et adjacente à une boucle ou à un *bulge* (139, 414). Par conséquent, les mutations mentionnées ci-dessus devraient avoir un effet négatif sur la cinétique d'appariement de la séquence r de l'ADNss avec la séquence R de l'ARN 3'UTR.

Grâce aux résultats que nous avons obtenus, mon équipe d'accueil envisage des études *ex vivo* en collaboration avec une équipe de l'unité. Une de nos hypothèses est que le repliement de l'ADNss *ex vivo* est identique ou très proche de celui qu'il adopte *in vitro* car la structure secondaire est conservée parmi les différents isolats et groupes du VIH-1. A ce jour, cette hypothèse n'a pas été démontrée. La NC ne déstabilise pas des duplexes stables, c'est-à-dire ayant en général une longueur supérieure à huit paires de bases consécutives (64). En accord avec cette observation, notre analyse phylogénétique prédit que dans l'ADNss les tiges n'excèdent pas huit paires de bases Watson-Crick consécutives chez les différents isolats et groupes du VIH-1. Par conséquent, si *ex vivo* les processus d'hybridation dépendent principalement de l'action de la NC, la stabilisation des tiges devrait avoir un effet important sur ces processus. Ainsi, nous supposons que des mutations ponctuelles stabilisant les tiges

Conclusions et perspectives

des éléments cpoly(A) et cTAR devraient empêcher ou fortement diminuer la déstabilisation de ces tiges par la NC et par conséquent avoir un effet sur l'appariement r-R, c'est-à-dire sur le premier transfert de brin. De même, la copie complète de la région u5 par la RT devrait être empêchée ou fortement diminuée par des mutations ponctuelles stabilisant la tige-boucle présente dans cette région ; ces mutations devraient donc avoir un effet sur le second transfert de brin. Une étude récente, basée en partie sur un modèle mathématique de la transcription inverse, suggère que le désassemblage de la capsid (décapsidation) est déclenché après que ce soit produit le premier transfert de brin et avant le second transfert de brin (427). Des mutations dans l'ADNss qui bloquent la transcription inverse au niveau du premier ou du second transfert de brin devraient donc empêcher ou retarder le déclenchement de la décapsidation.

En conclusion, nos résultats actuels et les prochains devraient contribuer à identifier les mécanismes moléculaires qui sont impliqués dans les deux transferts et à approfondir donc les connaissances sur la transcription inverse. Une connaissance détaillée des deux transferts de brin pourrait conduire à la conception de nouveaux antirétroviraux ciblant la NC et l'ADNss. En effet, deux études ont montré que le premier transfert de brin est inhibé *in vitro* par l'actinomycine D qui en se fixant sur l'ADNss inhibe l'étape d'hybridation (383, 384). L'actinomycine D n'a cependant pas une activité antirétrovirale spécifique car elle est d'abord un inhibiteur de la transcription et peut même activer la réplication du VIH-1 dans certaines lignées cellulaires (385). De plus, ce composé est toxique pour les cellules. Toutefois, les résultats obtenus avec l'actinomycine D suggèrent qu'il est possible de concevoir des agents pharmacologiques inhibant spécifiquement le premier transfert de brin si nous disposons d'une connaissance précise des interactions impliquant l'ADNss, l'ARN 3'UTR et la NC. Finalement, le premier transfert de brin et les transferts de brin internes impliquent des mécanismes moléculaires qui sont similaires. Par conséquent, les résultats de futures études pourraient contribuer à obtenir de nouvelles connaissances sur les transferts de brin internes qui sont principalement responsables de la recombinaison chez le VIH-1, et permettre ainsi la stimulation et l'émergence de nouvelles études sur ces processus. La compréhension de la recombinaison, responsable en partie de la diversité génétique du VIH-1, pourrait en effet aider à concevoir des agents pharmacologiques qui, en ciblant le transfert de brin dépendant de la NC, devraient réduire fortement l'apparition de souches multi-résistantes aux antirétroviraux.

Articles publiés

ARTICLE 1

Structural Insights into the HIV-1 Minus-strand Strong-stop DNA*[§]

Received for publication, December 3, 2015 Published, JBC Papers in Press, December 14, 2015, DOI 10.1074/jbc.M115.708099

Yingying Chen^{‡§1,2}, Ouerdia Maskri^{‡1,3}, Françoise Chaminade[‡], Brigitte René[‡], Jessica Benkaroun[‡], Julien Godet[¶], Yves Mély[¶], Olivier Mauffret[‡], and Philippe Fosse^{‡4}

From the [‡]LBPA, ENS Cachan, CNRS, Université Paris-Saclay, 61 avenue du Président Wilson, 94235 Cachan cedex, France, the [§]School of Life Sciences, East China Normal University, 500 Dongchuan Road, Shanghai 200241, China, and the [¶]Laboratoire de Biophotonique et Pharmacologie, UMR-CNRS 7213, Faculté de Pharmacie, Université de Strasbourg, 74 route du Rhin, 67401 Illkirch-Cedex, France

An essential step of human immunodeficiency virus type 1 (HIV-1) reverse transcription is the first strand transfer that requires base pairing of the R region at the 3'-end of the genomic RNA with the complementary r region at the 3'-end of minus-strand strong-stop DNA (ssDNA). HIV-1 nucleocapsid protein (NC) facilitates this annealing process. Determination of the ssDNA structure is needed to understand the molecular basis of NC-mediated genomic RNA-ssDNA annealing. For this purpose, we investigated ssDNA using structural probes (nucleases and potassium permanganate). This study is the first to determine the secondary structure of the full-length HIV-1 ssDNA in the absence or presence of NC. The probing data and phylogenetic analysis support the folding of ssDNA into three stem-loop structures and the presence of four high-affinity binding sites for NC. Our results support a model for the NC-mediated annealing process in which the preferential binding of NC to four sites triggers unfolding of the three-dimensional structure of ssDNA, thus facilitating interaction of the r sequence of ssDNA with the R sequence of the genomic RNA. In addition, using gel retardation assays and ssDNA mutants, we show that the NC-mediated annealing process does not rely on a single pathway (zipper intermediate or kissing complex).

Two strand transfers are required for reverse transcription of the human immunodeficiency virus type 1 (HIV-1)⁵ genome

(1). These strand transfers are enhanced by the HIV-1 nucleocapsid protein (NC) (2–6). During the first strand transfer, which involves reverse transcriptase (RT) with its RNase H activity to promote degradation of the 5'-copied genomic RNA (donor RNA), the 3'-end of the minus-strand strong-stop DNA (ssDNA) is transferred to the 3'-end of the genomic RNA (acceptor RNA). This transfer can occur in an intra- (*i.e.* transfer of ssDNA to the original RNA template) or intermolecular (*i.e.* transfer of ssDNA to the other genomic RNA copy in the virion) manner (1). The first requirement for synthesis of ssDNA is the annealing of the human primer tRNA₃^{Lys} onto the primer binding site of the genomic RNA (Fig. 1A). The model of reverse transcription presumes that the first strand transfer occurs after synthesis of the full-length ssDNA. More precisely, the transfer is mediated by base pairing of the R region at the 3'-end of the genomic RNA with the complementary r region at the 3'-end of ssDNA. The effect of shortening the length of complementarity has been studied using murine leukemia virus-based vectors capable of only a single round of replication (7). This study showed that removing more than 80% of the length of 3' R, from 69 to 12 nucleotides, did not significantly affect the efficiency of the first strand transfer. These results cannot be extrapolated directly to HIV-1, because the R regions of murine leukemia virus and HIV-1 are different. The effect of shortening the length of 3' R has also been studied in HIV-1 (8). This study showed that production of viral particles was barely affected by a 3' R sequence of only 37 nucleotides. However, this result did not directly demonstrate that the first strand transfer was not altered, because multiple rounds of viral replications were allowed. Studies on Moloney murine leukemia virus (9), spleen necrosis virus (10), and HIV-1 (11) with mutations in R suggested that partial ssDNAs ("weak-stop" DNAs) can be transferred prior to complete reverse transcription of the R region. However, the transfer of weak-stop DNAs occurred at low frequency (10, 11). In addition, the results of a study performed with a series of mutations in the 5', 3', or both R sequences suggested that the great majority of first strand transfers in HIV-1 occur after the synthesis of the full-length ssDNA (12). Therefore, an anti-HIV-1 therapy based on inhibition of the first strand transfer requires knowledge of the structures of the full-length ssDNA and the complete 3' R. Consistent with this notion, actinomycin D inhibits the first strand transfer through direct interaction with the ssDNA (13, 14).

* This work was supported in part by Centre National de la Recherche Scientifique LBPA UMR 8113, Agence Nationale de Recherche sur le Sida, and the European Project THINPAD "Targeting the HIV-1 Nucleocapsid Protein to fight Antiretroviral Drug Resistance" FP7 Grant 601969. The authors declare that they have no conflicts of interest with the contents of this article.

[§] This article contains supplemental Fig. S1.

¹ Both authors contributed equally to this work.

² Supported by doctoral fellowships provided by the Ministère des Affaires Étrangères, China Scholarship Council, and East China Normal University.

³ Supported by a doctoral fellowship provided by l'école doctorale Sciences Pratiques.

⁴ To whom correspondence should be addressed. Tel.: 33-147405999; Fax: 33-147407671; E-mail: fosse@ens-cachan.fr.

⁵ The abbreviations used are: HIV-1, human immunodeficiency virus type 1; KMnO₄, potassium permanganate; MBN, mung bean nuclease; NC, nucleocapsid protein; poly(A), polyadenylation signal; cpoly(A), complementary DNA sequence to poly(A); R, terminal repeat; r, complementary DNA sequence to R; ssDNA, minus-strand strong-stop DNA; TAR, transactivator response element; cTAR, complementary DNA sequence to TAR; U5, sequence unique at the 5'-end; u5, complementary DNA sequence to U5; U3, sequence unique at the 3'-end.

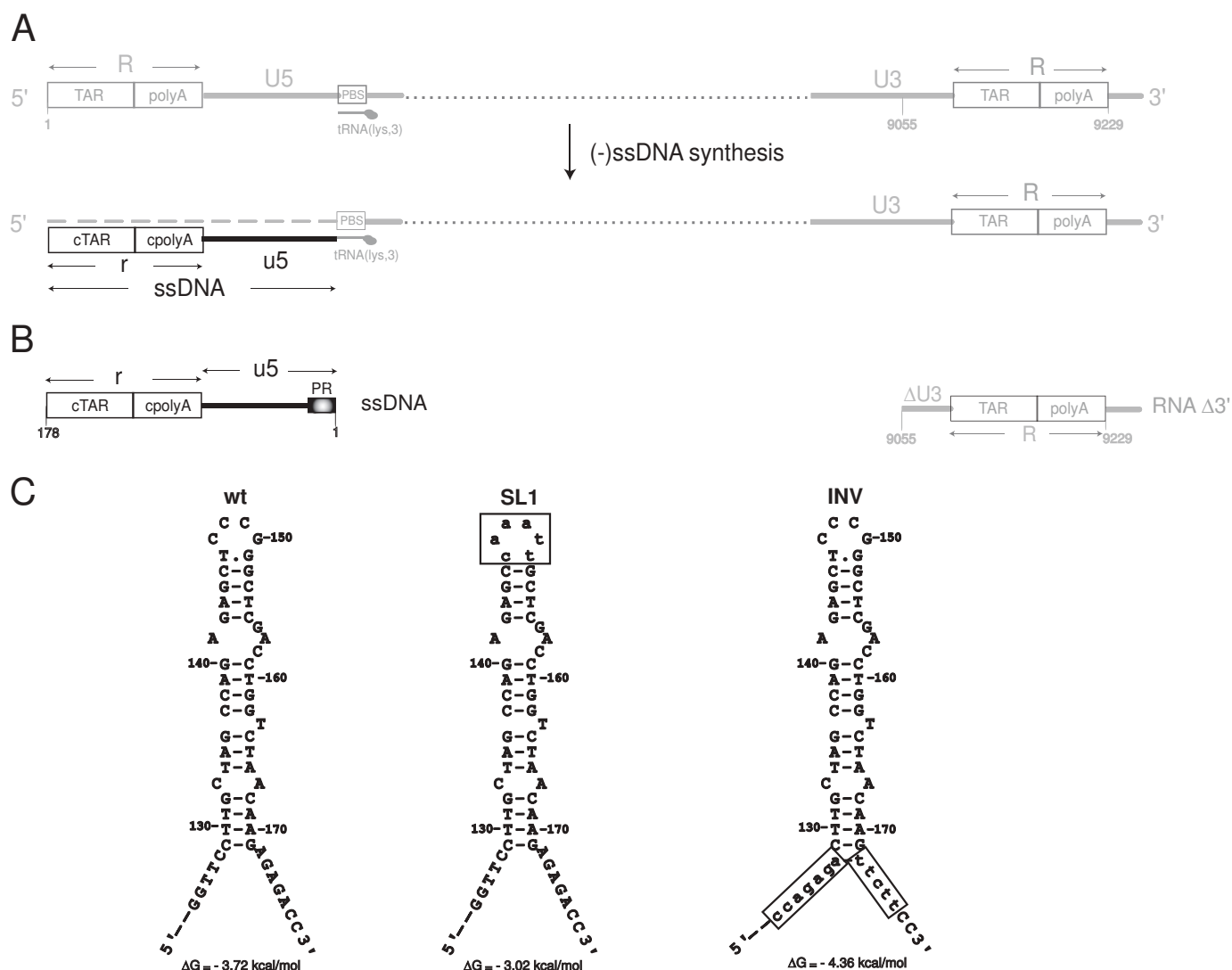


FIGURE 1. Strong-stop DNA analysis. *A*, schematic diagram of events preceding the first strand transfer. *R*, terminal repeat; *TAR*, transactivator response element; *poly(A)*, polyadenylation signal; *U5*, sequence unique at the 5'-end; *PBS*, primer binding site; *U3*, sequence unique at the 3'-end; *ssDNA*, minus-strand strong-stop DNA; *r*, complementary DNA sequence to *R*; *cTAR*, complementary DNA sequence to *TAR*; *cpoly(A)*, complementary DNA sequence to *poly(A)*; *u5*, complementary DNA sequence to *U5*. The 3' 18 nucleotides of the tRNA^{Lys} primer are annealed to the PBS, a complementary sequence in the viral genome (gray line); numbering is relative to the genomic RNA cap site (+1). Once annealed, RT catalyzes extension of the primer to form ssDNA (black line). The 5'-end of the genomic RNA is degraded (broken line) by the action of the RNase H activity of RT. The first strand transfer is mediated by base pairing of the *R* and *r* regions that are complementary. *B*, ssDNA and RNA used in this study. ssDNA (178 nucleotides) is the extension product of PR primer, numbering is relative to the first nucleotide of ssDNA, which is complementary to nucleotide 178 of the genomic RNA; RNA $\Delta 3'$ (200 nucleotides) contains the 3'-end of the *U3* sequence and the full-length *R* sequence. *C*, predicted secondary structures for the cTAR sequence in the wild-type and mutant ssDNAs. The DNA folding program of Zuker (43) was used to predict the most stable secondary structure for each cTAR sequence. The T-G base pair is supported by NMR analysis of the top half of cTAR (65). Mutations are shown as lowercase letters in boxes.

The *R* sequence contains the transactivator response element (*TAR*) and a portion of the polyadenylation signal (*poly(A)*) (15, 16). The 3'-*TAR* sequence folds into a hairpin in the complete HIV-1 RNA genome extracted from virions (17). The 3'-*poly(A)* sequence can form a hairpin (16, 18), but there is no data showing that this secondary structure is present in the complete HIV-1 RNA genome. The *r* sequence of ssDNA is predicted to fold into hairpins that are complementary to the *TAR* and *poly(A)* RNA sequences and are therefore named *cTAR* and *cpoly(A)*, respectively (19, 20). To date, the secondary structure of the full-length ssDNA has not been determined. Thus, there is no data showing that the *cTAR* and *cpoly(A)* sequences form the predicted hairpins in the context of the full-length ssDNA.

NC promotes annealing of ssDNA to the 3'-end of the genomic RNA because it possesses nucleic acid chaperone activity that facilitates the rearrangement of nucleic acids into the most thermodynamically stable structures containing the maximum number of base pairs (21, 22). This chaperone function depends on three properties (2): 1) ability to aggregate nucleic acids; 2) weak duplex destabilization activity; and 3) rapid on-off binding kinetics. The ability of NC to increase the rate of the first strand transfer is due at least in part to its stimulatory effect on the rate of annealing of the complementary hairpins (23, 24). Unfolding of the complementary hairpins is thought to be rate-limiting in the annealing process leading to the first strand transfer (23). As mentioned above, mutational analysis of the HIV-1 *R* sequence suggests that the great major-

ity of first strand transfers occur after completion of ssDNA synthesis, *i.e.* the complete cTAR sequence would be required for efficient strand transfer *in vivo* (12). Several studies performed with the cTAR sequence (55 to 59 nucleotides) support a dynamic structure of the cTAR hairpin, involving equilibrium between both the closed conformation and the partially open “Y” conformation (20, 25–28). Recently, we showed that NC slightly destabilizes the lower stem that is adjacent to the internal loop and shifts the equilibrium toward the Y conformation exhibiting at least 12 unpaired nucleotides in its lower part (29). We also showed that the apical and internal loops of cTAR are weak and strong binding sites for NC, respectively (29). To date, the interactions between NC and the full-length ssDNA have not been studied.

Because mutations in the TAR apical loop decrease the first strand transfer *in vitro*, Berkhout *et al.* (19) suggested that this process involves a “kissing” complex formed by the apical loops of TAR and cTAR hairpins. Consistent with this hypothesis, we found that efficient annealing of cTAR DNA to the 3′-end of the genomic RNA relies on sequence complementarities between TAR and cTAR apical loops (18). Studies using TAR RNA and cTAR DNA hairpins suggest that both the apical loops and the 3′/5′ termini of complementary hairpins are the initiation sites for the annealing reaction under subsaturating concentrations of NC (30, 31). In contrast, cTAR DNA-TAR RNA annealing in the presence of saturated NC depends only on nucleation through the 3′/5′ termini, resulting in the formation of a “zipper” intermediate (30, 32). Note that the two annealing pathways have not been demonstrated for the annealing reaction between the full-length ssDNA and the 3′-end of the genomic RNA.

In this study, we investigated the annealing process using ssDNA and RNA Δ3′ representing the 3′-end of the genomic RNA (Fig. 1B). Using chemical and enzymatic probes, our study is the first and the only one, to our knowledge, which determined the secondary structure of the ssDNA and the NC binding sites within this DNA.

Experimental Procedures

NC and Oligonucleotides—Full-length NC (NC(1–55)) was synthesized by the Fmoc (*N*-(9-fluorenyl)methoxycarbonyl)/pentafluorophenyl ester chemical method and purified to homogeneity by HPLC (33). NC was dissolved at concentrations of 1 to 2 mg/ml in a buffer containing 25 mM HEPES (pH 6.5), 50 mM NaCl, and 2.2 mol of ZnCl₂/mol of peptide. DNA oligonucleotides were purchased from Eurogentec. In the following oligonucleotide sequences, the uppercase letters indicate the bases that are complementary to the DNA sequence of the HIV-1 MAL isolate and the positions of the HIV-1 sequence targeted by the oligonucleotides are in parentheses: O1, gcagaattctggagtaatacactactataGGTCTCTCTTGTAGAC (1–17); O2, gcagcgtcgacGGTATCAC-TTCTGGGC (876–861); O3, gcaatacGGTCTCTCTTGTAGACAGGTCTGAGCAatttGCTCTGGCTAGC (1–47); O4, ccgccgtgactaccgcgagaccac; O5, gctcgaaCCCGGGCTC GACCTGGTCTAACAAGtcttCCtatagtgagtcgtattac (27–7); O6, catcaCCCGGGAGCTCTCTGGCTAGCAAGtctctgg-CACTGCTTAAGCCTC (32–70); O7, gaagaattctggagtaatacga-

ctcactataGGAGGCGTAACTTGGG (9055–9070); O8, atcgataccgtcgacc; O9, gcagaattctggagtaatacactactataGGAGCTTT-TGATCTCAG (8635–9229); Oligonucleotides O1, O5, O7, and O9 contain the T7 RNA polymerase promoter. PR, CTG-CTAGAGATTTTACACCG (178–158); PU3b1M, TCCGG-ATGTTGTTCTC (8990–8975).

Construction of Plasmids—Standard procedures were used for restriction enzyme digestion and plasmid construction (34). Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs. The expand high fidelity PCR system was from Roche Applied Science. Cloned sequences and mutations were verified by DNA sequencing. Plasmids pHIVCG-4 and pHIVCG8.6 contain DNA fragments of the HIV-1 genome derived from the MAL isolate (35, 36). Plasmid pYC5′ was generated by PCR amplification of linearized pHIVCG-4 with EcoRI using oligonucleotides O1 and O2. The resulting PCR product was digested with EcoRI and SalI and ligated into pHIVCG-4 digested with the same enzymes. Using oligonucleotides O3 and O4 and pYC5′ linearized with EcoRI, plasmid pBCSL1 was generated by PCR amplification. The resulting PCR product was digested with BsaI and ligated into pYC5′ digested with the same enzyme. Using oligonucleotides O5 and O6 and pYC5′ linearized with AvaI, plasmid pCCinv was generated by PCR amplification. The resulting PCR product was then digested with AvaI and ligated intramolecularly. Plasmid pFC3′-2 was generated by PCR amplification of linearized pHIVCG8.6 with EcoRI using oligonucleotides O7 and O8. The resulting PCR product was digested with EcoRI and XhoI and ligated into pCG44 (37) digested with the same enzymes. Plasmid pFC3′UTR was generated by PCR amplification of linearized pHIVCG8.6 with EcoRI using oligonucleotides O8 and O9. The resulting PCR product was digested with EcoRI and XhoI and ligated into pCG44 digested with the same enzymes.

In Vitro RNA Synthesis and Purification—Plasmids pYC5′, pBCSL1, and pCCinv were digested with HaeIII to generate templates for *in vitro* synthesis of RNAs 1–415wt, 1–415SL1, and 1–415inv, respectively. These RNA transcripts start and end with authentic HIV sequences, *i.e.* they do not contain additional sequences resulting from DNA plasmid construction. Plasmid pFC3′-2 was digested with XhoI to generate the template for *in vitro* synthesis of RNA Δ3′ (200 nucleotides) that contains at the 3′-end a poly(A) tail (22 adenine residues) and five nucleotides corresponding to the XhoI site. Plasmid pFC3′UTR was digested with XhoI to generate the template for *in vitro* synthesis of RNA 3′ UTR (620 nucleotides). Five μg of the cleaved plasmids was transcribed with bacteriophage T7 RNA polymerase under the conditions stipulated by the RiboMAX™ large scale RNA production system (Promega). RNAs were purified by electrophoresis on a denaturing 5% polyacrylamide gel as described (34).

Synthesis, Labeling, and Purification of the ssDNAs and DNA Size Markers—The DNA oligonucleotides PR and PU3b1M were used as primers. For synthesis of 5′-end labeled ssDNAs and DNA size marker (M, 356 nucleotides), the primers were 5′-end labeled using T4 polynucleotide kinase (New England Biolabs) and [γ -³²P]ATP (PerkinElmer Life Sciences). Ninety pmol of 5′-end labeled primer (3×10^5 cpm/pmol) and 60 pmol

of RNAs 1–415 (templates for the wild-type and mutant ssDNAs), or 3' UTR (template for the DNA size marker, M) in 150 μ l of water were heated at 90 °C for 3 min and frozen for 5 min on dry ice/ethanol bath. Then, 138 μ l of reaction buffer (final concentrations: 78 mM KCl, 1 mM DTT, and 52 mM Tris-HCl, pH 7.8) containing 30 units of HIV-1 reverse transcriptase (Worthington) was added and the sample was preincubated at 37 °C for 10 min before the primer extension reaction was initiated by 6 μ l of MgCl₂ and 6 μ l of dNTPs. The final reaction contained 50 mM Tris-HCl (pH 7.8), 75 mM KCl, 2 mM MgCl₂, 1 mM DTT, and 300 nM dNTPs. The reaction was incubated at 37 °C for 1 h and terminated by addition of 200 μ l of 0.3 M NaOH, 0.05 M EDTA and heating at 90 °C for 15 min. Then, the sample was extracted by phenol/chloroform followed by ethanol precipitation and the dried pellet was resuspended in 40 μ l of loading buffer A (7 M urea, 0.03% (w/v) bromophenol blue, and 0.03% (w/v) xylene cyanol). The 5'-end labeled ssDNAs and M were purified by electrophoresis on 6% denaturing polyacrylamide gels and isolated by elution followed by ethanol precipitation. The purified ssDNAs and M were dissolved in water and checked for purity and integrity on a 6% denaturing polyacrylamide gel.

Analysis of ssDNA Dimerization—The assays were carried out in a final volume of 10 μ l. The 5'-end labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) was dissolved in 7.2 μ l of water, heated at 90 °C for 2 min, and chilled for 2 min on ice. Then 0.8 μ l of the renaturation buffer was added (final concentrations: 75 mM KCl, 0.2 mM MgCl₂, and 50 mM Tris-HCl, pH 7.8) and the sample was incubated at 37 °C for 45 min. The reaction mixtures were then incubated at 37 °C for 15 min in the absence or presence of NC at various concentrations. The incubations were stopped by extraction with phenol/chloroform and each aqueous phase was mixed with 4 μ l of loading buffer B (50% (w/v) glycerol, 0.05% (w/v) bromophenol blue, 0.05% (w/v) xylene cyanol). The heat-denatured control of 5'-end labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol in 10 μ l of water) was performed by heating at 90 °C for 2 min and chilling for 2 min on ice, and mixing with 4 μ l of loading buffer B. Formation of homoduplexes was analyzed by electrophoresis on a 2% agarose (QA-Agarose™, Qbiogene) gel at 4 °C in 0.5 $\times$ TBM (45 mM Tris borate (pH 8.3), 0.1 mM MgCl₂). After electrophoresis, the gel was fixed, dried, and autoradiographed as described (37).

Gel-shift Annealing Assay—The annealing assay was carried out in a final volume of 12 μ l. The 5'-end labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) in 4 μ l of water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 1.0 μ l of renaturation buffer (final concentrations: 75 mM KCl, 0.2 mM MgCl₂, and 50 mM Tris-HCl, pH 7.8) and 1 μ l of NC (1.48 or 3.81 pmol) were added and the sample was incubated at 37 °C for 15 min. Unlabeled RNA Δ 3' (0.45 pmol) underwent the same renaturation treatment with 1 μ l of NC (5 or 12.85 pmol) and then added to the refolded ssDNA. The protein to nucleotide molar ratios were 1:18 or 1:7. These ratios refer to total nucleotide concentration. The reaction mixture was then incubated at 37 °C for 1, 3, 9, 15, 30, 60, or 120 min. At the end of incubations, the assays were phenol/chloroform extracted and each aqueous phase was mixed with 4 μ l of loading buffer B. The

samples were analyzed by electrophoresis on a 2% agarose (QA-Agarose™, Qbiogene) gel at 25 °C in 0.5 $\times$ TBE (45 mM Tris borate (pH 8.3), 1 mM EDTA). After electrophoresis, the gel was fixed, dried, and autoradiographed as described (37). The monomeric (m) and heteroduplex (hd) forms of ssDNA, and the high molecular weight DNA-RNA complexes (hmw) were quantified using a Typhoon™ TRIO (GE Healthcare) and ImageQuant software. The percent of heteroduplex was determined as $100 \times (\text{hd}/(\text{hd} + \text{m} + \text{hmw}))$.

Structural Probing of ssDNAs—Potassium permanganate and piperidine were purchased from Sigma. Mung bean nuclease and DNase I were purchased from New England Biolabs and Promega, respectively. Structural probing of ssDNA was carried out in a final volume of 10 μ l. The 5'-end labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) in 7.2 μ l of water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 0.8 μ l of renaturation buffer (final concentrations: 75 mM KCl, 0.2 mM MgCl₂, and 50 mM Tris-HCl (pH 7.8) for probing with KMnO₄ or DNase I; 75 mM KCl, 0.2 mM MgCl₂, and 50 mM sodium cacodylate (pH 6.5) for probing with mung bean nuclease) was added and the sample was incubated at 37 °C for 45 min. The reaction mixtures were then incubated at 37 °C for 15 min in the absence or presence of NC at various concentrations. The samples were then incubated with 0.125, 0.25, or 0.5 units of mung bean nuclease for 15 min at 37 °C or with 0.1 or 0.15 units of DNase I for 7 min at 37 °C. These cleavage reactions were stopped by phenol/chloroform extraction followed by ethanol precipitation. The dried pellets were resuspended in 7 μ l of loading buffer A. For potassium permanganate probing, ssDNA was treated with 0.25, 0.5, 0.75, or 1 mM KMnO₄ for 1 min at 37 °C. The treatment was stopped by adding 40 μ l of the termination buffer (0.7 M β -mercaptoethanol, 0.4 M NaOAc (pH 7.0), 10 mM EDTA, 25 μ g/ml of tRNA). DNA was then extracted with phenol/chloroform, ethanol precipitated, and dried. DNA was subjected to piperidine cleavage by resuspension of the dried pellet in 100 μ l of freshly diluted 1 M piperidine and heating at 90 °C for 30 min. The samples were lyophilized, resuspended in 20 μ l of water, and lyophilized again. After a second lyophilization from 15 μ l of water, the samples were resuspended in 7 μ l of loading buffer A. Sequence markers of the labeled ssDNA were produced by the Maxam-Gilbert method (38). To identify all cleavage sites at the nucleotide level, the samples were analyzed by long and short migration times on denaturing 6–8 and 12–14% polyacrylamide gels, respectively.

Alignment of the ssDNA Sequences—The ssDNA sequence alignments were performed using the HIV sequence database from the Los Alamos National Lab HIV. A position-weighted matrix was then computed using the open-source statistical software R from the Comprehensive R Archive Network (59). The sequence logo (39) was plotted using the seqLogo package.

Results

Synthesis, Purification, and Analysis of ssDNA—Commercial chemical synthesis of ssDNA was unattainable due to substrate length. Therefore, we synthesized the full-length ssDNA derived from the MAL isolate using HIV-1 reverse transcriptase. The 5'-end labeled ssDNA was synthesized and purified as described under “Experimental Procedures” section. Previous

Secondary Structure of the HIV-1 Strong-stop DNA

studies showed that both loose and tight duplexes can be characterized by native agarose gel electrophoresis at 4 °C in the TBM buffer (40–42). The labeled ssDNA was monomeric after incubation with or without NC and analyzed under the electrophoretic conditions described above (Fig. 2). Therefore, the annealing assays and the structural analyses were not complicated by the presence of dimeric forms of ssDNA.

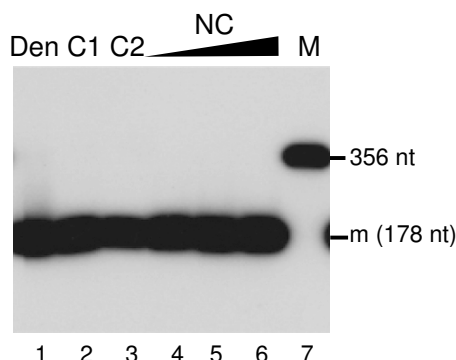


FIGURE 2. Analysis of ssDNA dimerization. The 5'-end labeled ssDNA in the absence (lanes C1 and C2) or presence of NC (lanes 4–6) was incubated at 37 °C as described under “Experimental Procedures.” The samples were analyzed by 2% agarose gel electrophoresis at 4 °C in TBM buffer and visualized by autoradiography. Den, heat-denatured ssDNA. Lane C1, control without phenol/chloroform extraction. Lane C2, control with phenol/chloroform extraction. The protein to nucleotide molar ratios were 1:18 (lane 4), 1:12 (lane 5), and 1:7 (lane 6). Lane M, size marker (356 nucleotides). Monomeric form of ssDNA is indicated by m (178 nucleotides).

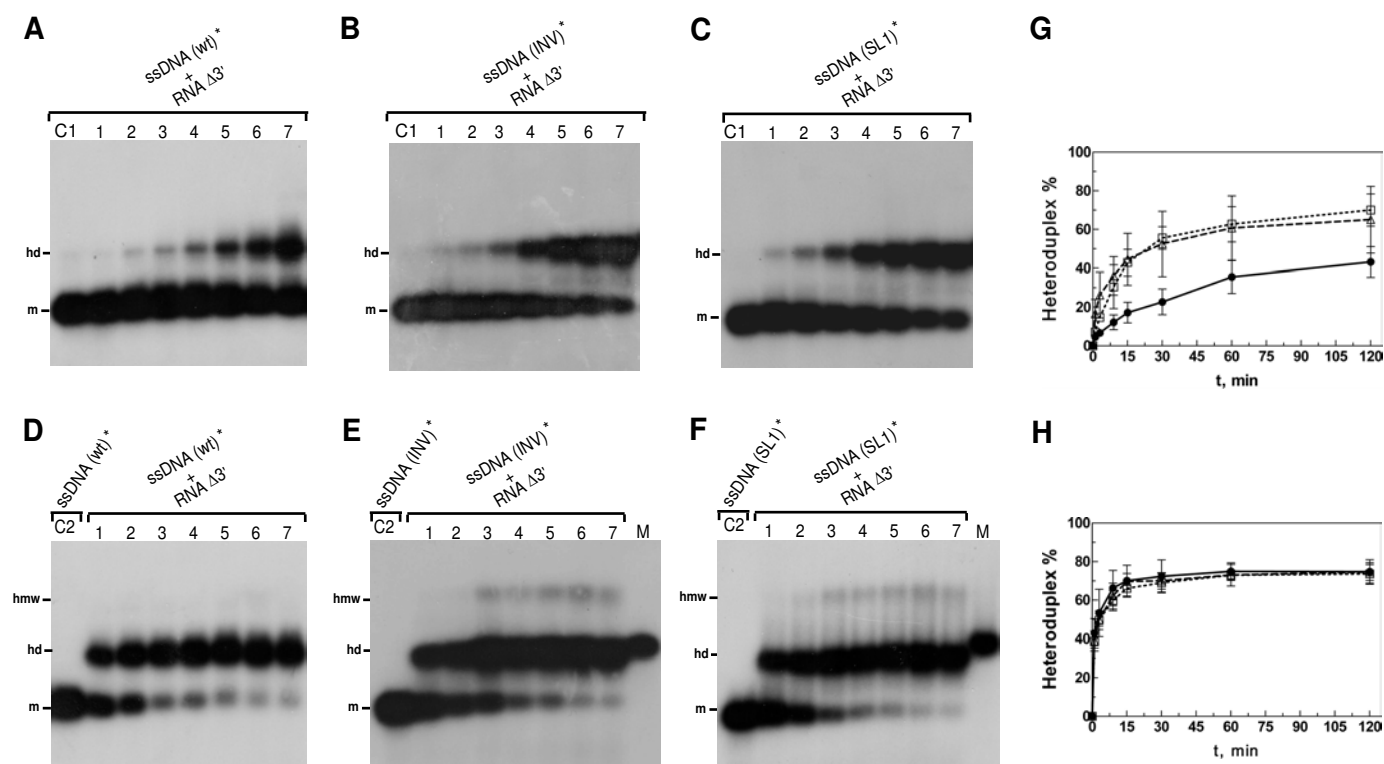


FIGURE 3. Mutational analysis of ssDNA-RNA $\Delta 3'$ annealing in the presence of NC. The NC-mediated annealing assays were performed as described under “Experimental Procedures.” The samples were analyzed by 2% agarose gel electrophoresis at 25 °C in the TBE buffer. A–C, lanes C1, the 5'-end labeled ssDNA (ssDNA*) was incubated with RNA $\Delta 3'$ for 120 min in the absence of NC. Lanes 1–7, ssDNA* and RNA $\Delta 3'$ were incubated for 1, 3, 9, 15, 30, 60, or 120 min in the presence of NC at a protein to nucleotide molar ratio of 1:18. The monomeric and heteroduplex forms of ssDNA are indicated by m and hd, respectively. D–F, lanes C2, ssDNA* was incubated for 120 min in the presence of NC at a protein to nucleotide molar ratio of 1:7. Lanes 1–7, ssDNA* and RNA $\Delta 3'$ were incubated for 1, 3, 9, 15, 30, 60, or 120 min in the presence of NC at a protein to nucleotide molar ratio of 1:7. Lane M, size marker (356 nucleotides). High molecular weight complexes are indicated by hmw. G and H, percentage of heteroduplex in the presence of NC at a protein to nucleotide molar ratio of 1:18 (G) or 1:7 (H). All experiments were repeated at least three times and error bars show standard deviations. Black circle, wild-type; open triangle, SL1; open square, INV.

