Determination of the secondary structure of minus strong-stop DNA and the mechanism of annealing involved in the first strand transfer in HIV-1

Yingying Chen

THESE DE DOCTORAT

DE L'ECOLE NORMALE SUPERIEURE DE CACHAN

Présentée par

Mademoiselle Yingying CHEN

Pour obtenir le grade de

DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN

Domaine :
Science de la vie et de la santé

Sujet de la thèse :

Determination of the secondary structure of minus strong-stop DNA and the mechanism of annealing involved in the first strand transfer in HIV-1

Thèse présentée et soutenue à Cachan le 14 septembre 2012
devant le jury composé de :

Dr. Eric LE CAM	Directeur de Recherches	Rapporteur
Dr. Marylène MOUGEL	Directrice de Recherches	Rapporteur
Dr. Jean-François RIOU	Professeur	Examinateur
Dr. Carine TISNÉ	Directrice de Recherches	Examinatrice
Dr. Philippe FOSSÉ	Chargé de Recherches	Directeur de thèse
Pr. Liansheng HOU	Professeur	Co-directeur de thèse

Laboratoire de biologie et de pharmacologie appliquée (LBPA)
ENS CACHAN (CNRS UMR 8113)
61, avenue du Président Wilson, 94235 CACHAN CEDEX (France)

REMERCIEMENTS

J'ai effectué mon travail de thèse dans l'équipe « Structures et Interactions des Acides Nucléiques » du Laboratoire de Biologie et de Pharmacologie Appliquée, UMR 8113 du CNRS. Je voudrais d'abord remercier Jean-François Mouscadet et Malcolm Buckle, qui ont assuré successivement la direction du laboratoire, pour m'avoir accueillie au sein de leur unité de recherche et m'avoir permis de travailler dans un environnement scientifique aussi exceptionnel.

Je suis très sensible à l'honneur que m'ont fait les membres du jury de thèse en acceptant de lire et d'évaluer ce travail. Je les remercie d'avoir consacré du temps à l'expertise de mon travail de thèse. Je tiens à remercier particulièrement Mme Mougel et M. Le Cam qui par leurs suggestions et remarques constructives ont permis l'amélioration de ce mémoire.

Je tiens également à adresser toute ma gratitude à Philippe Fossé pour son accueil chaleureux au sein de son groupe. Merci de ton soutien, de ton aide et de tes conseils scientifiques. Tu m'as offert l'opportunité d'entrer dans le monde de la structures des ARN et ADN. Les expériences que j'ai vécues au sein de ton groupe sont une richesse inestimable, et je t'en serai toujours reconnaissante.

Je tiens à remercier particulièrement Françoise Chaminade pour son aide précieuse et sa gentillesse ainsi que pour de m'avoir transmis l'ensemble des techniques couramment employées dans le laboratoire. Je te remercie aussi pour tous tes encouragements.

Je remercie Serge Fermandjian, Olivier Mauffret et Brigitte Hartmann pour leurs conseils scientifiques et leurs encouragements.

Je remercie Brigitte René et Loussiné Zargarian pour leurs gentillesses et leurs sympathies.

Je tiens également à remercier tous les membres passés et présents de l'équipe qui m'ont soutenue tout au long de cette thèse. Merci à Xiaoqian, Farah pour leurs gentillesses et leurs amitiés. Merci à Céline, Jessica, Thina et Marwa pour leurs gentillesses et leurs compagnies.

Merci aux anciens thésards, Ali Bazzi et Zeina Hobeika pour leurs conseils et amitiés.

Résumé

L'étude du premier transfert de brin est importante pour mieux comprendre la transcription inverse et la recombinaison qui dépend du transfert de brin interne. La recombinaison joue un rôle important dans la variabilité génétique du VIH-1 et l'émergence de résistances aux antirétroviraux. Le premier transfert de brin, processus crucial, se produit pendant la transcription inverse lorsqu'une partie de l'ADN "strong stop" simple-brin (ADNss) s'apparie avec la séquence R en 3' de l'ARN viral ; cette séquence peut former les tiges-boucles TAR et polyA. La protéine de nucléocapside du VIH-1 (NC) stimule ce processus par son activité chaperonne des acides nucléiques. La partie de l'ADNss, qui est complémentaire de R, est supposée former les tiges-boucles cTAR et cpolyA. Donc, le transfert repose probablement sur l'hybridation de deux acides nucléiques structurés. Bien que des études suggèrent un rôle des structures ADN et ARN dans le premier transfert de brin, elles n'ont pas ou très peu abordé directement l'aspect structural. Le rôle de la structure de l'ADNss dans l'hybridation responsable du premier transfert de brin avait été uniquement étudié avec un oligonucléotide (ADN cTAR) correspondant à peine au tiers de l'ADNss. A ce jour, la structure secondaire de l'ADNss n'a jamais été déterminée. L'objectif principal de ma thèse a été d'identifier les interactions et structures gouvernant l'hybridation de l'ADNss avec l'extrémité 3' de l'ARN génomique du VIH-1 (3'UTR). Les outils de la biologie moléculaire et trois sondes de structure ciblant l'ADN [permanganate de potassium, DNase I et mung bean nucléase (MB)] ont été utilisés pour atteindre cet objectif.

Premièrement, la structure secondaire de l'ADN cTAR a été déterminée en absence ou en présence de NC. L'analyse structurale au moyen des sondes de structure montre que l'ADN cTAR se replie sous la forme de deux tige-boucles qui sont en équilibre en absence de NC. Une conformation est nommée fermée car les extrémités 5' et 3' sont appariées tandis que l'autre conformation est dite en 'Y' car les extrémités 5' et 3' ne sont pas appariées. La NC déplace l'équilibre vers la conformation 'Y'. Au moyen d'empreintes à la MB, nous avons montré qu'en présence de 7 mM MgCl$_2$ (concentration optimale pour la transcription inverse et le transfert de brin *in vitro*) la NC se fixe plus fortement sur la boucle interne que sur la boucle apicale de la tige-boucle cTAR. Cette fixation préférentielle n'a pas été observée en présence de 0.2 mM MgCl$_2$ (concentration intracellulaire).

Deuxièmement, l'hybridation de l'ADNss sauvage et trois de ses mutants avec le 3' UTR a été étudiée en absence ou en présence de NC dans 0,2 mM et 2 mM MgCl$_2$ (concentration minimale pour la transcription inverse et le transfert de brin *in vitro*). Nous avons étudié deux ADNss : l'ADNss-L représente l'ADNss qui ne s'apparie pas avec la région PBS ; l'ADNss-S représente l'ADNss qui s'apparie avec la région PBS. Nous avons montré que la NC est exigée pour former un hétéroduplex constitué de l'intégralité de l'ADNss et du 3' UTR. Nos résultats suggèrent que l'hybridation de l'ADNss avec le 3' UTR peut être initiée à partir de plusieurs sites en présence de 0,2 mM MgCl$_2$. En revanche, nos résultats suggèrent que l'initiation de l'hybridation *via* les boucles apicales des tiges-boucles TAR et cTAR joue un rôle important dans 2 mM MgCl$_2$.

Finalement, les structures secondaires des ADNss-L et ADNss-S ont été déterminées en absence ou en présence de NC. Nos résultats suggèrent que l'ADNss forme principalement une seule conformation dans 2 mM MgCl$_2$. L'analyse des ADNss sur gel de polyacrylamide non-dénaturant suggère que l'ADNss adopte deux conformations en équilibre dans 0,2 mM MgCl$_2$. Ces deux conformations de l'ADNss dépendent de la séquence cTAR qui peut former une tige-boucle longue ou une tige-boucle courte dans l'ADNss. Les conformations fermées et en 'Y' de l'ADN cTAR ne sont pas formées dans l'ADNss. La NC se fixe préférentiellement au niveau de la région simple-brin qui relie les tiges-boucles cTAR et cpolyA. Cette fixation joue probablement un rôle important dans l'hybridation des tiges-boucles ARN et ADN complémentaires.

Abstract

The reverse transcription of the HIV-1 genome consists in a succession of steps allowing the conversion of the single stranded RNA genome in a double stranded DNA molecule. In the laboratory, we investigate the first strand transfer during which the strong stop DNA (ssDNA) migrates from the 5' end of the genome to the 3' end. Analysis of this step is important to gain insights into the reverse transcription process and the associated genetic recombination. Indeed, the strand transfer process is responsible for the recombination events that produce resistance to antiretroviral drugs that constitute a major problem in the anti-HIV-1 therapies. The main DNA and RNA sequences involved in the first strand transfer have been identified: these are the RNA sequences TAR and polyA and the complementary DNA sequences cTAR and cpolyA. The first strand transfer is facilitated by the HIV-1 nucleocapsid protein (NC). During the strand transfer, NC destabilizes the nucleic acid secondary structures and promotes their association. Nevertheless, the annealing process of the full-length ssDNA to the 3' end of the genomic RNA (3' UTR) is not known at the molecular and structural level. The main aim of my thesis is to better understand this annealing process. The tools of molecular biology and three DNA-targeted probes [potassium permanganate, DNase I and mung bean nuclease (MB)] will be use to achieve this goal.

Firstly, the cTAR secondary structure was determined in the absence or in the presence of NC. Structural analysis using structural probes showed that the cTAR DNA folds into two stem-loops in equilibrium in the absence of NC. One conformation is named closed because the 5' and 3' ends of the cTAR DNA are paired, while the other conformation is named 'Y' conformation because the 5' and 3' ends of cTAR are unpaired. NC slightly destabilizes the lower stem and shifts the equilibrium toward the 'Y' conformation. The MB footprinting results showed that in the presence of 7 mM $MgCl_2$ (optimal concentration for reverse transcription and strand transfer *in vitro*) NC binds more strongly the internal loop than the apical loop of the cTAR hairpin. However, this preferential binding site has not been observed in the presence of 0.2 mM $MgCl_2$ (intracellular concentration).

Secondly, the annealing of the full-length wild-type ssDNA and its three mutations to the 3' UTR was investigated in the absence or in the presence of NC and in 0.2 mM and 2 mM $MgCl_2$ (concentration required for reverse transcription and strand transfer *in vitro*). We have designed two full-length ssDNAs: ssDNA-L represents the ssDNA that is not annealed to the PBS region; ssDNA-S represents the ssDNA that is paired with the PBS region. We have shown that NC is required for the formation of heteroduplex of the full-length ssDNA and 3' UTR. Our results suggest that the annealing of ssDNA to 3' UTR can be initiated from different sites in the presence of 0.2 mM $MgCl_2$, whereas the initiation of annealing *via* the apical loops of TAR and cTAR hairpins plays an important role in the presence of 2 mM $MgCl_2$.

Finally, we have determined the secondary structures of the full-length ssDNA-S and ssDNA-L in the absence or in the presence of NC. Our results suggest that ssDNA folds mainly into one conformation in 2 mM $MgCl_2$. Analysis of ssDNA by non-denaturing polyacrylamide gel electrophoresis suggests that ssDNA adopts two conformations in equilibrium in 0.2 mM $MgCl_2$. The formation of two conformers of ssDNA may be due to the cTAR sequence that can form a long stem-loop or a short stem-loop in ssDNA. The closed and 'Y' conformations of cTAR DNA are not formed in ssDNA. NC preferentially binds to the single-stranded region between the cTAR and cpolyA hairpins in ssDNA. This binding site probably plays an important role in the annealing of complementary DNA and RNA hairpins.