Annealing of ssDNA to the 3'-R Region of the Genome—As mentioned under the Introduction, the annealing reaction between the full-length ssDNA and the 3'-end of the genomic RNA has not yet been studied. Here, we investigated the annealing process using NC, the full-length ssDNA, and RNA $\Delta 3'$ representing the 3'-end of the genomic RNA (Fig. 1B). The annealing assays were performed as described under “Experimental Procedures.” Annealing of RNA $\Delta 3'$ to the wild-type ssDNA was barely detectable after 120 min incubation in the absence of NC (Fig. 3A, lane C1). Because the genomic RNA and ssDNA are probably folded and associated with NC during the first strand transfer *in vivo*, RNA $\Delta 3'$ and ssDNA were renatured in the presence of NC before being mixed together. Annealing time courses were performed at low (Fig. 3, A–C) and high (Fig. 3, D–F) NC concentrations. Annealing of RNA $\Delta 3'$ to the wild-type ssDNA was significantly increased in the presence of NC at a protein to nucleotide molar ratio of 1:18 (Fig. 3A). The average yield of heteroduplex product was 43% after 120 min incubation (Fig. 3G, black circles). NC at a protein to nucleotide molar ratio of 1:7 facilitated greatly the annealing of RNA $\Delta 3'$ -wild-type ssDNA (Fig. 3D). The average yield of the heteroduplex product was 75% after 120 min incubation (Fig. 3H, black circles). These results show that the wild-type ssDNA synthesized and renatured *in vitro* is competent for the annealing reaction and the chaperone activity of NC is required for efficient annealing of ssDNA to the 3'-end of the genomic RNA.

To determine the role of the apical loop and the lower stem of the cTAR hairpin in the NC-mediated annealing process, two mutants were investigated (Fig. 1C). The SL1 mutant was designed such that the cTAR domain in ssDNA can form the stem-loop structure, but its apical loop cannot base pair with the apical loop of the TAR RNA hairpin. The INV mutant was designed such that the cTAR domain in ssDNA can form only the Y conformation and its lower part cannot base pair with the lower part of the TAR RNA hairpin. A band of very weak intensity migrating between the monomer and heterodimer forms was observed with the two mutants (Fig. 3, B and C). The position of this band is consistent with a heteroduplex that is shorter than the full-length heteroduplex. A likely explanation is that spontaneous cleavage occurred in a very small fraction of RNA $\Delta 3'$ molecules at the level of mismatched sites during the annealing process, mutant ssDNAs were thus annealed to a truncated form of RNA $\Delta 3'$. Surprisingly, the rate of annealing in the presence of NC at a protein to nucleotide molar ratio of 1:18 was significantly increased with the mutants, compared with the wild-type (Fig. 3, A-C and G). The average yield of heteroduplex products was at least 65% after 120 min incubation with these mutants (Fig. 3G). The mfold program predicts that the cTAR hairpin is more stable in the INV mutant than in the wild-type (Fig. 1C). However, the results suggest that mutations in the apical loop and the lower part of the cTAR hairpin facilitated the annealing process by decreasing the stability of mutant ssDNAs. In other words, the ssDNA structure of mutants was more destabilized in the presence of a low concentration of NC than that of the wild-type. Consistent with this notion, the wild-type and mutants exhibited similar kinetic profiles in the presence of a high concentration of NC (Fig. 3H). High molecular weight complexes were clearly observed with the mutants (Fig. 3, E and F), suggesting formation of DNA-RNA multimers composed of at least three molecules. Taken together, the results show that NC-mediated annealing of ssDNA to RNA $\Delta 3'$ does not rely on a single pathway (kissing or zipper). The results are also consistent with the notion that the NC-mediated annealing process could be initiated through complementary sequences that do not involve the apical loop and the 3'/5' termini of the cTAR and TAR hairpins.

Strategy for Secondary Structure Determination of ssDNA—To determine the secondary structure of the full-length wild-type ssDNA, we used structural probes and the mfold program (43). Thus, the full-length ssDNA was probed with DNase I, mung bean nuclease (MBN) and potassium permanganate (KMnO_4). DNase I is a double-strand specific endonuclease that produces single-strand nicks (44). MBN is highly selective for single-stranded nucleic acids and single-stranded regions in double-stranded nucleic acids (45). Note that nucleotide mismatches in double-stranded DNA are poor substrates for MBN cleavage at 37 °C (45, 46). Due to the bulky size of DNase I and MBN, the absence of cuts in the folded DNA structure by these enzymes may also be a result of steric hindrance. In contrast, KMnO_4 is a small probe that can be used to detect all regions in the folded DNA structure that are unpaired or distorted (47, 48): it is an oxidizing agent that preferentially attacks the 5,6-double bond of thymine. In B-DNA, this bond is shielded by base stacking interactions and, thus, the T residues in such

DNA duplexes are relatively resistant to oxidation. After treatment of ssDNA with piperidine, the DNA backbone was cleaved at the site of the modified thymines. The cleavage fragments generated by the nucleases and the KMnO_4 /piperidine treatment were analyzed by electrophoresis on denaturing polyacrylamide gels as described under "Experimental Procedures." Running Maxam-Gilbert sequence markers of ssDNA on the same gels in parallel allowed the identification of cleavage sites. Note that the nucleases cleave the phosphodiester bond and generate a 3'-hydroxyl terminus on 5'-end labeled DNA. In contrast, the KMnO_4 /piperidine treatment and Maxam-Gilbert reactions generate a 3'-phosphorylated terminus on 5'-end labeled DNA (38, 49). The electrophoretic mobility of Maxam-Gilbert sequence markers is therefore slightly greater than that of fragments produced by nucleases. This slight difference is observable only with short DNA fragments.

Analysis of the ssDNA Secondary Structure in the Absence of NC—Representative examples of probing experiments without NC are shown in Fig. 4. The results of a series of independent experiments are summarized in the secondary structure of ssDNA that is the most consistent with the probing data (Fig. 5A). In the absence of NC, the DNase I cleavage pattern (Fig. 4, A and B) is consistent with the secondary structure that we propose for ssDNA (Fig. 5A). Indeed, except moderate DNase I cleavage between G97-C98 occurring in an apical loop (Fig. 4B), all moderate and strong DNase I cleavages occurred within stems or at the ends of stems (Fig. 5A). Moderate DNase I cleavage between G97-C98 suggests the existence of transient base pairs in the apical loop. The ssDNA secondary structure is also supported by the MBN cleavage pattern (Fig. 4, C and D) because all moderate and strong MBN cleavages occurred within single-stranded regions and loops (Fig. 5A). Finally, the KMnO_4 probing results are also consistent with our secondary structure model. Thus, except the weak sensitivity of T115 to KMnO_4 , all paired thymine bases within the stems were barely (T5, T41, T45, T92, T108, T114, T130, and T131) or not reactive (T43, T80, T83, T84, T122, T154, T160, and T165) to KMnO_4 (Fig. 4, E and F). The weak sensitivity of T115 to KMnO_4 suggests that the G85-T115 mismatched base pair is not involved in strong stacking interactions with the surrounding base pairs. Consistent with this notion, Gogos *et al.* (50) showed that KMnO_4 interacts with the thymine residue of T-G mismatches. Residues T11, T32, T134, and T146, which are at the ends of stems or adjacent to a mismatch, exhibited moderate sensitivity to KMnO_4 because at least one side of the plane of the heterocyclic ring of these thymine residues is exposed. As expected, the reactivity level of all unpaired thymine bases toward KMnO_4 was high or moderate.

Analysis of the ssDNA Secondary Structure in the Presence of NC—To identify destabilized regions and protections induced by NC in ssDNA, we compared the enzymatic and KMnO_4 probing patterns of ssDNA in the absence or presence of increasing concentrations of NC (Fig. 6). The results of a series of independent experiments are summarized in the secondary structure of ssDNA that is the most consistent with the probing data (Fig. 5B). The DNase I cleavage patterns in the absence and presence of NC were similar (Fig. 6, A and B), indicating that NC did not change deeply the secondary structure of ssDNA.

Secondary Structure of the HIV-1 Strong-stop DNA

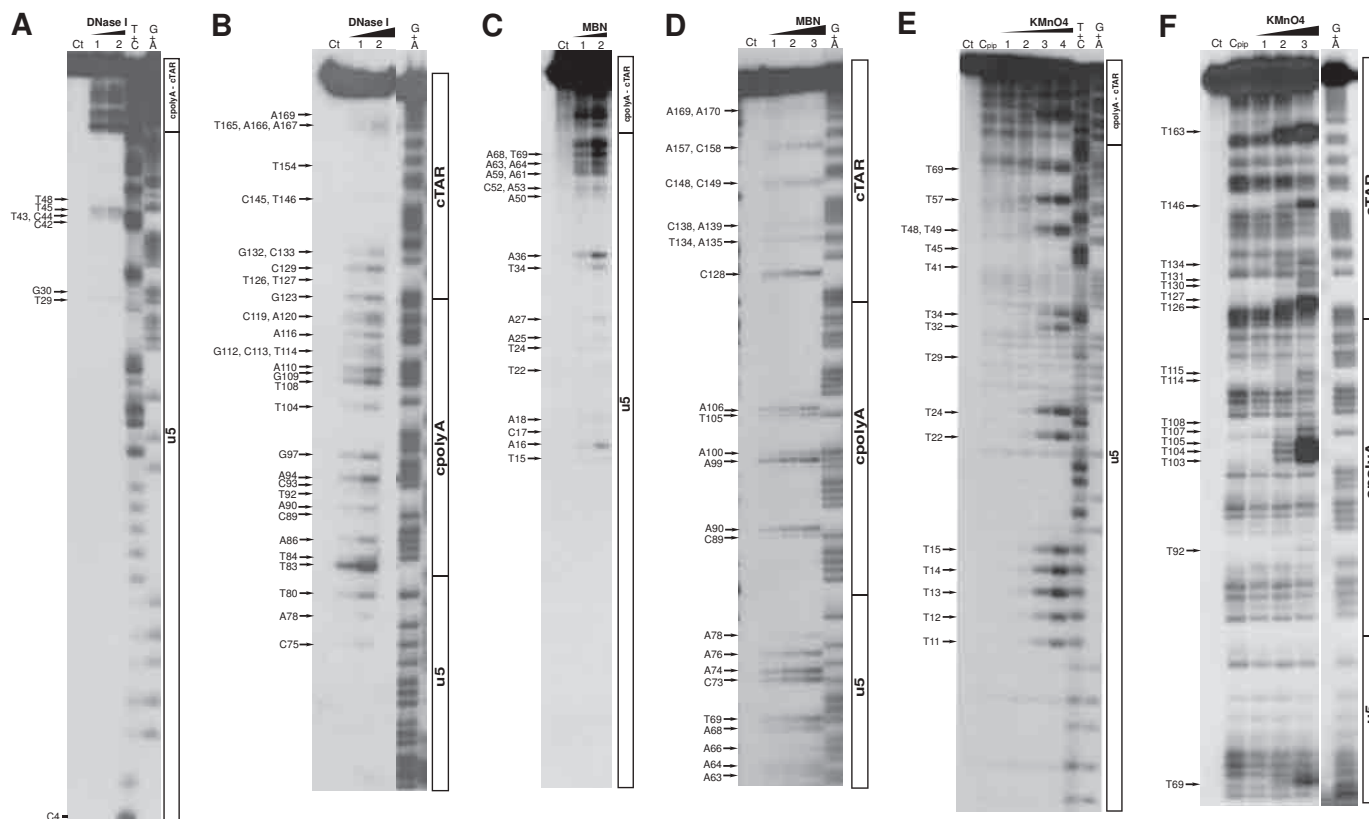


FIGURE 4. Analysis of the ssDNA secondary structure in the absence of NC. Probing experiments were performed and analyzed by electrophoresis on denaturing 14% (A, C, and E), 6% (B), and 8% (D and F) polyacrylamide gels as described under "Experimental Procedures." A and B, the 5'-end labeled ssDNA was incubated with DNase I (0.1 and 0.15 units, lanes 1 and 2). C and D, the 5'-end labeled ssDNA was incubated with mung bean nuclease (MBN) (0.125, 0.25, and 0.5 units, lanes 1–3). E and F, the 5'-end labeled ssDNA was incubated with KMnO_4 (0.25, 0.5, 0.75, and 1 mM, lanes 1–4). Lanes Ct are controls without any treatment. Lane C_{pip} is the control without KMnO_4 treatment but with piperidine treatment. T + C and G + A refer to Maxam-Gilbert sequence markers. Arrows indicate the cleavage sites and the reactive thymine residues. Schematic of the ssDNA is shown alongside the gel, indicating the u5, cpoly(A), and cTAR domains.

However, DNase I cleavage at the level of almost all sites increased in the presence of NC at a protein to nucleotide molar ratio of 1:18. A likely explanation is that binding of NC to single-stranded regions changed the three-dimensional folding of ssDNA and therefore the accessibility of double-stranded regions. In the presence of NC at a protein to nucleotide molar ratio of 1:3.5, DNase I cleavage at the level of all sites decreased strongly (Fig. 6, A and B, lanes 5). These results are consistent with the notion that NC at high concentrations binds the double-stranded regions of nucleic acids nonspecifically through electrostatic interactions of the basic residues with the phosphodiester backbone (18, 36, 51). Interestingly, NC at a protein to nucleotide molar ratio of 1:18 induced specific protection against DNase I at the level of C129 (Fig. 6B, lane 2). This suggests that the nucleotides encompassing residue C129 constitute a strong binding site for NC or undergo a local rearrangement upon NC binding.

MBN cleavage at the level of 14 unpaired nucleotides (T15, A16, C17, T22, T24, A25, A27, A50, C52, A53, A59, A61, A63, and A64) did not change in the presence of NC at a protein to nucleotide molar ratio of 1:18 (Fig. 6, C and D, lanes 2), indicating that these unpaired nucleotides did not constitute strong binding sites for NC. In contrast, MBN cleavage at the level of 17 unpaired nucleotides (T34, A36, A66, A68, T69, C73, A74, A76, C89, A90, A99, A100, T105, A106, C128, A157, and C158)

decreased in the presence of NC at a protein to nucleotide molar ratio of 1:18, showing that these unpaired nucleotides were less accessible to MBN. The strongest protections occurred at the level of T34, A68, T69, C73, A74, A76, A99, and C128, suggesting that these unpaired nucleotides constitute strong binding sites for NC. In the presence of NC at a protein to nucleotide molar ratio of 1:3.5, MBN cleavage at the level of all sites decreased strongly (Fig. 6, C and D, lanes 5), suggesting that the single-stranded regions were almost completely covered by NC. MBN cleavage at the level of G38, A94, A95, A141, C148, and C149 increased in the presence of NC at a protein to nucleotide molar ratio of 1:18. As mentioned above, it is likely that binding of NC to single-stranded regions changes the three-dimensional folding of ssDNA and therefore the accessibility of these nucleotides. MBN cleavage at the level of A40 was slightly increased with NC at a protein to nucleotide molar ratios of 1:18, suggesting that NC destabilized the T11-A40 base pair that is adjacent to the internal loop. Surprisingly, NC induced moderate MBN cleavages at the level of residues A116, C138, and A139 (Fig. 6D, lane 2) that are within stems (Fig. 5B). The nucleotides encompassing these MBN cleavage sites may adopt alternative conformations. However, DNase I and KMnO_4 probing data do not support this hypothesis. A possible explanation is that binding of NC to the G85-T115 mismatched base pair and the internal loop (sequence between

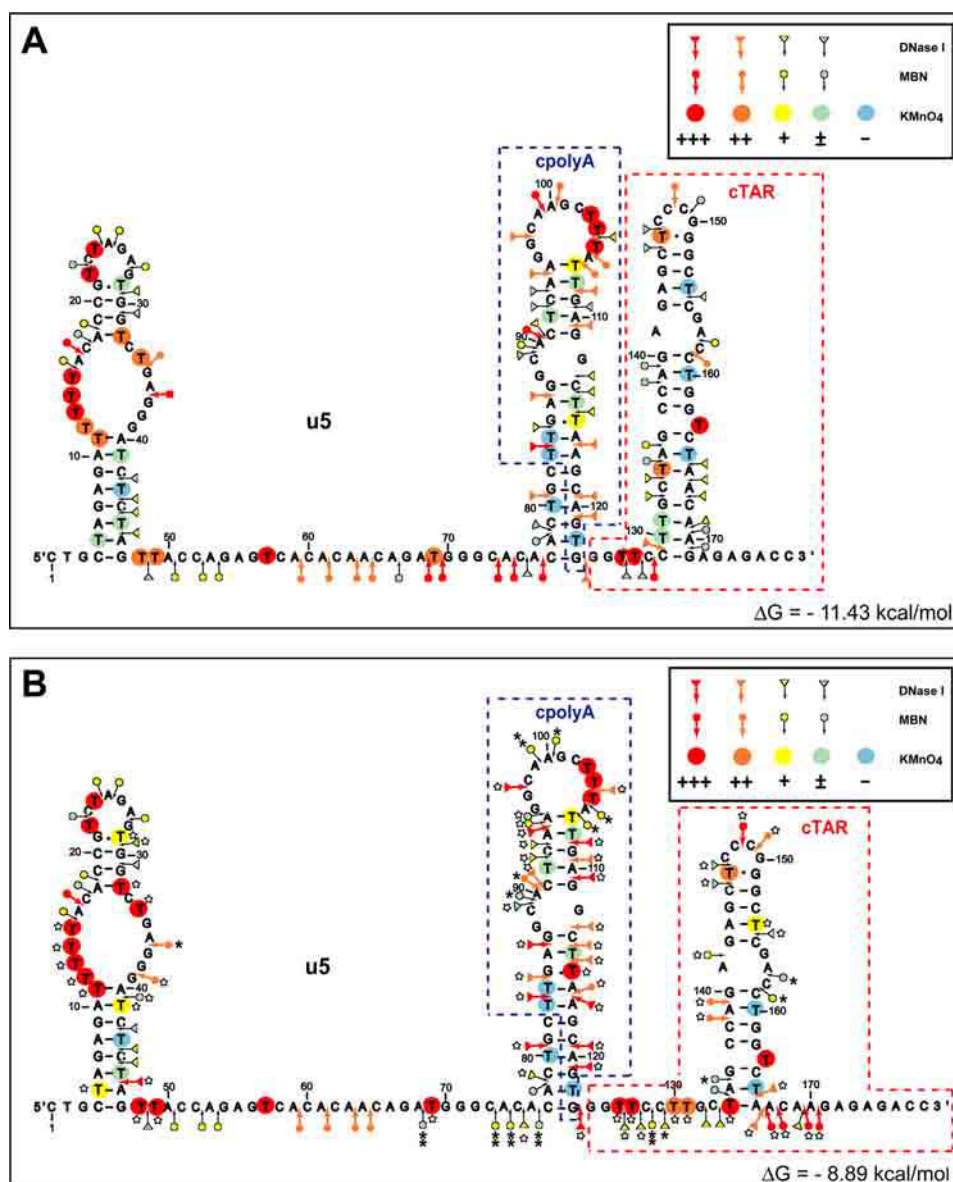


FIGURE 5. Secondary structure for ssDNA in the absence (A) and presence (B) of NC. The color code used for the reactivities is indicated in the *insets*. ΔG values were predicted by mfold (43). The *asterisks* indicate protections induced by NC at the protein to nucleotide molar ratio of 1:18. The *double asterisks* indicate strong protections induced by NC at the protein to nucleotide molar ratio of 1:18. The *stars* indicate the cleavage sites and the thymine residues where the reactivity is increased by NC at the protein to nucleotide molar ratio of 1:18.

156 and 158) increased the “breathing” process of adjacent base pairs. Note that MBN can cleave double-stranded DNA within AT-rich regions that exhibit “structural breathing” (45, 52). Interestingly, NC at a protein to nucleotide molar ratio of 1:18 induced a strong increase in MBN sensitivity for residues A166, A167, A169, and A170, indicating that these nucleotides were unpaired (Fig. 6D, lane 2).

As previously reported (29), KMnO₄ is expected to interact with the nucleobases of unpaired thymine residues that interact with the N-terminal zinc finger of NC. Indeed, a close examination of the structures of the NC:DNA complexes (53, 54) shows that the 5,6-double bond of the nucleobase thymine in these complexes is accessible to KMnO₄. Therefore, in the presence of NC at a protein to nucleotide molar ratio of 1:18, the sensitivity of thymine residues to KMnO₄ did not decrease (Fig. 6, E and F, lanes 2). Unpaired residues T12, T13, T14, T15, T34,

T48, T49, T69, T126, and T127 became more reactive to KMnO₄ in the presence of NC (Fig. 5B), suggesting that these thymine residues were involved in transient and weak base stacking interactions that were destabilized by NC. Consistent with the formation of stable stems, T43, T45, T80, T83, T84, T92, T107, T108, T114, T122, T160, and T165 did not become more reactive to KMnO₄ in the presence of NC (Fig. 5B). NC induced a slight increase in KMnO₄ sensitivity for residues T5, T41, and T154, indicating that at least one side of the plane of the heterocyclic ring of each thymine residue was exposed. This suggests that NC increased the fraying of the adjacent base pairs that are located at the ends of stems (Fig. 5B). Consistent with this notion, T11, T29, and T32 became more reactive to KMnO₄ in the presence of NC. Residue T115 became markedly reactive to KMnO₄ in the presence of NC (Fig. 6F), indicating that NC increased the accessibility of the G85-T115 mis-

Secondary Structure of the HIV-1 Strong-stop DNA

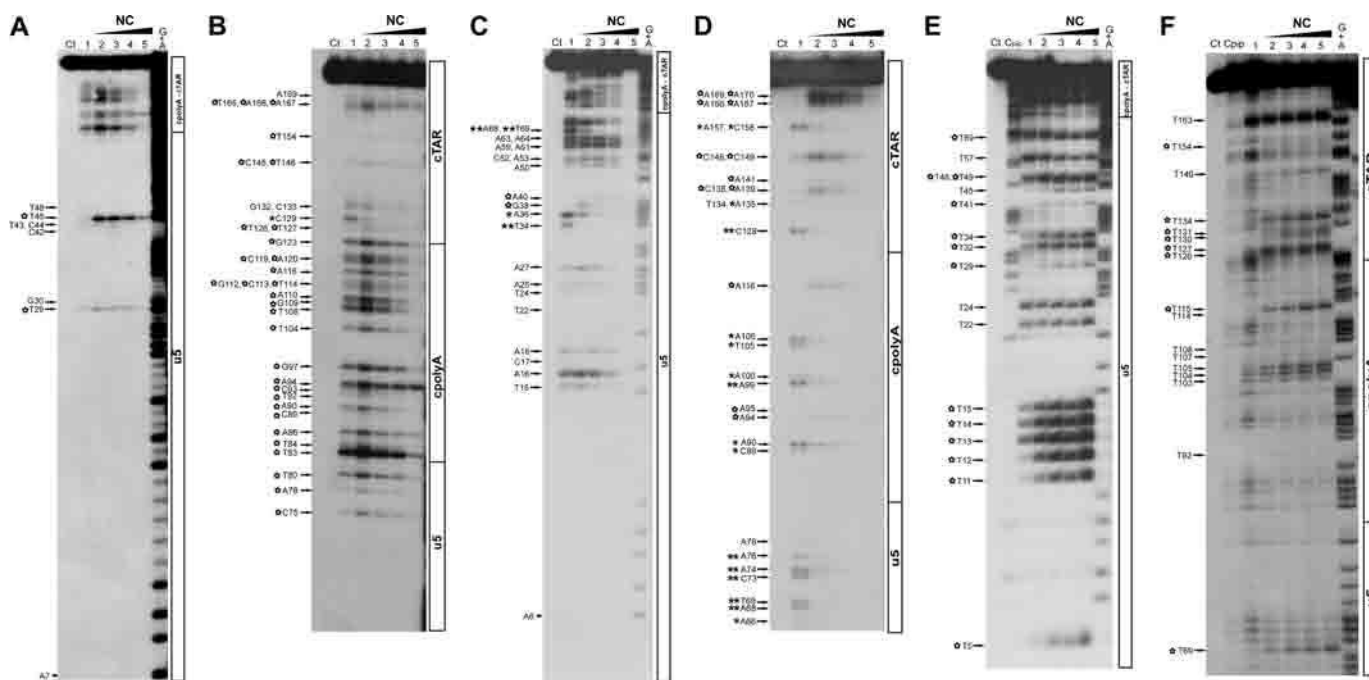


FIGURE 6. Analysis of the ssDNA secondary structure in the presence of NC. Probing experiments were performed and analyzed by electrophoresis on denaturing 14% (A, C, and E), 6% (B), and 8% (D and F) polyacrylamide gels as described under "Experimental Procedures." A and B, the 5'-end labeled ssDNA was incubated with 0.1 unit of DNase I in the absence (lanes 1) or presence of NC (lanes 2–5). C and D, the 5'-end labeled ssDNA was incubated with 0.25 unit of mung bean nuclease (MBN) in the absence (lanes 1) or presence of NC (lanes 2–5). MBN cleavage sites between positions 166 and 170 were well characterized by long migration times on a denaturing 6% polyacrylamide gel. E and F, the 5'-end labeled ssDNA was incubated with 0.5 mM KMnO_4 in the absence (lanes 1) or presence of NC (lanes 2–5). Lanes Ct are controls without any treatment. Lanes Cpip are controls without KMnO_4 treatment but with piperidine treatment. The protein to nucleotide molar ratios were 1:18 (lanes 2), 1:12 (lanes 3), 1:7 (lanes 4), and 1:3.5 (lane 5). The asterisks indicate protections induced by NC at the protein to nucleotide molar ratio of 1:18. The stars indicate the cleavage sites and the thymine residues where the reactivity is increased by NC at the protein to nucleotide molar ratio of 1:18. Schematic of the ssDNA is shown alongside the gel, indicating the u5, cpol(A), and cTAR domains.

matched base pair. Interestingly, NC induced a strong increase in KMnO_4 sensitivity for residues T130, T131, and T134 that are within the lower stem of the cTAR stem-loop (Figs. 5A and 6F). These results associated with the MBN probing data strongly suggest that the lower stem of the cTAR hairpin is open in the presence of NC (Fig. 5B).

Conservation of the ssDNA Sequence and Structure—Our study was performed with the ssDNA sequence of the MAL isolate. The HIV-1 NL4-3 isolate has been extensively used in studies dealing with the reverse transcription process (see Ref. 2 and references therein). Interestingly, ssDNA of the HIV-1 NL4-3 isolate can adopt a secondary structure that is close to that of the HIV-1 MAL isolate (Fig. 7). To provide phylogenetic support for the secondary structure of ssDNA in the presence of NC, we searched for its conservation among different HIV-1 groups. First, we determined a consensus sequence for each HIV-1 group, using the LANL HIV subtype reference database. The sequences retrieved from the QuickAlign procedure were used to calculate a position weight matrix from which a sequence logo was derived for each HIV-1 group (supplemental Fig. S1). A sequence logo is drawn such that the overall height of the stack indicates the information content at each position (derived from the Shannon entropy), whereas the height of the base symbols within the stack indicates their relative frequency (39, 55). The information content, which can reach a maximal theoretical value of 2 bits for nucleic acids, is usually interpreted as an indicator of evolutionary conservation. The ssDNA

sequence is relatively well conserved among different isolates and groups of HIV-1 (supplemental Fig. S1). Nucleotides 15–30 (numbering is relative to HIV-1 group M) located in the u5 domain correspond to the less conserved part of ssDNA. However, this domain displays highly conserved sequences between positions 1–14, 33–48, and 73–80. There are also highly conserved sequences between positions 97–112 and 115–122 in the cpol(A) domain. In the cTAR domain, the most highly conserved sequences correspond to nucleotides 144–151, 153–157, 163–166, and 177–181. Conservation of residues G36, G71, and G102 supports the hypothesis that these nucleotides are within the high-affinity binding sites for NC. The sequence logos were used to predict the ssDNA secondary structures of HIV-1 groups M, N, and O (Fig. 7). Only one full-length sequence was available to predict the ssDNA secondary structure of HIV-1 group P. Interestingly, the ssDNA sequences of four groups can fold into secondary structures that are close to that of the HIV-1 MAL isolate. In all secondary structures, there are three putative strong NC binding sites that are located in single-stranded regions containing at least a guanine residue. In HIV-1 groups M, N, and O, the putative site involving residue G36 is not in a loop. In HIV-1 group M, the base pairing interaction between nucleotides C18–A19 and T35–G36 is probably not stable because it is between an internal loop and a A–A mismatch. Therefore, the TG motif could be recognized by NC. In HIV-1 groups N and O, the strong NC binding site could be in the u5 apical loop containing a TG motif.

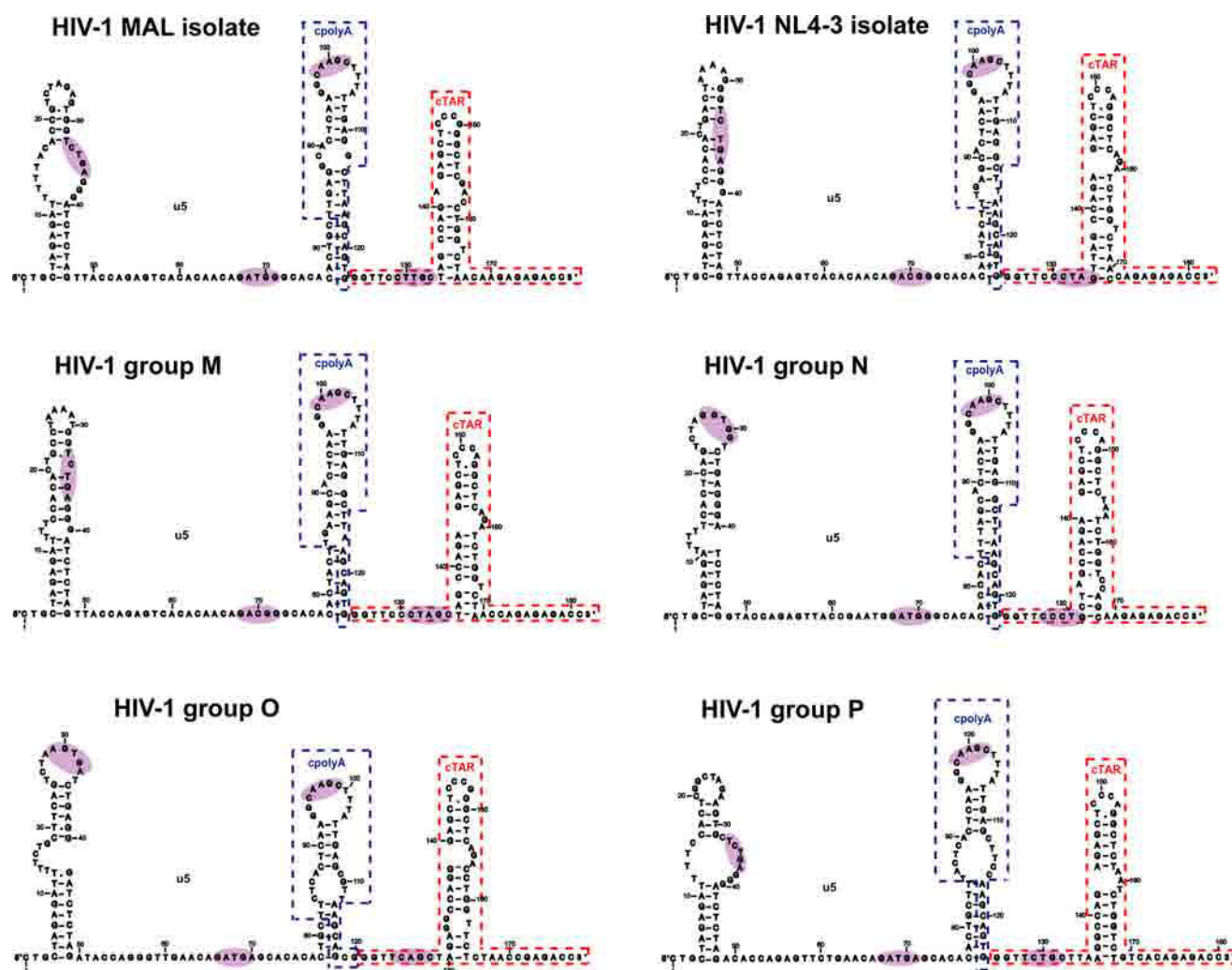


FIGURE 7. **Proposed secondary structures for the ssDNAs of different isolates and groups of HIV-1.** The mfold program (43) was used to predict the secondary structures. The ssDNA secondary structures of HIV-1 groups M, N, and O were generated from the sequence logos presented in supplemental Fig. S1. For HIV-1 group P, only one full-length ssDNA sequence is available (accession number GU111555). The putative high-affinity binding sites for NC are highlighted in purple.