Index

ABBREVIATIONS

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
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	*Cis*-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
FL	Full-length
ER	Endoplasmic reticulum
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	Lymphoadenopathy 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-κB	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)P$_2$	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
TRIM5α	Tripartite motif 5α
USE	Upstream enhancer element
UTR	Untranslated region
WDSV	Walleye dermal sarcoma virus
ZF	Zinc finger

INTRODUCTION

Chapter 1. Retroviruses

1. Characteristics of retroviruses

The retrovirus is an RNA virus that infects the vertebrates and insects, and causes different diseases, such as tumours, neurological disorders and the acquired immunodeficiency syndrome (AIDS). The majority of retroviruses are constituted of the viral RNA, except the spumaviruses that contain an infectious DNA genome (260). The retrovirus carries two copies of genomic RNA (gRNA), positive sense and single stranded, in each viral particle so that it is considered as diploid RNA virus, a feature that strongly potentiates the mechanism of recombination. During the early stage of the replication, one copy of gRNA is reverse transcribed by the viral reverse transcriptase (RT) to a double-stranded DNA (dsDNA) in the cytoplasm of the host cell. This dsDNA is then integrated into the host's genome by the viral integrase (IN), at which point the retroviral DNA is referred to a provirus. A special variant of retroviruses are endogenous retroviruses which are integrated into the genome of the host and inherited across generations. Around 8% of the human genome is derived from sequences with similarity to infectious retrovirus, which can be easily recognized because all infectious retroviruses contain at least three genes, including *gag* (encoding structural proteins), *pol* (viral enzymes) and *env* (surface envelop proteins), as well as long terminal repeats (LTRs) **(Figure 1)** (153).
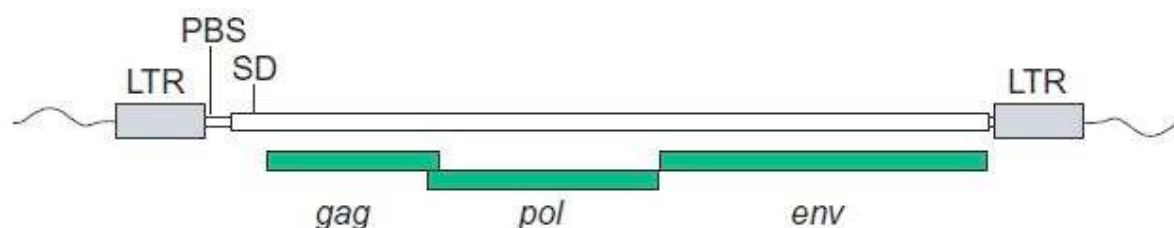
Figure 1. Structure of infectious retroviruses (153).

2. Classification of retroviruses

The reference of classification is from the International Committee of Taxonomy of Viruses (ICTV, http://www.ictvonline.org/virusTaxonomy.asp). The affiliation of one family is defined by common taxonomic denominators: genetic information as well as the replicative properties. Certain structure elements, as the presence of envelope and pathogenesis are equally counted in some cases. Retroviruses are further divided into seven groups defined by evolutionary relatedness, each with the taxonomic rank of genus. Five of these groups represent retroviruses with oncogenic potential (formerly referred to as oncoviruses), and the

other two groups are the lentiviruses and the spumaviruses **(Figure 2)**. Retroviruses is broadly divided into two categories - simple and complex - distinguishable by the organization of their genomes (279). All oncogenic members except the human T-cell leukemia virus-bovine leukemia virus (HTLV-BLV) genus are simple retroviruses. HTLV-BLV, the lentiviruses and spumaviruses are complex.
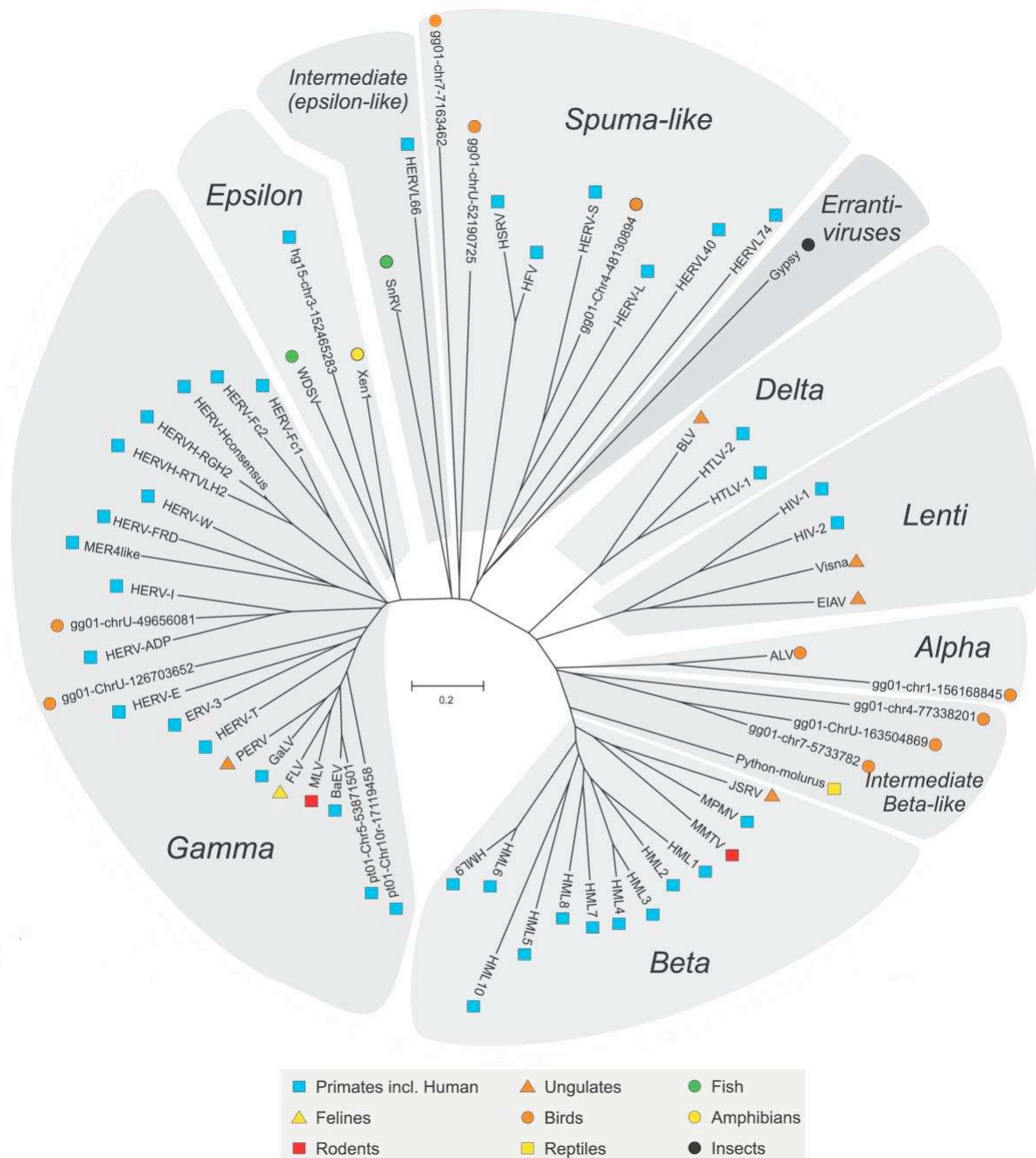
Figure 2. Dendrogram of retroviruses (197). ALV: Avian Leukosis Virus; BLV: Bovine Leukemia Virus; HERV: Human Endogenous Retrovirus; HFV: Human Foamy Virus; HIV: Human Immunodeficiency Virus; HTLV: Human T-Lymphotropic Virus; MLV: Murine Leukemia Virus; MMTV Mouse Mammary Tumor Virus; WDSV: Walleye Dermal Sarcoma Virus.

Chapter 2. General properties of HIV

1 Identification of HIV from AIDS

Nearly 30 years after the first reported cases of AIDS and the discovery of the etiologic agent, effective control of the AIDS pandemic, which is arguably the most serious infectious disease, remains elusive. In 2010, it was estimated that 34 million people are living with HIV, and in that year alone, 2.7 million people newly infected with HIV **(Figure 3)**.

Global summary of the AIDS epidemic | 2010

Number of people living with HIV	Total	34.0 million [31.6 million–35.2 million]
	Adults	30.1 million [28.4 million–31.5 million]
	Women	16.8 million [15.8 million–17.6 million]
	Children (<15 years)	3.4 million [3.0 million–3.8 million]
People newly infected with HIV in 2010	Total	2.7 million [2.4 million–2.9 million]
	Adults	2.3 million [2.1 million–2.5 million]
	Children (<15 years)	390 000 [340 000–450 000]
AIDS deaths in 2010	Total	1.8 million [1.6 million–1.9 million]
	Adults	1.5 million [1.4 million–1.6 million]
	Children (<15 years)	250 000 [220 000–290 000]

World Health Organization   UNAIDS

Figure 3: Global summary of the HIV/AIDS epidemic, December 2010. (from World Health Organization and UNAIDS) http://www.who.int/hiv/data/en/

AIDS was first recognized in the United States in 1981, following an increase in the incidence of usually rare opportunistic infections (such as the pneumonia caused by *Pneumocystis carinii*) in homosexual men that were caused by a general immune deficiency. The common characteristic of all the patients is a dramatic diminution of CD4+ lymphocyte cells in peripheral blood. HIV was first isolated in 1983 by the team of L. Montagnier of Pasteur Institute at Paris (18). From a sample of patient with lymphoadenopathy, the activity of reverse transcription was identified. The visualization with electronic microscopy confirmed the presence of a retrovirus that was initially named as Lymphoadenopathy Associated Virus (LAV). Quickly, some teams identified two viruses that considered as the pathogenic agent responsible of AIDS, HTLV-III and AIDS Associated-

Virus (64, 129). Until 1985 the scientific community concluded that these three viruses are identical. In 1986, the name of HIV was taken by the discovery of the second similar virus with slightly different genome structures, isolated from two West-African patients with AIDS and then called as HIV-2 (65). By its genomic sequences and its proteins, HIV-2 is different from the LAV-I/HTLV-III, isolated from U.S.A., Europe and Central Africa. It differs also from STLV-III, isolated from Rhesus Macaques with AIDS, but displays an antigenic relationship with the latter virus, at the level of its external envelope protein (64).

2 <u>Origin of HIV</u>

The key to understanding the origin of HIV was the discovery that closely related viruses-the simian immunodeficiency viruses (SIVs)-were present in a wide variety of African primates. Collectively, HIV and SIV comprise the primate lentiviruses, and SIVs have been isolated in more than 20 African primate species.

The evolutionary history of HIV-1 and HIV-2 has been reconstructed in great detail by inferring phylogenetic trees of the primate lentiviruses. It was discovered that the two human viruses are related to different SIVs and therefore have different evolutionary origins **(Figure 4 and Table 1)**. Specifically, HIV-1 is most closely related to SIVcpz, which is found in some sub-species of chimpanzee that inhabit parts of equatorial Western and Central Africa, respectively (133, 325). This HIV-1 progenitor probably was passed from chimpanzees to human hunters through blood borne transmission. HIV-2 closely resembles SIVsm that is found in the West African sooty mangabey and the Taï Forest in Côte d'Ivoire. HIV-2 ancestor may be passed from monkeys in the Sierra Leone and Liberia where is the sooty mangabey's natural habitat through the bushmeat trade (88).

Table 1. Comparison of HIV species

Species	Virulence	Infectivity	Prevalence	Inferred origin
HIV-1	High	High	Global	Commun Chimpanzee
HIV-2	Lower	Low	West Africa	Sooty Mangabey

0.05 substitutions per site
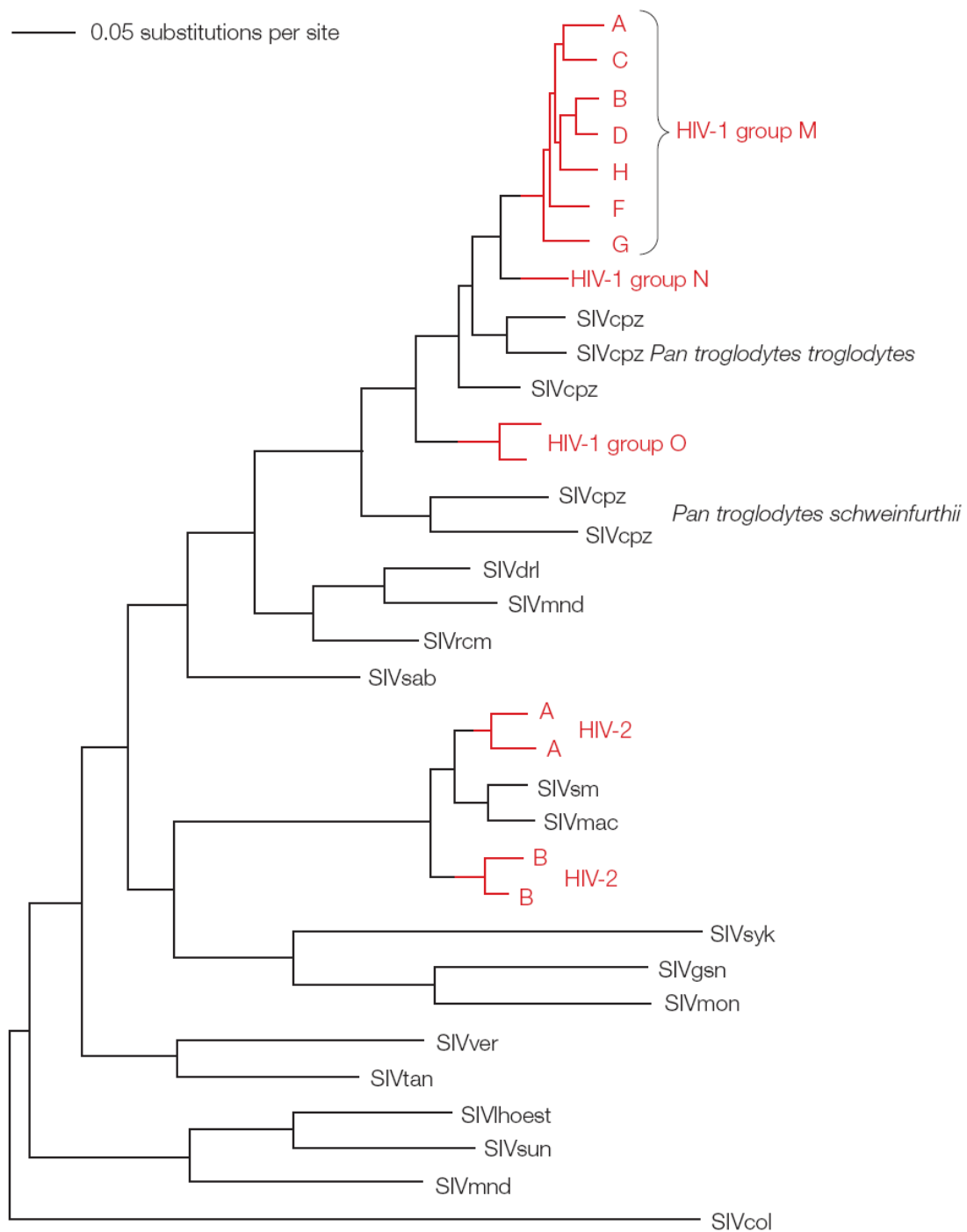
Figure 4. Evolutionary history of the HIV-1 and HIV-2 (325). Because both HIV-1 and HIV-2 lineages (red branches) fall within the simian immunodeficiency viruses (SIVs) that are isolated from other primates, they represent independent cross-species transmission events. The tree and other evidence also indicate that HIV-1 groups M, N and O represent separate transfers from chimpanzees (SIVcpz), again because there is a mixing of the HIV-1 and SIV lineages. Similarly, HIV-2 seems to have been transferred from sooty mangabey monkeys (SIVsm) on many occasions. For clarity, only some subtypes of HIV-1 and HIV-2 are shown.

2.1 HIV-1

Phylogenetic analysis of HIV-1 and related viruses from non-human primates suggest that three independent transmission events early in the 20th century spawned three HIV-1 groups: M, N and O. Although strains related to the M and N groups have been found in chimpanzees, recent evidence suggests that HIV-1 O group may have originated in gorillas, in which the closest relatives of this group have been identified (367). It is speculated that the virus then spread among humans along the Congo River into Kinshasa, Zaire, where the earliest documented case of HIV-1 infection (with group M strain) in humans has been traced to a blood sample from 1959 (422). The group N and O remain essentially restricted to West Africa, whereas the M group is responsible for the global HIV-1 pandemic **(Figure 5)**. HIV-1 M group spread first throughout Africa where it has differentiated into several lineages called subtypes. Group M lineages include subtypes and circulating recombinant forms (CRFs). The HIV-1 M group is divided into 11 subtypes (A1, A2, B, C, D, F1, F2 G, H, J and K) and 48 forms of CRFs. Among the HIV-1 isolates studied in different laboratories, HXB2, NL4-3 and LAI are subtype B while MAL isolate is a mixed form of A1-D-K (http://www.hiv.lanl.gov/).

2.2 HIV-2

HIV-2 closely resembles SIVsm from the West African sooty mangabey (*Cercocebus torquatus atys*) (134), a simian virus that is thought to have entered the human population on at least eight separate occasions, yielding eight distinct HIV-2 groups of which only groups A and B are endemic; with the remainder being single-person infections (78, 160) HIV-2 is present throughout West Africa and the highest prevalence of HIV-2 reported two decades ago (in Guinea Bissau) was 8% in adults and up to 20% in individuals over 40 years of age (321). Most countries are now experiencing adecline in HIV-2 prevalence while HIV-1 infection increases in the younger population (76).
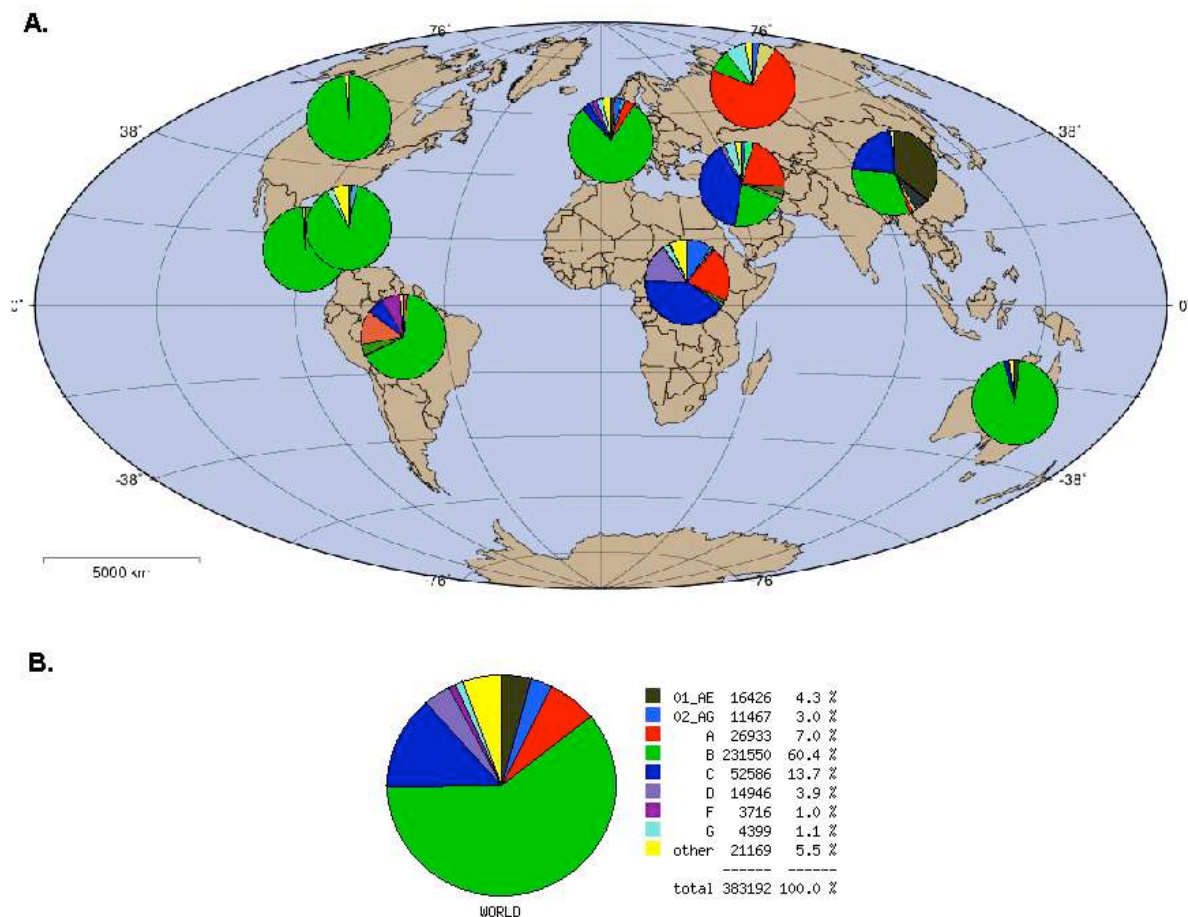
Figure 5. Global distribution of HIV-1 subtype group M in 2011. CRF: Circulating recombinant form. (source from Los Alamos national laboratory: http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp)

3 <u>Pathogenesis of HIV</u>

HIV infects primarily vital cells in the human immune system such as CD4+ T cells, macrophages, and dendritic cells (DCs). HIV infection leads to low levels of CD4+ T cells through three main mechanisms: First, direct viral killing of infected cells; second, increased rates of apoptosis in infected cells; and third, killing of infected CD4+ T cells by CD8 cytotoxic lymphocytes that recognize infected cells. When CD4+ T cell numbers decline below a critical level, cell-mediated immunity is lost, and the body becomes progressively more susceptible to opportunistic infections. The typical course of HIV infection is presented in **Figure 6**.
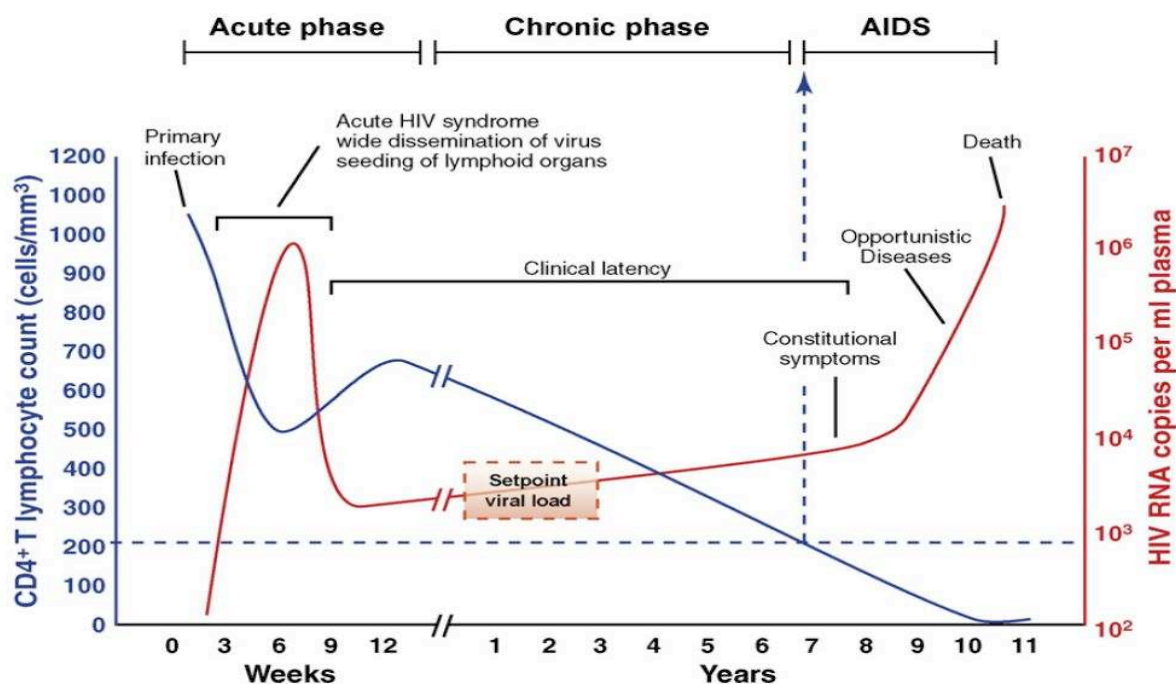
Figure 6. Typical course of HIV infection. (4)