Discussion

The first strand transfer is essential for HIV-1 replication (1). This transfer requires a base pairing interaction between the R sequence at the 3'-end of the genomic RNA and the r sequence at the 3'-end of the ssDNA (Fig. 1A). Therefore, the HIV-1 full-length ssDNA (176–181 nucleotides in length) is a key actor in the annealing process leading to the first strand transfer. The mechanism of annealing has been investigated using model nucleic acid substrates containing at least portions of the complementary R/r sequences. The first model systems were based on annealing of an RNA molecule (81 or 148 nucleotides representing a portion of the 3'-end of the genomic RNA) to a DNA molecule (81 or 131 nucleotides) corresponding to a portion of the 3'-end of the ssDNA (23, 56). Short DNAs representing the cTAR sequence (55 or 59 nucleotides) or the mini-cTAR sequence (26 or 27 nucleotides) were used in many studies investigating the annealing process (18, 20, 28–32, 57). Only one study examined the secondary structure of ssDNA using truncated ssDNAs (50, 63, and 128 nucleotides) (58). Thus, all *in vitro* studies deal-

ing with the ssDNA structure and the annealing process were performed with single-stranded DNAs representing truncated forms of ssDNA. Therefore, there is no data showing that the cTAR and cpoly(A) sequences form the predicted long stem-loops (19) in the context of the full-length ssDNA. Moreover, the mfold program (43) predicts that some of the most stable secondary structures of ssDNA (MAL and NL4-3 isolates) do not form the TAR stem-loop structure (data not shown). Our study is the first to investigate the secondary structure of the full-length ssDNA and its interaction with the 3'-end of the genomic RNA.

We used structural probes (DNase I, mung bean nuclease, and potassium permanganate) to directly determine the secondary structure of ssDNA in the absence or presence of NC. Our probing data support the folding of ssDNA into three stem-loop structures in the absence of NC (Fig. 5A). Note that a large part of the u5 sequence (29 consecutive nucleotides) is single-stranded. It is important to emphasize that the r region (sequences cpoly(A) and cTAR) does not fold into an independent domain, because six base pairs are formed between the u5

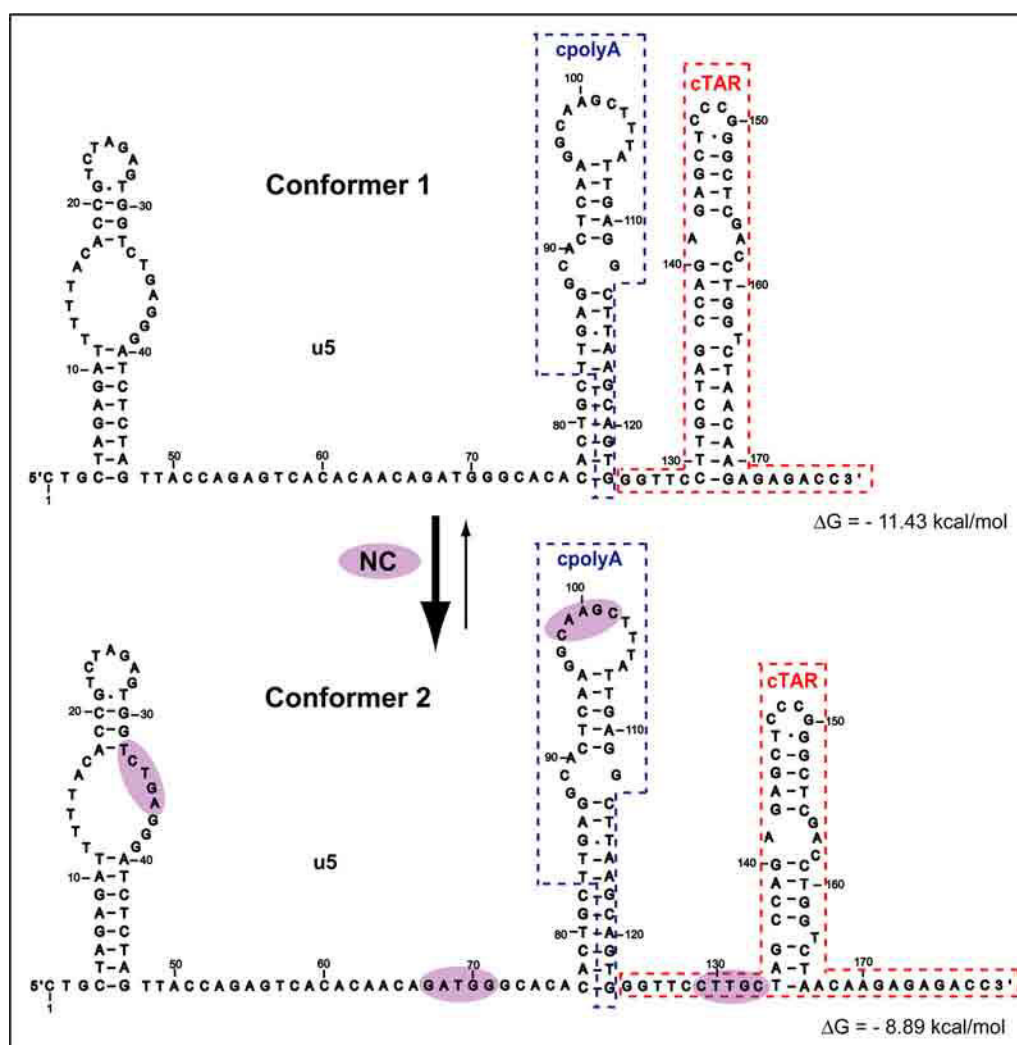


FIGURE 8. **Proposed model for conformational switch of ssDNA induced by NC binding.** ΔG values were predicted by the mfold program (43). The high-affinity binding sites for NC are highlighted in purple.

and cpoly(A) sequences (Fig. 8). Our results show that the cpoly(A) and the cTAR hairpins are formed in the full-length ssDNA. Previous studies showed that short DNAs corresponding to the cTAR sequence (55 or 59 nucleotides) adopt two alternative hairpin conformations in the absence of NC: the closed form and the partially open Y form (20, 27, 29). Interestingly, our probing data indicate that the cTAR sequence in ssDNA adopts the partially open Y form but not the closed form (conformer 1 in Fig. 8). Consistent with this finding, the mfold program predicts that the ssDNA containing the Y form is more stable than the ssDNA containing the closed form ($\Delta G_Y = -11.43$ versus $\Delta G_C = -7.54$ kcal/mol). Furthermore, the three-dimensional folding of ssDNA could prevent the formation of the closed form. Taken together the data indicate that the cTAR DNA model adopts a secondary structure (closed form) that is not formed in ssDNA.

Interestingly, a low concentration of NC (protein to nucleotide molar ratio of 1:18) increases the accessibility of the cpoly(A) and cTAR domains to DNase I (Fig. 5B), suggesting that the three-dimensional folding of ssDNA is destabilized upon binding of NC to its strong sites. Destabilization of the three-dimensional structure of ssDNA could be a first step in

the NC-mediated annealing process. In other words, NC could facilitate interaction of the R sequence of ssDNA with the R sequence of the genomic RNA by increasing the accessibility of these complementary sequences. Our probing data show that NC does not change deeply the secondary structure of ssDNA. Thus, the global architecture of the u5 and cpoly(A) stem-loops is not altered in the presence of NC. The KMnO_4 and MBN reactivities indicate that NC increases the fraying of base pairs that are located at the ends of stems in the u5 stem-loop. There is no evidence that a population of ssDNA molecules is totally melted in the presence of NC at a high concentration (protein to nucleotide molar ratio of 1:3.5). Indeed, T43, T80, T83, T84, T122, T160, and T165 were unreactive to KMnO_4 (Fig. 6, E and F). However, the lower stem of the cpoly(A) hairpin is probably destabilized during the NC-mediated annealing process, *i.e.* in the presence of the poly(A) RNA hairpin. Furthermore, the long stem (11 base pairs) of the cpoly(A) hairpin should be destabilized to allow synthesis of full-length plus-strand DNA. The G85-T115 mismatched base pair is probably involved in DNA helix destabilization by NC, because the protein increases the reactivity of residue T115 to KMnO_4 (Fig. 5B). Godet *et al.* (59)

showed that mismatched base pairs play an important role in DNA helix destabilization by NC.

Interestingly, the cTAR region undergoes a significant conformational change upon NC binding. Indeed, the lower part of the cTAR hairpin is more open in the presence of NC than in its absence (Fig. 8). This open form is supported by strong MBN cleavage sites between positions 166 and 171 and the reactivities of T126, T127, T130, T131, and T134 toward KMnO_4 (Fig. 5B). These latter MBN data indicate that the 3'-end of cTAR is accessible and could therefore interact with TAR. Consistent with the loss of four base pairs in the most open form, C129 is a weak DNase I cleavage site in the presence of NC (Fig. 5B), whereas it is a moderate DNase I cleavage site in its absence (Fig. 5A). Our results are in agreement with the hypothesis that NC induces the O2 conformation of the cTAR hairpin exhibiting 22 unpaired nucleotides in its lower part (20). This conformation is consistent with the hypothesis that the 3'/5' termini of the cTAR and TAR hairpins form a zipper intermediate that initiates the annealing reaction (30, 32). However, we showed that mutations preventing the formation of the zipper intermediate do not affect the annealing of ssDNA to the 3'-end of the genomic RNA (Fig. 3). Mutations within the first 10 nucleotides of the 5' R sequence, *i.e.* the 10 last nucleotides of the cTAR sequence, produce virions that are markedly defective for reverse transcription (12). Our results suggest that this reverse transcription defect is not due to inhibition of the annealing reaction. A kissing complex formed by the apical loops of TAR and cTAR could initiate the annealing reaction (18, 19, 29, 30). The accessibility of the cTAR apical loop to MBN increases in the presence of NC (Fig. 5B), indicating that this loop should be more accessible to the TAR apical loop. However, we show that mutations preventing the loop-loop interaction do not affect the annealing process (Fig. 3). Finally, our study suggests that the NC-mediated annealing process does not rely on a single pathway (zipper intermediate or kissing complex). It is likely that the virus uses the two pathways and perhaps a third involving the poly(A) and cpoly(A) hairpins. Indeed, our results are consistent with the "acceptor invasion" model proposed by Bambara and co-workers (60, 61). In this model, RT pausing at the base of the 5' TAR hairpin initiates RNases H cleavages in the 5' poly(A) hairpin creating a gap for the invasion of the 3' poly(A) hairpin (acceptor RNA) and interaction with the cpoly(A) hairpin. This initiates the strand exchange by a branch migration process until the 3' terminus of ssDNA is fully transferred.

Identification of NC binding sites in the full-length ssDNA contributes to an understanding of the annealing mechanism. To identify the ssDNA nucleotides coated by NC, we investigated the NC-induced protection of ssDNA to enzymatic digestion. To this end, we performed a footprinting analysis, using MBN and DNase I (Fig. 6, A–D). The DNase I and MBN probing data indicate that ssDNA is almost completely covered by NC at a high concentration (protein to nucleotide molar ratio of 1:3.5). NC at a low concentration (protein to nucleotide molar ratio of 1:18) induced the strongest protections against MBN at the level of five sites (T34, A68–T69, 73C–A–C–A76, A99, and C128). The internal loop (nucleotides 156–158) of the cTAR domain was also protected by NC, but the protection was

lower than the five protections mentioned above. In a previous report, we also showed that this internal loop is a binding site for NC in a short DNA (55 nucleotides) that folds into the cTAR hairpin (29). Protection at the level of the 73C–A–C–A76 sequence is probably not due to a direct interaction with NC, because this sequence does not display the features of a NC binding site. In contrast, four of the five protected sites are within single-stranded regions containing at least a guanine residue or the TG motif that are preferential binding sites for NC (62–65). Note that the binding site size of NC is 5 to 8 nucleotides (24). Taken together, the data support the model presented in Fig. 8, in which NC interacts strongly with four sites at a protein to nucleotide molar ratio of 1:18. Our study is the first to identify the high-affinity binding sites for NC in the full-length ssDNA. Interestingly, a strong binding site for NC containing residue G132 is located between the cpoly(A) and cTAR hairpins. A recent study performed with a modified version of the cTAR sequence (57 nucleotides) also suggests that residue G132 is preferentially recognized by NC (59). In contrast to this study, we did not find that residue G171 is a strong binding site for NC. Therefore, our results show that the properties of the cTAR DNA model are not identical to those of the cTAR domain in the full-length ssDNA. Our probing data are consistent with the notion that the cTAR region is more open in the presence of NC than in its absence (Fig. 5, compare A and B). We propose that the great majority of ssDNA molecules in the absence of NC fold into conformer 1 and the minority into conformer 2 (Fig. 8). An attractive hypothesis is that NC binding to the 129C–T–T–G–C133 sequence in conformer 2 shifts the equilibrium toward this conformer (Fig. 8).

Phylogenetic analysis supports the notion that ssDNAs of different HIV-1 groups can adopt similar secondary structures containing three stem-loops and three single-stranded regions (Fig. 7). Because the base pairing interaction between nucleotides 4–11 and 41–58 (numbering is relative to HIV-1 group M) is completely conserved (supplemental Fig. S1), it may play a key role during synthesis of minus- and plus-strand DNAs. Note that the counterpart RNA of this base pairing interaction, the primer binding site 3 helix, is also conserved and may be required for efficient initiation of reverse transcription from the tRNA₃^{Lys} primer (66). Consistent with the hypothesis that the first strand transfer is facilitated by two loop-loop interactions (19), the upper parts of the cpoly(A) (nucleotides 92–112) and cTAR (139–165) hairpins are highly conserved (Fig. 7). The TAR and poly(A) RNA hairpins play important roles in HIV-1 transcription and polyadenylation at the 3'-end of HIV-1 mRNAs, respectively (15, 16). Therefore, we cannot exclude the possibility that conservation of the cTAR and cpoly(A) DNA hairpins represent conservation at the level of RNA structure, or both. Indeed, Berkhout (67) showed that the upper part of the TAR RNA hairpin is well conserved and its group also showed conservation of the poly(A) RNA hairpin (16). Consistent with these findings, there is a strong evolutionary pressure to restore proper folding of the TAR and poly(A) hairpins upon mutational disruption (68, 69). An attractive hypothesis, therefore, is that the TAR and poly(A) hairpins have been conserved at both levels, DNA and RNA, for opti-

mal HIV-1 replication. NC is a weak duplex destabilizer that cannot destabilize long fully base paired nucleic acid helices, it can destabilize short base paired regions (4–8 base pairs) flanked by duplex ends, loops, bulges, or mismatches (24, 70). In agreement with the weak duplex destabilizing activity of NC, the stems of ssDNAs do not exceed eight consecutive Watson-Crick base pairs (Fig. 7). This suggests a coevolutionary relationship between the ssDNA sequence and the NC activity to promote the first strand transfer and synthesis of full-length plus-strand DNA. Interestingly, previous studies support the notion of a coevolutionary relationship between the cTAR structure and the NC activity (59, 71). Finally, our phylogenetic analysis suggests conservation of the four high-affinity binding sites for NC that we identified in ssDNA of the MAL isolate.

Our study is the first to determine the secondary structure of the full-length HIV-1 ssDNA in the absence or presence of NC. To our knowledge, the secondary structures of ssDNAs of other retroviruses have not been determined. Our results support a model for the NC-mediated annealing process in which the first step is the preferential binding of NC to four sites. These binding events trigger unfolding of the three-dimensional structure of ssDNA, thus facilitating interaction of the r sequence of ssDNA with the R sequence of the genomic RNA. The second step is simultaneous molecular crowding of nucleic acids and weak destabilization of secondary structures leading to the DNA-RNA heteroduplex. Finally, our results suggest that initiation of the annealing process does not rely on a single pathway (zipper intermediate or kissing complex).

Author Contributions—P. F. conceived and designed the experiments. Y. C. and O. M. performed all experiments presented in the paper. F. C. and J. B. were involved in plasmids and RNAs preparation. P. F., Y. C., and O. M. analyzed the results of experiments. P. F. and J. G. performed phylogenetic analysis. P. F., O. M., B. R., and Y. M. wrote the paper.

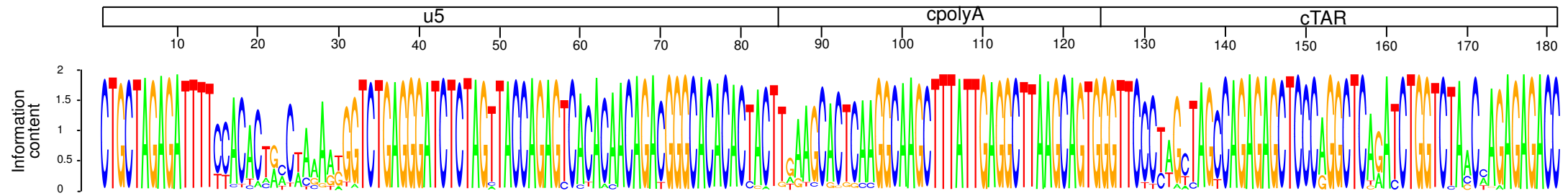
References

- Basu, V. P., Song, M., Gao, L., Rigby, S. T., Hanson, M. N., and Bambara, R. A. (2008) Strand transfer events during HIV-1 reverse transcription. *Virus Res.* **134**, 19–38
- Levin, J. G., Mitra, M., Mascarenhas, A., and Musier-Forsyth, K. (2010) Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* **7**, 754–774
- Darlix, J. L., Vincent, A., Gabus, C., de Rocquigny H., and Roques, B. (1993) Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C. R. Acad. Sci. III* **316**, 763–771
- Peliska, J. A., Balasubramanian, S., Giedroc, D. P., and Benkovic, S. J. (1994) Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* **33**, 13817–13823
- Guo, J., Henderson, L. E., Bess, J., Kane, B., and Levin, J. G. (1997) Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J. Virol.* **71**, 5178–5188
- Chen, Y., Balakrishnan, M., Roques, B. P., and Bambara, R. A. (2003) Steps of the acceptor invasion mechanism for HIV-1 minus-strand strong stop transfer. *J. Biol. Chem.* **278**, 38368–38375
- Dang, Q., and Hu, W. S. (2001) Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. *J. Virol.* **75**, 809–820
- Berkhout, B., van Wamel, J., and Klaver, B. (1995) Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. *J. Mol. Biol.* **252**, 59–69
- Lobel, L. I., and Goff, S. P. (1985) Reverse transcription of retroviral genomes: mutations in the terminal repeat sequences. *J. Virol.* **53**, 447–455
- Ramsey, C. A., and Panganiban, A. T. (1993) Replication of the retroviral terminal repeat sequence during *in vivo* reverse transcription. *J. Virol.* **67**, 4114–4121
- Klaver, B., and Berkhout, B. (1994) Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res.* **22**, 137–144
- Ohi, Y., and Clever, J. L. (2000) Sequences in the 5' and 3' R elements of human immunodeficiency virus type 1 critical for efficient reverse transcription. *J. Virol.* **74**, 8324–8334
- Guo, J., Wu, T., Bess, J., Henderson, L. E., and Levin, J. G. (1998) Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer *in vitro* and endogenous reverse transcriptase assays. *J. Virol.* **72**, 6716–6724
- Jeeninga, R. E., Huthoff, H. T., Gulyaev, A. P., and Berkhout, B. (1998) The mechanism of actinomycin D-mediated inhibition of HIV-1 reverse transcription. *Nucleic Acids Res.* **26**, 5472–5479
- Feng, S., and Holland, E. C. (1988) HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* **334**, 165–167
- Berkhout, B., Klaver, B., and Das, A. T. (1995) A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses. *Virology* **207**, 276–281
- Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**, 711–716
- Kanevsky, I., Chaminade, F., Ficheux, D., Moumen, A., Gorelick, R., Negroni, M., Darlix, J. L., and Fossé, P. (2005) Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J. Mol. Biol.* **348**, 1059–1077
- Berkhout, B., Vastenhout, N. L., Klasens, B. I., and Huthoff, H. (2001) Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **7**, 1097–1114
- Bernacchi, S., Stoylov, S., Piémont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mély, Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.* **317**, 385–399
- Darlix, J. L., Godet, J., Ivanyi-Nagy, R., Fossé, P., Mauffret, O., and Mély, Y. (2011) Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J. Mol. Biol.* **410**, 565–581
- Thomas, J. A., and Gorelick, R. J. (2008) Nucleocapsid protein function in early infection processes. *Virus Res.* **134**, 39–63
- You, J. C., and McHenry, C. S. (1994) Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J. Biol. Chem.* **269**, 31491–31495
- Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.* **80**, 217–286
- Godet, J., and Mély, Y. (2010) Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol.* **7**, 687–699
- Azoulay, J., Clamme, J. P., Darlix, J. L., Roques, B. P., and 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**, 691–700
- Hong, M. K., Harbron, E. J., O'Connor, D. B., Guo, J., Barbara, P. F., Levin, J. G., and Musier-Forsyth, K. (2003) Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-

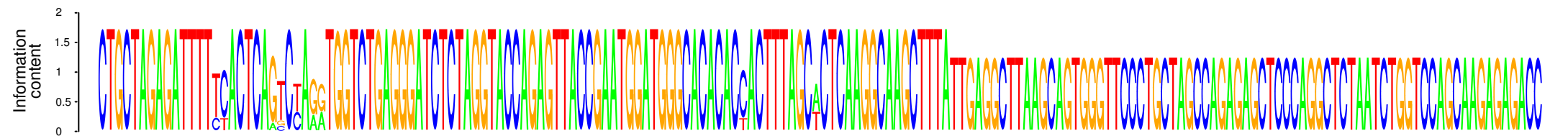
- priming in minus-strand transfer. *J. Mol. Biol.* **325**, 1–10
28. Liu, H. W., Zeng, Y., Landes, C. F., Kim, Y. J., Zhu, Y., Ma, X., Vo, M. N., Musier-Forsyth, K., and Barbara, P. F. (2007) Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5261–5267
 29. Kanevsky, I., Chaminade, F., Chen, Y., Godet, J., René, B., Darlix, J. L., Mély, Y., Mauffret, O., and Fossé, P. (2011) Structural determinants of TAR RNA-DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *Nucleic Acids Res.* **39**, 8148–8162
 30. Vo, M. N., Barany, G., Rouzina, I., and Musier-Forsyth, K. (2009) HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop “kissing” to “zipper”. *J. Mol. Biol.* **386**, 789–801
 31. Liu, H. W., Cosa, G., Landes, C. F., Zeng, Y., Kovaleski, B. J., Mullen, D. G., Barany, G., Musier-Forsyth, K., and Barbara, P. F. (2005) Single-molecule FRET studies of important intermediates in the nucleocapsid-protein-chaperoned minus-strand transfer step in HIV-1 reverse transcription. *Biophys. J.* **89**, 3470–3479
 32. Godet, J., de Rocquigny H., Raja, C., Glasser, N., Ficheux, D., Darlix, J. L., and Mély, Y. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.* **356**, 1180–1192
 33. de Rocquigny, H., Ficheux, D., Gabus, C., Fournié-Zaluski, M. C., Darlix, J. L., and Roques, B. P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.* **180**, 1010–1018
 34. Polge, E., Darlix, J. L., Paoletti, J., and Fossé, P. (2000) Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J. Mol. Biol.* **300**, 41–56
 35. Darlix, J. L., Gabus, C., Nugeyre, M. T., Clavel, F., and 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**, 689–699
 36. Lapadat-Tapolsky, M., De Rocquigny, H., Van Gent, D., Roques, B., Plasterk, R., and Darlix, J. L. (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.* **21**, 831–839
 37. Fossé, P., Motté, N., Roumier, A., Gabus, C., Muriaux, D., Darlix, J. L., and Paoletti, J. (1996) A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. *Biochemistry* **35**, 16601–16609
 38. Maxam, A. M., and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560
 39. Schneider, T. D., and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100
 40. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and 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**, 4945–4949
 41. Laughrea, M., and Jetté, L. (1996) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248–271 are dispensable for dimer formation. *Biochemistry* **35**, 1589–1598
 42. Ali, M. B., Chaminade, F., Kanevsky, I., Ennifar, E., Josset, L., Ficheux, D., Darlix, J. L., and Fossé, P. (2007) Structural requirements for nucleocapsid protein-mediated dimerization of avian leukosis virus RNA. *J. Mol. Biol.* **372**, 1082–1096
 43. Zuker, M. (2003) mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415
 44. Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J., and Fox, K. R. (2007) Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods* **42**, 128–140
 45. Desai, N. A., and Shankar, V. (2003) Single-strand-specific nucleases. *FEMS Microbiol. Rev.* **26**, 457–491
 46. Till, B. J., Burtner, C., Comai, L., and Henikoff, S. (2004) Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.* **32**, 2632–2641
 47. Holstege, F. C., and Timmers, H. T. (1997) Analysis of open complex formation during RNA polymerase II transcription initiation using heteroduplex templates and potassium permanganate probing. *Methods* **12**, 203–211
 48. Craig, M. L., Tsodikov, O. V., McQuade, K. L., Schlax, P. E., Jr., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1998) DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.* **283**, 741–756
 49. Rubin, C. M., and Schmid, C. W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.* **8**, 4613–4619
 50. Gogos, J. A., Karayiorgou, M., Aburatani, H., and Kafatos, F. C. (1990) Detection of single base mismatches of thymine and cytosine residues by potassium permanganate and hydroxylamine in the presence of tetralkylammonium salts. *Nucleic Acids Res.* **18**, 6807–6814
 51. Wang, H., Yeh, Y. S., and Barbara, P. F. (2009) HIV-1 nucleocapsid protein bends double-stranded nucleic acids. *J. Am. Chem. Soc.* **131**, 15534–15543
 52. Johnson, P. H., and Laskowski, M., Sr. (1970) Mung bean nuclease I: II. resistance of double stranded deoxyribonucleic acid and susceptibility of regions rich in adenosine and thymidine to enzymatic hydrolysis. *J. Biol. Chem.* **245**, 891–898
 53. Bourbigot, S., Ramalanjaona, N., Boudier, C., Salgado, G. F., Roques, B. P., Mély, Y., Bouaziz, S., and Morellet, N. (2008) How the HIV-1 nucleocapsid protein binds and destabilises the (–)primer binding site during reverse transcription. *J. Mol. Biol.* **383**, 1112–1128
 54. Bazzi, A., Zargarian, L., Chaminade, F., Boudier, C., De Rocquigny H., René, B., Mély, Y., Fossé, P., and Mauffret, O. (2011) Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyribose in the binding polarity of NC. *Nucleic Acids Res.* **39**, 3903–3916
 55. Schneider, T. D. (2002) Consensus sequence Zen. *Appl. Bioinformatics* **1**, 111–119
 56. Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Levin, J. G. (2000) Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J. Virol.* **74**, 8980–8988
 57. Vo, M. N., Barany, G., Rouzina, I., and Musier-Forsyth, K. (2006) Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J. Mol. Biol.* **363**, 244–261
 58. Heilman-Miller, S. L., Wu, T., and Levin, J. G. (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**, 44154–44165
 59. Godet, J., Kenfack, C., Przybilla, F., Richert, L., Duportail, G., and Mély, Y. (2013) Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res.* **41**, 5036–5048
 60. Song, M., Balakrishnan, M., Chen, Y., Roques, B. P., and Bambara, R. A. (2006) Stimulation of HIV-1 minus strand strong stop DNA transfer by genomic sequences 3' of the primer binding site. *J. Biol. Chem.* **281**, 24227–24235
 61. Piekna-Przybylska, D., and Bambara, R. A. (2011) Requirements for efficient minus strand strong-stop DNA transfer in human immunodeficiency virus 1. *RNA Biol.* **8**, 230–236
 62. Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.* **72**, 1902–1909
 63. Vuilleumier, C., Bombarda, E., Morellet, N., Gérard, D., Roques, B. P., and Mély, Y. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **38**, 16816–16825
 64. Avilov, S. V., Godet, J., Piémont, E., and Mély, Y. (2009) Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. *Biochemistry* **48**, 2422–2430
 65. Bazzi, A., Zargarian, L., Chaminade, F., De Rocquigny, H., René, B., Mély,

Secondary Structure of the HIV-1 Strong-stop DNA

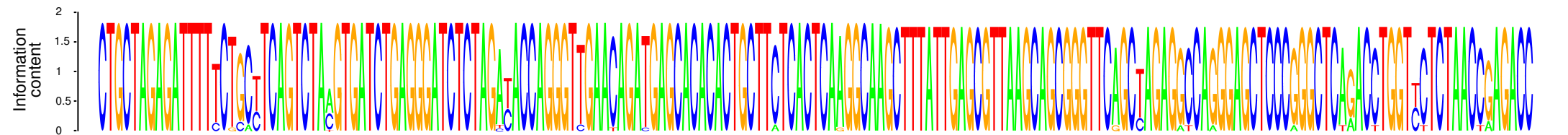
- Y., Fossé, P., and Mauffret, O. (2012) Intrinsic nucleic acid dynamics modulates HIV-1 nucleocapsid protein binding to its targets. *PLoS One* **7**, e38905
66. Damgaard, C. K., Andersen, E. S., Knudsen, B., Gorodkin, J., and Kjems, J. (2004) RNA interactions in the 5' region of the HIV-1 genome. *J. Mol. Biol.* **336**, 369–379
67. Berkhout, B. (1992) Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res.* **20**, 27–31
68. Klaver, B., and Berkhout, B. (1994) Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *EMBO J.* **13**, 2650–2659
69. Berkhout, B., Klaver, B., and Das, A. T. (1997) Forced evolution of a regulatory RNA helix in the HIV-1 genome. *Nucleic Acids Res.* **25**, 940–947
70. Beltz, H., Piémont, E., Schaub, E., Ficheux, D., Roques, B., Darlix, J. L., and Mély, Y. (2004) Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **338**, 711–723
71. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J. P., Ficheux, D., Roques, B., Darlix, J. L., and Mély, Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.* **328**, 95–108



Logo HIV-1 group M



Logo HIV-1 group N



Logo HIV-1 group O

Supplemental Figure 1. **Sequence logos for ssDNA sequence alignments.** Sequence logos resulted from: 1805 aligned ssDNA sequences of HIV-1 group M, 10 aligned ssDNA sequences of HIV-1 group N and 26 aligned ssDNA sequences of HIV-1 group O. Numbering is relative to HIV-1 group M.

ARTICLE 2

Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein

ANISSA BELFETMI,¹ LOUSSINÉ ZARGARIAN,¹ CARINE TISNÉ,² DONA SLEIMAN,² NELLY MORELLET,³ EWEN LESCOP,³ OUERDIA MASKRI,¹ BRIGITTE RENÉ,¹ YVES MÉLY,⁴ PHILIPPE FOSSÉ,¹ and OLIVIER MAUFFRET¹

¹LBPA, ENS de Cachan, CNRS, Université Paris-Saclay, 94235 Cachan Cedex, France

²Laboratoire de Cristallographie et RMN biologiques, Université Paris Descartes, CNRS UMR 8015, 75006 Paris Cedex, France

³Centre de Recherches de Gif, Institut de Chimie des Substances Naturelles, CNRS UPR 2301, 91190 Gif sur Yvette Cedex, France

⁴Laboratoire de Biophotonique et Pharmacologie, CNRS UMR 7213, Faculté de Pharmacie, Université de Strasbourg, 67401 Illkirch Cedex, France

ABSTRACT

The mature HIV-1 nucleocapsid protein NCp7 (NC) plays a key role in reverse transcription facilitating the two obligatory strand transfers. Several properties contribute to its efficient chaperon activity: preferential binding to single-stranded regions, nucleic acid aggregation, helix destabilization, and rapid dissociation from nucleic acids. However, little is known about the relationships between these different properties, which are complicated by the ability of the protein to recognize particular HIV-1 stem-loops, such as SL1, SL2, and SL3, with high affinity and without destabilizing them. These latter properties are important in the context of genome packaging, during which NC is part of the Gag precursor. We used NMR to investigate destabilization of the full-length TAR (*trans* activating response element) RNA by NC, which is involved in the first strand transfer step of reverse transcription. NC was used at a low protein:nucleotide (nt) ratio of 1:59 in these experiments. NMR data for the imino protons of TAR identified most of the base pairs destabilized by NC. These base pairs were adjacent to the loops in the upper part of the TAR hairpin rather than randomly distributed. Gel retardation assays showed that conversion from the initial TAR–cTAR complex to the fully annealed form occurred much more slowly at the 1:59 ratio than at the higher ratios classically used. Nevertheless, NC significantly accelerated the formation of the initial complex at a ratio of 1:59.

Keywords: HIV-1 nucleocapsid protein; nucleic acid dynamics; imino protons; NMR spectroscopy; kinetic model

INTRODUCTION

The mature HIV-1 nucleocapsid protein NCp7 (NC), which is generated by proteolytic cleavage of the Gag precursor, is a small retroviral protein consisting of a poorly folded highly basic N-terminal region and two highly conserved CCHC zinc finger motifs connected by a basic linker displaying some flexibility. Several forms of this protein are observed during the course of the replication cycle. NC exists as a domain of the Gag precursor protein and the partially matured products NCp15 and NCp9, which are generated by the retroviral protease (Mirambeau et al. 2010). NC is involved in many steps in the HIV-1 replication cycle, including reverse transcription, selective genome packaging, genome dimerization, genome protection, and integration (Levin et al. 2005, 2010; Darlix et al. 2007, 2011, 2014; Sleiman et al. 2012). During reverse transcription, the nucleic acid chaperone activity of NC facilitates the rearrangement of nucleic acids into their most thermodynamically stable struc-

tures. This activity results from three different properties of NC: (i) nucleic acid aggregation; (ii) duplex destabilization; (iii) rapid on-off kinetics (Cruceanu et al. 2006a,b; Stewart-Maynard et al. 2008; Godet and Mély 2010; Levin et al. 2010). Nucleic acid aggregation is thought to be dependent on the N-terminal basic part of NC, whereas duplex destabilization is associated with the zinc fingers (Beltz et al. 2005; Levin et al. 2005, 2010; Vo et al. 2009a,b; Godet and Mély 2010). However, some studies have suggested a role of zinc fingers in nucleic acid aggregation (Krishnamoorthy et al. 2003; Levin et al. 2005; Mirambeau et al. 2006; Mitra et al. 2013) and others have implicated basic residues in nucleic acid destabilization (Wu et al. 2014).

The first-strand transfer occurring during reverse transcription of the HIV-1 genome has been widely studied (Basu et al. 2008; Godet and Mély 2010; Levin et al. 2010). After the initiation of reverse transcription at the PBS

Corresponding author: olivier.mauffret@lbpa.ens-cachan.fr

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.054445.115>.

© 2016 Belfetmi et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

(primer binding site) with tRNA^{lys} (Levin et al. 2010; Sleiman et al. 2012, 2013), the minus-strand strong-stop DNA [(−)ssDNA] is synthesized from the 5′ terminal part of the RNA genome. Continuation of the reverse transcription process requires the translocation of (−)ssDNA from the 5′-end to the 3′-end of the RNA genome. This translocation requires the presence of the R (Repeat) sequence at both ends of the genome. The R sequence consists of the TAR (*trans* activating response element) and poly(A) elements, both of which can self-fold into stem-loop structures in their RNA and DNA forms (Berkhout et al. 2001). The upper part of the TAR stem-loop is required for efficient transcription of the integrated HIV-1 genome. The lower part of the TAR stem-loop also plays an important role in HIV-1 replication (Klaver and Berkhout 1994). NC promotes the annealing of the TAR RNA hairpin with the cTAR DNA hairpin. This is an essential step in the first strand transfer process and has thus been the subject of many studies (Godet et al. 2006; Liu et al. 2007; Vo et al. 2009a,b; Godet and Mély 2010). The NC-mediated destabilization of the secondary structures of TAR and cTAR stem-loops is one of the initial steps in the annealing process. NC destabilizes the cTAR DNA to a much greater extent than the more stable TAR RNA (Bernacchi et al. 2002). More generally, the destabilizing activity of NC has been shown to be inversely correlated with oligonucleotide stability (Bernacchi et al. 2002; Beltz et al. 2003, 2005). It has been shown, by time-resolved fluorescence spectroscopy, that there are preferentially destabilized segments in the lower part of the cTAR DNA molecule (Godet et al. 2013). The TAR RNA hairpin appears to be only weakly destabilized by NC, but precise data are lacking. A precise investigation of the destabilization process is thus required to improve our understanding of the destabilizing activity of NC and the initial steps of the NC-promoted annealing of cTAR and TAR.

The ability of NC to destabilize secondary structures in nucleic acids is a key component of its chaperone activity. It has been shown that this protein significantly accelerates the annealing of complementary sequences only if these sequences display a significant level of secondary structures (You and McHenry 1994; Golinelli and Hughes 2003; Heilman-Miller et al. 2004). However, despite some recent advances in the field (Bourbigot et al. 2008), the details of the destabilization mechanism and its structural and/or dynamic basis remain unknown. Globally, NC appears to be a weak duplex destabilizer able to destabilize only segments of <8 or 11 base pairs (bp) in length, depending on the study considered (Beltz et al. 2004; Levin et al. 2005). NC thus destabilizes low-stability stem-loop structures, such as the cTAR hairpin. Numerous bulges and internal loops in the TAR and cTAR hairpins are required to lower the stability of the neighboring double-stranded segments and permit their destabilization by NC (Beltz et al. 2004). Phylogenetic comparisons of TAR RNA structures in different HIV-1 isolates have demonstrated a conservation of bulges and loops (Berkhout 1992;

Zarudyana et al. 2014). It appears likely that the stabilities of the TAR and cTAR stem-loops coevolved with the destabilizing activity of NC (Beltz et al. 2004).

The relationship between nucleic acid binding affinity and nucleic acid secondary structure destabilization is complex. One previous study suggested that the strong NC-interaction domains do not correspond to the domains of the HIV-1 genome destabilized by NC (Wilkinson et al. 2008). Consistent with this observation, structural studies have shown that NC binding to the SL2 and SL3 loops does not significantly destabilize the adjacent stems (De Guzman et al. 1998; Amarasinghe et al. 2000; Johnson et al. 2000). The high affinity of binding may prevent the rapid on-off binding kinetics of NC required for its chaperone activity (Cruceanu et al. 2006a; Levin et al. 2010). Consistent with this hypothesis, Gag and NCp15 proteins have higher affinities than NC for nucleic acids but their chaperone activity is weaker (Cruceanu et al. 2006b; Wu et al. 2010; Wang et al. 2014).

In this context, we used NMR spectroscopy to investigate the NC-induced destabilization of the full-length TAR RNA stem-loop (59 nt) (Fig. 1A) to improve our understanding of the mode of action of NC in first-strand transfer. The TAR RNA has been the subject of many NMR studies, but these previous investigations focused on the upper part of the hairpin and the dynamic aspects (Puglisi et al. 1992; Aboul-ela et al. 1995, 1996; Al-Hashimi et al. 2002; Pitt et al. 2004; Musselman et al. 2007; Dethoff et al. 2008; Bailor et al. 2010; Dethoff and Al-Hashimi 2011). We analyzed the NC-mediated destabilization process by monitoring imino proton resonance shifts and line broadening after the addition of NC. We followed the destabilization events as a function of time using very fast two-dimensional NMR methods (Schanda and Brutscher 2005; Schanda et al. 2005; Farjon et al. 2009).

RESULTS

NMR analysis of the TAR RNA (59 nt)

The imino region of the ¹H-¹⁵N TROSY spectrum of the TAR RNA obtained at 10°C is shown in Figure 1B. Imino protons were identified with a combination of ¹H-¹⁵N TROSY and NOESY spectra (Fig. 1B,C). Twenty-two imino protons were assigned, whereas 26 signals were expected on the basis of secondary structure predictions (taking into account the expected two signals for each G:U base pair) (Fig. 1A). No NMR signals were observed for the G1–C59, U4–A56, U6–A55, and A22–U40 imino groups corresponding to base pairs that are located at the ends of stems. Some weaker and broader additional resonances were observed close to 12.7–12.8 ppm in the proton dimension, but could not be assigned due to the absence or weakness of associated correlations on 2D NOESY (Fig. 1). With the exception of the terminal base pair G1–C59, most of the missing imino

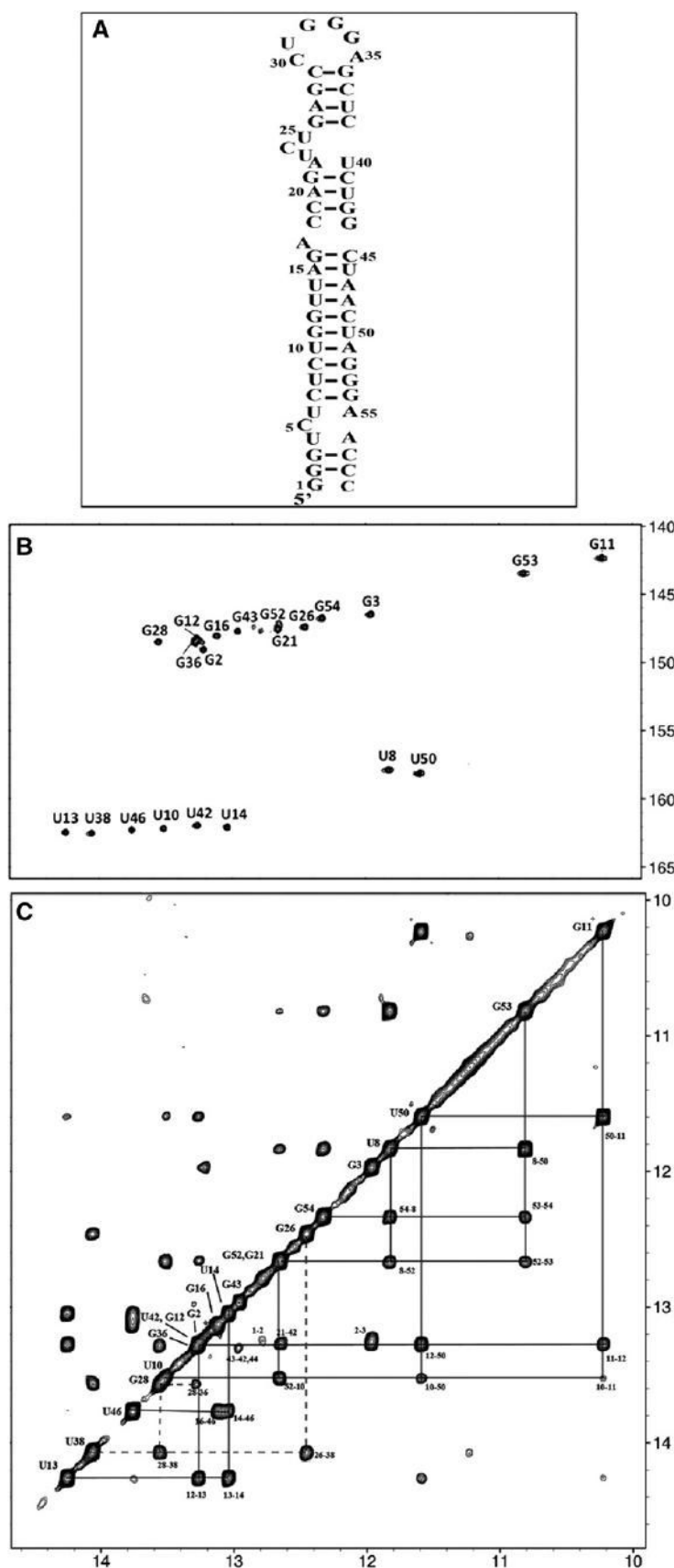


FIGURE 1. (Legend on next page)

resonances corresponded to A–U base pairs located near a bulge (Fig. 1A; the status of the C18–G44 imino proton is difficult to assess due to overlap). Thus, the imino protons were not observed for any of the A–U base pairs adjacent to a bulge, i.e., U4–A56, U6–A55, and A22–U40 (Fig. 1B,C).

TAR–cTAR annealing at a low NC:nt ratio

We then added NC to the TAR RNA at an NC:RNA ratio of 1:1 (1 NC:59 nucleotides [nt]). We chose to use this low ratio to prevent aggregation effects and excessive broadening of resonances leading to the loss of large amounts of information. It was important to check that the TAR–cTAR annealing was accelerated by NC at this low NC:nt ratio. These conditions are very different from those observed *in vivo* (Darlix et al. 2011), but they make it possible to investigate the role of NC chaperone activity in the various steps of the annealing process. Gel retardation assays displayed a striking pattern of biexponential behavior, with fast and slow components (Fig. 2). We analyzed the results obtained with a kinetic model initially developed for a similar experimental system. This model (described in the “Materials and Methods” section) involves the rapid formation of a transient duplex that is gradually converted to the fully annealed TAR–cTAR duplex (Vo et al. 2006, 2009b). The first step of the reaction was analyzed in the framework of a bimolecular and reversible reaction, because it was not possible to use pseudo-first order simplification here (see “Materials and Methods” section). Indeed, the determination of rate constants at a low NC:nt ratio requires the use of high DNA and RNA concentrations. The rate constants are reported in Table 1. We can compare the values obtained at the 1:14 ratio common to our study and that of Vo and colleagues (Fig. 9 in Vo et al. [2009b]), the values for k_1 , k_{-1} , and k_2 were very similar

($k_1 = 4.4 \times 10^3$ versus $2.5 \times 10^3 \text{ M}^{-1} \text{ sec}^{-1}$; $k_{-1} = 4 \times 10^{-3}$ versus $5 \times 10^{-3} \text{ sec}^{-1}$ and $k_2 = 3.0 \times 10^{-4}$ versus $5 \times 10^{-4} \text{ sec}^{-1}$).

When the NC:nt ratio decreases from 1:14 to 1:29, 1:59, and 0:59 (no NC), k_1 decreases by factors of 2, 5, and 27, respectively, whereas the k_{-1} remains constant. Moreover, the k_2 elementary constant decreases by more than an order of magnitude from 1:14 to 0:29. Our data thus indicate that the transition from the initial RD^* duplex (R and D stand for RNA and DNA, respectively, and RD^* is the intermediate RNA:DNA complex, see “Materials and Methods” section for more information) to the extended duplex is extremely slow at NC:nt ratios below 1:14. Interestingly, the decrease in k_1 value as the ratio decreases from 1:59 to 0:59 was significant (factor of 5). Thus, even at this low ratio (1:59), NC significantly increases the rate of formation of the initial RD^* duplex. Therefore, from the results of this kinetic analysis, we would expect an NMR study of the interaction of NC with TAR at a ratio of 1:59 to identify the TAR destabilization events important for formation of the initial TAR-cTAR (RD^*) duplex.

NMR analysis of NC-induced TAR secondary structure destabilization

The addition of NC to TAR RNA at a 1:1 ratio (NC:nt ratio of 1:59) induced considerable line broadening for all the imino proton resonances (Fig. 3). However, some resonances were affected to a larger degree than others and some even completely disappeared. Assignments of resonances were mostly deduced from analysis of 2D ^1H - ^{15}N TROSY and SOFAST-HMQC experiments (Figs. 3, 4). We classified the imino proton resonances into three categories: (i) those broadened to such an extent that they were not observable (G26, G28, and U38) (Figs. 3, 4); (ii) those broadened but still observable (G16, G21, U42, and G43), although there was some ambiguity in the assignment for G21 and G52, which partly overlap, and (iii) those displaying little broadening (G3, U10, G11, U13, U14, U46, G52, G53, and G54) (Figs. 3, 4). The imino protons not mentioned above were difficult to assess due to overlapping. In addition, the resonances of the second group (G21, U42, and G43) displayed slight chemical

FIGURE 1. Schematic diagram of TAR with the detected imino protons indicated, together with the associated NMR spectra. (A) Diagram of the TAR molecule with black marks indicating the base pairs for which imino protons were successfully detected and assigned in NMR experiments. The observation of G44 was ambiguous, and therefore no black mark was assigned to this base pair. (B) ^1H - ^{15}N TROSY spectrum of the imino groups of labeled ^{15}N , ^{13}C -labeled TAR at 10°C and 950 MHz. The peaks are labeled with the name of the residue bearing the imino proton. (C) The imino group part of the 2D NOESY spectrum of ^{15}N , ^{13}C -labeled TAR at 10°C and 950 MHz. Imino protons are labeled with the name of the residue bearing this proton. The NOE connectivities involving the different parts of TAR are indicated with lines of different types: solid lines for NOE connectivities in the long stem (base pairs C7–G54 to G16–C45); dotted lines for the upper stem (base pairs G26–C39 to C29–G36). The connectivities of the intermediate stem (base pairs C18–G44 to G21–C39) are not indicated by lines, because some ambiguity remains for the detection of G44 (with U42) near 13.25 ppm (spectral overlap). A corresponding cross-peak labeled “43–42,44” indicates the possible cross-peak within this stem. Note that the distance corresponding to the separation of G23 and U42 imino protons would be expected to exceed 5 Å in a standard A-RNA helix.

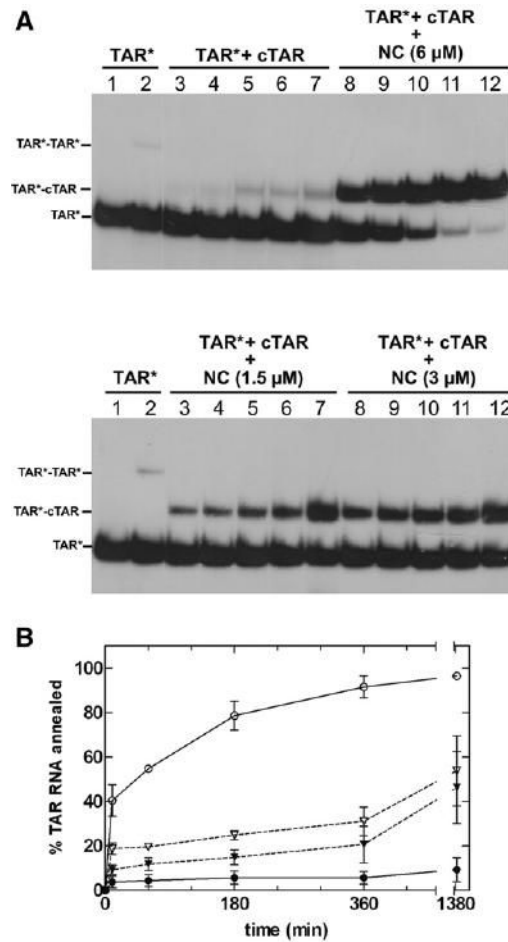


FIGURE 2. (A) Time course of TAR RNA annealing with cTAR DNA in the presence or absence of various concentrations of NC. Lane 1, heat-denatured TAR ^{32}P -RNA. TAR ^{32}P -RNA alone (lanes 2) or mixed with cTAR DNA (lanes 3–12) was incubated at 20°C for 10 min (lanes 2,3,8), 1 h (lanes 4,9), 3 h (lanes 5,10), 6 h (lanes 6,11), or 24 h (lanes 7,12). The monomeric form of TAR is indicated by TAR*. The TAR RNA-cTAR DNA duplex is indicated by TAR*-cTAR. In the absence of cTAR and NC (lanes 2), the band (TAR*-TAR*) above the monomer probably corresponds to the dimeric form of TAR described elsewhere (Andersen et al. 2004). (B) The graph was derived from the experiments shown in A. Filled circles, in the absence of NC; filled triangle, NC (1.5 μM); open triangles, NC (3 μM); open circles, NC (6 μM).

shift variations, whereas no such variations were observed for the resonances of the third group (except G3).

The three different classes of imino protons described above are located in different parts of the molecule (Fig. 5A): the most affected resonances (first group) involved the 4-bp stem G26–C39 to C29–G36 located between the apical loop and the three-pyrimidine bulge (Fig. 5A). Three imino protons in this stem were not observable in the presence of NC, and

TABLE 1. Kinetic parameters for the annealing of TAR-cTAR hairpins in the presence and absence of NC at various protein:nucleotide ratios

	NC(0:59)	NC(1:14)	NC(1:29)	NC(1:59)
k_1 ($M^{-1} \text{ sec}^{-1}$)	$0.15 \pm 0.2 \times 10^3$	$4.1 \pm 1.5 \times 10^3$	$1.5 \pm 0.4 \times 10^3$	$0.8 \pm 0.3 \times 10^3$
k_{-1} (sec^{-1})	$4.0 \pm 3 \times 10^{-3}$	$4.0 \pm 1 \times 10^{-3}$	$4.0 \pm 1 \times 10^{-3}$	$4.0 \pm 1 \times 10^{-3}$
k_2 (sec^{-1})	$0.25 \pm 0.2 \times 10^{-4}$	$3.0 \pm 1 \times 10^{-4}$	$0.3 \pm 0.2 \times 10^{-4}$	$0.4 \pm 0.2 \times 10^{-4}$
χ^2/N	1.1	1.6	1.1	1.3

χ^2/N is the total χ^2 value divided by the number of measurements.

the signal of the fourth guanine (G36) may be overlapped so it was therefore difficult to take it into account. The imino protons of the second group belong to the short stem located in the middle of the hairpin (base pairs C18–G44 to G21–C41). Those of the third group belong to the large central stem extending from base pairs C7–G54 to G16–C45 and to a 2 bp stem at the extreme end of the molecule. Nine imino protons (of 13 in total) from the long central stem were still clearly observable. They displayed a weak overlap and no shift relative to the other resonances of the TAR molecule.

The specific broadening observed for particular imino protons is consistent with a preferential destabilization of these base pairs, corresponding to an initial phase of TAR destabilization by NC. We therefore suggest that NC preferentially destabilizes the apical stem of the TAR hairpin, having a much lesser effect on the long stem in the center of this molecule. The short stem at the 5'/3' end was only moderately affected by NC, because the G3 imino proton remained clearly observable. Overall, the NMR data indicated a selective destabilization of the apical stem adjacent to the UCU bulge.

DISCUSSION

We provide here the first description of NMR studies on the full-length TAR hairpin in the absence and presence of NC in a 1:1 ratio. TAR is probably the RNA molecule most widely studied by NMR, but previous studies focused exclusively on the upper part of the TAR hairpin. Moreover, most previous studies on the upper part of TAR focused on the relative motion of the two upper helices. These helices, separated by the UCU bulge, which acts as a hinge, display independent and collective motions at the nanosecond timescale (Zhang et al. 2006, 2007; Salmon et al. 2013). The bottom part of the molecule has also been shown to have an important biological role (Klaver and Berkhout 1994). However, destabilization of the full-length TAR stem-loop by NC has never been investigated by NMR spectroscopy.

We assigned the imino protons of the TAR RNA and, therefore, determined base-pairing status for the whole molecule. The assignments of the imino protons of the upper part of the TAR hairpin were carefully compared with those obtained in previous studies dealing with this portion of the

molecule. The assignments were similar, except for the G43 residue, which has a different sequence context due to the truncation of the TAR sequence in the previous studies (Hennig and Williamson 2000; Dethoff et al. 2008). The assigned resonances and the inter-base pair NOE networks were consistent with the self-folding of TAR proposed in Figure 1A. Interestingly, the imino protons of 3 bp were not observable. All

these base pairs were U.A base pairs adjacent to a bulge (in Fig. 1A, no imino protons are indicated for C18–G44 due to overlap, resulting in ambiguity, and G1–C59 in the terminal base pair). The A22–U40 imino proton which was not observed here, was also unobservable in previous studies dealing with the upper half of the TAR hairpin (Puglisi et al. 1992; Aboul-ela et al. 1996; Zhang et al. 2006).

Overall, our experimental data are consistent with the structure predicted by *mfold* (Zuker 2003), but the imino protons of several A–U base pairs adjacent to a bulge were not observed, revealing significant fraying at these locations (Fig. 1A). Two of these A–U base pairs are located in the U4–U6 region, resulting in an unpaired U4C5U6 motif similar to the U23C24U25 bulge. These findings are consistent with chemical probing data showing some reactivity for the residues of the U4C5U6 motif, but less than for the U23C24U25 bulge and the apical loop (Kanevsky et al. 2005). Significant motions, like those already described for the U23C24U25 bulge (Zhang et al. 2006, 2007) probably

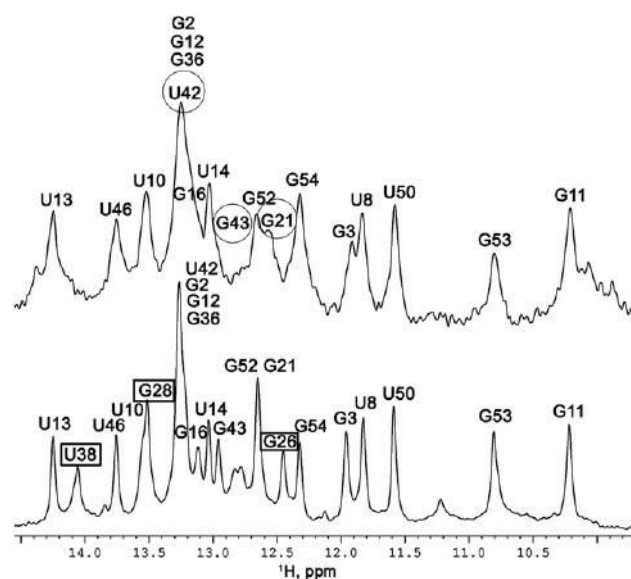


FIGURE 3. Imino proton region of ^{15}N labeled TAR at 10°C . (Bottom) Free TAR, (top) TAR with NC in a 1:1 ratio (1:59 NC:nt). The resonances that disappear totally are indicated by rectangles in spectrum at the bottom; the resonances showing significant broadening and/or chemical shifts are indicated by circles in the top spectrum.

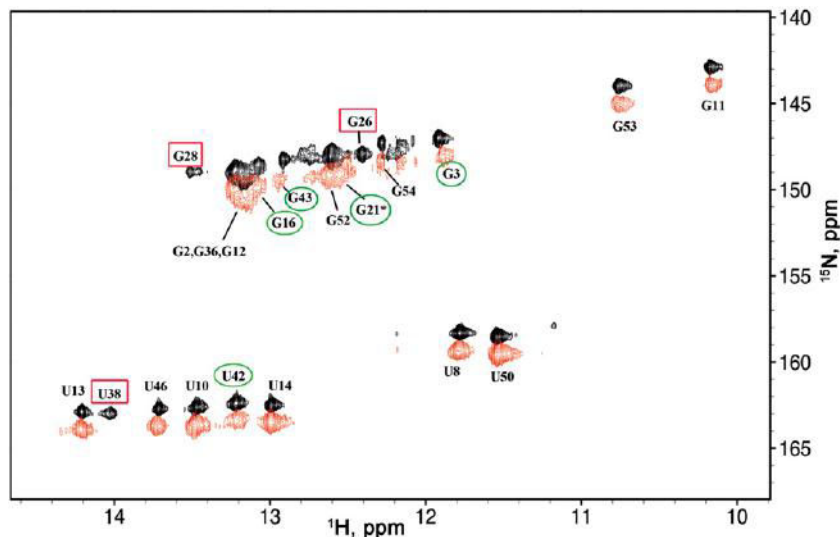


FIGURE 4. 2D ^1H - ^{15}N SOFAST-HMQC spectra at 10°C and 950 MHz of free TAR (black) and TAR-NC in a 1:1 ratio (1:59 NC:nt) (red) at 10°C and 950 MHz. The imino groups are labeled with the name of the residue. The resonances that disappear totally are indicated by red rectangles and those that display significant broadening and/or chemical shifts are indicated by green circles. The two spectra have been shifted to facilitate visualization of the sets of resonances.

occur for the U4C5U6 motif, increasing thus the potential complexity and flexibility of the TAR hairpin.

NC has nucleic acid destabilizing activity associated with chaperone properties. It is thought to exert this activity by binding preferentially to single-stranded nucleic acid regions and destabilizing the adjacent double-stranded regions (Levin et al. 2005). This activity would be expected to result in line broadening for imino protons, revealing changes to the dynamics of the base pair. Other factors may also contribute to NMR line broadening: (i) the higher correlation time associated with complex formation; (ii) fast- to intermediate-regime exchange associated with free and bound forms of the oligonucleotide, as previously described (Barraud et al. 2008). Many biophysical studies on DNA and RNA oligonucleotides have demonstrated destabilization of the nucleic acid secondary structure by NC (Beltz et al. 2005; Levin et al. 2005; Godet and Mély 2010; Godet et al. 2013) and some NMR studies described the direct impact on base pair dynamics (Johnson et al. 2000; Tisé et al. 2001). Line broadening preferentially affected the short stem surrounded by two single-stranded regions (apical loop and bulge U23C24U25). There are probably broadening effects linked to NC binding (chemical exchange), but we suggest that most of these broadening effects are due to the direct modification of base pair dynamics in the double-stranded region following the binding of NC to the adjacent single-stranded region (Fig. 5A). In contrast, the long central stem (nucleotides 6–16/45–55) was only slightly affected. The imino protons of the G16–C45 to G21–C41 base pairs displayed intermediate behavior, with significantly more broadening than those of the long central stem but less than for the apical

stem. The short stem at the TAR end was only mildly affected by NC, because the G3 imino proton was detected in the NC:TAR complex.

Several studies have reported data for NC binding sites within the TAR stem-loop. Isothermal calorimetry titrations showed that TAR contained one high affinity site (110 nM), possibly with a second site of intermediate affinity (Heng et al. 2012). Kanevsky et al. (2005) showed that the apical loop G32 and G33 residue constitute a binding site for NC, this is probably not the case for G34, which has been shown to be involved in a transient cross-loop base pair (Huthoff et al. 2004; Lee et al. 2014). This finding was supported by the results of biophysical studies on the dynamic properties of the TAR apical loop demonstrating that G32 was highly dynamic (Dethoff et al. 2008), constituting a favorable binding site for NC.

Indeed several studies have shown that NC binds preferentially to dynamic, accessible guanine residues (Sun et al. 2007; Bazzi et al. 2012). These data converge to suggest that there is a moderate- or high-affinity site in the apical loop. In this study, it was not possible to check for the presence of this site with ^1H - ^{15}N NMR data due to the absence of observable imino protons in the apical loop.

However, the presence of a putative strong binding site in the apical loop was not necessarily correlated with destabilization of the adjacent stem. Indeed, one key property of the chaperone activity of NC is the rapid on/off kinetics of nucleic acid binding (Cruceanu et al. 2006a). This property suggests that NC does not need to bind to high-affinity sites to destabilize DNA/RNA structures. Indeed, its ability to recognize specific high-affinity stem-loops is important for the selective recognition of the viral RNA genome by the NC domain of the Gag protein during the packaging process, but not for reverse transcription. Several structural studies on NC–nucleic acid complexes (De Guzman et al. 1998; Amarasinghe et al. 2000; Johnson et al. 2000; Bourbigot et al. 2008) have provided support for this assertion, by showing that the strong binding of NC to the apical loops of the SL2 and SL3 elements does not destabilize the adjacent stems. In contrast, NC binding to stem-loops with lower affinities than SL2 and SL3 induces destabilization of the adjacent stems (Johnson et al. 2000; Bourbigot et al. 2008). Consistent with these findings, the sequential processing of Gag by protease enhances the nucleic acid chaperone activity of the resulting polypeptides (NCp15, NCp9, NCp7), and this improvement is associated with a lower affinity and faster binding/dissociation kinetics with nucleic acids (Cruceanu et al. 2006b; Rein 2010; Wu et al. 2013; Wang et al. 2014).

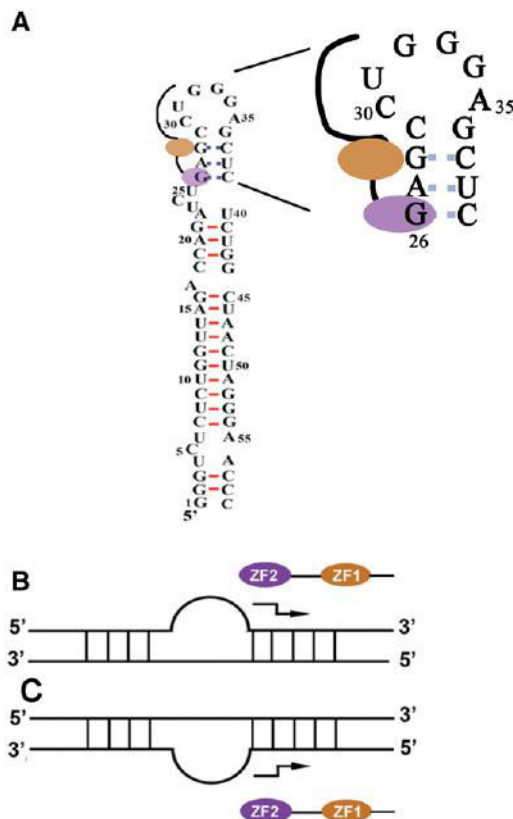


FIGURE 5. Hypothetical diagrams of RNA-DNA binding and destabilization by NC. (A) View of the TAR hairpin with a report of the observed imino protons (red) observed in the NC-TAR complex and those observable in the free RNA but not in the NC-RNA complex (dashed blue lines). NC is depicted with its two zinc fingers (ZF1: orange ellipse; ZF2: violet ellipse), the linker between the two zinc fingers and the N-terminal basic region are indicated by solid lines. We show also an expanded view of the destabilized region highlighting the hypothetical binding site at the G26 base for the ZF2 domain, whereas the ZF1 domain is more involved in the destabilization process. This image is consistent with both the data presented here and the work of McCauley et al. (2015). (B) Diagram of the destabilization of RNA secondary structure mediated by NC. The example shows a double-stranded RNA interrupted by a bulge. The binding polarity of the zinc fingers along the 5'-3' nucleic acid chain is as observed in the three solved structures of RNA-NC complexes, with the N- to C-terminal polarity of the peptide chain antiparallel to the 5'-3' direction of the nucleic acid chain (Bazzi et al. 2011). The NC is thought to bind preferentially to the strand containing single-stranded regions, and a staggered arrow indicates the preferentially destabilized region. This orientation of the destabilization process results from the ordering of the two domains of NC along the nucleic acid chain, with ZF1 responsible for stem destabilization (see "Discussion" section). (C) Diagram of the destabilization of DNA secondary structure mediated by NC. This example is similar to that shown in B, but with an opposite binding polarity deduced from that of known DNA-NC complexes (Bazzi et al. 2011). The position of the bulge is reversed relative to B as it is the case for the cTAR hairpin. The mechanisms involved in RNA and DNA destabilization converge to destabilize the same region of the two molecules, to the right of the bulge.

The destabilized stems are probably adjacent to binding sites of low or moderate affinity for NC. There should be a balance between the preferential binding of NC to exposed

guanines residues and the need for rapid on/off kinetics. Consistent with this, a SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) study of the HIV-1 NL4-3 RNA genome suggested that the strong NC interaction domains and the domains destabilized by NC are not located in the same regions (Wilkinson et al. 2008). Indeed, Wilkinson and colleagues showed that the strong NC interaction domains in the SL1, SL2, and SL3 stem-loops were characterized by single-stranded regions containing unpaired G residues, whereas the destabilized domains contained single-stranded A/U-rich motifs adjacent to a duplex in which the first base pair included a guanosine nucleotide. The UCU bulge followed by the G26-C39 pair of the TAR hairpin is one of the six destabilized sites identified in the HIV-1 genome (Wilkinson et al. 2008). SHAPE analysis detected no strong NC binding sites within the TAR element.

Previous studies have shown that the destabilization of the TAR RNA hairpin by NC(12-55) is weaker than that of the cTAR DNA hairpin by the same peptide (Bernacchi et al. 2002; Azoulay et al. 2003; Beltz et al. 2005). Bernacchi et al. (2002) showed that destabilization of the TAR RNA hairpin by NC(12-55) involved only 1 bp at a peptide to nucleotide molar ratio of 1:7.5. TAR destabilization in our conditions was limited because NC destabilizes only three consecutive base pairs with high efficiency. However, the NC:nt ratio was significantly different in the two studies (1:7.5 versus 1:59) (Bernacchi et al. 2002) and NC(1-55) has a greater destabilization capacity than NC(12-55) (Levin et al. 2005; Vo et al. 2009b; Mitra et al. 2013). Finally, the extensive broadening of imino proton resonance does not necessarily indicate a loss of the corresponding base pair. It may, instead, indicate an increase in the proton exchange rate of the observed base pair.

One interesting finding of this study was the difference in behavior of the two stems adjacent to the internal UCU bulge, the upper stem (nucleotides 26-29/36-39) being significantly more destabilized by NC than the lower stem (nucleotides 18-22, 40-44). The SHAPE data (Wilkinson et al. 2008) indicated that the UCU bulge was less reactive after the removal of NC, suggesting a partial destabilization of one or both adjacent stems by the protein. We suggest that there may be an initial binding event at the G26 position adjacent to the internal loop, as the structure of this site is close to the consensus site already identified by SHAPE with a guanine located in a double-stranded segment (Fig. 11 in Wilkinson et al. 2008). The breathing of the G26-C39 base pair adjacent to a bulge might facilitate an interaction of NC with G26, thereby inducing destabilization of the four base pairs of this stem from G26 to C29 (Fig. 5A). This mechanism is very similar to that described for RNA helicases, in which a first binding event allows the adjacent base pairs to be destabilized, thereby increasing their fraying (Woodson 2010; Jankowsky 2011; Mustoe et al. 2014). We suggest that the UCU bulge and the apical loop constitute binding sites of moderate or low

affinity for NC, allowing fast dissociation kinetics, with the protein sliding toward the double-stranded segments. According to this hypothesis, the upper stem would be preferentially destabilized because it is located between two single-stranded regions, whereas the lower stem is close to only one single-stranded region (bulge UCU). Alternatively, the upper stem may be preferentially destabilized because the NC functions with a definite polarity. The ability of NC to bind nucleic acids with a defined polarity has been suggested in a number of studies (Fig. 5B,C; Vuilleumier et al. 1999; Bourbigot et al. 2008; Bazzi et al. 2011; Darlix et al. 2011). In the three RNA–NC complexes for which high-resolution structures have been obtained (De Guzman et al. 1998; Amarasinghe et al. 2000; Spriggs et al. 2008), the protein is oriented anti-parallel to the 5′–3′ direction of the nucleic acid chain. As a result, the N-terminal zinc finger (ZF1) binds a residue 3′ to the residue binding to the C-terminal zinc finger (ZF2) (Fig. 5B). This binding polarity may be responsible for the polarity of destabilization observed for the adjacent stems (Fig. 5B). This hypothesis is consistent with the SHAPE data for the 5′ UTR region of the HIV-1 genome, suggesting that the destabilized stems are located on the 3′-side of single-stranded regions (Wilkinson et al. 2008). The relationship between binding polarity and the polarized destabilization of stem-loops may result from the lack of equivalence of the two zinc fingers, ZF1 and ZF2, with ZF1 more involved in the destabilization than ZF2. In contrast, ZF2 is involved in the specific recognition of guanine residues (Guo et al. 2002; Heath et al. 2003; Beltz et al. 2005; Narayanan et al. 2006; Hergott et al. 2013; Mitra et al. 2013). Binding polarity may therefore be responsible for the localization of ZF1 close to a double-stranded region. The binding of NC to G26 may position the ZF1 toward the upper stem (Fig. 5A) triggering the destabilization of this stem (Fig. 5B).