4 Characteristics of HIV-1

4.1 Structure of HIV-1 particle

The HIV-1 particle is different in structure from other retroviruses. It is roughly spherical with a diameter of about 120 nm, around 60 times smaller than a red blood cell, yet large for a virus **(Figure 7)**. In immature HIV particles, the major structural protein, Gag, is arranged in a radial fashion, with the N-terminal matrix (MA, p17) domain associated with the viral membrane, a lipid bilayer decorated with the products of the *env* gene. And it is followed by the internal capsid (CA, p24) domain and the C-terminal nucleocapsid (NC) domain pointed toward the center (126, 363, 398). The protein Env, consists of a cap made of three molecules called glycoprotein (gp) 120, and a stem consisting of three gp41 molecules that anchor the structure into the viral envelope. This glycoprotein complex enables the virus to attach to and fuse with target cells to initiate the infectious cycle. Both surface proteins, especially gp120, have been considered as targets of future treatments or vaccines against HIV-1. During or shortly after budding, Gag is cleaved by the viral protease (PR), leading to virus maturation, which is reflected in a dramatic morphological change required for infectivity (363, 388).
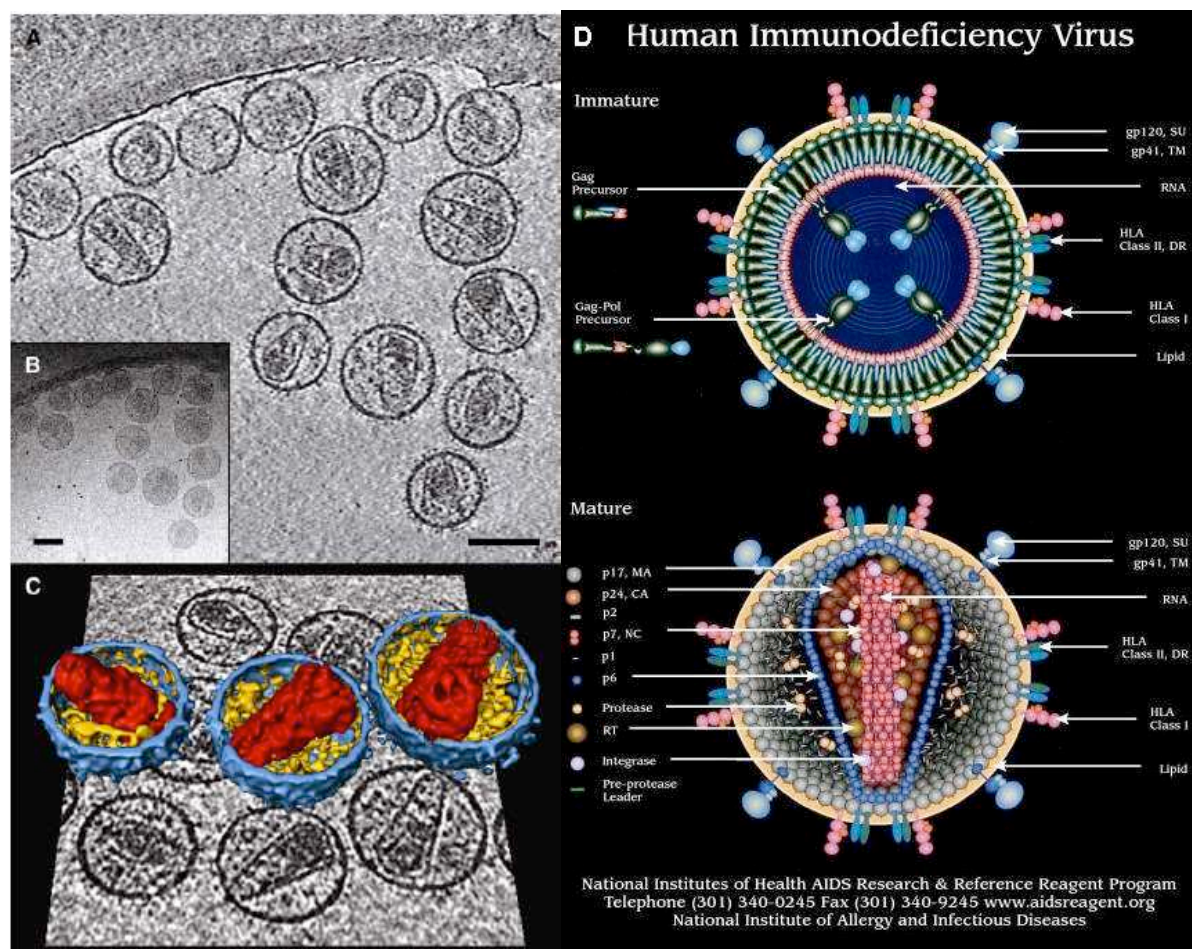
Figure 7. Structures of HIV-1 Virions (43). (A) A central slice through one-quarter of the field of view in one of the cryo-electron tomograms. The scale bar is 100 nm. (B) A single untilted image of the same area illustrating the even distribution of gold beads in the sample. The scale bar is 100 nm. (C) 3D rendering of three sample virions from the tomogram. Blue, viral membrane; yellow, density between the membrane and the core; red, viral capsid. (D) structures of immature and mature HIV-1 particles. (source from: https://www.aidsreagent.org/program_info.cfm#5)

4.2 Genetic organization

The RNA genome consists of at least seven structural landmarks (LTR, TAR, RRE, PE, SLIP, CRS, and INS) (http://www.hiv.lanl.gov/), and nine genes (*gag*, *pol*, *env*, *tat*, *rev*, *nef*, *vif*, *vpr*, *vpu* and sometimes a tenth *tev*, which is a fusion of *tat, env* and *rev*), encoding 15 proteins **(Figure 8) (152).**
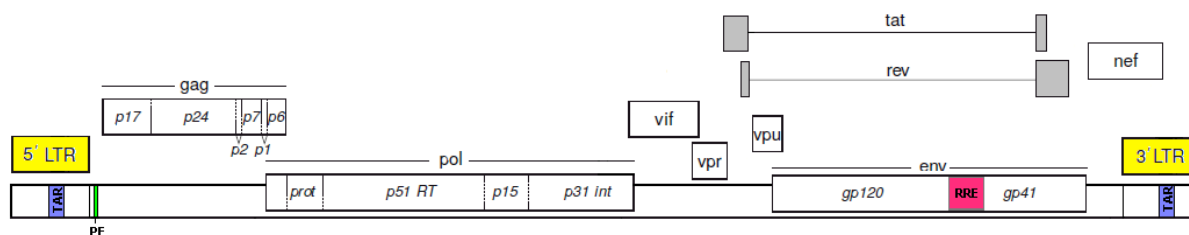
Figure 8. Genomic organization of HIV-1. Open reading frames (ORFs) are shown as rectangle. 5' LTR and 3' LTR are shown in yellow, TAR in blue, PE in green and RRE in pink.

LTR (long terminal repeat) is the DNA sequence flanking the genome of integrated proviruses. It contains important regulatory regions, especially those for transcription initiation and polyadenylation.

TAR (transactivator response element) is the binding site for Tat protein and for cellular proteins; consists of approximately the first 57 nucleotides of the viral mRNAs in HIV-1. TAR RNA forms a hairpin stem-loop structure with a side bulge; the bulge is necessary for Tat binding and function.

RRE (Rev responsive element) is an RNA element encoded within the env region of HIV-1. It consists of approximately 200 nucleotides (positions 7327 to 7530 from the start of transcription in HIV-1, spanning the border of gp120 and gp41). The RRE is necessary for Rev function; it contains a high affinity site for Rev; in all, approximately seven binding sites for Rev exist within the RRE RNA (221) (http://www.hiv.lanl.gov/).

PE (Psi element) is a set of 4 stem-loop (SL) structures preceding and overlapping the Gag start codon which are the sites recognized by the conserved motif with the canonical sequence $CysX_2CysX_4HisX_4Cys$, identified as NC zinc fingers (ZFs). The PE is present in unspliced genomic transcripts but absent from spliced viral mRNAs. The Psi (Ψ) site is formed with SL1 and SL3.

SLIP is a TTTTTT slippery site, followed by a stem-loop structure, is responsible for regulating the -1 ribosomal frameshift out of the Gag reading frame into the Pol reading frame (221).

CRS (*cis*-acting repressive sequences) postulated to inhibit structural protein expression in the absence of Rev. One such site was mapped within the *pol* region of HIV-1. The exact function has not been defined; splice sites have been postulated to act as CRS sequences (221).

INS (inhibitory/instability RNA sequences) found within the structural genes of HIV-1 and of other complex retroviruses. Multiple INS elements exist within the genome and can act independently. The INS elements have been defined by functional assays as elements that inhibit expression posttranscriptionally. Mutation of the RNA elements was shown to lead to INS inactivation and up-regulation of gene expression (http://www.hiv.lanl.gov/).

5 Replication cycle

The HIV-1 replication cycle (schematically shown in **Figure 9**) can be summarised in two phases.

- Early phase: 1) attachment to the host cell and uncoating of the virion; 2) reverse transcription of gRNA; 3) provirus integration;

- Late phase: 4) transcription and processing of gRNA; 5) export of viral RNAs; 6) synthesis of viral proteins, packaging of gRNA and assembly of viral; 7) maturation and budding of the viral particle.
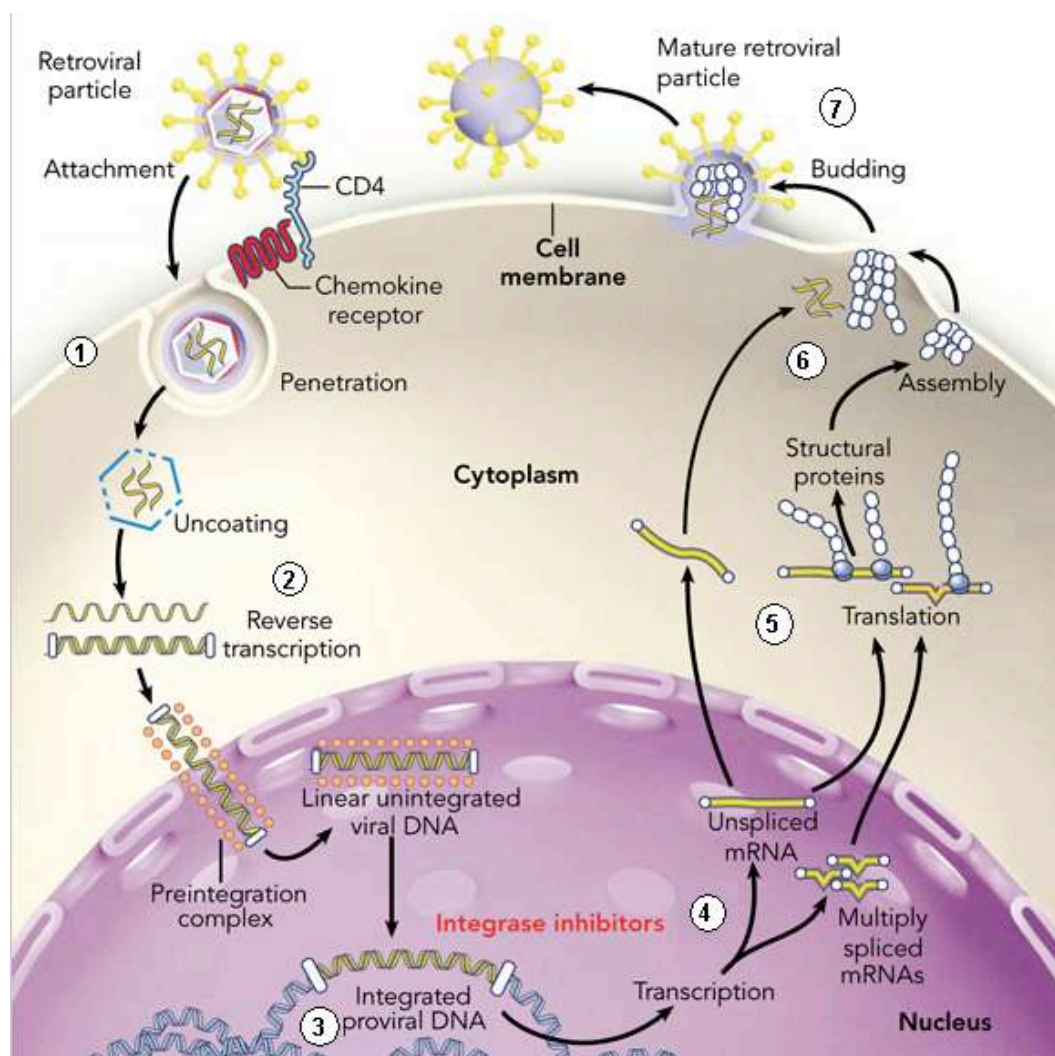
Figure 9. HIV-1 life cycle (Adapted from (57)).