Our results are very similar to those of a recent single-molecule stretching paper investigating TAR destabilization by NC (McCauley et al. 2015). This recent study, like ours, identified specific targeted sites corresponding to guanines adjacent to unstable stem regions, such as mismatches, loops, and bulges. G26 was identified as a key site for the initiation of destabilization in both studies. The two studies differed in the NC:nt ratio used, with a ratio of 1:4 in the single-molecule study and 1:59 here. We used a low ratio to detect the initial destabilizing event preceding the binding of NC to other sites as its concentration increases (Darlix et al. 2011; McCauley et al. 2015).

At the low NC:nt ratio (1:59) used in our NMR study, destabilization was limited to the upper part of TAR. Thus, the long stem involving nucleotides 6–16/45–55 and the short terminal stem (nucleotides 1–4/56–59) displayed no significant destabilization. Additional NC molecules are probably required to destabilize these stems. Our results provide an opportunity to compare rate constants from the different parts of the reactional scheme (see Equation 2). The formation of the initial TAR–cTAR binary complex appears

to be the first event after the addition of NC, because the associated bimolecular rate k_1 displays considerable variation, by a factor of almost 30 when the NC:nt ratio is increased from 0 to 1:14 (Table 1). In contrast, under the same conditions, the dissociation rate k_{-1} displayed no variation. The strand-exchange rate k_2 increased by a factor of 10 when the NC:nt ratio is increased from 0 to 1:14. Finally, our data are consistent with the destabilizing activity of NC being important only for the steps of the annealing reaction requiring duplex disruption. Indeed, the formation of the initial binary complex, involving the melting of the upper 4-bp stem, is highly dependent on the destabilizing activity of NC, whereas the reciprocal reaction does not require any structure melting. The final conversion into the extended duplex is also highly dependent on NC, but to a lesser extent than binary complex formation, and this step requires higher NC concentrations. In conclusion, our results are consistent with the formation of an extended kissing loop complex TAR–cTAR as suggested, in previous studies performed at low NC:nt ratios. However, they do not exclude the possibility that the “zipper” pathway is preferred at high NC:nt ratios (Godet et al. 2006; Liu et al. 2007; Vo et al. 2009b).

MATERIALS AND METHODS

Samples

The $^{15}\text{N}/^{13}\text{C}$ labeled TAR RNA sequence was obtained from Silantes GmbH (Germany). The TAR sequence used was that of the NL4-3 isolate. The NC(1–55) sequence (NL4-3 isolate) was overexpressed and the product was purified, as previously described (Barraud et al. 2007).

Samples were dissolved in 20 mM sodium phosphate buffer, pH 6.5, supplemented 50 mM NaCl and 0.1 mM EDTA in a 300 μL Shigemi tube. The RNA strand concentration in the tube was 340 μM , TAR RNA, and NC(1–55) were mixed in a 1:1 molar ratio (NC:nt 1:59).

NMR

NMR spectra were recorded on a Bruker Avance 950 MHz (TGIR, Gif sur Yvette, France) equipped with a triple-resonance cryoprobe with z-axis field gradient. All experiments were performed in 90% H_2O and 10% $^2\text{H}_2\text{O}$, at 283 K. For NOESY experiments, two mixing times of 150 and 200 msec were used, and the spectral widths were 20 833 Hz in both dimensions. 512 t1 experiments were recorded in the States-TPPI mode, each with 4029 data points per experiment. ^1H – ^{15}N TROSY sensitivity enhanced gradient (Pervushin et al. 1997; Nietlispach 2005) experiments were carried out with 256 points in t1 and 1024 data points per experiment. ^1H – ^{15}N 2D SOFAST-HMQC experiments were recorded with a PC9 and REBURP pulses for the band selective excitation and refocusing of imino protons (Schanda and Brutscher 2005; Farjon et al. 2009) and for recording the changes in the complexes over time. The relaxation delay was set to 200 msec and 200 experiments were recorded in the ^{15}N dimension; the total experiment time per 2D was 295 sec.

Gel retardation assays

A heat-denatured TAR (7.5 pmol of TAR 32 P-RNA at 3×10^3 cpm/pmol in 10 μ L of double-distilled water) was used as a control. It was prepared by heating at 90°C for 2 min and chilling for 2 min on ice, and mixing with 4 μ L of loading buffer (50% w/v glycerol, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanol). The annealing assays were carried out in a final volume of 10 μ L. TAR 32 P-RNA (7.5 pmol at 3×10^3 cpm/pmol) in 3.6 μ L of water was heated at 90°C for 2 min and chilled for 2 min on ice. We then added 0.9 μ L of annealing buffer (final concentrations: 30 mM NaCl, 0.2 mM MgCl₂, and 25 mM Tris-HCl pH 7.5) and the sample was incubated at 20°C for 30 min. Unlabeled DNA (7.5 pmol) was subjected to the same renaturation treatment and was then added to refolded TAR 32 P-RNA. The reaction mixture was incubated at 20°C in the absence or presence of NC (1.5, 3, or 6 μ M) for various periods of time. Reactions were stopped by adding SDS (at a final concentration of 0.5%) and by proteinase K digestion (2 μ g) at 20°C for 30 min. At the end of the incubation period, we added 4 μ L of loading buffer to the reaction mixture. The samples were analyzed by electrophoresis on a 12% polyacrylamide gel (37.5:1 [w/w], acrylamide/bisacrylamide) at 25°C in 1 $\times$ TBE buffer (90 mM Tris-borate [pH 8.3], 2 mM EDTA). After electrophoresis, the gel was fixed, dried and autoradiographed. Unannealed and annealed TAR RNAs were quantified with a Typhoon TRIO (GE Healthcare) and ImageQuant software. The percent annealed TAR RNA was determined as $100 \times [\text{annealed}/(\text{annealed} + \text{unannealed})]$.

Kinetic analysis of the gel-retardation assays

We performed gel retardation assays to follow the extended duplex formation over time. Extended duplex formation at various NC:nt ratios appeared to follow biexponential kinetics with two processes occurring at very different rates. This behavior could be described by Equation 1, where $P(t)$ is the percentage of annealed molecules (Vo et al. 2009b). The kinetic phases could therefore be described by two components with rate constants k_f for the fast component and k_s for the slow component, where f is the fraction of the fast component and corresponds to the probability of intermediate complex RD^* formation:

$$P(t) = P_{\text{plateau}}(f \cdot (1 - e^{-k_f t}) + (1 - f) \cdot (1 - e^{-k_s t})). \quad (1)$$

This analysis is similar to those used in previous studies (Vo et al. 2006, 2009b). We used this approach because it was developed for the same experimental system (analysis by gel-retardation assays of TAR-cTAR annealing in the presence of NC). The TAR and cTAR hairpins were assumed to anneal via a two-step pathway:



This reaction scheme describes a first rapid step in which a productive and transient complex is formed. This complex is slowly converted to the extended duplex corresponding to the fully annealed state. The RD^* and RD complexes (R and D indicate RNA and DNA strands, respectively) cannot be distinguished on the basis of gel electrophoresis data.

Unlike Vo et al. (2009b), we were unable to make use of an approximation to pseudo-first-order conditions because $[D]_0$ and $[R]_0$ concentrations were identical. The reasons for the choice of

these concentrations are explained below. The first step of the reaction (2) corresponds to a second-order bimolecular reversible reaction. Some simplifications can be used to describe this reaction: $[D] = [R]$ and $[RD^*] = [D]_0 - [D]$ due to our initial conditions, $[D]_0 = [R]_0$ and the stoichiometry of the reaction. In these conditions

$$d \frac{[RD^*]}{dt} = -d \frac{[D]}{dt} = k_1 [D][R] - k_{-1} [RD^*] \quad (3)$$

and

$$- \frac{d[D]}{dt} = k_1 [D]^2 + k_{-1} [D] - k_{-1} [D]_0. \quad (4)$$

After integration [of the form $dx/(a + bx + cx^2)$], the time dependence of the concentration of free cTAR was obtained as follows (Baker and Weng 1992):

$$[D] = \frac{(k_{-1} + k_{\text{obs}})Ce^{-k_{\text{obs}}t} - k_{-1} + k_{\text{obs}}}{2k_1(1 - Ce^{-k_{\text{obs}}t})} \quad (5)$$

where

$$k_{\text{obs}} = (4k_1k_{-1}[D]_0 + k_{-1}^2)^{1/2},$$

and

$$C = \frac{2k_1[D]_0 + k_{-1} - k_{\text{obs}}}{2k_1[D]_0 + k_{-1} + k_{\text{obs}}}.$$

Using the expression of $[D]$ and $[RD^*] = [D]_0 - [D]$ as a function of the time, we can then deduce the fractions of the fast (f) component throughout the reaction, this fraction corresponding to the probability of intermediate formation:

$$f = \frac{[RD^*]}{[D]_0}. \quad (6)$$

Equation 1 can be therefore written for all values of time, using $k_f = k_{\text{obs}}$. According to which the fast component assigned to the first step of the reaction is related to the rate constant k_{obs} described in Equation 5 and the slow component rate constant corresponding to the second and slower component of Equation 2 is proportional to the fraction of the preannealed intermediate RD^* such that $k_s = f \cdot k_2$ (Vo et al. 2009b). We simulated the values of Equation 1 using Equations 5 and 6 and numerical values for k_1 , k_{-1} , and k_2 (k_{-2} was not taken into account as its value was very low) and retained the values giving the lowest χ^2 value. Statistical procedures were used to estimate the uncertainty on the various rate constants. Using the uncertainty of the various experimental measurements, we generated files of synthetic experimental data, using a Gaussian distribution and the Box-Muller method (Bevington and Robinson 2003). The global χ^2 calculations were used to identify the best values for the rate constants. The criterion for selecting uncertainty was a doubling of the χ^2 values.

Choice of nucleic acid concentrations

In the gel-retardation experiments described above, the kinetic analysis was carried out with a second-order bimolecular and reversible mechanism. The associated equations are much more complex than those used when pseudo-first order conditions can be assumed (i.e., one of the components [DNA] is in large excess over the other [RNA]). We chose to work with equal DNA and

RNA concentrations and with an NC:RNA ratio (1:59) similar to those of the NMR studies. These conditions are probably also more similar to those in vivo with similar DNA and RNA concentrations. These conditions necessitated the use of an analysis method appropriate for a bimolecular association mechanism. In addition to the choice concerning stoichiometry, we also worked at higher DNA concentrations than in previous studies. This point was discussed by Vo et al. (2006, 2009b) when they described the limitations of the experimental conditions associated with pseudo-first order conditions and the concentrations used. Indeed, the elementary rates k_1 and k_{-1} could be determined only for NC:nt > 1:16, because these rates can be determined only if there is a large fast component (f). However, at the concentrations used by these authors ($[D]_0 = 150$ nM), the f fraction, the fraction corresponding to the fast component, could be deduced from pseudo-first order conditions as follows:

$$f = \frac{k_1 D}{k_1 D + k_{-1}} \quad (7)$$

From Equation 7 it is clear that in the conditions used, $k_1 D \ll k_{-1}$. The f fraction is small, at ~5%. However, using the same NC:nt ratio and therefore the same k_1 and k_{-1} rate values but with a larger DNA concentration (a larger D term), $k_1 D$ is larger and the relationship $k_1 D \ll k_{-1}$ no longer holds; this implies the existence of a significant f fraction, as calculated below. A second limitation is that, at a high NC:nt ratio (>1:7), the fast component is dominant and the elementary rate constant associated with the second step (k_2) cannot be determined. Indeed, under the conditions used by Vo et al. (2006, 2009b), k_2 could be only determined at ratios of 1:16 to 1:8, a rather limited range.

These limitations can be overcome and rate constants determined correctly even at low NC:nt ratios in the case of a high DNA concentration ($[D]_0 = 750$ nM) is used, indeed in these conditions, $k_1 D$ became close from k_{-1} . In such conditions, the fraction f is 30%, 18%, 9%, and 2% at the 1:14, 1:29, 1:59 and 0:59 ratios, respectively. The fast component is therefore clearly observable (except at the 0:59 ratio) as shown by the experimental data (Fig. 4). Furthermore, as the fast component is never present in a large excess, it was also possible to determine k_2 .

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support received from the TGIR-RMN-THC Fr3050 CNRS grant. We thank ANRS (Agence Nationale Française de Recherche contre le SIDA et les Hépatites) for financial support for our research.

Received September 16, 2015; accepted December 6, 2015.

REFERENCES

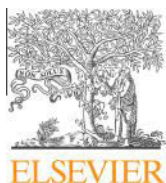
Aboul-ela F, Karn J, Varani G. 1995. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol* **253**: 313–332.
 Aboul-ela F, Karn J, Varani G. 1996. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res* **24**: 3974–3981.
 Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ. 2002. Concerted motions in HIV-1 TAR RNA may allow access to

bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol* **315**: 95–102.
 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 Ψ -RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301**: 491–511.
 Andersen ES, Contera SA, Knudsen B, Damgaard CK, Bsenbacher F, Kjems J. 2004. Role of the trans-activation response element in dimerization of HIV-1 RNA. *J Biol Chem* **279**: 22243–22249.
 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**: 691–700.
 Bailor MH, Sun XY, Al-Hashimi HM. 2010. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**: 202–206.
 Baker GM, Weng LC. 1992. Apparent first-order behavior under second-order kinetic conditions: a general concept illustrated by the reversible binding of hydrogen peroxide to cytochrome C oxidase. *J Theor Biol* **158**: 221–229.
 Barraud P, Gaudin C, Dardel F, Tisné C. 2007. New insights into the formation of HIV-1 reverse transcription initiation complex. *Biochimie* **89**: 1204–1210.
 Barraud P, Dardel F, Tisné C. 2008. Optimizing HSQC experiment for the observation of exchange broadened signals in RNA–protein complexes. *Cr Chim* **11**: 474–479.
 Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Res* **134**: 19–38.
 Bazzi A, Zargarian L, Chaminade F, Boudier C, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O. 2011. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyriboses in the binding polarity of NC. *Nucleic Acids Res* **39**: 3903–3916.
 Bazzi A, Zargarian L, Chaminade F, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O. 2012. Intrinsic nucleic acid dynamics modulates HIV-1 nucleocapsid protein binding to its targets. *PLoS One* **7**: e38905.
 Beltz H, Azoulay J, Bernacchi S, Clamme JP, Ficheux D, Roques B, Darlix JL, Mély Y. 2003. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* **328**: 95–108.
 Beltz H, Piémont E, Schaub E, Ficheux D, Roques B, Darlix JL, Mély Y. 2004. Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* **338**: 711–723.
 Beltz H, Clauss C, Piémont E, Ficheux D, Gorelick RJ, Roques B, Gabus C, Darlix JL, 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**: 1113–1126.
 Berkhout B. 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res* **20**: 27–31.
 Berkhout B, Vastenhout NL, Klasens BI, Huthoff H. 2001. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **7**: 1097–1114.
 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**: 385–399.
 Bevington PR, Robinson DK. 2003. *Data reduction and error analysis for the physical sciences*, 3rd ed. McGraw-Hill, NY.
 Bourbigot S, Ramalanjaona N, Boudier C, Salgado GF, 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**: 1112–1128.
 Cruceanu M, Gorelick RJ, Musier-Forsyth K, Rouzina I, Williams MC. 2006a. Rapid kinetics of protein–nucleic acid interaction is a major

- component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol* **363**: 867–877.
- Cruceanu M, Urbaneja MA, Hixson CV, Johnson DG, Datta SA, Fivash MJ, Stephen AG, Fisher RJ, Gorelick RJ, Casas-Finet JR, et al. 2006b. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **34**: 593–605.
- Darlix JL, Garrido JL, Morellet N, Mély Y, de Rocquigny H. 2007. Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Adv Pharmacol* **55**: 299–346.
- Darlix JL, Godet J, Ivanyi-Nagy R, Fossé P, Mauffret O, Mély Y. 2011. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J Mol Biol* **410**: 565–581.
- Darlix JL, de Rocquigny H, Mauffret O, Mély Y. 2014. Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus Res* **193**: 2–15.
- 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**: 384–388.
- Dethoff EA, Al-Hashimi HM. 2011. Characterization of ps-Ms dynamics in the TAR apical loop by NMR. *Biophys J* **100**: 238a.
- Dethoff EA, Hansen AL, Musselman C, Watt ED, Andricioaei I, Al-Hashimi HM. 2008. Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**: 3906–3915.
- Farjon J, Boisbouvier J, Schanda P, Pardi A, Simorre JP, Brutscher B. 2009. Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *J Am Chem Soc* **131**: 8571–8577.
- Godet J, Mély Y. 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol* **7**: 687–699.
- Godet J, de Rocquigny H, Raja C, Glasser N, Fichoux D, Darlix JL, Mély Y. 2006. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol* **356**: 1180–1192.
- Godet J, Kenfack C, Przybilla F, Richert L, Duportail G, Mély Y. 2013. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res* **41**: 5036–5048.
- Golinelli MP, Hughes SH. 2003. Secondary structure in the nucleic acid affects the rate of HIV-1 nucleocapsid-mediated strand annealing. *Biochemistry* **42**: 8153–8162.
- Guo J, Wu T, Kane BF, Johnson DG, Henderson LE, Gorelick RJ, Levin JG. 2002. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **76**: 4370–4378.
- Heath MJ, Derebail SS, Gorelick RJ, DeStefano JJ. 2003. Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing. *J Biol Chem* **278**: 30755–30763.
- 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**: 44154–44165.
- Heng X, Kharytonchik S, Garcia EL, Lu K, Divakaruni SS, LaCotti C, Edme K, Telesnitsky A, Summers MF. 2012. Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J Mol Biol* **417**: 224–239.
- Hennig M, Williamson JR. 2000. Detection of N-H...N hydrogen bonding in RNA via scalar couplings in the absence of observable imino proton resonances. *Nucleic Acids Res* **28**: 1585–1593.
- Hergott CB, Mitra M, Guo J, Wu T, Miller JT, Iwatani Y, Gorelick RJ, Levin JG. 2013. Zinc finger function of HIV-1 nucleocapsid protein is required for removal of 5'-terminal genomic RNA fragments: a paradigm for RNA removal reactions in HIV-1 reverse transcription. *Virus Res* **171**: 346–355.
- Huthoff H, Girard F, Wijmenga SS, Berkhout B. 2004. Evidence for a base triple in the free HIV-1 TAR RNA. *RNA* **10**: 412–423.
- Jankowsky E. 2011. RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36**: 19–29.
- Johnson PE, Turner RB, Wu ZR, Hairston L, Guo J, Levin JG, Summers MF. 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **39**: 9084–9091.
- Kanevsky I, Chaminade F, Fichoux D, Moumen A, Gorelick R, Negroni M, Darlix JL, Fossé P. 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J Mol Biol* **348**: 1059–1077.
- Klaver B, Berkhout B. 1994. Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *EMBO J* **13**: 2650–2659.
- Krishnamoorthy G, Roques B, Darlix JL, Mély Y. 2003. DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection. *Nucleic Acids Res* **31**: 5425–5432.
- Lee J, Dethoff EA, Al-Hashimi HM. 2014. Invisible RNA state dynamically couples distant motifs. *Proc Natl Acad Sci* **111**: 9485–9490.
- Levin JG, Guo J, Rouzina I, Musier-Forsyth K. 2005. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80**: 217–286.
- Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **7**: 754–774.
- Liu HW, Zeng Y, Landes CF, Kim YJ, Zhu Y, Ma X, Vo MN, Musier-Forsyth K, Barbara PF. 2007. Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc Natl Acad Sci* **104**: 5261–5267.
- McCauley MJ, Rouzina I, Manthei KA, Gorelick RJ, Musier-Forsyth K, Williams MC. 2015. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc Natl Acad Sci* **112**: 13555–13560.
- Mirambeau G, Lonnais S, Coulaud D, Hameau L, Lafosse S, Jusset 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**: 496–511.
- Mirambeau G, Lonnais S, Gorelick RJ. 2010. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol* **7**: 724–734.
- Mitra M, Wang W, Vo MN, Rouzina I, Barany G, Musier-Forsyth K. 2013. The N-terminal zinc finger and flanking basic domains represent the minimal region of the human immunodeficiency virus type-1 nucleocapsid protein for targeting chaperone function. *Biochemistry* **52**: 8226–8236.
- Musselman C, Al-Hashimi HM, Andricioaei I. 2007. IRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys J* **93**: 411–422.
- Mustoe AM, Brooks CL, Al-Hashimi HM. 2014. Hierarchy of RNA functional dynamics. *Annu Rev Biochem* **83**: 441–466.
- Narayanan N, Gorelick RJ, DeStefano JJ. 2006. Structure/function mapping of amino acids in the N-terminal zinc finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity. *Biochemistry* **45**: 12617–12628.
- Nietispach D. 2005. Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J Biomol NMR* **31**: 161–166.
- Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T₂ relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci* **94**: 12366–12371.
- Pitt SW, Majumdar A, Serganov A, Patel DJ, Al-Hashimi HM. 2004. Arginamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg²⁺-induced conformational stabilization. *J Mol Biol* **338**: 7–16.

- Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**: 76–80.
- Rein A. 2010. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol* **7**: 700–705.
- Salmon L, Bascom G, Andricioaei I, Al-Hashimi HM. 2013. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for inter-helical motions revealed. *J Am Chem Soc* **135**: 5457–5466.
- Schanda P, Brutscher B. 2005. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* **127**: 8014–8015.
- Schanda P, Kupce E, Brutscher B. 2005. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**: 199–211.
- Sleiman D, Goldschmidt V, Barraud P, Marquet R, Paillart JC, Tisné C. 2012. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res* **169**: 324–339.
- Sleiman D, Barraud P, Brachet F, Tisné C. 2013. The interaction between tRNA^{Lys} and the primer activation signal deciphered by NMR spectroscopy. *PLoS One* **8**: e64700.
- Spriggs S, Garyu L, Connor R, Summers MF. 2008. Potential intra- and intermolecular interactions involving the unique-5' region of the HIV-1 5'-UTR. *Biochemistry* **47**: 13064–13073.
- Stewart-Maynard KM, Cruceanu M, Wang F, Vo MN, Gorelick RJ, Williams MC, Rouzina I, Musier-Forsyth K. 2008. Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J Virol* **82**: 10129–10142.
- Sun XY, Zhang Q, Al-Hashimi HM. 2007. Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg²⁺ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**: 1698–1713.
- Tisné C, Roques BP, Dardel F. 2001. Heteronuclear NMR studies of the interaction of tRNA^{Lys} with HIV-1 nucleocapsid protein. *J Mol Biol* **306**: 443–454.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2006. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol* **363**: 244–261.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2009a. Effect of Mg²⁺ and Na⁺ on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J Mol Biol* **386**: 773–788.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2009b. HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop “kissing” to “zipper”. *J Mol Biol* **386**: 789–801.
- 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**: 16816–16825.
- 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**: 7145–7159.
- Wilkinson KA, 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**: e96.
- Woodson SA. 2010. Taming free energy landscapes with RNA chaperones. *RNA Biol* **7**: 677–686.
- Wu T, Datta SA, 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**: 556–567.
- Wu H, Mitra M, McCauley MJ, Thomas JA, Rouzina I, Musier-Forsyth K, Williams MC, Gorelick RJ. 2013. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res* **42**: 2525–2537.
- 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**: 2525–2537.
- You JC, McHenry CS. 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem* **269**: 31491–31495.
- Zarudyan MI, Potyahaylo AL, Kolomiets IM, Hovorun DM. 2014. Phylogenetic study on structural elements of HIV-1 poly(A) region. 2. USE domain and TAR hairpin. *Biopolym Cell* **30**: 29–36.
- Zhang Q, Sun X, Watt ED, Al-Hashimi HM. 2006. Resolving the motional modes that code for RNA adaptation. *Science* **311**: 653–656.
- Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**: 1263–1267.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

ARTICLE 3



Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

RNA SHAPE chemistry with aromatic acylating reagents



Laura Nodin^a, Olivier Noël^a, Françoise Chaminade^b, Ouerdia Maskri^b, Vincent Barbier^c, Olivier David^c, Philippe Fossé^{b,*}, Juan Xie^{a,*}

^a PPSM, CNRS, Institut d'Alembert, ENS de Cachan, 61 Avenue du P^e Wilson, F-94235 Cachan, France

^b LBPA, CNRS, Institut d'Alembert, ENS de Cachan, 61 Avenue du P^e Wilson, F-94235 Cachan, France

^c Institut Lavoisier, CNRS, Université de Versailles St. Quentin-en-Yvelines, 45 avenue des Etats-Unis, 78035 Versailles, France

ARTICLE INFO

Article history:

Received 22 October 2014

Revised 4 December 2014

Accepted 8 December 2014

Available online 15 December 2014

Keywords:

Active ester

Acyl fluoride

Acylating reagent

RNA

SHAPE chemistry

Symmetric anhydride

ABSTRACT

As chemical methods for RNA secondary structure determination, SHAPE chemistry (selective 2'-hydroxyl acylation analyzed by primer extension) has been developed to specifically target flexible nucleotides (often unpaired nucleotides) independently to their purine or pyrimidine nature. In order to improve the specificity of acylating reagents towards unpaired nucleotides, we have explored the reactivity of symmetric anhydrides, acyl fluorides, active esters like succinimidyl ester and cyanomethyl esters for 2'-O-acylation reaction. Among the tested compounds, only the acyl fluoride **4** showed a low reactivity (compared to NMIA). However, this study is the first to show that nucleophilic catalysts like DMAP greatly improved the selective 2'-hydroxyl acylation by symmetric anhydrides, acyl fluorides and succinimidyl ester, with the 2-fluorobenzoic anhydride **5** being the most reactive.

© 2014 Elsevier Ltd. All rights reserved.

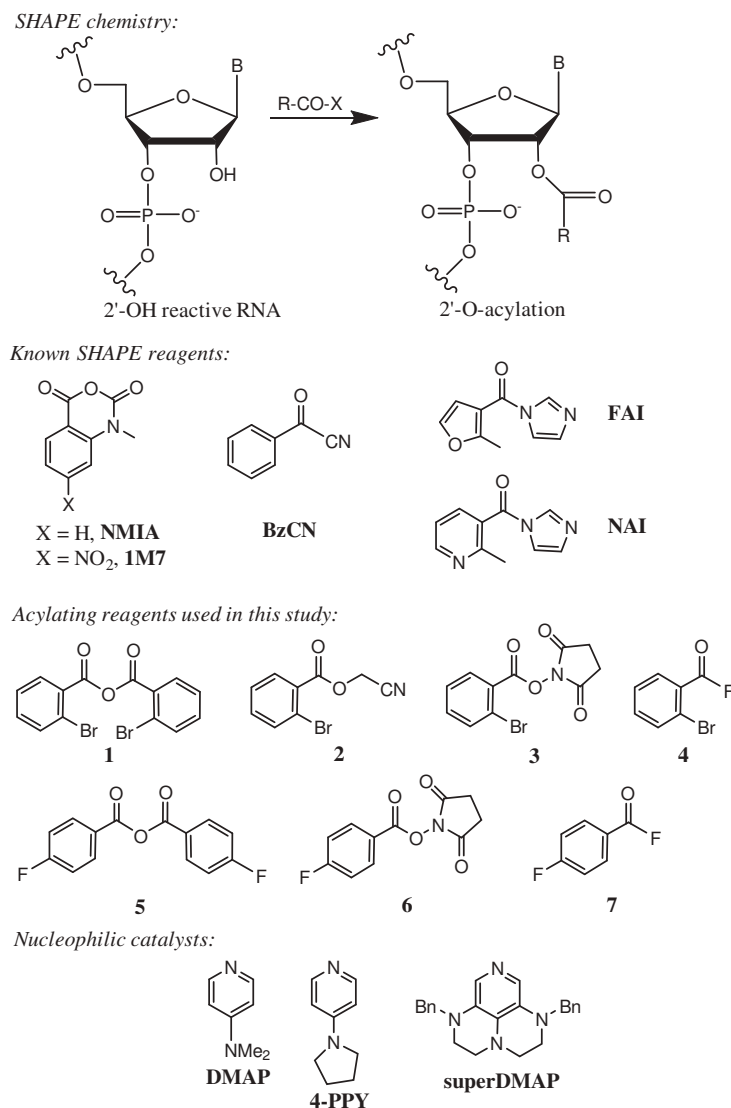
RNA is an essential macromolecule with diverse functions that mainly depend on its ability to fold into higher order structures. RNA serves as the genome for many pathogenic viruses. In addition to its role as messenger between DNA and the ribosome in bacteria and eukaryotes, RNA is involved in numerous fundamental cellular processes such as gene regulation and host defense. Note that many RNA functions are still unknown. This point can be illustrated by the recent discovery of thousands of long non-coding RNAs (lncRNAs, from 200 to 10,000 nucleotides in general) in mammalian species.¹ A study estimates approximately 15,000 lncRNAs in humans.² Most lncRNAs play key roles in development, epigenetics, cancer, brain function and hereditary diseases³ but fundamental questions regarding the functions of lncRNAs remain unanswered.

RNAs show a hierarchical organization in which the primary sequence determines the secondary structure (the pattern of base pairing) that governs the tertiary folding. RNA functions may imply sequence, secondary and tertiary structures, or a combination of these features. However, secondary structures, usually more stable than tertiary interactions, are usually considered as the key point for RNA functions. An extensive understanding of the functions

of RNAs in general and lncRNAs in particular requires an important effort to improve our knowledge of their structures, folding and dynamics, especially concerning their secondary structure. The crystallographic and NMR approaches have led to determine the three-dimensional structures of several functional RNAs but the high-throughput or applicability of these methods is limited, especially for long RNAs. An exciting way to map RNA secondary structures is the use of base-selective chemical reagents, an approach that has the potential to overcome the shortcomings associated to X-ray crystallography or NMR. Ideally, high-throughput determination of RNA secondary structure should rely on one chemical reagent able to interrogate all nucleotides simultaneously, in a single experiment. However, during the last thirty years, multiple reagents need to be used in separate experiments to obtain accurate results. Since 2005, selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry has been developed to specifically target flexible nucleotides (often unpaired nucleotides) independently to their purine or pyrimidine nature. This approach uses acylating reagents such as NMIA (N-methylisatoic anhydride),⁴ 1M7 (1-methyl-7-nitroisatoic anhydride)⁵ or BzCN (benzoyl cyanide)⁶ (Scheme 1). Recently, two other electrophilic reagents, FAI (2-methyl-3-furoic acid imidazolide) and NAI (2-methylnicotinic acid imidazolide) were developed to unravel RNA secondary structure inside living cells.⁷ The SHAPE chemistry is based on the acylation of the ribose 2'-hydroxyl group and its molecular mechanisms are only partially understood.^{4,8,9}

* Corresponding authors. Tel.: +33 1 47405999; fax: +33 1 47407671 (P.F.), tel.: +33 1 47405586; fax: +33 1 47402454 (J.X.).

E-mail addresses: philippe.fosse@ens-cachan.fr (P. Fossé), joanne.xie@ens-cachan.fr (J. Xie).



Scheme 1. Scheme and reagents for SHAPE chemistry.

Although SHAPE is potentially a powerful approach for determining secondary structure and dynamics of long RNAs at high-throughput, the results obtained by this approach remain ambiguous because the electrophilic reagents used are in fact only partially specific for unpaired nucleotides.¹⁰ Recent studies were performed by associating an acylating reagent to other chemical or enzymatic probes to improve the information content of experimental data.^{11–13} This latter approach is more accurate but much more time-consuming and therefore not compatible with high-throughput determination of RNA secondary structures in a single experiment.

For the acylation of hydroxyl group, the carboxyl components can be activated as acyl halides, acyl azides, acylimidazoles (as for FAI and NAI), anhydrides (like NMIA and 1M7), esters, etc., allowing the modulation of the carbonyl electrophilicity. Therefore, there is still plenty of room to find new efficient reagents. However, to date, only thirteen compounds have been studied for the 2'-hydroxyl acylation and only five (NMIA, 1M7, BzCN, FAI and NAI) can be used in RNA SHAPE analysis.^{4–7} Note that these latter reagents exhibit similar RNA 2'-hydroxyl acylation patterns.^{5,7}

Among the reported SHAPE reagents (Scheme 1), BzCN and 1M7 showed higher reactivity than NMIA, NAI and FAI, in accordance with their hydrolysis rate, with a half-life of 0.25 s, 14 s, 8 min, 33 min and 73 min, respectively. Since these reagents are partially specific for unpaired nucleotides, we therefore examined other acylating reagents to contribute to: (i) a better understanding of the SHAPE mechanisms; (ii) the design of a new acylating reagent more specific for unpaired nucleotides.

Considering the moderate acylating activity of NMIA, we have synthesized seven reagents (Scheme 1) that might exhibit a better acylating activity than NMIA. Symmetric anhydrides **1**¹⁴ and **5**¹⁵ bearing electron withdrawing halogen atom on *ortho* or *para* position, succinimidyl ester **3**,¹⁶ **6**¹⁷ as well as the cyanomethyl ester **2**¹⁸ have been prepared. Acyl fluorides, less reactive than acyl chlorides, have shown promising reactivity in coupling reaction.¹⁹ Acyl fluorides **4**²⁰ and **7**²¹ were also synthesized according to the reported procedure. However, due to the volatile nature, compound **7** was obtained as a solution in petroleum ether. It is therefore difficult to determine its concentration. Moreover, this compound is not stable and transforms slowly into the anhydride **5**.

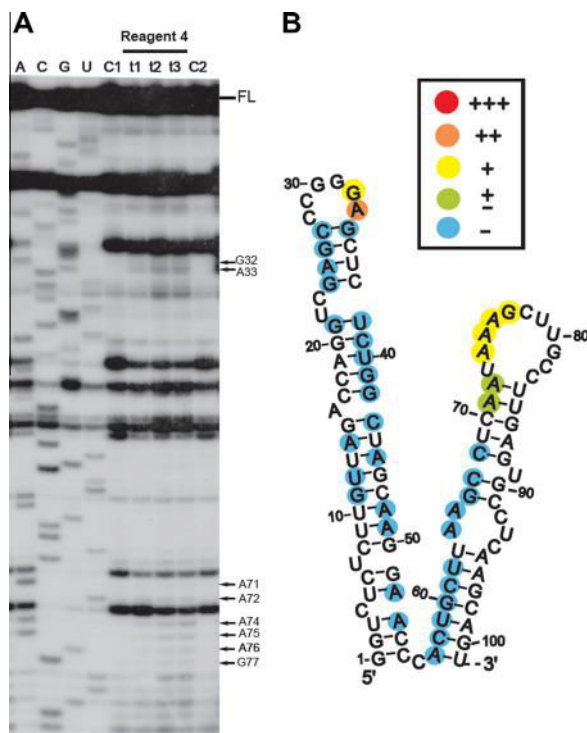


Figure 1. SHAPE analysis of the HIV-1 R region using reagent **4**. (A) SHAPE experiments were performed and analyzed by electrophoresis on a denaturing 8% polyacrylamide gel as described.²⁵ Control lanes C1 and C2: 5' RNA (2 pmol) was incubated without reagent at 37 °C for 1 min and 5 min, respectively. 5' RNA (2 pmol) was incubated with reagent **4** (50 nmol) at 37 °C for 1 min (lane t1), 5 min (lane t2) or 15 min (lane t3). A, C, G and U are sequencing reactions performed using dideoxy terminating nucleotides. Note that reverse transcriptase stops dNTP incorporation at the nucleotide preceding the site of acylation, that is, exactly 1 nucleotide shorter than the corresponding sequencing lane. FL indicates the full-length product of primer extension. Arrows indicate the acylated nucleotides. (B) Superposition of reagent **4** reactivity on the secondary structure of the R region. The colour code used for the reactivity is indicated in the inset. The absence of colour indicates nucleotides that are not analyzable because they are close to the 5' end or primer binding site or because of nicks and pauses in the RNA template during reverse transcription.

RNA stem-loop structures are good models to test the acylating activity of reagents because they contain paired and unpaired nucleotides. Since the secondary structure of the R region of the human immunodeficiency virus type 1 (HIV-1) forms two stem-loop structures,²² the acylating activity of compounds was tested with the 5' RNA containing this region.^{22,23} Sites of 2'-O-adduct formation have been identified using primer extension approach²⁴ and electrophoresis on a denaturing polyacrylamide gel.²⁵ No reactivity information is obtained at the level of: (i) nucleotides that bind the primer; (ii) 20–50 nucleotides that are located upstream of the primer binding site; (iii) nucleotides that are close to the 5'-end.⁶ Representative examples of SHAPE experiments are shown in Figures 1–5 and the results of a series of independent experiments are summarized in Table 1. 5' RNA was incubated with acylating reagents for 1, 5 and 15 min. Longer incubation (30, 60 or 120 min) did not lead to further acylation compared to 15 min reaction, due to their fast hydrolysis.

Among the synthesized acylating reagents, only the acyl fluoride **4** was able to modify the 5' RNA. Further experiment has been realized with compound **4** (Fig. 1A). However, the reagent **4** was less efficient than NMIA (Table 1). Interestingly, the modification pattern was consistent with the secondary structure of the R region

(Fig. 1B). Indeed, almost all paired bases were unreactive, whereas unpaired nucleotides, especially purine nucleotides, in the two apical loops were reactive. Consistent with the observation that nucleotides adjacent to a single-stranded, bulged or mismatched nucleotide can exhibit low reactivity towards SHAPE reagents,²⁴ A71 and A72 were very weakly reactive to **4**.

Employing nucleophilic catalysts like DMAP for the esterification of alcohols by acid anhydrides is a well-known procedure.²⁶ Activity of compounds **1–6** was therefore studied in the presence of 6% DMAP (Fig. 2 and Table 1). As shown for reagent **5** (Fig. 2A), RNA modification with reagents **1**, **3**, **5** and **6** occurred only in the presence of the catalyst (Table 1), while the compound **2** remained inactive which might be explained by the lower electrophilicity of the cyanomethyl ester. The activation order (**5** > **1**, **4** > **3**, **6**) by DMAP could be explained by the electrophilicity of the carbonyl to facilitate the nucleophilic addition of DMAP to the acyl donor (**1**, **4**, **5** > **3**, **6**) and the basicity of the leaving groups to deprotonate the alcohol nucleophile in the rate-determining transition state²⁶ (**5** > **1**, **4**, with $pK_a = 4.14$, 3.2 and 2.85 for 4-F-C₆H₄CO₂⁻/4-F-C₆H₄CO₂H, 2-Br-C₆H₄CO₂⁻/2-Br-C₆H₄CO₂H and F⁻/HF, respectively). Interestingly, the very low acylating activity of compound **6** was greatly increased in the presence of 50% DMAP (Fig. 3), with similar activity as NMIA.

Structurally constrained derivatives 4-PPY²⁷ and superDMAP²⁸ (Scheme 1) have been shown to be more efficient nucleophilic catalysts than DMAP. Activity of reagent **5** was therefore studied in the presence of these catalysts (Fig. 2A). RNA modification with reagent **5** was slightly less efficient in the presence of 6% 4-PPY than in the presence of 6% DMAP. In the presence of 0.6% superDMAP, the amount of full-length product was strongly decreased, indicating an excessive modification of RNA. The superDMAP proved therefore to be a powerful catalyst. Unfortunately, assays by decreasing the quantity of superDMAP failed to improve the selectivity.

Next, we compared the modification pattern of NMIA with those of reagents **1**, **4** and **5** (Fig. 4) that are our most active compounds. The patterns were similar, supporting the notion that the mechanism of action of our compounds is close to that of NMIA, that is, selective 2'-hydroxyl acylation. However, NMIA was more efficient than reagents **1**, **4** and **5** and it does not require the presence of 6% DMAP (Table 1). To confirm that reagent **5** was the most efficient synthesized compound whatever the RNA substrate, we performed SHAPE analysis of the HCV SL1(–) region using reagents **1** and **5** (Fig. 5). As expected, reagent **5** was more efficient than reagent **1** and it modified unpaired nucleotides in the apical loop and one bulge.

In summary, we have synthesized symmetric anhydrides, acyl fluorides, active esters like succinimidyl ester and cyanomethyl esters as aromatic acylating reagents to evaluate their reactivity towards RNA SHAPE chemistry. Only the acyl fluoride reagent **4** was able to modify the unpaired nucleotides, with a lower reactivity than NMIA. However, we have shown that nucleophilic catalyst like DMAP (6%) greatly improved the reactivity of acylating reagents, with the 2-fluorobenzoic anhydride **5** more reactive than 2-bromobenzoic anhydride **1** and 2-bromobenzoyl fluoride **4** which are more reactive than the succinimidyl esters **3** and **6**. It should be emphasized that the acylating activity of inactive compound **6** in the presence of 50% DMAP is similar to that of NMIA (compare Fig. 3 and Fig. 4). The order of DMAP activation could be explained by the electrophilicity of the carbonyl group as well as the basicity of the leaving group of the acylating reagents to deprotonate the alcohol nucleophile of RNA in the rate-limiting step. This result is interesting, opening the perspective to the future design of SHAPE reagents and to modulate its reactivity by nucleophilic catalysis.

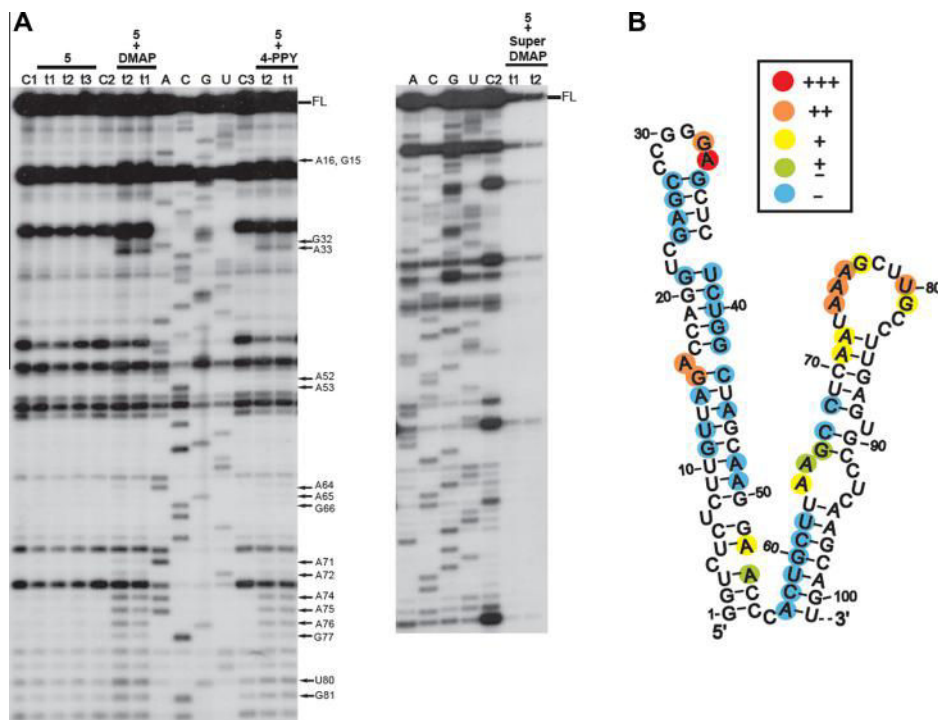


Figure 2. Nucleophilic catalysts are required for the acylation activity of compound **5**. (A) Control lanes C1, C2 and C3: 5' RNA (2 pmol) was incubated without reagent at 37 °C for 1 min, 5 min and 15 min, respectively. 5' RNA (2 pmol) was incubated with compound **5** (50 nmol) at 37 °C for 1 min (lane t1), 5 min (lane t2) or 15 min (t3) in the absence or presence of 6% DMAP, 6% 4-PPY or 0.6% superDMAP. A, C, G and U are sequencing reactions performed using dideoxy terminating nucleotides. A16 and G15 were well characterized by long migration times on a denaturing 8% polyacrylamide gel (data not shown). (B) Superposition of reagent **5** reactivity in the presence of 6% DMAP on the secondary structure of the R region.

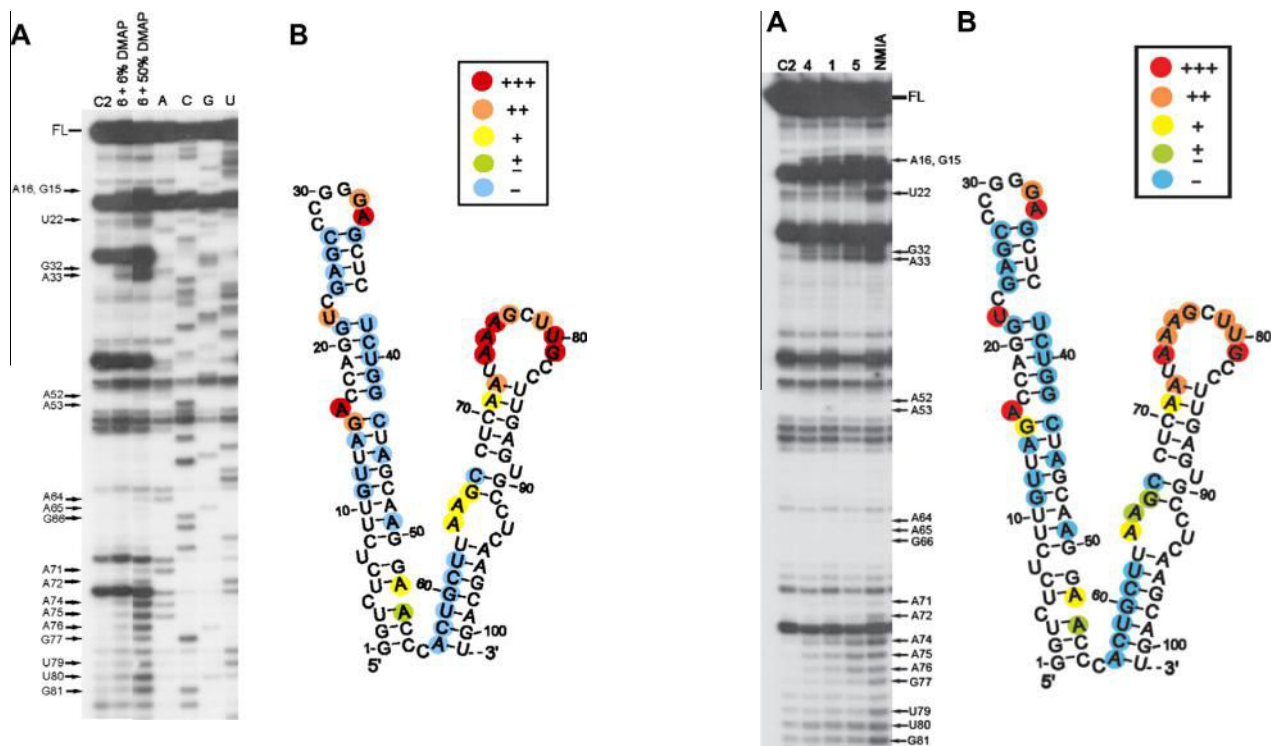


Figure 3. SHAPE analysis of the HIV-1 R region using reagent **6** in the presence of 6% or 50% DMAP. (A) Control lane C2: 5' RNA (2 pmol) was incubated without reagent at 37 °C for 5 min. 5' RNA (2 pmol) was incubated with compound **6** (50 nmol) at 37 °C for 5 min in the presence of 6% or 50% DMAP. (B) Superposition of reagent **6** reactivity in the presence of 50% DMAP on the secondary structure of the R region.

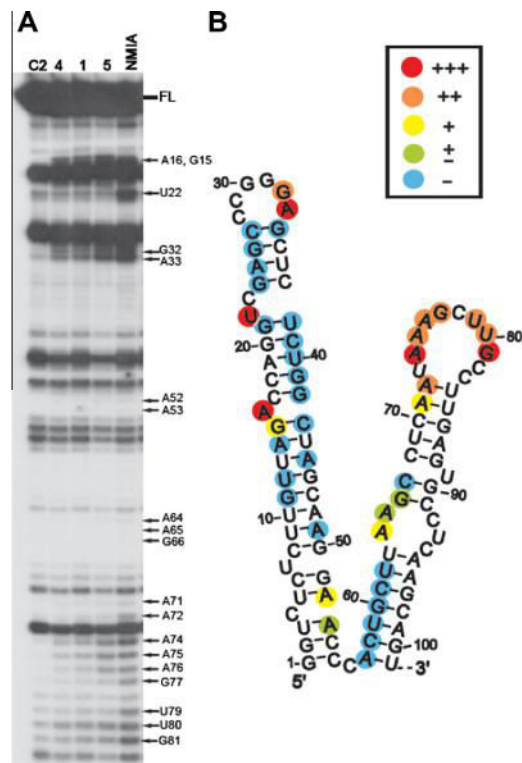


Figure 4. Comparison of SHAPE experiments performed using NMIA or reagents **1**, **4** and **5**. (A) Control lane C2: 5' RNA (2 pmol) was incubated without reagent at 37 °C for 5 min. 5' RNA (2 pmol) was incubated at 37 °C for 5 min in the presence of 6% DMAP and with NMIA (50 nmol) or compounds **1**, **4** and **5** (50 nmol). (B) Superposition of NMIA reactivity on the secondary structure of the R region.

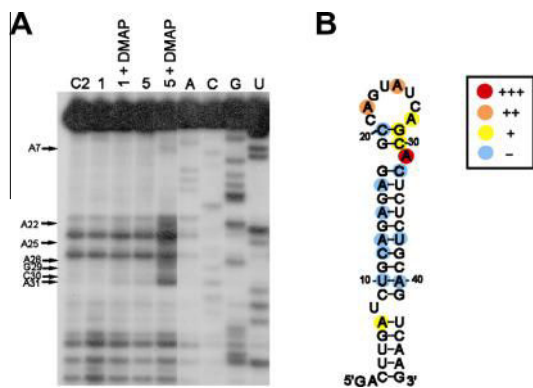


Figure 5. SHAPE analysis of the HCV SL1(–) region using reagents **1** and **5**. (A) Control lane C2: X(–) RNA (2 pmol) was incubated without reagent at 37 °C for 5 min. X(–) RNA (2 pmol) was incubated with compounds **1** or **5** (50 nmol) at 37 °C for 5 min in the absence or presence of 6% DMAP. A, C, G and U are sequencing reactions performed using dideoxy terminating nucleotides. (B) Superposition of reagent **5** reactivity on the secondary structure of the SL1(–) region.

Table 1
RNA modification activity of compounds **1–6** (50 nmol) in the absence or presence of 6% DMAP

Compound	1	2	3	4	5	6	NMIA
–DMAP	–	–	–	+	–	–	++++
+DMAP	++	–	+	++	+++	+	++++

References and notes

- Xie, C.; Yuan, J.; Li, H.; Li, M.; Zhao, G.; Bu, D.; Zhu, W.; Wu, W.; Chen, R.; Zhao, Y. *Nucleic Acids Res.* **2014**, *42*, D98.
- Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D. G.; Lagarde, J.; Veeravalli, L.; Ruan, X.; Ruan, Y.; Lassmann, T.; Carninci, P.; Brown, J. B.; Lipovich, L.; Gonzalez, J. M.; Thomas, M.; Davis, C. A.; Shiekhattar, R.; Gingeras, T. R.; Hubbard, T. J.; Notredame, C.; Harrow, J.; Guigó, R. *Genome Res.* **2012**, *9*, 1775.
- Novikova, I. V.; Hennelly, S. P.; Tung, C. S.; Sanbonmatsu, K. Y. *J. Mol. Biol.* **2013**, *425*, 3731.
- Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223.
- Mortimer, S. A.; Weeks, K. M. *J. Am. Chem. Soc.* **2007**, *129*, 4144.
- Mortimer, S. A.; Weeks, K. M. *Nat. Protoc.* **2009**, *4*, 1413.
- Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y. *Nat. Chem. Biol.* **2013**, *9*, 18.
- McGinnis, J. L.; Dunkle, J. A.; Cate, J. H.; Weeks, K. M. *J. Am. Chem. Soc.* **2012**, *134*, 6617.
- Hiratsuka, H. *J. Biol. Chem.* **1982**, *257*, 13354.
- Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. *Biochemistry* **2011**, *50*, 804.
- Novikova, I. V.; Hennelly, S. P.; Sanbonmatsu, K. Y. *Nucleic Acids Res.* **2012**, *40*, 5034.
- Lusvardi, S.; Sztuba-Solinska, J.; Purzycka, K. J.; Pauly, G. T.; Rausch, J. W.; Grice, S. F. *Nucleic Acids Res.* **2013**, *41*, 6637.

- Todd, G. C.; Walter, N. G. *RNA* **2013**, *19*, 685.
- Kita, Y.; Akai, S.; Yogishi, M.; Nakajima, Y.; Yasuda, H.; Tamura, Y. *Tetrahedron Lett.* **1984**, *25*, 6027.
- Li, Y.; Xue, D.; Wang, C.; Liu, Z. T.; Xiao, J. *Chem. Commun.* **2012**, 1320.
- Schulze, A.; Giannis, A. *Adv. Synth. Catal.* **2004**, *346*, 252.
- Wang, G.; Yu, Q. Y.; Wang, J.; Wang, S.; Chan, S. Y.; Yu, X. Q. *RSC Adv.* **2013**, *3*, 21306.
- Setzer, P.; Forcher, G.; Boeda, F.; Pearson-Long, M. S. M.; Bertus, P. *Eur. J. Org. Chem.* **2014**, 171.
- Montalbetti, C. A. G. N.; Falque, V. *Tetrahedron* **2005**, *61*, 10827.
- Olah, G. A.; Welch, J. J. *Am. Chem. Soc.* **1978**, *100*, 5396.
- Birell, J. A.; Desrosiers, J. N.; Jacobsen, E. N. *J. Am. Chem. Soc.* **2011**, *133*, 13872.
- Kanevsky, I.; Chaminade, F.; Fichoux, D.; Moumen, A.; Gorelick, R.; Negroni, M.; Darlix, J. L.; Fossé, P. *J. Mol. Biol.* **2005**, *348*, 1059.
- In vitro RNA synthesis, purification and renaturation: The 5' RNA and the wild-type X(–) RNA were synthesized and purified as previously described (Ref. 22 and Ivanyi-Nagy, R.; Kanevsky, I.; Gabus, C.; Laverne, J. P.; Fichoux, D.; Penin, F.; Fossé, P.; Darlix, J.-L. *Nucleic Acids Res.* **2006**, *34*, 2618.). RNA (2.0 pmol) in 7.8 µL of double-distilled water was heated at 90 °C for 2 min and chilled on ice. Then, 2 µL of renaturation buffer (final concentrations: 50 mM HEPES-KOH (pH 8), 200 mM AcOK and 5 mM MgCl₂), was added and the sample was incubated for 30 min at 37 °C before to be chilled on ice for 5 min. At the end of this renaturation step, 0.20 µL of a tRNA mixture (5 µg/µL) from *E. coli* MRE 600 (Roche Molecular Biochemicals) was added to the sample before the addition of the acylating reagent.
- Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. *Nat. Protoc.* **2006**, *1*, 1610.
- SHAPE analysis: Anhydrous DMSO, NMIA and Et₃N were purchased from Sigma–Aldrich. All solutions were freshly prepared. RNA modification was carried out in a final volume of 12 µL. 1.5 µL of renaturation buffer was added to the renatured RNA (2 pmol in 10 µL) before incubation at 37 °C for 1, 5 or 15 min in the presence of 0.5 µL of neat DMSO (control) or in the presence of 0.5 µL of acylating reagent (100 mM in DMSO). One microliter of renaturation buffer was added to the renatured RNA (10 µL) when the reaction was performed at 37 °C for 1 or 5 min in the presence of 0.5 µL of acylating reagent (100 mM in DMSO) and 0.5 µL of DMAP or 4-PPY (6 mM in DMSO) or 0.5 µL of superDMAP (0.6 mM in DMSO). In the case of RNA modification with 0.5 µL of compound **5** (100 mM in DMSO), 0.5 µL of Et₃N (100 mM in DMSO) was added to the reaction mixture in the absence or in the presence of a catalyst. At the end of incubation, RNA was recovered by ethanol precipitation and the dried pellet was resuspended in 6.5 µL of double-distilled water. Oligodeoxyribonucleotides complementary to nucleotides 132–117 of 5' RNA, and 99–83 of the wild-type X(–) RNA were used as primers and were 5'-end-labeled using T4 polynucleotide kinase (New England Biolabs) and [γ -³²P] ATP (Perkin Elmer). The acylated nucleotides were identified by primer extension with avian myeloblastosis reverse transcriptase (RT) (Invitrogen) as described below. One microliter of the 5'-end-labeled primer (10⁵ cpm/µL) was added to the resuspended RNA and the mixture was heated at 90 °C for 2 min, chilled for 15 s in liquid nitrogen and defrosted at room temperature. Then, 7.5 µL of the reverse transcription mixture (final concentrations: 50 mM Tris-HCl (pH 8.5), 8 mM MgCl₂, 30 mM KCl, 1 mM DTT, 250 mM dNTP, 0.13 U/µL of RT) was added and the reaction mixture was incubated at 37 °C for 30 min. At this end of incubation, the RNA template was degraded by addition of 10 µL of 0.3 M NaOH–0.05 M EDTA and heating at 90 °C for 10 min. DNA was recovered by ethanol precipitation and the dried pellet was resuspended in 7 µL of loading buffer (98% (v/v) deionized formamide, 10 mM EDTA, 0.025% (w/v) bromophenol blue and 0.025% (w/v) xylene cyanol). Primer extension of controls that were not treated with an acylating reagent was run in parallel. The samples were analysed by electrophoresis on an 8% denaturing polyacrylamide gel.
- Spivey, A. C.; Arseniyadis, S. *Angew. Chem., Int. Ed.* **2004**, *43*, 5436.
- Hassner, A.; Krepski, L. R.; Alexanian, V. *Tetrahedron* **1978**, *34*, 2069.
- De Rycke, N.; Couty, F.; David, O. R. P. *Chem. Eur. J.* **2011**, *17*, 12852.

Références Bibliographiques

1. **Meiering CD, Linial ML.** 2001. Historical Perspective of Foamy Virus Epidemiology and Infection. *Clin Microbiol Rev* 14:165–176.
2. **Griffiths DJ.** 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol* 2.
3. **Murphy FA.** 1995. On target: “Virus Taxonomy. Classification and Nomenclature of Viruses — Sixth Report of the International Committee on Taxonomy of Viruses.” *Arch Virol* 140:391–392.
4. **Jern P, Sperber GO, Blomberg J.** 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* 2:50.
5. **Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220:868–871.
6. **Clavel F, Brun-Vézinet F, Guétard D, Chamaret S, Laurent A, Rouzioux C, Rey M, Katlama C, Rey F, Champelinaud JL, et. al.** 1986. [LAV type II: a second retrovirus associated with AIDS in West Africa]. *Comptes Rendus Académie Sci Sér III Sci Vie* 302:485–488.
7. **Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey MA, Santos-Ferreira MO, Laurent AG, Dauguet C, Katlama C, Rouzioux C, Al E.** 1986. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 233:343–346.
8. **Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH.** 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes* troglodytes. *Nature* 397:436–441.
9. **Rambaut A, Posada D, Crandall KA, Holmes EC.** 2004. The causes and consequences of HIV evolution. *Nat Rev Genet* 5:52–61.
10. **Chen Z, Luckay A, Sodora DL, Telfer P, Reed P, Gettie A, Kanu JM, Sadek RF, Yee J, Ho DD, Zhang L, Marx PA.** 1997. Human immunodeficiency virus type 2 (HIV-2) seroprevalence and characterization of a distinct HIV-2 genetic subtype from the natural range of simian immunodeficiency virus-infected sooty mangabeys. *J Virol* 71:3953–3960.
11. **Plantier J-C, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F.** 2009. A new human immunodeficiency virus derived from gorillas. *Nat Med* 15:871–872.
12. **Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, Fruteau C, Noë R, Peeters M, Brookfield JFY, Shaw GM, Sharp PM, Hahn BH.** 2005. Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (*Cercocebus atys atys*) from the Taï Forest, Côte d’Ivoire: Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2. *J Virol* 79:12515–12527.
13. **Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, Mbanya D, Kaptué L, Ndemi N, Gürtler L, Devare S, Brennan CA.** 2011. Confirmation of Putative HIV-1 Group P in Cameroon. *J Virol* 85:1403–1407.
14. **Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, Ngolle EM, Sharp PM, Shaw GM, Delaporte E, Hahn BH, Peeters M.** 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* 444:164.
15. **Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J, Posada D, Peeters M, Pybus OG, Lemey P.** 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346:56–61.
16. **Chen Z, Telfer P, Gettie A, Reed P, Zhang L, Ho DD, Marx PA.** 1996. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J Virol* 70:3617–3627.
17. **MacNeil A, Sarr AD, Sankalé J-L, Meloni ST, Mboup S, Kanki P.** 2007. Direct Evidence of Lower Viral Replication Rates In Vivo in Human Immunodeficiency Virus Type 2 (HIV-2) Infection than in HIV-1 Infection. *J Virol* 81:5325–5330.
18. **Kannangai R, Sridharan G, David S.** 2012. Human immunodeficiency virus type-2-A milder, kinder virus: An update. *Indian J Med Microbiol* 30:6.
19. **Hemelaar J.** 2012. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* 18:182–192.
20. **Fauci A, Desrosiers R.** 1997. Pathogenesis of HIV and SIV. *Retroviruses*.

Références Bibliographiques

21. **Briggs JAG, Wilk T, Welker R, Kräusslich H-G, Fuller SD.** 2003. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J* 22:1707–1715.
22. **Ganser-Pornillos BK, Yeager M, Sundquist WI.** 2008. The structural biology of HIV assembly. *Curr Opin Struct Biol* 18:203–217.
23. **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* 11:672–675.
24. **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* 113:11537–11542.
25. **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* 35:2695–2704.
26. **Feng S, Holland EC.** 1988. HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* 334:165–167.
27. **Jackson PE, Tebit DM, Rekosh D, Hammarskjöld M-L.** 2016. Rev-RRE Functional Activity Differs Substantially Among Primary HIV-1 Isolates. *AIDS Res Hum Retroviruses* 32:923–934.
28. **Mailler E, Bernacchi S, Marquet R, Paillart J-C, Vivet-Boudou V, Smyth RP.** 2016. The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses* 8:1–19.
29. **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 BA, Salemi M, Telesnitsky A, Summers MF.** 2015. RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* 348:917–921.
30. **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.
31. **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* 12:866–872.
32. **Foley B, Apetrei C, Mizrahi I, Rambaut A, Korber B, Leitner T, Hahn B, Mullins J, Wolinsky S.** 2016. HIV Sequence Compendium 2016.
33. **Garcia-Miranda P, Becker JT, Benner BE, Blume A, Sherer NM, Butcher SE.** 2016. Stability of HIV Frameshift Site RNA Correlates with Frameshift Efficiency and Decreased Virus Infectivity. *J Virol* 90:6906–6917.
34. **Schwartz S, Campbell M, Nasioulas G, Harrison J, Felber BK, Pavlakis GN.** 1992. Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent gag expression. *J Virol* 66:7176–7182.
35. **Wolff H, Brack-Werner R, Neumann M, Werner T, Schneider R.** 2003. Integrated functional and bioinformatics approach for the identification and experimental verification of RNA signals: application to HIV-1 INS. *Nucleic Acids Res* 31:2839–2851.
36. **Lobritz MA, Ratcliff AN, Arts EJ.** 2010. HIV-1 Entry, Inhibitors, and Resistance. *Viruses* 2:1069–1105.
37. **Miyauchi K, Kim Y, Latinovic O, Morozov V, Melikyan GB.** 2009. HIV Enters Cells via Endocytosis and Dynamin-Dependent Fusion with Endosomes. *Cell* 137:433–444.
38. **Zheng Y-H, Lovsin N, Peterlin BM.** 2005. Newly identified host factors modulate HIV replication. *Immunol Lett* 97:225–234.
39. **Wyatt R, Sodroski J.** 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280:1884–1888.
40. **Hiscott J, Kwon H, Génin P.** 2001. Hostile takeovers: viral appropriation of the NF- κ B pathway. *J Clin Invest* 107:143–151.
41. **Cary DC, Fujinaga K, Peterlin BM.** 2016. Molecular mechanisms of HIV latency. *J Clin Invest* 126:448–454.
42. **Maldarelli F.** 2016. The role of HIV integration in viral persistence: no more whistling past the proviral graveyard. *J Clin Invest* 126:438–447.

Références Bibliographiques

43. **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* 40:10345–10355.
44. **Valente ST, Gilmartin GM, Venkataraman K, Arriagada G, Goff SP.** 2009. HIV-1 mRNA 3' End Processing Is Distinctively Regulated by eIF3f, CDK11, and Splice Factor 9G8. *Mol Cell* 36:279–289.
45. **Gilmartin GM, Fleming ES, Oetjen J, Graveley BR.** 1995. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev* 9:72–83.
46. **Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ.** 2008. Transcription-Dependent Gene Looping of the HIV-1 Provirus Is Dictated by Recognition of Pre-mRNA Processing Signals. *Mol Cell* 29:56–68.
47. **McLaren M, Marsh K, Cochrane A.** 2008. Modulating HIV-1 RNA processing and utilization. *Front Biosci J Virtual Libr* 13:5693–5707.
48. **Berkhout B, Silverman RH, Jeang K-T.** 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* 59:273–282.
49. **Das AT, Harwig A, Berkhout B.** 2011. The HIV-1 Tat Protein Has a Versatile Role in Activating Viral Transcription. *J Virol* 85:9506–9516.
50. **Li G, Piamongsant S, Faria NR, Voet A, Pineda-Peña A-C, Khouri R, Lemey P, Vandamme A-M, Theys K.** 2015. An integrated map of HIV genome-wide variation from a population perspective. *Retrovirology* 12:18–18.
51. **Purcell DF, Martin MA.** 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* 67:6365–6378.
52. **Taniguchi I, Mabuchi N, Ohno M.** 2014. HIV-1 Rev protein specifies the viral RNA export pathway by suppressing TAP/NXF1 recruitment. *Nucleic Acids Res* 42:6645–6658.
53. **Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C.** 2010. Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. *FEBS J* 277:867–876.
54. **Chen J, Nikolaitchik O, Singh J, Wright A, Bencsics CE, Coffin JM, Ni N, Lockett S, Pathak VK, Hu W-S.** 2009. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci* 106:13535–13540.
55. **Paillart J-C, Shehu-Xhilaga M, Marquet R, Mak J.** 2004. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* 2:461–472.
56. **Olson ED, Cantara WA, Musier-Forsyth K.** 2015. New Structure Sheds Light on Selective HIV-1 Genomic RNA Packaging. *Viruses* 7:4826–4835.
57. **Berkowitz R, Fisher J, Goff SP.** 1996. RNA packaging. *Curr Top Microbiol Immunol* 214:177–218.
58. **Muriaux D, Mirro J, Nagashima K, Harvin D, Rein A.** 2002. Murine Leukemia Virus Nucleocapsid Mutant Particles Lacking Viral RNA Encapsidate Ribosomes. *J Virol* 76:11405–11413.
59. **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* 460:711–716.
60. **Didierlaurent L, Racine PJ, Houzet L, Chamontin C, Berkhout B, Mougél 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* 39:8915–8927.
61. **Das AT, Vrolijk MM, Harwig A, Berkhout B.** 2012. Opening of the TAR hairpin in the HIV-1 genome causes aberrant RNA dimerization and packaging. *Retrovirology* 9:59–59.
62. **Kuzembayeva M, Dilley K, Sardo L, Hu W-S.** 2014. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* 454–455:362–370.
63. **Lu K, Heng X, Summers MF.** 2011. Structural Determinants and Mechanism of HIV-1 Genome Packaging. *J Mol Biol* 410:609–633.
64. **Levin JG, Guo J, Rouzina Ioulia, Musier-Forsyth K.** 2005. Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. *Prog Nucleic Acid Res Mol Biol* 80:217–286.
65. **Paillart J-C, Dettenhofer M, Yu X, Ehresmann C, Ehresmann B, Marquet R.** 2004. First Snapshots of the HIV-1 RNA Structure in Infected Cells and in Virions. *J Biol Chem* 279:48397–48403.