5.1 Attachment to the host cell and uncoating of the virion

HIV-1 can infect DCs that express two receptors: the principal receptor CD4 and an auxiliary co-receptor, which derives from the chemokine receptor family, typically either CCR5 or CXCR4 (244). DCs are one of the first cells encountered by the virus during sexual transmission. They are currently thought to play an important role by transmitting HIV to T-cells when the virus is captured in the mucosa by DCs (318). HIV-1 enters cells through a pH-independent membrane fusion event, which results in release of the core particle into the cytoplasm. However, recent studies now describe HIV-1 entry and membrane fusion

following endocytosis (77, 269). The principal virus protein involved in entry is the envelope glycoprotein. The envelope gene encodes a protein Env that measures 160 kD when fully glycosylated and is divided into two regions: the surface unit gp120 and the transmembrane region gp41 (408). Entry to the host cell begins through interaction of the trimeric envelope complex (gp160 spike) and CD4-CCR5 or CD4-CXCR4 on the cell surface **(Figure 10A)** (54, 408). The first step in fusion involves the high-affinity attachment of the CD4 binding domains of gp120 to CD4. Once gp120 is bound with the CD4 protein, the envelope complex undergoes a structural change **(Figure 10B)**, exposing the chemokine binding domains of gp120 (V3-loop, shown in **Figure 10B et 10C**) and allowing them to interact with the target chemokine receptor **(Figure 10C)** (54, 408). This makes a more stable two-pronged attachment, which allows the N-terminal fusion peptide gp41 to penetrate the cell membrane (54, 408). Heptad repeat (HR) sequences in gp41, HR1 (blue in **Figure 10D**), and HR2 (red in **Figure 10D**) then interact, causing the collapse of the extracellular portion of gp41 into a hairpin **(Figure 10E)**. This loop structure brings the virus and cell membranes close together, allowing fusion of the membranes and subsequent entry of the viral capsid **(Figure 10F)** (54, 408). After HIV-1 has bound to the target cell, the genomic RNA and various enzymes, including reverse transcriptase (RT), integrase (IN) and protease, are injected into the cell (54, 408).
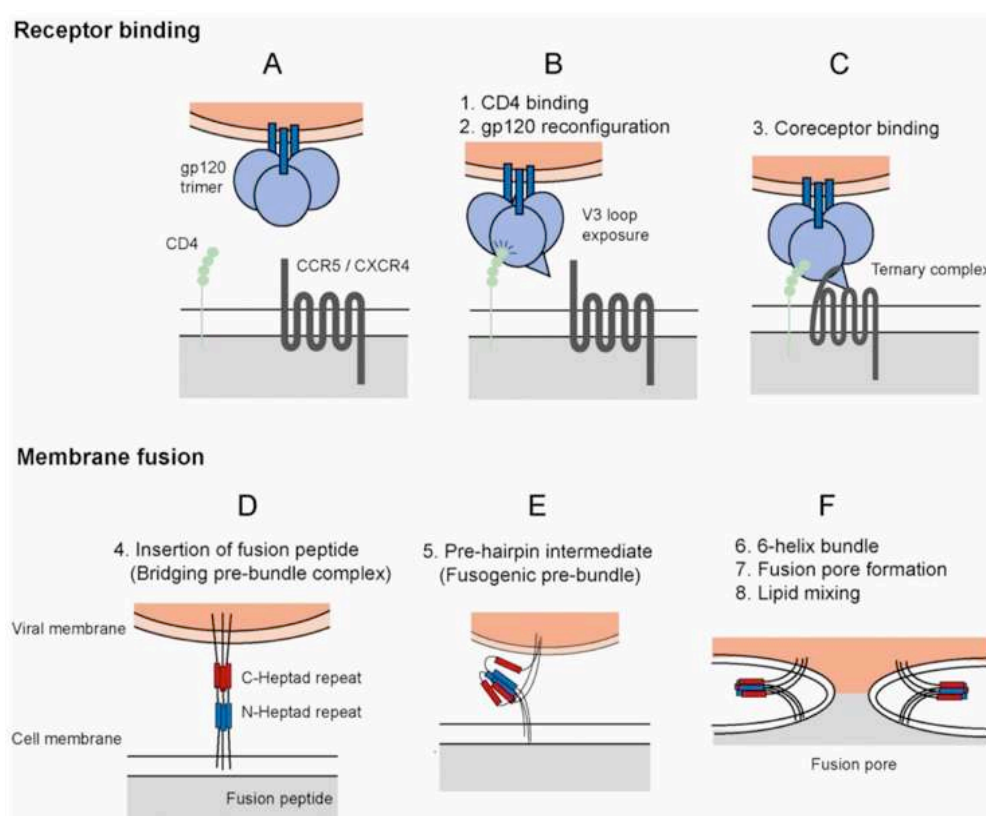
Figure 10. Entry of the HIV-1 into a host cell (244).

5.2 Reverse transcription of genomic RNA

Shortly after the viral capsid enters the cell, the viral RT copies the single-stranded (+) RNA genome into a dsDNA molecule (420). The process of reverse transcription is error-prone, and the resulting mutations may cause drug resistance or allow the virus to evade the body's immune system. The reverse transcriptase also has ribonuclease H (RNase H) activity that degrades the gRNA during the synthesis of cDNA. The reverse transcription of the HIV-1 genome will be developed in details in Chapter 4.

5.3 Provirus integration

At the end of the reverse transcription process, the viral dsDNA is transported into the cell nucleus. The integration of the viral DNA into the host cell's genome is carried out by the viral IN **(Figure 11)** (420). This integrated viral DNA may then lie dormant, in the latent stage of HIV infection (420). To actively produce the virus, certain cellular transcription factors need to be present, the most important of which is NF-κB, which is up-regulated when T-cells become activated (172). This means that those cells most likely to be killed by HIV-1 are those currently fighting infection.
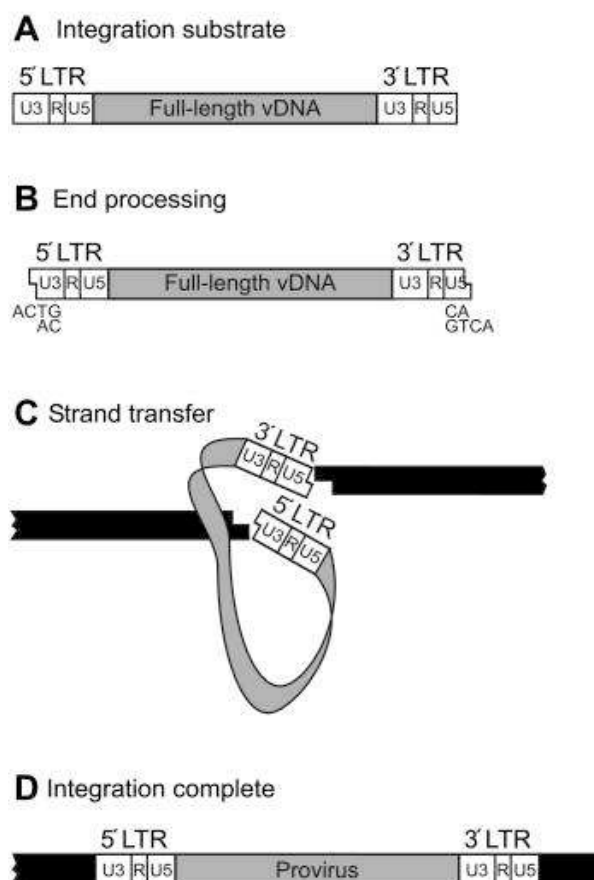
Figure 11. Integration of viral DNA into the host chromosome (368)

5.4 Synthesis and processing of viral RNA

During the late phase of the viral replication, the integrated DNA provirus is transcribed into full-length mRNA. Most HIV-1 strains use four different splice donors and eight different acceptors to produce more than 40 different spliced mRNA species in infected cells (210) **(Figure 12)**.
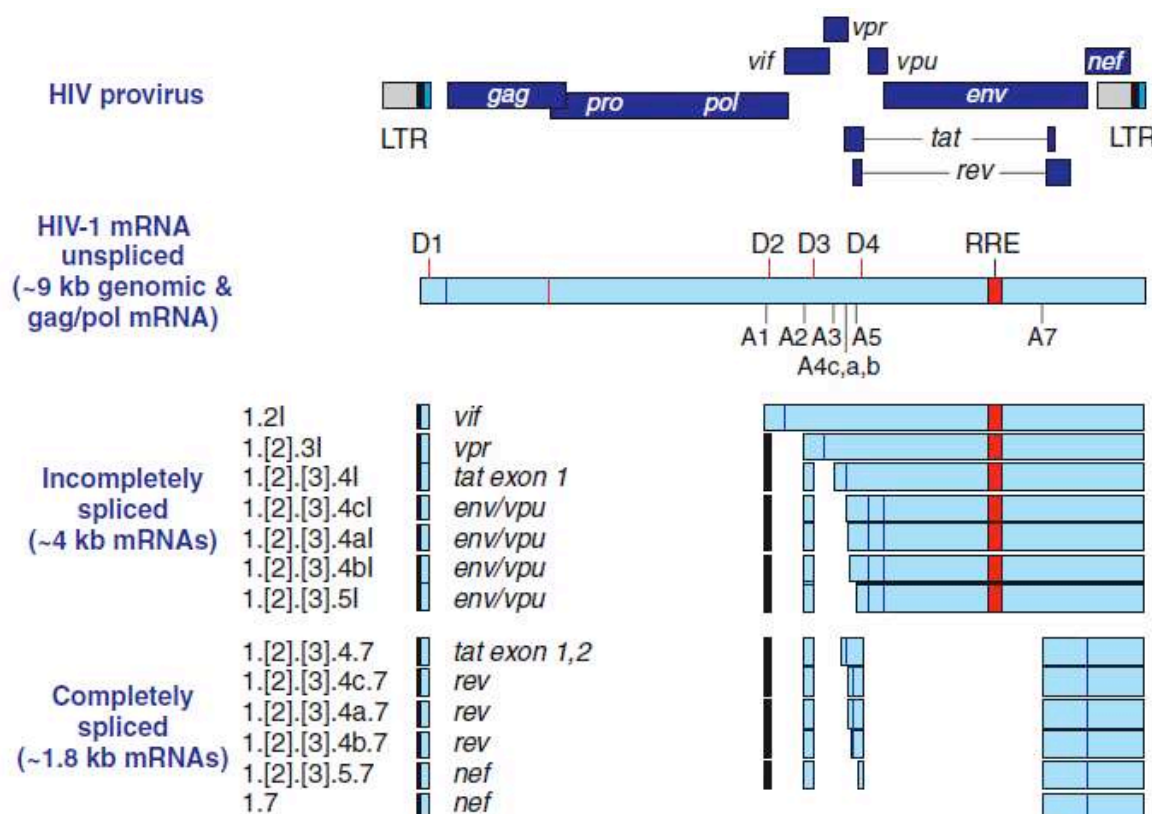
Figure 12. Locations of splice sites in the HIV-1 genome (Adapted from (210)). (Top) Schematic diagram of HIV-1 genome. The dark blue rectangles indicate ORFs and are labeled with the gene names. The LTRs are shown at each edge of the genome: U3-gray, R-black, U5-light blue. Full-length RNA transcripts begin at the 5' end of the R region of the 5' LTR (left) and 3' processing and poly(A) addition takes place at the 3' end of the R region in the 3' LTR (right). (Middle) Locations of 5 splice donors (red bars) and 3 splice acceptors (black bars) in the HIV-1 genome. The splice sites D5 and A6 exist only in the HXB-2 isolate. The location of the RRE is shown by the red rectangle. (Bottom) Incompletely and completely spliced mRNA generated by the splice sites indicated in the unspliced mRNA. The locations of the AUG codons used to initiate protein synthesis are shown as purple bars.