Références Bibliographiques

66. **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* 303:152–163.
67. **Heng X, Kharytonchyk S, Garcia EL, Lu K, Divakaruni SS, LaCotti C, Edme K, Telesnitsky A, Summers MF.** 2012. Identification of a Minimal Region of the HIV-1 5'-Leader Required for RNA Dimerization, NC Binding, and Packaging. *J Mol Biol* 417:224–239.
68. **Deforges J, Chamond N, Sargueil B.** 2012. Structural investigation of HIV-1 genomic RNA dimerization process reveals a role for the Major Splice-site Donor stem loop. *Biochimie* 94:1481–1489.
69. **Muriaux D, Mirro J, Harvin D, Rein A.** 2001. RNA is a structural element in retrovirus particles. *Proc Natl Acad Sci* 98:5246–5251.
70. **Martin-Serrano J, Neil SJD.** 2011. Host factors involved in retroviral budding and release. *Nat Rev Microbiol* 9:519–531.
71. **Demirov DG, Freed EO.** 2004. Retrovirus budding. *Virus Res* 106:87–102.
72. **Strack B, Calistri A, Craig S, Popova E, Göttlinger HG.** 2003. AIP1/ALIX Is a Binding Partner for HIV-1 p6 and EIAV p9 Functioning in Virus Budding. *Cell* 114:689–699.
73. **von Schwedler UK, Stuchell M, Müller B, Ward DM, Chung H-Y, Morita E, Wang HE, Davis T, He G-P, Cimbara DM, Scott A, Kräusslich H-G, Kaplan J, Morham SG, Sundquist WI.** 2003. The Protein Network of HIV Budding. *Cell* 114:701–713.
74. **Hellmund C, Lever AML.** 2016. Coordination of Genomic RNA Packaging with Viral Assembly in HIV-1. *Viruses* 8:1–13.
75. **Mattei S, Schur FK, Briggs JA.** 2016. Retrovirus maturation — an extraordinary structural transformation. *Curr Opin Virol* 18:27–35.
76. **Vogt V.** 1997. Retroviral Virions and Genomes. *Retroviruses*.
77. **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* 68:5013–5018.
78. **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 polypeptide, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology* 407:225–236.
79. **Bukrinskaya AG.** 2004. HIV-1 assembly and maturation. *Arch Virol* 149:1067–1082.
80. **Jiang J, Aiken C.** 2006. Maturation of the viral core enhances the fusion of HIV-1 particles with primary human T cells and monocyte-derived macrophages. *Virology* 346:460–468.
81. **Peng C, Ho BK, Chang TW, Chang NT.** 1989. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J Virol* 63:2550–2556.
82. **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* 68:6782–6786.
83. **Chertova E, Chertov O, Coren LV, 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* 80:9039–9052.
84. **Pettit SC, Moody MD, Wehbie RS, Kaplan AH, Nantermet PV, 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* 68:8017–8027.
85. **Briggs JAG, Kräusslich H-G.** 2011. The Molecular Architecture of HIV. *J Mol Biol* 410:491–500.
86. **Cruceanu M, Urbaneja MA, Hixson CV, 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* 34:593–605.
87. **Kessl JJ, Kutluay SB, Townsend D, Rebersburg 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* 166:1257–1268.e12.
88. **Engelman A, Cherepanov P.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol* 10:279–290.

Références Bibliographiques

89. **Nisole S, Stoye JP, Saïb A.** 2005. TRIM family proteins: retroviral restriction and antiviral defence. *Nat Rev Microbiol* 3:799–808.
90. **Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J.** 2004. The cytoplasmic body component TRIM5 α restricts HIV-1 infection in Old World monkeys. *Nature* 427:848–853.
91. **Sayah DM, Sokolskaja E, Berthoux L, Luban J.** 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 430:569–573.
92. **Nakayama EE, Shioda T.** 2012. TRIM5 α and Species Tropism of HIV/SIV. *Front Microbiol* 3:13.
93. **Richardson MW, Guo L, Xin F, Yang X, Riley JL.** 2014. Stabilized human TRIM5 α protects human T cells from HIV-1 infection. *Mol Ther J Am Soc Gene Ther* 22:1084–1095.
94. **Jia X, Zhao Q, Xiong Y.** 2015. HIV suppression by host restriction factors and viral immune evasion. *Curr Opin Struct Biol* 31:106–114.
95. **Langelier CR, Sandrin V, Eckert DM, Christensen DE, Chandrasekaran V, Alam SL, Aiken C, Olsen JC, Kar AK, Sodroski JG, Sundquist WI.** 2008. Biochemical Characterization of a Recombinant TRIM5 α Protein That Restricts Human Immunodeficiency Virus Type 1 Replication. *J Virol* 82:11682–11694.
96. **Li X, Sodroski J.** 2008. The TRIM5 α B-Box 2 Domain Promotes Cooperative Binding to the Retroviral Capsid by Mediating Higher-Order Self-Association. *J Virol* 82:11495–11502.
97. **Mrosek M, Meier S, Ucurum-Fotiadis Z, von Castelmur E, Hedbom E, Lustig A, Grzesiek S, Labeit D, Labeit S, Mayans O.** 2008. Structural Analysis of B-Box 2 from MuRF1: Identification of a Novel Self-Association Pattern in a RING-like Fold. *Biochemistry* 47:10722–10730.
98. **Nepveu-Traversy M-E, Bérubé J, Berthoux L.** 2009. TRIM5 α and TRIMCyp form apparent hexamers and their multimeric state is not affected by exposure to restriction-sensitive viruses or by treatment with pharmacological inhibitors. *Retrovirology* 6:100–100.
99. **Sastri J, Campbell EM.** 2011. Recent Insights into the Mechanism and Consequences of TRIM5 α Retroviral Restriction. *AIDS Res Hum Retroviruses* 27:231–238.
100. **Turelli P, Mangeat B, Jost S, Vianin S, Trono D.** 2004. Inhibition of hepatitis B virus replication by APOBEC3G. *Science* 303:1829.
101. **Sheehy AM, Gaddis NC, Malim MH.** 2003. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* 9:1404–1407.
102. **Sheehy AM, Gaddis NC, Choi JD, Malim MH.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418:646–650.
103. **Iwatani Y, Chan DSB, Wang F, Maynard KS, Sugiura W, Gronenborn AM, Rouzina I, Williams MC, Musier-Forsyth K, Levin JG.** 2007. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res* 35:7096–7108.
104. **Chaurasiya KR, McCauley MJ, Wang W, Qualley DF, Wu T, Kitamura S, Geertsema H, Chan DSB, Hertz A, Iwatani Y, Levin JG, Musier-Forsyth K, Rouzina I, Williams MC.** 2014. Oligomerization transforms human APOBEC3G from an efficient enzyme to a slowly dissociating nucleic acid-binding protein. *Nat Chem* 6:28–33.
105. **Kobayashi M, Takaori-Kondo A, Miyauchi Y, Iwai K, Uchiyama T.** 2005. Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. *J Biol Chem* 280:18573–18578.
106. **An P, Winkler CA.** 2010. Host genes associated with HIV/AIDS: advances in gene discovery. *Trends Genet* 26:119–131.
107. **Jin X, Wu H, Smith H.** 2007. APOBEC3G levels predict rates of progression to AIDS. *Retrovirology* 4:20.
108. **Biasin M, Piacentini L, Lo Caputo S, Kanari Y, Magri G, Trabattoni D, Naddeo V, Lopalco L, Clivio A, Cesana E, Fasano F, Bergamaschi C, Mazzotta F, Miyazawa M, Clerici M.** 2007. Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G: a possible role in the resistance to HIV of HIV-exposed seronegative individuals. *J Infect Dis* 195:960–964.
109. **Gillick K, Pollpeter D, Phalora P, Kim E-Y, Wolinsky SM, Malim MH.** 2013. Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination. *J Virol* 87:1508–1517.

Références Bibliographiques

110. **Wang Y, Kinlock BL, Shao Q, Turner TM, Liu B.** 2014. HIV-1 Vif inhibits G to A hypermutations catalyzed by virus-encapsidated APOBEC3G to maintain HIV-1 infectivity. *Retrovirology* 11:89–89.
111. **Wissing S, Galloway NLK, Greene WC.** 2010. HIV-1 Vif versus the APOBEC3 cytidine deaminases: An intracellular duel between pathogen and host restriction factors. *Mol Aspects Med* 31:383–397.
112. **Goldstone DC, Ennis-Adeniran V, Hedden JJ, Groom HCT, Rice GI, Christodoulou E, Walker PA, Kelly G, Haire LF, Yap MW, de Carvalho LPS, Stoye JP, Crow YJ, Taylor IA, Webb M.** 2011. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* 480:379–382.
113. **Lahouassa H, Daddacha W, Hofmann H, Ayinde D, Logue EC, Dragin L, Bloch N, Maudet C, Bertrand M, Gramberg T, Pancino G, Priet S, Canard B, Laguette N, Benkirane M, Transy C, Landau NR, Kim B, Margottin-Goguet F.** 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat Immunol* 13:223–228.
114. **Ahn J.** 2016. Functional organization of human SAMHD1 and mechanisms of HIV-1 restriction. *Biol Chem* 397:373–379.
115. **Votteler J, Sundquist WI.** 2013. Virus budding and the ESCRT pathway. *Cell Host Microbe* 14:232–241.
116. **Garrus JE, von Schwedler UK, Pornillos OW, Morham SG, Zavitz KH, Wang HE, Wettstein DA, Stray KM, Côté M, Rich RL, Myszka DG, Sundquist WI.** 2001. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107:55–65.
117. **Morita E, Sundquist WI.** 2004. Retrovirus budding. *Annu Rev Cell Dev Biol* 20:395–425.
118. **Demirov DG, Ono A, Orenstein JM, Freed EO.** 2002. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proc Natl Acad Sci* 99:955–960.
119. **Bendjennat M, Saffarian S.** 2016. The Race against Protease Activation Defines the Role of ESCRTs in HIV Budding. *PLoS Pathog* 14:1–25.
120. **Bagaya BS, Vega JF, Tian M, Nickel GC, Li Y, Krebs KC, Arts EJ, Gao Y.** 2015. Functional bottlenecks for generation of HIV-1 intersubtype Env recombinants. *Retrovirology* 12:44–44.
121. **Ward AB, Wilson IA.** 2015. Insights into the trimeric HIV-1 envelope glycoprotein structure. *Trends Biochem Sci* 40:101–107.
122. **Hwang SS, Boyle TJ, Lysterly HK, Cullen BR.** 1992. Identification of envelope V3 loop as the major determinant of CD4 neutralization sensitivity of HIV-1. *Science* 257:535–537.
123. **Kieslich CA, Tamamis P, Guzman YA, Onel M, Floudas CA.** 2016. Highly Accurate Structure-Based Prediction of HIV-1 Coreceptor Usage Suggests Intermolecular Interactions Driving Tropism. *PLoS ONE* 11:1–15.
124. **Conte MR, Matthews S.** 1998. Retroviral Matrix Proteins: A Structural Perspective. *Virology* 246:191–198.
125. **Hill CP, Worthylake D, Bancroft DP, Christensen AM, Sundquist WI.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc Natl Acad Sci* 93:3099–3104.
126. **Mercredi PY, Bucca N, Loeliger B, Gaines CR, Mehta M, Bhargava P, Tedbury PR, Charlier L, Floquet N, Muriaux D, Favard C, Sanders CR, Freed EO, Marchant J, Summers MF.** 2016. Structural and Molecular Determinants of Membrane Binding by the HIV-1 Matrix Protein. *J Mol Biol* 428:1637–1655.
127. **Tedbury PR, Freed EO.** 2014. The role of matrix in HIV-1 envelope glycoprotein incorporation. *Trends Microbiol* 22:372–378.
128. **Perilla JR, Gronenborn AM.** 2016. Molecular Architecture of the Retroviral Capsid. *Trends Biochem Sci* 41:410–420.
129. **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* 87:1285–1294.
130. **Liu C, Perilla JR, Ning J, Lu M, Hou G, Ramalho R, Himes BA, Zhao G, Bedwell GJ, Byeon I-J, Ahn J, Gronenborn AM, Prevelige PE, Rousso I, Aiken C, Polenova T, Schulten K, Zhang P.** 2016.

Références Bibliographiques

- Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat Commun* 7:10714.
131. **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.
132. **Godet J, Mély Y.** 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol* 7:687–699.
133. **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* 7:724–734.
134. **Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K.** 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* 7:754–774.
135. **Pachulska-Wieczorek K, Stefaniak AK, Purzycka KJ.** 2014. Similarities and differences in the nucleic acid chaperone activity of HIV-2 and HIV-1 nucleocapsid proteins in vitro. *Retrovirology* 11:54–54.
136. **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* 42:2525–2537.
137. **Tsuchihashi Z, Brown PO.** 1994. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* 68:5863–5870.
138. **Darlix J-L, Godet J, Ivanyi-Nagy R, Fossé P, Mauffret O, Mély Y.** 2011. Flexible Nature and Specific Functions of the HIV-1 Nucleocapsid Protein. *J Mol Biol* 410:565–581.
139. **Belfetmi A, Zargarian L, Tisné C, Sleiman D, Morellet N, Lescop E, Maskri O, René B, Mély Y, Fossé P, Mauffret O.** 2016. Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. *RNA* 22:506–517.
140. **Pontius BW, Berg P.** 1991. Rapid renaturation of complementary DNA strands mediated by cationic detergents: a role for high-probability binding domains in enhancing the kinetics of molecular assembly processes. *Proc Natl Acad Sci* 88:8237–8241.
141. **Pontius BW, Berg P.** 1990. Renaturation of complementary DNA strands mediated by purified mammalian heterogeneous nuclear ribonucleoprotein A1 protein: implications for a mechanism for rapid molecular assembly. *Proc Natl Acad Sci* 87:8403–8407.
142. **Koculi E, Lee N-K, Thirumalai D, Woodson SA.** 2004. Folding of the Tetrahymena Ribozyme by Polyamines: Importance of Counterion Valence and Size. *J Mol Biol* 341:27–36.
143. **Heilman-Miller SL, Thirumalai D, Woodson SA.** 2001. Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations1. *J Mol Biol* 306:1157–1166.
144. **Tisné C, Roques BP, Dardel F.** 2004. The Annealing Mechanism of HIV-1 Reverse Transcription Primer onto the Viral Genome. *J Biol Chem* 279:3588–3595.
145. **Tisné C, Roques BP, Dardel F.** 2003. Specific recognition of primer tRNA^{Lys3} by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension. *Biochimie* 85:557–561.
146. **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* 279:384–388.
147. **Godet J, Boudier C, Humbert N, Ivanyi-Nagy R, Darlix J-L, Mély Y.** 2012. Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus Res* 169:349–360.
148. **Bazzi A, Zargarian L, Chaminade F, Boudier C, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O.** 2011. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyribose in the binding polarity of NC. *Nucleic Acids Res* 39:3903–3916.
149. **Darlix J-L, de Rocquigny H, Mauffret O, Mély Y.** 2014. Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus Res* 193:2–15.
150. **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* 42:7145–7159.
151. **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* 301:491–511.

Références Bibliographiques

152. **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* 72:1902–1909.
153. **Williams MC, Rouzina I, Wenner JR, Gorelick RJ, Musier-Forsyth K, Bloomfield VA.** 2001. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc Natl Acad Sci* 98:6121–6126.
154. **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* 69:2101–2109.
155. **Baudin F, Marquet R, Isel C, Darlix J-L, 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* 229:382–397.
156. **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* 66:4144–4153.
157. **Lever A, Gottlinger H, Haseltine W, Sodroski J.** 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* 63:4085–4087.
158. **Kanevsky I, Chaminade F, Chen Y, Godet J, René B, Darlix J-L, Mély Y, Mauffret O, Fossé P.** 2011. Structural determinants of TAR RNA-DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *Nucleic Acids Res* 39:8148–8162.
159. **Kanevsky I, Chaminade F, Ficheux D, Moumen A, Gorelick R, Negroni M, Darlix J-L, Fossé P.** 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J Mol Biol* 348:1059–1077.
160. **Zargarian L, Kanevsky I, Bazzi A, Boynard J, Chaminade F, Fossé P, Mauffret O.** 2009. Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. *Nucleic Acids Res* 37:4043–4054.
161. **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* 383:1112–1128.
162. **Mori M, Kovalenko L, Lyonnais S, Antaki D, Torbett BE, Botta M, Mirambeau G, Mély Y.** 2015. Nucleocapsid Protein: A Desirable Target for Future Therapies Against HIV-1. *Curr Top Microbiol Immunol* 389:53–92.
163. **Miller MD, Farnet CM, Bushman FD.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71:5382–5390.
164. **Welker R, Hohenberg H, Tessmer U, Huckhagel C, Kräusslich H-G.** 2000. Biochemical and Structural Analysis of Isolated Mature Cores of Human Immunodeficiency Virus Type 1. *J Virol* 74:1168–1177.
165. **Zaitseva L, Myers R, Fassati A.** 2006. tRNAs promote nuclear import of HIV-1 intracellular reverse transcription complexes. *Plos Biol* 4:e332–e332.
166. **Yu KL, Lee SH, Lee ES, You JC.** 2016. HIV-1 nucleocapsid protein localizes efficiently to the nucleus and nucleolus. *Virology* 492:204–212.
167. **Thomas JA, Gorelick RJ.** 2008. Nucleocapsid protein function in early infection processes. *Virus Res* 134:39–63.
168. **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* 329:411–421.
169. **Lyonnais S, Gorelick RJ, Mergny J-L, Cam EL, Mirambeau G.** 2003. G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. *Nucleic Acids Res* 31:5754–5763.
170. **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* 364:496–511.

Références Bibliographiques

171. **Arhel NJ, Souquere-Besse S, Munier S, Souque P, Guadagnini S, Rutherford S, Prévost M-C, Allen TD, Charneau P.** 2007. HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *EMBO J* 26:3025–3037.
172. **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.
173. **Gao K, Gorelick RJ, Johnson DG, Bushman F.** 2003. Cofactors for Human Immunodeficiency Virus Type 1 cDNA Integration In Vitro. *J Virol* 77:1598–1603.
174. **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* 74:8980–8988.
175. **El Meshri SE, Dujardin D, Godet J, Richert L, Boudier C, Darlix JL, Didier P, Mély Y, de Rocquigny H.** 2015. Role of the Nucleocapsid Domain in HIV-1 Gag Oligomerization and Trafficking to the Plasma Membrane: A Fluorescence Lifetime Imaging Microscopy Investigation. *J Mol Biol* 427:1480–1494.
176. **Cimarelli A, Sandin S, Höglund S, Luban J.** 2000. Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA. *J Virol* 74:3046–3057.
177. **Poon DT, Wu J, Aldovini A.** 1996. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. *J Virol* 70:6607–6616.
178. **Bell NM, Lever AML.** 2013. HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol* 21:136–144.
179. **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* 375:592–610.
180. **Muriaux D, Rocquigny HD, Roques B-P, Paoletti J.** 1996. NCp7 Activates HIV-1Lai RNA Dimerization by Converting a Transient Loop-Loop Complex into a Stable Dimer. *J Biol Chem* 271:33686–33692.
181. **Ali MB, Chaminade F, Kanevsky I, Ennifar E, Josset L, Ficheux D, Darlix J-L, Fossé P.** 2007. Structural requirements for nucleocapsid protein-mediated dimerization of avian leukosis virus RNA. *J Mol Biol* 372:1082–1096.
182. **Weeks KM, Crothers DM.** 1991. RNA recognition by Tat-derived peptides: Interaction in the major groove? *Cell* 66:577–588.
183. **Easley R, Carpio L, Dannenberg L, Choi S, Alani D, Van Duyne R, Guendel I, Klase Z, Agbottah E, Kehn-Hall K, Kashanchi F.** 2010. Transcription through the HIV-1 nucleosomes: Effects of the PBAF complex in Tat activated transcription. *Virology* 405:322–333.
184. **Musinova YR, Sheval EV, Dib C, Germini D, Vassetzky YS.** 2015. Functional roles of HIV-1 Tat protein in the nucleus. *Cell Mol Life Sci* 73:589–601.
185. **Kao S-Y, Calman AF, Luciw PA, Peterlin BM.** 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* 330:489–493.
186. **Pollard VW, Malim MH.** 1998. The HIV-1 Rev protein. *Annu Rev Microbiol* 52:491–532.
187. **Biswas N, Wang T, Ding M, Tumne A, Chen Y, Wang Q, Gupta P.** 2012. ADAR1 is a novel multi targeted anti-HIV-1 cellular protein. *Virology* 422:265–277.
188. **Doria M, Neri F, Gallo A, Farace MG, Michienzi A.** 2009. Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Res* 37:5848–5858.
189. **Horenkamp FA, Breuer S, Schulte A, Lülff S, Weyand M, Saksela K, Geyer M.** 2011. Conformation of the dileucine-based sorting motif in HIV-1 Nef revealed by intermolecular domain assembly. *Traffic Cph Den* 12:867–877.
190. **Alvarado JJ, Tarafdar S, Yeh JI, Smithgall TE.** 2014. Interaction with the Src Homology (SH3-SH2) Region of the Src-family Kinase Hck Structures the HIV-1 Nef Dimer for Kinase Activation and Effector Recruitment. *J Biol Chem* 289:28539–28553.
191. **Kärkkäinen S, Hiipakka M, Wang J-H, Kleino I, Vähä-Jaakkola M, Renkema GH, Liss M, Wagner R, Saksela K.** 2006. Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. *EMBO Rep* 7:186–191.

Références Bibliographiques

192. **Suzu S, Harada H, Matsumoto T, Okada S.** 2005. HIV-1 Nef interferes with M-CSF receptor signaling through Hck activation and inhibits M-CSF bioactivities. *Blood* 105:3230–3237.
193. **Mwimanzi P, Markle TJ, Ogata Y, Martin E, Tokunaga M, Mahiti M, Kuang XT, Walker BD, Brockman MA, Brumme ZL, Ueno T.** 2013. Dynamic range of Nef functions in chronic HIV-1 infection. *Virology* 439:74–80.
194. **Desai TM, Marin M, Sood C, Jiong Shi, Nawaz F, Aiken C, Melikyan GB.** 2015. Fluorescent protein-tagged Vpr dissociates from HIV-1 core after viral fusion and rapidly enters the cell nucleus. *Retrovirology* 12:1–20.
195. **Pham TNQ, Lukhele S, Hajjar F, Routy J-P, Cohen ÉA.** 2014. HIV Nef and Vpu protect HIV-infected CD4⁺ T cells from antibody-mediated cell lysis through down-modulation of CD4 and BST2. *Retrovirology* 11:15.
196. **Schubert U, Bour S, Ferrer-Montiel AV, Montal M, Maldarell F, Strebel K.** 1996. The two biological activities of human immunodeficiency virus type 1 Vpu protein involve two separable structural domains. *J Virol* 70:809–819.
197. **González ME.** 2015. Vpu Protein: The Viroporin Encoded by HIV-1. *Viruses* 1999-4915 7:4352–4368.
198. **Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA.** 1987. The HIV A (sor) gene product is essential for virus infectivity. *Nature* 328:728–730.
199. **Zhang W, Du J, Evans SL, Yu Y, Yu X-F.** 2012. T-cell differentiation factor CBF- β regulates HIV-1 Vif-mediated evasion of host restriction. *Nature* 481:376–379.
200. **Zhang S, Zhong L, Chen B, Pan T, Zhang X, Liang L, Li Q, Zhang Z, Chen H, Zhou J, Luo H, Zhang H, Bai C.** 2015. Identification of an HIV-1 replication inhibitor which rescues host restriction factor APOBEC3G in Vif–APOBEC3G complex. *Antiviral Res* 122:20–27.
201. **Tang C, Louis JM, Aniana A, Suh J-Y, Clore GM.** 2008. Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* 455:693–696.
202. **Counts CJ, Ho PS, Donlin MJ, Tavis JE, Chen C.** 2015. A Functional Interplay between Human Immunodeficiency Virus Type 1 Protease Residues 77 and 93 Involved in Differential Regulation of Precursor Autoprocessing and Mature Protease Activity. *PLoS ONE* 10:1–25.
203. **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* 427:2360–2378.
204. **Zheng R, Jenkins TM, Craigie R.** 1996. Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci* 93:13659–13664.
205. **McColl DJ, Chen X.** 2010. Strand transfer inhibitors of HIV-1 integrase: Bringing IN a new era of antiretroviral therapy. *Antiviral Res* 85:101–118.
206. **Quashie PK, Han Y-S, Hassounah S, Mesplède T, Wainberg MA.** 2015. Structural Studies of the HIV-1 Integrase Protein: Compound Screening and Characterization of a DNA-Binding Inhibitor. *PLoS ONE* 10:1–18.
207. **Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet J-F, Brochon J-C, Deprez E.** 2006. Relationship between the Oligomeric Status of HIV-1 Integrase on DNA and Enzymatic Activity. *J Biol Chem* 281:22707–22719.
208. **Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, Dar MJ, KewalRamani VN, Engelman A.** 2010. The Requirement for Cellular Transportin 3 (TNPO3 or TRN-SR2) during Infection Maps to Human Immunodeficiency Virus Type 1 Capsid and Not Integrase. *J Virol* 84:397–406.
209. **Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A.** 2014. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* 42:10209–10225.
210. **Nanni RG, Ding J, Jacobo-Molina A, Hughes SH, Arnold E.** 1993. Review of HIV-1 reverse transcriptase three-dimensional structure: Implications for drug design. *Perspect Drug Discov Des* 1:129–150.
211. **Miller MT, Tuske S, Das K, DeStefano JJ, Arnold E.** 2016. Structure of HIV-1 reverse transcriptase bound to a novel 38-mer hairpin template-primer DNA aptamer. *Protein Sci* 25:46–55.
212. **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

Références Bibliographiques

- complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci* 90:6320–6324.
213. **Steitz TA.** 1999. DNA Polymerases: Structural Diversity and Common Mechanisms. *J Biol Chem* 274:17395–17398.
214. **Huang H, Chopra R, Verdine GL, Harrison SC.** 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282:1669–1675.
215. **Mansky LM, Temin HM.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 69:5087–5094.
216. **Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582–1586.
217. **Menéndez-Arias L, Betancor G, Matamoros T.** 2011. HIV-1 reverse transcriptase connection subdomain mutations involved in resistance to approved non-nucleoside inhibitors. *Antiviral Res* 92:139–149.
218. **Marchand B, Tchesnokov EP, Götte M.** 2007. The Pyrophosphate Analogue Foscarnet Traps the Pre-translocational State of HIV-1 Reverse Transcriptase in a Brownian Ratchet Model of Polymerase Translocation. *J Biol Chem* 282:3337–3346.
219. **Ng H-L, Kopka ML, Dickerson RE.** 2000. The structure of a stable intermediate in the A ↔ B DNA helix transition. *Proc Natl Acad Sci* 97:2035–2039.
220. **Wöhrl BM, Krebs R, Goody RS, Restle T.** 1999. Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate1. *J Mol Biol* 292:333–344.
221. **Van Cor-Hosmer SK, Daddacha W, Kelly Z, Tsurumi A, Kennedy EM, Kim B.** 2012. The impact of molecular manipulation in residue 114 of human immunodeficiency virus type-1 reverse transcriptase on dNTP substrate binding and viral replication. *Virology* 422:393–401.
222. **Ganai RA, Johansson E.** 2016. DNA Replication—A Matter of Fidelity. *Mol Cell* 62:745–755.
223. **Joyce CM, Benkovic SJ.** 2004. DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* 43:14317–14324.
224. **Steitz TA.** 1998. Structural biology: A mechanism for all polymerases. *Nature* 391:231–232.
225. **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* 385:693–713.
226. **Furfine ES, Reardon JE.** 1991. Reverse transcriptase.RNase H from the human immunodeficiency virus. Relationship of the DNA polymerase and RNA hydrolysis activities. *J Biol Chem* 266:406–412.
227. **Corona A, Meleddu R, Esposito F, Distinto S, Bianco G, Masaoka T, Maccioni E, Menéndez-Arias L, Alcaro S, Le Grice SFJ, Tramontano E.** 2016. Ribonuclease H/DNA Polymerase HIV-1 Reverse Transcriptase Dual Inhibitor: Mechanistic Studies on the Allosteric Mode of Action of Isatin-Based Compound RMNC6. *PLoS ONE* 11:1–18.
228. **Champoux JJ, Schultz SJ.** 2009. Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription. *FEBS J* 276:1506–1516.
229. **Palaniappan C, Fuentes GM, Rodríguez-Rodríguez L, Fay PJ, Bambara RA.** 1996. Helix Structure and Ends of RNA/DNA Hybrids Direct the Cleavage Specificity of HIV-1 Reverse Transcriptase RNase H. *J Biol Chem* 271:2063–2070.
230. **Yan J, Wu H, Tom T, Brodsky O, Maegley K.** 2011. Targeting divalent metal ions at the active site of the HIV-1 RNase H domain: NMR studies on the interactions of divalent metal ions with RNase H and its inhibitors. *Am J Anal Chem* 639.
231. **Cowan JA, Ohyama T, Howard K, Rausch JW, Cowan SML, Grice SFJL.** 2000. Metal-ion stoichiometry of the HIV-1 RT ribonuclease H domain: evidence for two mutually exclusive sites leads to new mechanistic insights on metal-mediated hydrolysis in nucleic acid biochemistry. *JBIC J Biol Inorg Chem* 5:67–74.
232. **Rogolino D, Carcelli M, Sechi M, Neamati N.** 2012. Viral enzymes containing magnesium: Metal binding as a successful strategy in drug design. *Coord Chem Rev* 256:3063–3086.

Références Bibliographiques

233. **Fedoroff OYu null, Salazar M, Reid BR.** 1993. Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J Mol Biol* 233:509–523.
234. **Bachelin M, Hessler G, Kurz G, Hacia JG, Dervan PB, Kessler H.** 1998. Structure of a stereoregular phosphorothioate DNA/RNA duplex. *Nat Struct Mol Biol* 5:271–276.
235. **Szyperski T, Götte M, Billeter M, Perola E, Cellai L, Heumann H, Wüthrich K.** 1999. NMR structure of the chimeric hybrid duplex r(gcaguggc).r(gcca)d(CTGC) comprising the tRNA-DNA junction formed during initiation of HIV-1 reverse transcription. *J Biomol NMR* 13:343–355.
236. **Schultz SJ, Champoux JJ.** 2008. RNase H activity: Structure, specificity, and function in reverse transcription. *Virus Res* 134:86–103.
237. **Pullen KA, Ishimoto LK, Champoux JJ.** 1992. Incomplete removal of the RNA primer for minus-strand DNA synthesis by human immunodeficiency virus type 1 reverse transcriptase. *J Virol* 66:367–373.
238. **Poeschla E.** 2013. The importance of becoming double-stranded: Innate immunity and the kinetic model of HIV-1 central plus strand synthesis. *Virology* 441:1–11.
239. **Schatz O, Cromme FV, Grüninger-Leitch F, Le Grice SF.** 1989. Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett* 257:311–314.
240. **Beilhartz GL, Wendeler M, Baichoo N, Rausch J, Le Grice S, Götte M.** 2009. HIV-1 Reverse Transcriptase Can Simultaneously Engage Its DNA/RNA Substrate at Both DNA Polymerase and RNase H Active Sites: Implications for RNase H Inhibition. *J Mol Biol* 388:462–474.
241. **Hergott CB, Mitra M, Guo J, Wu T, Miller JT, Iwatani Y, Gorelick RJ, Levin JG.** 2013. Zinc finger function of HIV-1 nucleocapsid protein is required for removal of 5'-terminal genomic RNA fragments: A paradigm for RNA removal reactions in HIV-1 reverse transcription. *Virus Res* 171:346–355.
242. **Kim J, Roberts A, Yuan H, Xiong Y, Anderson KS.** 2012. Nucleocapsid Protein Annealing of a Primer–Template Enhances (+)-Strand DNA Synthesis and Fidelity by HIV-1 Reverse Transcriptase. *J Mol Biol* 415:866–880.
243. **Lyonnais S, Gorelick RJ, Heniche-Boukhalfa F, Bouaziz S, Parissi V, Mouscadet J-F, Restle T, Gatell JM, Le Cam E, Mirambeau G.** 2013. A protein ballet around the viral genome orchestrated by HIV-1 reverse transcriptase leads to an architectural switch: From nucleocapsid-condensed RNA to Vpr-bridged DNA. *Virus Res* 171:287–303.
244. **Lener D, Tanchou V, Roques BP, Grice SFJL, Darlix J-L.** 1998. Involvement of HIV-I Nucleocapsid Protein in the Recruitment of Reverse Transcriptase into Nucleoprotein Complexes Formed in Vitro. *J Biol Chem* 273:33781–33786.
245. **Abbondanzieri EA, Bokinsky G, Rausch JW, Zhang JX, Le Grice SFJ, Zhuang X.** 2008. Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* 453:184–189.
246. **Liu S, Harada BT, Miller JT, Le Grice SFJ, Zhuang X.** 2010. Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat Struct Mol Biol* 17:1453–1460.
247. **Liu S, Abbondanzieri EA, Rausch JW, Grice SFJL, Zhuang X.** 2008. Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science* 322:1092–1097.
248. **Trono D.** 1992. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. *J Virol* 66:4893–4900.
249. **Zhang J, Temin HM.** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* 259:234–238.
250. **Didierlaurent L, Houzet L, Morichaud Z, Darlix J-L, Mougél M.** 2008. The conserved N-terminal basic residues and zinc-finger motifs of HIV-1 nucleocapsid restrict the viral cDNA synthesis during virus formation and maturation. *Nucleic Acids Res* 36:4745–4753.
251. **Mougél M, Cimarelli A, Darlix J-L.** 2010. Implications of the Nucleocapsid and the Microenvironment in Retroviral Reverse Transcription. *Viruses* 2:939–960.
252. **Mougél M, Houzet L, Darlix J-L.** 2009. When is it time for reverse transcription to start and go? *Retrovirology* 6:24–24.
253. **Thomas JA, Gorelick RJ.** 2008. Nucleocapsid protein function in early infection processes. *Virus Res* 134:39–63.

Références Bibliographiques

254. **Zhang H, Dornadula G, Alur P, Laughlin MA, Pomerantz RJ.** 1996. Amphipathic domains in the C terminus of the transmembrane protein (gp41) permeabilize HIV-1 virions: a molecular mechanism underlying natural endogenous reverse transcription. *Proc Natl Acad Sci* 93:12519–12524.
255. **Zhang H, Dornadula G, Pomerantz RJ.** 1996. Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenvironments: an important stage for viral infection of nondividing cells. *J Virol* 70:2809–2824.
256. **Zhang H, Dornadula G, Pomerantz RJ.** 1998. Natural endogenous reverse transcription of HIV-1. *J Reprod Immunol* 41:255–260.
257. **Hooker CW, Harrich D.** 2003. The first strand transfer reaction of HIV-1 reverse transcription is more efficient in infected cells than in cell-free natural endogenous reverse transcription reactions. *J Clin Virol* 26:229–238.
258. **Warrilow D, Meredith L, Davis A, Burrell C, Li P, Harrich D.** 2008. Cell Factors Stimulate Human Immunodeficiency Virus Type 1 Reverse Transcription In Vitro. *J Virol* 82:1425–1437.
259. **Abbink TEM, Berkhout B.** 2008. HIV-1 reverse transcription initiation: A potential target for novel antivirals? *Virus Res* 134:4–18.
260. **Wilkinson KA, 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:e96–e96.
261. **Isel C, Westhof E, Massire C, Le Grice SF, Ehresmann B, Ehresmann C, Marquet R.** 1999. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J* 18:1038–1048.
262. **Beerens N, Groot F, Berkhout B.** 2001. Initiation of HIV-1 Reverse Transcription Is Regulated by a Primer Activation Signal. *J Biol Chem* 276:31247–31256.
263. **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* 26:3667–3676.
264. **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* 67:3246–3253.
265. **Sleiman D, Goldschmidt V, Barraud P, Marquet R, Paillart J-C, Tisné C.** 2012. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res* 169:324–339.
266. **Das AT, Klaver B, Berkhout B.** 1995. Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *J Virol* 69:3090–3097.
267. **Mak J, Jiang M, Wainberg MA, Hammarskjöld ML, Rekosh D, Kleiman L.** 1994. Role of Pr160gag-pol in mediating the selective incorporation of tRNA(Lys) into human immunodeficiency virus type 1 particles. *J Virol* 68:2065–2072.
268. **Mak J, Kleiman L.** 1997. Primer tRNAs for reverse transcription. *J Virol* 71:8087–8095.
269. **Arts EJ, Stetor SR, Li X, Rausch JW, Howard KJ, Ehresmann B, North TW, Wöhrl BM, Goody RS, Wainberg MA, Grice SF.** 1996. Initiation of (–) strand DNA synthesis from tRNA(3Lys) on lentiviral RNAs: implications of specific HIV-1 RNA-tRNA(3Lys) interactions inhibiting primer utilization by retroviral reverse transcriptases. *Proc Natl Acad Sci* 93:10063–10068.
270. **Berkhout B, Das AT.** 2015. On the primer binding site mutation that appears and disappears during HIV and SIV replication. *Retrovirology* 12:1–4.
271. **Bampi C, Jacquenet S, Lener D, Décimo D, Darlix J-L.** 2004. The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance. *Int J Biochem Cell Biol* 36:1668–1686.
272. **Huang Y, Wang J, Shalom A, Li Z, Khorchid A, Wainberg MA, Kleiman L.** 1997. Primer tRNA^{3Lys} on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. *J Virol* 71:726–728.
273. **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* 12:40–40.

Références Bibliographiques

274. **Roldan A, Warren OU, Russell RS, Liang C, Wainberg MA.** 2005. A HIV-1 Minimal Gag Protein Is Superior to Nucleocapsid at in Vitro Annealing and Exhibits Multimerization-induced Inhibition of Reverse Transcription. *J Biol Chem* 280:17488–17496.
275. **Galindo-Murillo R, Davis DR, Cheatham III TE.** 2016. Probing the influence of hypermodified residues within the tRNA³Lys anticodon stem loop interacting with the A-loop primer sequence from HIV-1. *Biochim Biophys Acta BBA - Gen Subj* 1860:607–617.
276. **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* 268:25269–25272.
277. **Graham WD, Barley-Maloney L, Stark CJ, Kaur A, Stolyarchuk K, Sproat B, Leszczynska G, Malkiewicz A, Safwat N, Mucha P, Guenther R, Agris PF.** 2011. Functional Recognition of the Modified Human tRNA^{Lys}3UUU Anticodon Domain by HIV's Nucleocapsid Protein and a Peptide Mimic. *J Mol Biol* 410:698–715.
278. **Saadatmand J, Niu M, Kleiman L, Guo F.** 2009. The contribution of the primer activation signal to differences between Gag- and NCp7-facilitated tRNA^{Lys}3 annealing in HIV-1. *Virology* 391:334–341.
279. **Huthoff H, Bugala K, Barciszewski J, Berkhout B.** 2003. On the importance of the primer activation signal for initiation of tRNA(lys3)-primed reverse transcription of the HIV-1 RNA genome. *Nucleic Acids Res* 31:5186–5194.
280. **Ooms M, Cupac D, Abbink TEM, Huthoff H, Berkhout B.** 2007. The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. *Nucleic Acids Res* 35:1649–1659.
281. **Iwatani Y, Rosen AE, Guo J, Musier-Forsyth K, Levin JG.** 2003. Efficient initiation of HIV-1 reverse transcription in vitro. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. *J Biol Chem* 278:14185–14195.
282. **Goldschmidt V, Rigourd M, Ehresmann C, Grice SFJL, Ehresmann B, Marquet R.** 2002. Direct and Indirect Contributions of RNA Secondary Structure Elements to the Initiation of HIV-1 Reverse Transcription. *J Biol Chem* 277:43233–43242.
283. **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* 279:35923–35931.
284. **Isel C, Lanchy JM, Grice SFL, Ehresmann C, Ehresmann B, Marquet R.** 1996. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA³Lys. *EMBO J* 15:917.
285. **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* 247:236–250.
286. **Coffin JM, Haseltine WA.** 1977. Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. *Proc Natl Acad Sci U S A* 74:1908.
287. **Lori F, Veronese F di M, Vico AL de, Lusso P, Reitz MS, Gallo RC.** 1992. Viral DNA carried by human immunodeficiency virus type 1 virions. *J Virol* 66:5067–5074.
288. **Post K, Kankia B, Gopalakrishnan S, Yang V, Cramer E, Saladores P, Gorelick RJ, Guo J, Musier-Forsyth K, Levin JG.** 2009. Fidelity of plus-strand priming requires the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res* 37:1755–1766.
289. **Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA.** 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Res* 134:19–38.
290. **Varmus HE, Heasley S, Kung H-J, Oppermann H, Smith VC, Bishop JM, Shank PR.** 1978. Kinetics of synthesis, structure and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. *J Mol Biol* 120:55–82.
291. **Drummond JE, Mounts P, Gorelick RJ, Casas-Finet JR, Bosche WJ, Henderson LE, Waters DJ, Arthur LO.** 1997. Wild-type and mutant HIV type 1 nucleocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase. *AIDS Res Hum Retroviruses* 13:533–543.

Références Bibliographiques

292. **Anthony RM, DeStefano JJ.** 2007. In vitro Synthesis of Long DNA Products in Reactions with HIV-RT and Nucleocapsid Protein. *J Mol Biol* 365:310–324.
293. **Lanciault C, Champoux JJ.** 2005. Effects of unpaired nucleotides within HIV-1 genomic secondary structures on pausing and strand transfer. *J Biol Chem* 280:2413–2423.
294. **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* 41:301–312.
295. **Rausch JW, Le Grice SFJ.** 2004. “Binding, bending and bonding”: polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. *Int J Biochem Cell Biol* 36:1752–1766.
296. **Fabris D, Marino JP, Le Grice SFJ.** 2009. Revisiting Plus-Strand DNA Synthesis in Retroviruses and Long Terminal Repeat Retrotransposons: Dynamics of Enzyme: Substrate Interactions. *Viruses* 1:657–677.
297. **Sarafianos SG, Das K, Tantillo C, Clark AD Jr, Ding J, Whitcomb JM, Boyer PL, Hughes SH, Arnold E.** 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J* 20:1449–1461.
298. **Yi-Brunozzi HY, Brinson RG, Brabazon DM, Lener D, Le Grice SFJ, Marino JP.** 2008. High-resolution NMR analysis of the conformations of native and base analog substituted retroviral and LTR-retrotransposon PPT primers. *Chem Biol* 15:254–262.
299. **Beilhartz GL, Götte M.** 2010. HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors. *Viruses* 2:900–926.
300. **Jones FD, Hughes SH.** 2007. In vitro analysis of the effects of mutations in the G-tract of the human immunodeficiency virus type 1 polypurine tract on RNase H cleavage specificity. *Virology* 360:341–349.
301. **Smith CM, Smith JS, Roth MJ.** 1999. RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. *J Virol* 73:6573–6581.
302. **Wu T, Guo J, Bess J, Henderson LE, Levin JG.** 1999. Molecular Requirements for Human Immunodeficiency Virus Type 1 Plus-Strand Transfer: Analysis in Reconstituted and Endogenous Reverse Transcription Systems. *J Virol* 73:4794–4805.
303. **Guo J, Henderson LE, Bess J, Kane B, Levin JG.** 1997. Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA. *J Virol* 71:5178–5188.
304. **Zhang H, Bagasra O, Niikura M, Poesz BJ, Pomerantz RJ.** 1994. Intravirion reverse transcripts in the peripheral blood plasma on human immunodeficiency virus type 1-infected individuals. *J Virol* 68:7591–7597.
305. **Johnson PE, Turner RB, Wu ZR, Hairston L, Guo J, Levin JG, Summers MF.** 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* 39:9084–9091.
306. **Egelé C, Schaub E, Ramalanjaona N, Piémont E, Ficheux D, Roques B, Darlix J-L, Mély Y.** 2004. HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J Mol Biol* 342:453–466.
307. **Charneau P, Mirambeau G, Roux P, Paulous S, Buc H, Clavel F.** 1994. HIV-1 reverse transcription. A termination step at the center of the genome. *J Mol Biol* 241:651–662.
308. **Hu C, Saenz DT, Fadel HJ, Walker W, Peretz M, Poeschla EM.** 2010. The HIV-1 Central Polypurine Tract Functions as a Second Line of Defense against APOBEC3G/F. *J Virol* 84:11981–11993.
309. **Wurtzer S, Goubard A, Mammano F, Saragosti S, Lecossier D, Hance AJ, Clavel F.** 2008. Functional Central Polypurine Tract Provides Downstream Protection of the Human Immunodeficiency Virus Type 1 Genome from Editing by APOBEC3G and APOBEC3B. *J Virol* 82:5116.
310. **Hameau L, Jeusset J, Lafosse S, Coulaud D, Delain E, Unge T, Restle T, Le Cam E, Mirambeau G.** 2001. Human immunodeficiency virus type 1 central DNA flap: dynamic terminal product of plus-strand displacement dna synthesis catalyzed by reverse transcriptase assisted by nucleocapsid protein. *J Virol* 75:3301–3313.
311. **Hsu M, Wainberg MA.** 2000. Interactions between human immunodeficiency virus type 1 reverse transcriptase, tRNA primer, and nucleocapsid protein during reverse transcription. *J Hum Virol* 3:16–26.

Références Bibliographiques

312. **Kelleher CD, Champoux JJ.** 1998. Characterization of RNA strand displacement synthesis by Moloney murine leukemia virus reverse transcriptase. *J Biol Chem* 273:9976–9986.
313. **van Wamel JL, Berkhout B.** 1998. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology* 244:245–251.
314. **Panganiban AT, Fiore D.** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* 241:1064–1069.
315. **Jones KA, Peterlin BM.** 1994. Control of RNA initiation and elongation at the HIV-1 promoter. *Annu Rev Biochem* 63:717–743.
316. **Goodrich DW, Duesberg PH.** 1990. Retroviral recombination during reverse transcription. *Proc Natl Acad Sci* 87:2052–2056.
317. **Onafuwa-Nuga A, Telesnitsky A.** 2009. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev MMBR* 73:451–480, Table of Contents.
318. **Delviks-Frankenberry K, Galli A, Nikolaitchik O, Mens H, Pathak VK, Hu W-S.** 2011. Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* 3:1650–1680.
319. **Ritchie AJ, Cai F, Smith NMG, Chen S, Song H, Brackenridge S, Abdool Karim SS, Korber BT, McMichael AJ, Gao F, Goonetilleke N.** 2014. Recombination-mediated escape from primary CD8+ T cells in acute HIV-1 infection. *Retrovirology* 11:69–69.
320. **Sakuragi S, Shioda T, Sakuragi J.** 2015. Properties of human immunodeficiency virus type 1 reverse transcriptase recombination upon infection. *J Gen Virol* 96:3382–3388.
321. **Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM.** 2008. The Challenge of HIV-1 Subtype Diversity. *N Engl J Med* 358:1590–1602.
322. **Coffin JM.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol* 42:1–26.
323. **Svarovskaia ES, Delviks KA, Hwang CK, Pathak VK.** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. *J Virol* 74:7171–7178.
324. **Negróni M, Ricchetti M, Nouvel P, Buc H.** 1995. Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. *Proc Natl Acad Sci* 92:6971–6975.
325. **Nikolenko GN, Svarovskaia ES, Delviks KA, Pathak VK.** 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. *J Virol* 78:8761–8770.
326. **Hwang CK, Svarovskaia ES, Pathak VK.** 2001. Dynamic copy choice: Steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proc Natl Acad Sci* 98:12209–12214.
327. **Hanson MN, Balakrishnan M, Roques BP, Bambara RA.** 2005. Effects of Donor and Acceptor RNA Structures on the Mechanism of Strand Transfer by HIV-1 Reverse Transcriptase. *J Mol Biol* 353:772–787.
328. **Negróni M, Buc H.** 2000. Copy-choice recombination by reverse transcriptases: Reshuffling of genetic markers mediated by RNA chaperones. *Proc Natl Acad Sci* 97:6385–6390.
329. **Chen Y, Balakrishnan M, Roques BP, Bambara RA.** 2003. Steps of the Acceptor Invasion Mechanism for HIV-1 Minus Strand Strong Stop Transfer. *J Biol Chem* 278:38368–38375.
330. **Xie P.** 2009. A polymerase-site-jumping model for strand transfer during DNA synthesis by reverse transcriptase. *Virus Res* 144:65–73.
331. **Song M, Balakrishnan M, Chen Y, Roques BP, Bambara RA.** 2006. Stimulation of HIV-1 Minus Strand Strong Stop DNA Transfer by Genomic Sequences 3' of the Primer Binding Site. *J Biol Chem* 281:24227–24235.
332. **Negróni M, Buc H.** 2001. OPINION: Retroviral recombination: what drives the switch? *Nat Rev Mol Cell Biol* 2:151–155.
333. **Liang C, Rong L, Götte M, Li X, Quan Y, Kleiman L, Wainberg MA.** 1998. Mechanistic studies of early pausing events during initiation of HIV-1 reverse transcription. *J Biol Chem* 273:21309–21315.

Références Bibliographiques

334. **Shen W, Gao L, Balakrishnan M, Bambara RA.** 2009. A recombination hot spot in HIV-1 contains guanosine runs that can form a G-quartet structure and promote strand transfer in vitro. *J Biol Chem* 284:33883–33893.
335. **DeStefano JJ, Bambara RA, Fay PJ.** 1994. The mechanism of human immunodeficiency virus reverse transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates. *J Biol Chem* 269:161–168.
336. **DeStefano JJ, Mallaber LM, Rodriguez-Rodriguez L, Fay PJ, Bambara RA.** 1992. Requirements for strand transfer between internal regions of heteropolymer templates by human immunodeficiency virus reverse transcriptase. *J Virol* 66:6370–6378.
337. **DeStefano JJ, Buiser RG, Mallaber LM, Myers TW, Bambara RA, Fay PJ.** 1991. Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. *J Biol Chem* 266:7423–7431.
338. **Derebail SS, DeStefano JJ.** 2004. Mechanistic analysis of pause site-dependent and -independent recombinogenic strand transfer from structurally diverse regions of the HIV genome. *J Biol Chem* 279:47446–47454.
339. **Derebail SS, Heath MJ, DeStefano JJ.** 2003. Evidence for the differential effects of nucleocapsid protein on strand transfer in various regions of the HIV genome. *J Biol Chem* 278:15702–15712.
340. **Laughrea M, Jette 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* 13464.
341. **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* 91:4945–4949.
342. **Muriaux D, Girard P-M, 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* 270:8209–8216.
343. **Balakrishnan M, Fay PJ, Bambara RA.** 2001. The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region. *J Biol Chem* 276:36482–36492.
344. **Balakrishnan M, Roques BP, Fay PJ, Bambara RA.** 2003. Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis. *J Virol* 77:4710–4721.
345. **Andersen ES, Jeeninga RE, Damgaard CK, Berkhout B, Kjems J.** 2003. Dimerization and template switching in the 5' untranslated region between various subtypes of human immunodeficiency virus type 1. *J Virol* 77:3020–3030.
346. **Chin MPS, Rhodes TD, Chen J, Fu W, Hu W-S.** 2005. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci U S A* 102:9002–9007.
347. **Moore MD, Fu W, Nikolaitchik O, Chen J, Ptak RG, Hu W-S.** 2007. Dimer initiation signal of human immunodeficiency virus type 1: its role in partner selection during RNA copackaging and its effects on recombination. *J Virol* 81:4002–4011.
348. **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* 70:8348–8354.
349. **Hu WS, Temin HM.** 1990. Retroviral recombination and reverse transcription. *Science* 250:1227–1233.
350. **Klaver B, Berkhout B.** 1994. Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res* 22:137–144.
351. **Ohi Y, Clever JL.** 2000. Sequences in the 5' and 3' R elements of human immunodeficiency virus type 1 critical for efficient reverse transcription. *J Virol* 74:8324–8334.
352. **Peliska JA, Benkovic SJ.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science* 258:1112–1118.
353. **Luo GX, Taylor J.** 1990. Template switching by reverse transcriptase during DNA synthesis. *J Virol* 64:4321–4328.

Références Bibliographiques

354. **Fu TB, Taylor J.** 1992. When retroviral reverse transcriptases reach the end of their RNA templates. *J Virol* 66:4271–4278.
355. **DeStefano JJ, Mallaber LM, Fay PJ, Bambara RA.** 1993. Determinants of the RNase H cleavage specificity of human immunodeficiency virus reverse transcriptase. *Nucleic Acids Res* 21:4330–4338.
356. **Peliska JA, Balasubramanian S, Giedroc DP, Benkovic SJ.** 1994. Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* 33:13817–13823.
357. **Guo J, Wu T, Kane BF, Johnson DG, Henderson LE, Gorelick RJ, Levin JG.** 2002. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* 76:4370–4378.
358. **You JC, McHenry CS.** 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem* 269:31491–31495.
359. **Lapadat-Tapolsky M, Gabus C, Rau M, Darlix J-L.** 1997. Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability1. *J Mol Biol* 268:250–260.
360. **Driscoll MD, Golinelli MP, Hughes SH.** 2001. In vitro analysis of human immunodeficiency virus type 1 minus-strand strong-stop DNA synthesis and genomic RNA processing. *J Virol* 75:672–686.
361. **Golinelli MP, Hughes SH.** 2001. Self-priming of retroviral minus-strand strong-stop DNAs. *Virology* 285:278–290.
362. **Hong MK, Harbron EJ, O'Connor DB, Guo J, Barbara PF, Levin JG, Musier-Forsyth K.** 2003. Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer. *J Mol Biol* 325:1–10.
363. **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* 279:44154–44165.
364. **Zhan P, Liu X, Li Z, Pannecouque C, De Clercq E.** 2009. Design strategies of novel NNRTIs to overcome drug resistance. *Curr Med Chem* 16:3903–3917.
365. **Morphy R, Rankovic Z.** 2005. Designed multiple ligands. An emerging drug discovery paradigm. *J Med Chem* 48:6523–6543.
366. **Ilna T, Parniak MA.** 2008. Inhibitors of HIV-1 reverse transcriptase. *Adv Pharmacol San Diego Calif* 56:121–167.
367. **Meleddu R, Distinto S, Corona A, Bianco G, Cannas V, Esposito F, Artese A, Alcaro S, Matyus P, Bogdan D, Cottiglia F, Tramontano E, Maccioni E.** 2015. (3Z)-3-(2-[4-(aryl)-1,3-thiazol-2-yl]hydrazin-1-ylidene)-2,3-dihydro-1H-indol-2-one derivatives as dual inhibitors of HIV-1 reverse transcriptase. *Eur J Med Chem* 93:452–460.
368. **Beilhartz GL, Ngure M, Johns BA, DeAnda F, Gerondelis P, Götte M.** 2014. Inhibition of the ribonuclease H activity of HIV-1 reverse transcriptase by GSK5750 correlates with slow enzyme-inhibitor dissociation. *J Biol Chem* 289:16270–16277.
369. **Velthuisen EJ, Johns BA, Gerondelis P, Chen Y, Li M, Mou K, Zhang W, Seal JW, Hightower KE, Miranda SR, Brown K, Leesnitzer L.** 2014. Pyridopyrimidinone inhibitors of HIV-1 RNase H. *Eur J Med Chem* 83:609–616.
370. **Tramontano E, Di Santo R.** 2010. HIV-1 RT-Associated RNase H Function Inhibitors: Recent Advances in Drug Development (English). *Curr Med Chem* 17:2837–2853.
371. **Dorfman T, Luban J, Goff SP, Haseltine WA, Göttlinger HG.** 1993. Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* 67:6159–6169.
372. **Déméné H, Dong CZ, Ottmann M, Rouyez MC, Jullian N, Morellet N, Mely Y, Darlix JL, Fournié-Zaluski MC, Saragosti S.** 1994. 1H NMR structure and biological studies of the His23-->Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* 33:11707–11716.

Références Bibliographiques

373. **Wu H, Mitra M, McCauley MJ, Thomas JA, Rouzina I, Musier-Forsyth K, Williams MC, Gorelick RJ.** 2013. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res* 171:263–277.
374. **Rice WG, Schaeffer CA, Harten B, Villinger F, South TL, Summers MF, Henderson LE, Bess JW, Arthur LO, McDougal JS.** 1993. Inhibition of HIV-1 infectivity by zinc-ejecting aromatic C-nitroso compounds. *Nature* 361:473–475.
375. **Rice WG, Supko JG, Malspeis L, Buckheit RW, Clanton D, Bu M, Graham L, Schaeffer CA, Turpin JA, 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* 270:1194–1197.
376. **Arthur LO, Bess JW, Chertova EN, Rossio JL, Esser MT, Benveniste RE, Henderson LE, Lifson JD.** 1998. Chemical inactivation of retroviral infectivity by targeting nucleocapsid protein zinc fingers: a candidate SIV vaccine. *AIDS Res Hum Retroviruses* 14 Suppl 3:S311–319.
377. **Chertova EN, Kane BP, McGrath C, Johnson DG, Sowder RC, Arthur LO, Henderson LE.** 1998. Probing the topography of HIV-1 nucleocapsid protein with the alkylating agent N-ethylmaleimide. *Biochemistry* 37:17890–17897.
378. **Pannecouque C, Szafarowicz B, Volkova N, Bakulev V, Dehaen W, Mély Y, Daelemans D.** 2010. Inhibition of HIV-1 replication by a bis-thiadiazolbenzene-1,2-diamine that chelates zinc ions from retroviral nucleocapsid zinc fingers. *Antimicrob Agents Chemother* 54:1461–1468.
379. **Turpin JA, Schito ML, Jenkins LMM, Inman JK, Appella E.** 2008. Topical microbicides: a promising approach for controlling the AIDS pandemic via retroviral zinc finger inhibitors. *Adv Pharmacol San Diego Calif* 56:229–256.
380. **Jenkins LMM, Durell SR, Maynard AT, Stahl SJ, Inman JK, Appella E, Legault P, Omichinski JG.** 2006. Comparison of the specificity of interaction of cellular and viral zinc-binding domains with 2-mercaptobenzamide thioesters. *J Am Chem Soc* 128:11964–11976.
381. **Shvadchak V, Sanglier S, Rocle S, Villa P, Haiech J, Hibert M, Van Dorsselaer A, Mély Y, de Rocquigny H.** 2009. Identification by high throughput screening of small compounds inhibiting the nucleic acid destabilization activity of the HIV-1 nucleocapsid protein. *Biochimie* 91:916–923.
382. **Breuer S, Chang MW, Yuan J, Torbett BE.** 2012. Identification of HIV-1 inhibitors targeting the nucleocapsid protein. *J Med Chem* 55:4968–4977.
383. **Guo J, Wu T, Bess J, Henderson LE, Levin JG.** 1998. Actinomycin D inhibits human immunodeficiency virus type 1 minus-strand transfer in in vitro and endogenous reverse transcriptase assays. *J Virol* 72:6716–6724.
384. **Jeeninga RE, Huthoff HT, Gultyaev AP, Berkhout B.** 1998. The mechanism of actinomycin D-mediated inhibition of HIV-1 reverse transcription. *Nucleic Acids Res* 26:5472–5479.
385. **Imamichi T, Murphy MA, Adelsberger JW, Yang J, Watkins CM, Berg SC, Baseler MW, Lempicki RA, Guo J, Levin JG, Lane HC.** 2003. Actinomycin D induces high-level resistance to thymidine analogs in replication of human immunodeficiency virus type 1 by interfering with host cell thymidine kinase expression. *J Virol* 77:1011–1020.
386. **Sosic A, Frecentese F, Perissutti E, Sinigaglia L, Santagada V, Caliendo G, Magli E, Ciano A, Zagotto G, Parolin C, Gatto B.** 2013. Design, synthesis and biological evaluation of TAR and cTAR binders as HIV-1 nucleocapsid inhibitors. *MedChemComm* 4:1388–1393.
387. **Sosic A, Sinigaglia L, Cappellini M, Carli I, Parolin C, Zagotto G, Sabatino G, Rovero P, Fabris D, Gatto B.** 2016. Mechanisms of HIV-1 Nucleocapsid Protein Inhibition by Lysyl-Peptidyl-Anthraquinone Conjugates. *Bioconjug Chem* 27:247–256.
388. **Sodroski J, Trus M, Perkins D, Patarca R, Wong-Staal F, Gelmann E, Gallo R, Haseltine WA.** 1984. Repetitive structure in the long-terminal-repeat element of a type II human T-cell leukemia virus. *Proc Natl Acad Sci U S A* 81:4617–4621.
389. **Pfeiffer JK, Telesnitsky A.** 2001. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. *J Virol* 75:11263–11274.

Références Bibliographiques

390. **Dang Q, Hu WS.** 2001. Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. *J Virol* 75:809–820.
391. **Berkhout B, van Wamel J, Klaver B.** 1995. Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. *J Mol Biol* 252:59–69.
392. **Berkhout B, Das AT, van Wamel JL.** 1998. The native structure of the human immunodeficiency virus type 1 RNA genome is required for the first strand transfer of reverse transcription. *Virology* 249:211–218.
393. **Berkhout B, Vastenhouw NL, Klasens BI, Huthoff H.** 2001. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA N Y N* 7:1097–1114.
394. **Moumen A, Polomack L, Roques B, Buc H, Negroni M.** 2001. The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* 29:3814–3821.
395. **Jones CP, Cantara WA, Olson ED, Musier-Forsyth K.** 2014. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proc Natl Acad Sci U S A* 111:3395–3400.
396. **Harrich D, Hooker CW, Parry E.** 2000. The Human Immunodeficiency Virus Type 1 TAR RNA Upper Stem-Loop Plays Distinct Roles in Reverse Transcription and RNA Packaging. *J Virol* 74:5639.
397. **Harrich D, Ulich C, Gaynor RB.** 1996. A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. *J Virol* 70:4017–4027.
398. **Berkhout B, Klaver B, Das AT.** 1995. A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses. *Virology* 207:276–281.
399. **Miller WA, White KA.** 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annu Rev Phytopathol* 44:447–467.
400. **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* 279:22243–22249.
401. **Beerens N, Kjems J.** 2010. Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. *RNA N Y N* 16:1226–1235.
402. **Brulé F, Bec G, Keith G, Grice SFJL, Roques BP, Ehresmann B, Ehresmann C, Marquet R.** 2000. In vitro evidence for the interaction of tRNA^{3Lys} with U3 during the first strand transfer of HIV-1 reverse transcription. *Nucleic Acids Res* 28:634–640.
403. **Piekna-Przybylska D, DiChiacchio L, Mathews DH, Bambara RA.** 2010. A sequence similar to tRNA^{3Lys} gene is embedded in HIV-1 U3-R and promotes minus-strand transfer. *Nat Struct Mol Biol* 17:83–89.
404. **Ooms M, Abbink TEM, Pham C, Berkhout B.** 2007. Circularization of the HIV-1 RNA genome. *Nucleic Acids Res* 35:5253–5261.
405. **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* 317:385–399.
406. **Liu H-W, Zeng Y, Landes CF, Kim YJ, Zhu Y, Ma X, Vo M-N, Musier-Forsyth K, Barbara PF.** 2007. Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc Natl Acad Sci U S A* 104:5261–5267.
407. **Piekna-Przybylska D, Dykes C, Demeter LM, Bambara RA.** 2011. Sequences in the U3 region of human immunodeficiency virus 1 improve efficiency of minus strand transfer in infected cells. *Virology* 410:368–374.
408. **Piekna-Przybylska D, Bambara RA.** 2011. Requirements for efficient minus strand strong-stop DNA transfer in human immunodeficiency virus 1. *RNA Biol* 8:230–236.
409. **Vo M-N, Barany G, Rouzina I, Musier-Forsyth K.** 2009. HIV-1 Nucleocapsid Protein Switches the Pathway of Transactivation Response Element RNA/DNA Annealing from Loop–Loop “Kissing” to “Zipper.” *J Mol Biol* 386:789–801.
410. **Darlix JL, Vincent A, Gabus C, de Rocquigny H, Roques B.** 1993. Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *Comptes Rendus Académie Sci Sér III Sci Vie* 316:763–771.

Références Bibliographiques

- 411. **Godet J, de Rocquigny H, Raja C, Glasser N, Ficheux D, Darlix J-L, Mély Y.** 2006. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol* 356:1180–1192.
- 412. **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* 326:691–700.
- 413. **Beltz H, Azoulay J, Bernacchi S, Clamme J-P, Ficheux D, Roques B, Darlix J-L, Mély Y.** 2003. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* 328:95–108.
- 414. **McCauley MJ, Rouzina I, Manthei KA, Gorelick RJ, Musier-Forsyth K, Williams MC.** 2015. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc Natl Acad Sci* 112:13555–13560.
- 415. **Maxam AM, Gilbert W.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages, p. 499–560. In *Enzymology*, B-M in (ed.), Academic Press.
- 416. **Levin JG, Guo J, Rouzina Ioulia, Musier-Forsyth K.** 2005. Nucleic Acid Chaperone Activity of HIV-1 Nucleocapsid Protein: Critical Role in Reverse Transcription and Molecular Mechanism. *Prog Nucleic Acid Res Mol Biol* 80:217–286.
- 417. **Bazzi A, Zargarian L, Chaminade F, Boudier C, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O.** 2011. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyribose in the binding polarity of NC. *Nucleic Acids Res* 39:3903–3916.
- 418. **Bazzi A, Zargarian L, Chaminade F, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O.** 2012. Intrinsic Nucleic Acid Dynamics Modulates HIV-1 Nucleocapsid Protein Binding to Its Targets. *PLoS ONE* 7:1–12.
- 419. **Zuker M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31:3406.
- 420. **Avilov SV, Godet J, Piémont E, Mély Y.** 2009. Site-Specific Characterization of HIV-1 Nucleocapsid Protein Binding to Oligonucleotides with Two Binding Sites. *Biochemistry* 48:2422–2430.
- 421. **Chen Y, Maskri O, Chaminade F, René B, Benkaroun J, Godet J, Mély Y, Mauffret O, Fossé P.** 2016. Structural Insights into the HIV-1 Minus-strand Strong-stop DNA. *J Biol Chem* 291:3468–3482.
- 422. **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* 93:5572.
- 423. **Polge E, Darlix J-L, Paoletti J, Fossé P.** 2000. Characterization of loose and tight dimer forms of avian leukosis virus RNA1. *J Mol Biol* 300:41–56.
- 424. **Kanevsky I, Vasilenko N, Dumay-Odelot H, Fossé P.** 2003. In vitro characterization of a base pairing interaction between the primer binding site and the minimal packaging signal of avian leukosis virus genomic RNA. *Nucleic Acids Res* 31:7070–7082.
- 425. **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* 336:369–379.
- 426. **Berkhout B.** 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res* 20:27–31.
- 427. **Cosnefroy O, Murray PJ, Bishop KN.** 2016. HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology* 13:58–58.

Titre : Relations structure-fonction du premier transfert de brin chez le VIH-1

Mots clés : VIH-1, transcription inverse, transfert de brin, ADNss et protéine de nucléocapside.

Résumé : Les premier et second transferts de brin sont deux étapes essentielles de la transcription inverse du génome du virus de l'immunodéficience humaine de type 1 (VIH-1). De nombreuses études in vitro suggèrent que les transferts nécessitent l'action de la protéine de nucléocapside du VIH-1 (NC). Le premier transfert, se produisant de l'extrémité 5' vers l'extrémité 3' de l'ARN génomique du VIH-1, repose en grande partie sur un appariement ADN-ARN impliquant la région r de l'ADN strong-stop (ADNss) et la région R située à l'extrémité 3' de l'ARN viral (ARN 3'UTR). Les structures, interactions et dynamiques qui gouvernent cet appariement sont mal connues. Jusqu'à notre étude, la formation de l'hybride ADN-ARN n'avait été étudiée in vitro qu'avec des courts acides nucléiques qui ne reflètent que partiellement les structures formées par l'ADNss et l'ARN 3'UTR dans le virus en cours de réplication. L'objectif principal de ma thèse a été de caractériser in vitro les mécanismes moléculaires qui gouvernent l'hybridation de l'ADNss avec l'ARN 3'UTR. Pour atteindre cet objectif, j'ai utilisé des méthodes de la biologie moléculaire et j'ai analysé la structure secondaire de l'ADNss entier avec trois sondes de structure (DNase I, mung bean nuclease et permanganate de potassium) : i) en l'absence de NC ; ii) en présence de NC ; iii) en présence de l'ARN 3'UTR. Les résultats obtenus nous ont permis d'être les premiers à déterminer in vitro la structure secondaire de l'ADNss entier du VIH-1 et à identifier dans celui-ci quatre sites sur lesquels se fixe préférentiellement la NC. L'interaction de la NC avec chacun de ces sites a été analysée par retard sur gel. A notre connaissance, les structures secondaires des ADNss d'autres rétrovirus n'ont pas été déterminées. Nos données structurales sont en faveur d'une structure secondaire de l'ADNss du VIH-1 constituée de trois tiges-boucles (u5, cpoly(A) et cTAR) et trois régions simple-brin en l'absence ou présence de NC. Notre analyse phylogénétique suggère que la structure secondaire de l'ADNss et les sites forts NC sont conservés parmi les différents groupes du VIH-1. Nos résultats suggèrent aussi qu'une partie de la région u5 de l'ADNss établit des interactions faibles avec une partie de la région U3 de l'ARN 3'UTR en l'absence de NC. En réalisant des cinétiques d'hybridation et en utilisant deux ADNss mutés, nous avons montré que l'hybridation ADNss-ARN 3'UTR nécessite l'activité de la NC et que ce processus ne repose pas sur une seule voie d'initiation. Nos résultats supportent un modèle dans lequel la première étape est la fixation de la NC au niveau des quatre sites, ce qui va déclencher l'ouverture de la structure tridimensionnelle de l'ADNss et favoriser ainsi l'accessibilité de la région r ; la seconde étape étant la déstabilisation par la NC des structures secondaires ; la troisième étape étant l'appariement par la NC des régions complémentaires r et R. L'ensemble des résultats obtenus permet à mon équipe d'accueil d'initier de nouvelles études in vitro et ex vivo.



Title : Structure-function relationships of the first strand transfer in HIV-1.

Keywords : HIV-1, reverse transcription, strand transfer, ssDNA and nucleocapsid protein.

Abstract: An essential step of human immunodeficiency virus type 1 (HIV-1) reverse transcription is the first strand transfer that requires base pairing of the R region at the 3'-end of the genomic RNA with the complementary r region at the 3'-end of minus-strand strong-stop DNA (ssDNA). HIV-1 nucleocapsid protein (NC) facilitates this annealing process. Determination of the ssDNA structure is needed to understand the molecular basis of NC-mediated genomic RNA-ssDNA annealing. For this purpose, we investigated ssDNA using structural probes (nucleases and potassium permanganate). This study is the first to determine the secondary structure of the full-length HIV-1 ssDNA in the absence or presence of NC. Our probing data obtained in the absence of NC, suggest weak contacts between the u5 region of ssDNA and the U3 region the genomic RNA. The probing data and phylogenetic analysis support the folding of ssDNA into three stem-loop structures and the presence of four high-affinity binding sites for NC. Using the gel retardation assay, we analyzed the interaction of NC with each site. Taken together, our results support a model for the NC-mediated annealing process in which the preferential binding of NC to four sites triggers unfolding of the three-dimensional structure of ssDNA, thus facilitating interaction of the r sequence of ssDNA with the R sequence of the genomic RNA. In addition, using gel retardation assays and ssDNA mutants, we show that the annealing of ssDNA to the 3'-end of the genomic RNA requires NC activity and does not rely on a single pathway (zipper intermediate or kissing complex).