The 3' processing and polyadenylation of HIV-1 pre-mRNAs involves recognition of the AAUAAA and GU-rich motifs duplicated at the ends of the R sequences found in both the 5' and 3' LTRs. HIV-1 uses multiple regulatory elements to direct processing to the 3' LTR cleavage site. First, the HIV-1 U3 sequence, which is upstream of the 3' processing signal, but not associated with the 5' processing signal, contains upstream enhancer elements (USE) that act to facilitate binding of cleavage/polyadenylation specificity factor (CPSF) and

enhance polyadenylation at the 3' end of the HIV-1 transcripts (142). Another USE near the 5' end of the Nef gene binds the cellular SR protein 9G8, which recruits the 3' processing factor CF1m and CPSF (380). Second, the 5' and 3' LTR poly(A) processing sites are imbedded in a region of secondary structure called the poly(A) hairpin located immediately downstream from the TAR hairpin structure. Factors binding to sequences upstream of the AAUAAA site are believed to open up the poly(A) hairpin and allow preferential use of the 3' LTR poly(A) processing site (84). Finally, the splicing factor U1 snRNP acts to inhibit the 3' processing and poly(A) site in the 5' LTR by binding to the adjacent 5' splice donor D1 **(Figure 13)**. Mutations of D1 that weaken binding of U1 snRNP allow the usage of the normally silent 5' LTR poly(A) site (8, 9).
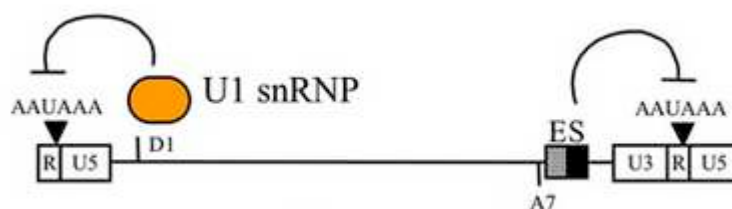
Figure 13. 5' polyadenylation signal is suppressed by interaction of U1 snRNP with D1. 3' polyadenylation site is regulated by the competing inhibitory action of ESS3 (S) and the action of ESE3 (E). (Adapted from (259))

The basal transcriptional activity of HIV-1 is very low and the viral Tat protein and host factors increase the transcription of the viral genome. Tat acts through a *cis*-acting RNA enhancer, the TAR located in the R region of the LTR (17, 42, 138). The HIV-1 RNA then undergoes complex multiple splicing to produce mRNAs for the regulatory/accessory and structural proteins. In the early phase, HIV-1 mRNA is multiply spliced to produce several splice variants ranging from 1.8 to 2 kb in size. They are mainly polycistronic, but produce preferentially Tat, Rev or Nef **(Figure 14A)** depending on the splice acceptor site used (324, 338). These mRNAs are constitutively exported to the cytoplasm and translated. In the late phase, in the presence of Rev, unspliced and incompletely spliced RNAs (4 kb) are exported to the cytoplasm to produce Gag, Pol, Env, Vif, Vpr and Vpu **(Figure 14B)**. Rev and cellular factors work in concert to bring these RNAs to the cytoplasm for translation (259).
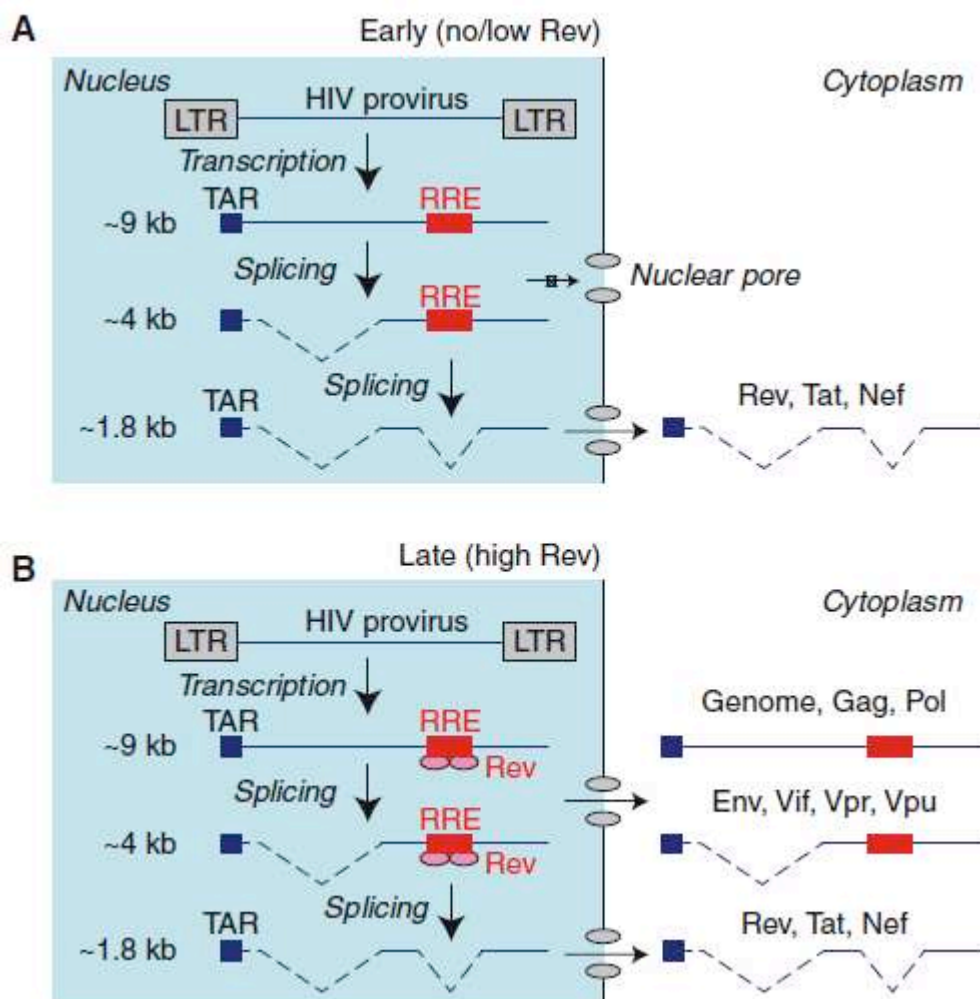
Figure 14. Early and late phases of HIV-1 mRNA expression (210). Full-length unspliced ~9-kb, incompletely spliced ~4-kb mRNA, and completely spliced ~1.8-kb mRNAs are expressed at both early and late times. (A) In the absence of Rev or when Rev is below the threshold necessary for it to function, the ~9-kb and ~4-kb mRNAs are confined to the nucleus and either spliced or degraded. Completely spliced ~1.8-kbmRNAs are constitutively exported to the cytoplasm and translated to yield Rev, Tat, and Nef. (B) When the levels of Rev (shown as a pink oval) in the nucleus exceed the threshold necessary for function, the ~9-kb and ~4-kb mRNAs are exported to the cytoplasm and translated. The RRE is shown as a red rectangle.

5.5 Packaging of genomic RNA

During the late phase of the viral replication cycle, the HIV-1 selectively and efficiently packages two copies of its positive strand, unspliced, 5'-capped, and 3'-polyadenylated RNA genome by a mechanism that has been extensively studied (75, 178, 298, 328). The packaging mechanism efficiently discriminates against the monomeric genome, the spliced viral mRNAs that encode for viral accessory envelope proteins, and the more highly abundant cellular mRNAs (35). Packaging is mediated by the retroviral Gag proteins, which can efficiently assemble in the absence of their native genomes by incorporating an equivalent amount of cellular RNAs (63, 277, 392). Although retroviruses can package essentially any RNA (some

mutants even package ribosomes (278)), RNAs containing the appropriate viral packaging signals are efficiently enriched in assembling virions.

Genome selection appears to proceed *via* the direct binding of the NC domain of Gag to conserved RNA Psi (Ψ) site, which are generally located near the 5' end of the gRNA **(Figure 15)**. As for most other retroviruses, the nucleotides that participate in HIV-1 genome selection appear to reside near the 5' end of the genome and primarily within the 5' UTR (75, 100, 178). Relatively short elements within the 5' UTR that are independently capable of directing heterologous RNAs into assembling virus-like particles (VLPs) have been identified for some retroviruses, but HIV-1 appears to require most of its 5' UTR (2, 238) as well as downstream nucleotides within the gag coding region (48, 248, 301) for optimal packaging efficiency.
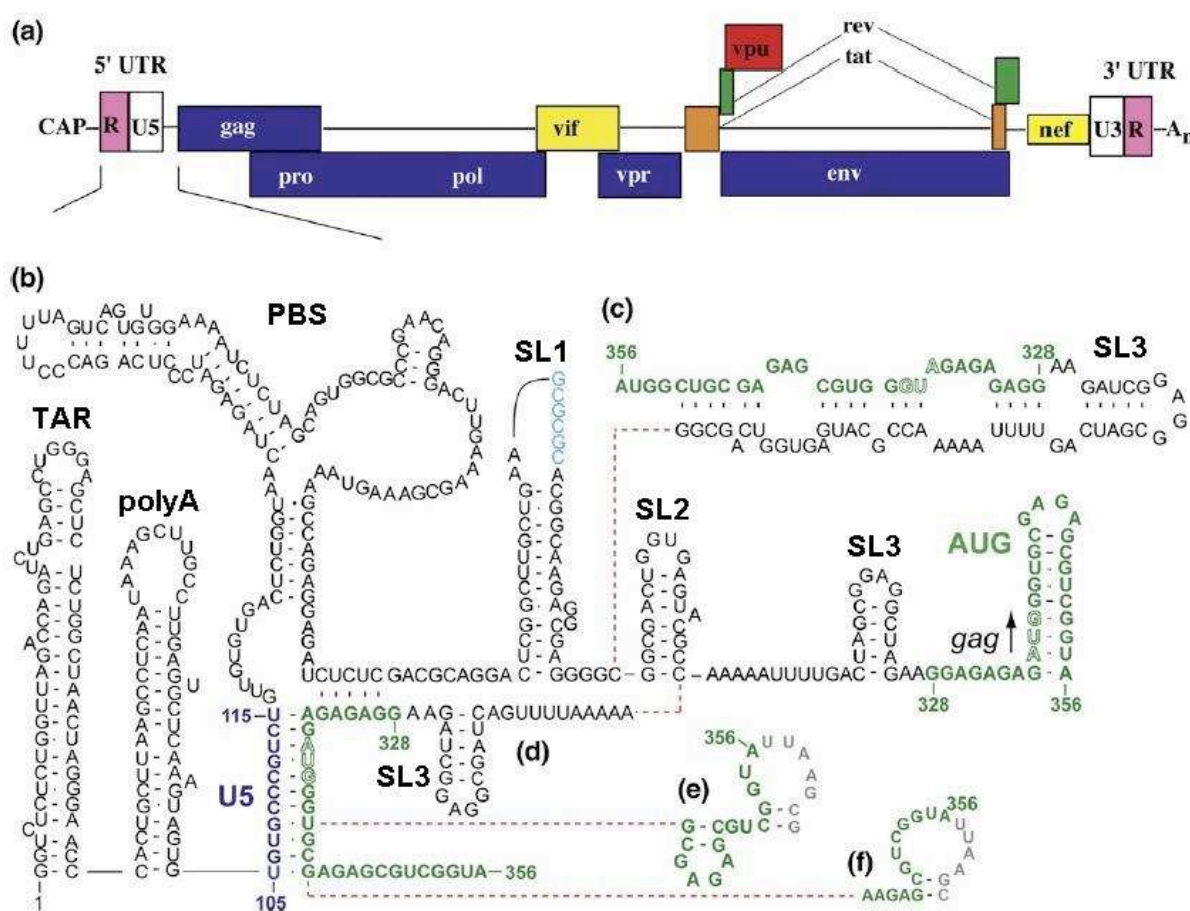
Figure 15. (a) Diagram of the composition of the HIV-1 genome showing locations of the 5' UTR and splice sites. (b–f) Representative secondary structures predicted for the HIV-1 5' UTR. In this figure, variations among the recent predictions for the AUG region (green) are shown. (Adapted from (247))

The 5' UTR is the most conserved region of the HIV-1 genome (220, 297), and in addition to promoting packaging, it also helps regulate or promote transcriptional activation, splicing, primer binding during reverse transcription, and dimerization of gRNA. Although

the dimerization initiation site (DIS, also known as SL1) appears to be a main determinant for RNA packaging, the U5-AUG interaction has been identified as a key player of regulation of genome packaging (1, 246, 350). More recently, Deforges *et al.* (89) proposed that the sequence in the major splice-site donor (SD, known as SL2) may also bring their contribution to this regulation *in vitro*. Furthermore, other studies found that the poly(A) in 5' UTR is a major packaging determinant for both spliced and unspliced viral RNA (100, 178). There is now considerable evidence that genome selection is mediated primarily by HIV-1 The NC domain of Gag **(Figure 16)**.
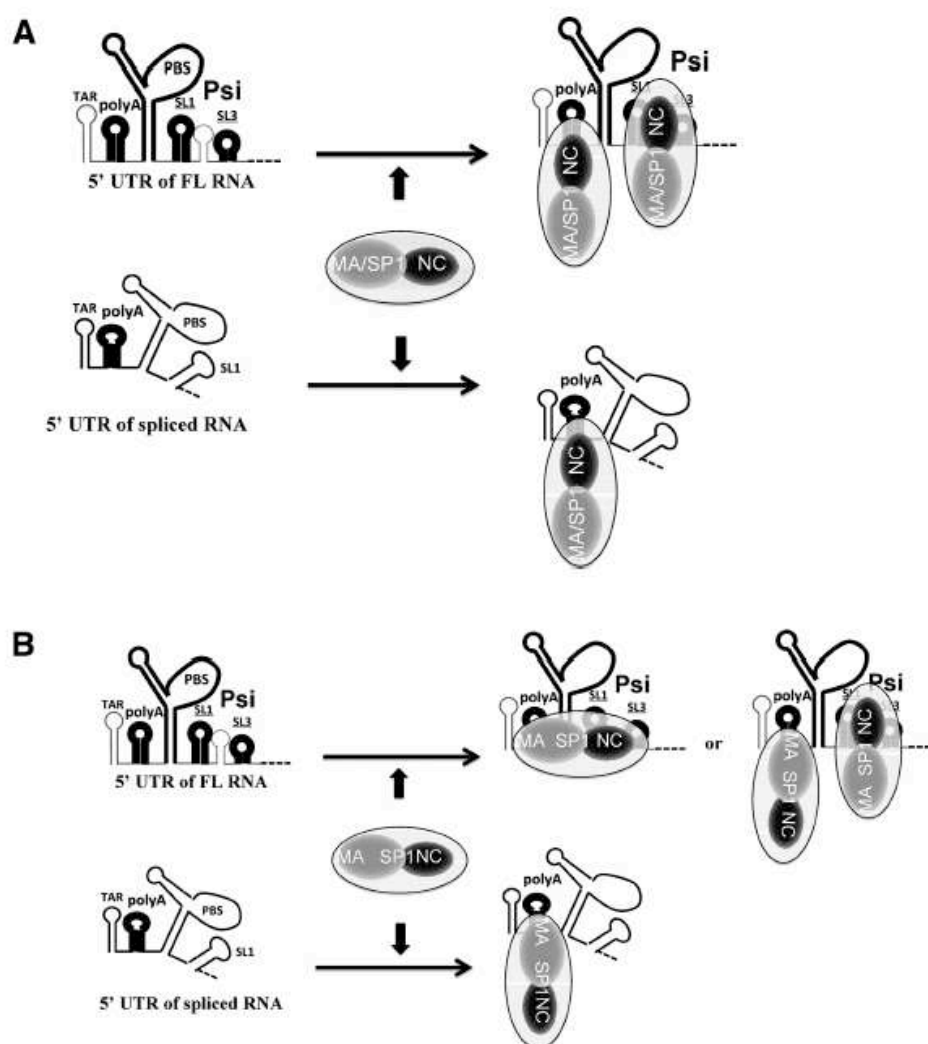
Figure 16. Models for selective encapsidation of viral unspliced and spliced RNAs (100). (A) The full-length RNA (FL RNA) includes the Psi site formed by SL1 and SL3 hairpins that is properly folded and specifically recognized by Gag. The NC domain of Gag is indicated in black and GagMA/SP1 domains in grey. (A) The poly(A) hairpin harbored by FL and spliced RNAs, could also be recognized by the NC domain of Gag, conferring to the FL RNA a second Gag binding site. Thus, FL RNA is more competitive than spliced RNAs, since the FL RNA through its polyA and Psi sites can bind more Gag molecules than spliced RNAs. (B) In this alternative model, GagMA/SP1 domains bind loosely to the poly(A), while the NC domain of Gag, as in (A), tightly binds to the Psi region. These two types of Gag:RNA interaction illustrate how FL RNA is more competitive that spliced RNAs for packaging.

19

5.6 Assembly of genomic RNA and virion proteins

The requirement for two genome molecules is intriguing since all other viruses contain only a single copy of their genetic material (388). Both RNA molecules are utilized for strand-transfer-mediated recombination during reverse transcription (180, 181), but only one DNA allele is generated, and retroviruses are therefore considered "pseudodiploid". Thus, strand transfer-mediated recombination from heterozygotes likely serves as a primary pathway for the rapid evolution of viruses that are resistant to antiretroviral therapies (291). Retroviral genomes exist as weak, non-covalently linked dimers in immature and young virus particles, and the stability of the RNA dimer increases with virus age, which might be important for subsequent reverse transcription events. The genome also appears to play a structural role in virus assembly, although this function can also be achieved by cellular RNAs (277).

The assembly of new HIV-1 virions **(Figure 17)**, begins at the plasma membrane of the host cell. The Env polyprotein (gp160) goes through the endoplasmic reticulum (ER) and is transported to the Golgi complex where it is cleaved by PR and processed into two HIV-1 envelope glycoproteins gp41 and gp120. They are then transported to the plasma membrane of the host cell where gp41 anchors the gp120 to the membrane of the infected cell. The Gag (Pr55) and Gag-Pol (Pr160) polyproteins also associate with the inner surface of the plasma membrane along with the HIV-1 gRNA as the forming virion begins to bud from the host cell.
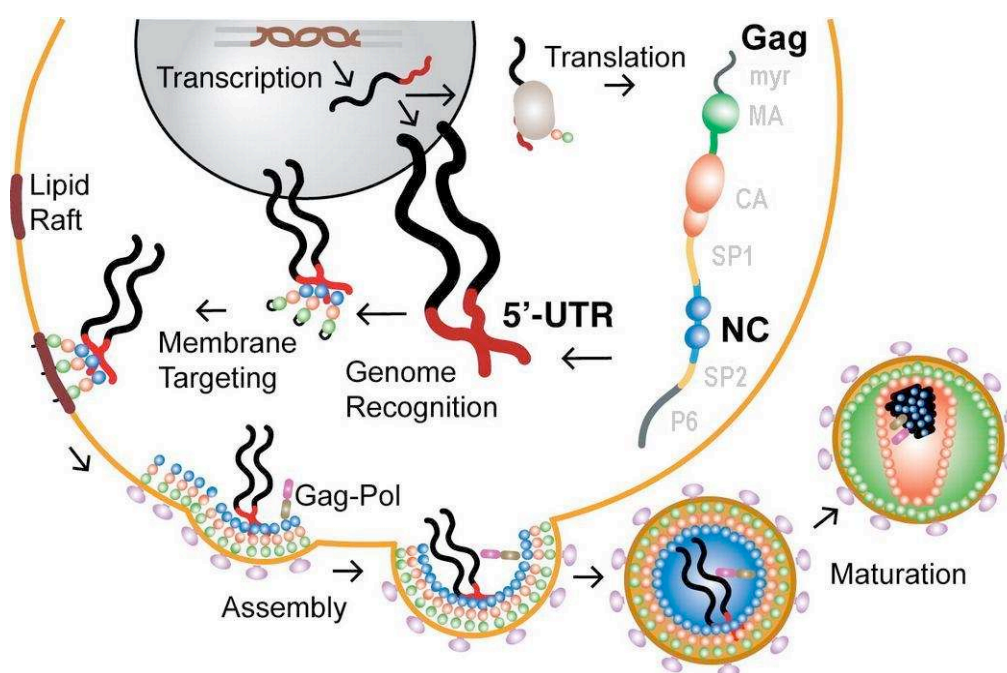
Figure 17. Genome packaging and assembly of HIV-1. (Adapted from (247))

5.7 Maturation of the viral particle

The construction of the spherical virion of HIV-1 driven by the viral *gag* proteins. The precursor proteins of Gag (Pr55) and Gag–Pol (Pr160) are expressed from an unspliced full-length viral RNA, which is targeted to the plasma membrane by their N-terminal myristoylation and the adequate accumulation of precursor proteins drive virion assembly beneath the host cell membrane (132). Many features of the virion assembly, budding, and release processes are largely unknown, but many host cell factors are suggested to support and/or actively contribute to these processes (255). ~1000 – 1500 copies of Gag precursors are required to construct the normal virion (45, 60, 131), but budded particle is not fully infectious at its current state. To acquire infectivity, the virion must undergo a maturation process, which is initiated by the activation of the viral protease (PR) (44). Virion maturation is believed to initiate and complete during or immediately after particle release (49, 207), although it still remains unclear what triggers viral PR activation. The viral PR cleaves the Gag polyproteins into individual functional proteins and enzymes which arrange the unstable gRNA into stable gRNA dimer.

The Pr55 polypeptide is composed of six proteins and peptides, with the N-to-C order as MA-CA-sp2-NC-sp1-p6, thus contains five proteolytic cleavage sites to be processed by PR. It was previously demonstrated that the processing rates of the five cleavage sites were not equal (307). The processing rate of the cleavage site between SP1 and NC (SP1/NC) is the fastest, while SP2/p6 and MA/CA are the second and the third, respectively. The processing rates of the two remaining sites (CA/SP2 and NC/SP1) are much slower than those of the three aforementioned sites **(Figure 18)**.

During the virion maturation process, stabilization of the RNA dimer primes during the primary cleavage (SP1/NC) of Pr55^{Gag}. However, the primary cleavage alone is insufficient and the ensuing cleavages are required for complete the uniform dimerization of viral RNA. During proteolysis, HIV-1 NC exists in two intermediate forms, NCp15 (partial cleavage product containing NC/SP2/p6) and NCp9 (partial cleavage product containing NC/SP2) and the fully processed form, NCp7. All three of these proteins exhibit nucleic acid chaperone activities (74). An extended structural model of the HIV-1 Gag polypeptide has been proposed from high-resolution structures of isolated domains **(Figure 22B)**. SP1-NC cleavage by PR rapidly separates the MA-CA shell and the nucleocapsid complex formed between RNA, NCp15, RT and IN. NCp15 processing by PR into NCp9 and finally NCp7 leads to NC:RNA co-aggregation/condensation within the confines of the capsid cone. NCp7 is

followed by the p6 domain. This later p6 domain is required for particle budding during which the viral particles pinch-off from the cellular membrane (90). The p6 domain contains a Proline-rich and a di-Leucine domains, which are the target of the cellular proteins Tsg101 and ALIX, respectively, involved in the cellular class E protein sorting pathway and the HIV-1 budding machinery (90, 357, 389).
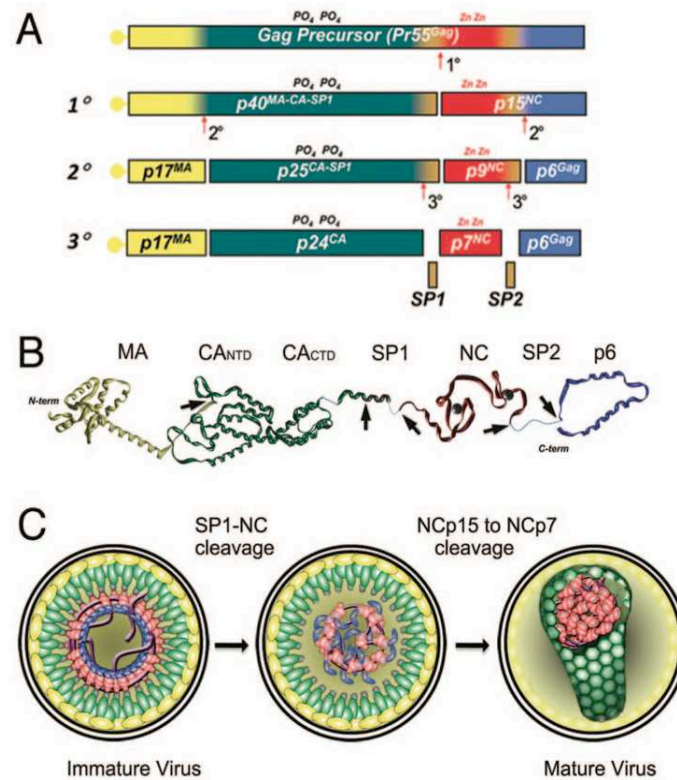
Figure 18. Maturation of Gag proteins in HIV-1 (Adapted from (267)).

Chapter 3. Proteins involved in HIV-1 DNA synthesis

1. <u>Cellular proteins</u>

Several cellular proteins have roles in the replication cycle of HIV-1 **(Figure 19)**. The infection begins when Env glycoprotein spikes engage the host receptor complex CD4-CCR5 (step 1), leading to fusion of the viral and cellular membranes and entry of the viral particle into the cell (step 2). Partial core shell uncoating (step 3) facilitates reverse transcription (step 4), which yields the pre-integration complex (PIC). Following import into the cell nucleus (step 5), PIC-associated integrase orchestrates the formation of the integrated provirus, aided by the host chromatin-binding protein lens epithelium-derived growth factor (LEDGF) (step 6). Proviral transcription (step 7), mediated by host RNA polymerase II and positive transcription elongation factor b (P-TEFb), yields viral mRNAs of different sizes, the larger of which require energy-dependent export to leave the nucleus via host protein CRM1 (step 8). These mRNAs serve as templates for protein production (step 9), and the full-length RNA is incorporated into viral particles with protein components (step 10). Viral-particle budding (step 11) and release (step 12) from the cell is mediated by ESCRT complexes and ALIX, and is accompanied or soon followed by PR-mediated maturation (step 13) to create an infectious viral particle. Each step in the HIV-1 replication cycle is a potential target for antiviral intervention (110).
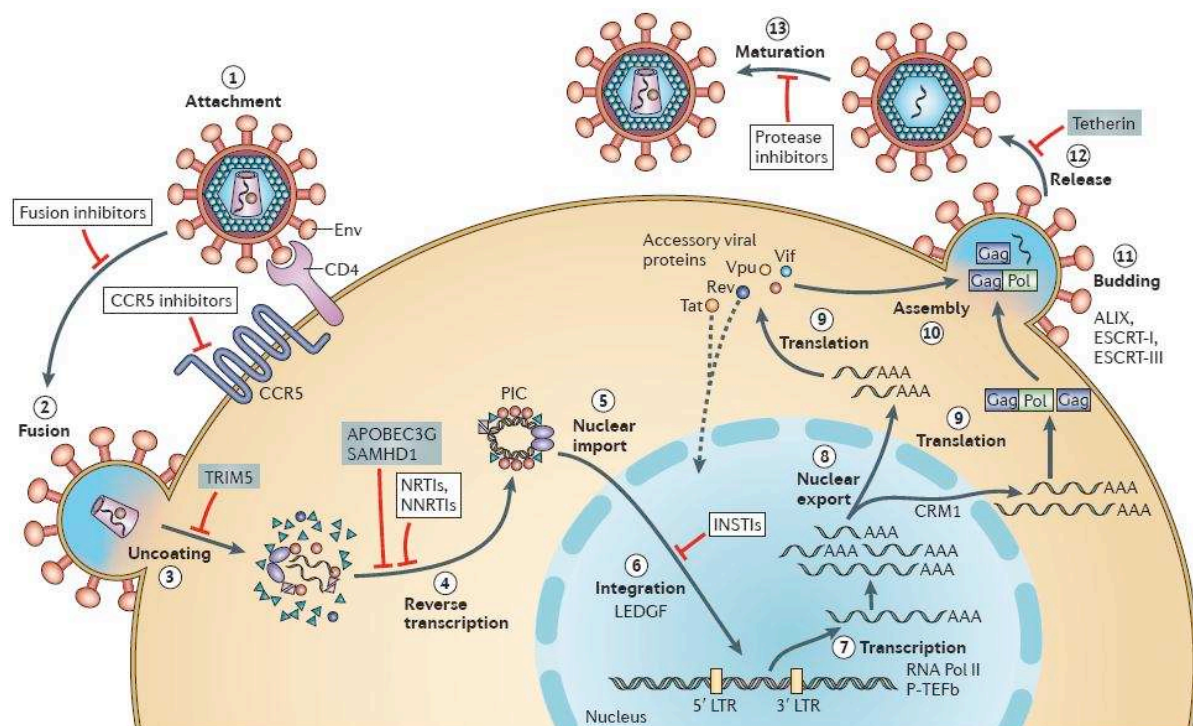
Figure 19. Roles of the host proteins in the replication cycle of HIV-1 (110). The sites of action of clinical inhibitors (white boxes) and cellular restriction factors (blue boxes) are indicated.

TRIM5

HIV-1 infects humans and chimpanzees but not Old World Monkeys (OWMs) such as the rhesus monkey (Rh) and cynomolgus monkey (CM) (359). HIV-1 effiently enters cells of OWMs but encounters a block before reverse transcription. Importantly, resistance against HIV-1 infection was shown to be dominant in heterokaryons between human and OWM cells, suggesting the presence of inhibitory factor(s) against HIV-1 infection but not for SIV in OWM cells (272). In 2004, the screening of an Rh cDNA library identified tripartite motif 5α (TRIM5α) as a species-specific cellular antiviral factor (359) **(Table 2)**. TRIM5 present in the cytoplasm recognizes motifs within the capsid proteins and interferes with the uncoating process, therefore preventing successful reverse transcription and transport to the nucleus of the viral genome (339, 360).

Table 2. Species-specific restriction by TRIM5α (281).

TRIM5α mediated viral restriction						"Yes" denotes restriction. "Weak" denotes weak restriction. "No" denotes no restriction. "N.D." denotes no result has yet been published. SIVmac, simian immunodeficiency virus isolated from a macaque; SIVagm, simian immunodeficiency virus isolated from an African green monkey; N-MLV, N-tropic murine leukemia virus; B-MLV, B-tropic murine leukemia virus.
TRIM5α	HIV-1	SIVmac	SIVagm	N-MLV	B-MLV	
Human	No	No	No	Yes	No	
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 is one of splicing variants produced by *TRIM5* gene and TRIM5 proteins are members of the TRIM family containing RING, B-box 2, and coiled-coil domains **(Figure 20)**. The RING domains possess E3 ubiquitin ligase activity (193). TRIM5α has been shown to form a dimer (208, 228). The B-box 2 domain mediates higher-order self-association of Rh-TRIM5α oligomers (98, 241). The coiled-coil domain of TRIM5α is important for the formation of homo-oligomers (268), and the homo-oligomerization of TRIM5α is essential for antiviral activity (196, 280). Among TRIM5 splicing variants, TRIM5α alone has an additional C-terminal PRYSPRY domain **(Figure 20)**. TRIM5α recognizes the multimerized CA proteins (viral core) of an incoming virus by its PRYSPRY domain and is thus believed to control retroviral infection.
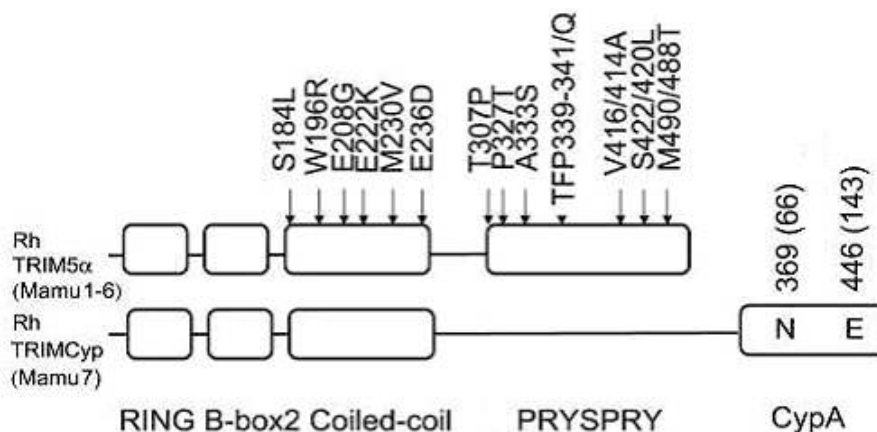
Figure 20. Domains of rhesus monkey (Rh) TRIM5α and TRIMCyp proteins (281). The RING, B-box2, Coiled-coil and PRYSPRY domains of Rh-TRIM5α are shown by squares. Polymorphisms are shown outside the squares. The numbers in parentheses show the amino acid positions counting from the initiation methionine codon of the CypA ORF.

It has also been reported that some Rh and CM individuals have retrotransposed cyclophilin A ORF in the *TRIM5* gene (290, 334), which produces TRIM5-cyclophilin A fusion protein (TRIMCyp). TRIMCyp was originally identified as an anti-HIV-1 factor of New World owl monkeys (NWM) (374). The interaction between HIV-1 CA and CypA can be inhibited by cyclosporine A. This is a very interesting example of a gain-of-function by retrotransposition. The owl monkey has been shown to express only TRIMCyp but not TRIM5α.

Even moderate overexpression of wild-type human TRIM5 provides substantial reduction of HIV-1 infection (212). It is conceivable that the alleles regulating TRIM5α protein levels can have a major effect on HIV-1 acquisition or HIV-1 replication. Although TRIM5 missense polymorphisms explain little of the variation in European descent populations tested to date, TRIM5 expression could be important in HIV acquisition.

APOBEC3

The mammalian Apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3 (APOBEC3) protein family has emerged as a key mediator of intrinsic restriction to retroviruses, including HIV, and to hepatitis B virus (61). APOBEC3 proteins, a class of cytidine deaminase enzymes, are incorporated into virions. APOBEC3G enzymes, with two zinc finger domains, edit newly synthesized viral DNA by deaminating dC to dU, resulting in lethal G-to-A hypermutations. The antiviral activity of APOBEC3 can also be deaminase-independent through interference with virus transcription or integration (38, 173, 288). HIV-1 has clearly out-maneuvered its host – HIV-1 encoded viral infectivity factor (vif) targets

APOBEC3G and 3F for proteasomal degradation through a ubiquitination pathway that involves the host proteins Cullin5, elongins B and C, and Rbx1 (412). Seven APOBEC3 family genes (A, B, C, DE, F, G and H) are located in a tandem array on chromosome 22q12–q13.2 that spans 150 kb. Their anti-HIV activity and sensitivity to degradation mediated by HIV vif is variable **(Table 3).**

Table 3. Anti-HIV-1 profile of APOBEC3 proteins (4).

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

Among all APOBEC3 proteins, APOBEC3G has arguably the strongest antiviral effect. The impact of the APOBEC3G expression levels on HIV-1 infection and the severity of HIV disease has been investigated. Most studies reported a positive correlation between expression levels and favorable outcomes (37, 101, 200, 296). Higher APOBEC3G expression in PBMCs and cervical tissues was observed in HIV-exposed seronegative women; their PBMCs were also resistant to infection with a macrophagotropic HIV-1 strain (R5) (37). APOBEC3G mRNA levels range from high to low in non-progressors > exposed uninfected > progressors, respectively, and were inversely correlated with virus load and positively correlated with CD4+ T-cell levels (199). Taken together, the evidence indicates that APOBEC3 genetic variation and expression levels modify HIV-1 acquisition and disease progression *in vivo*, suggesting that delivery of vif-resistant APOBEC3 protein to HIV-1 susceptible cells could potently restrict HIV-1 replication.

In cells infected with ΔVif HIV-1, APOBEC3G is not degraded and is effectively incorporated into the budding virus and transferred to the next target cell, where it exerts antiviral effects at multiple levels. APOBEC3G can inhibit the elongation of the RT in a deaminase-independent manner **(Figure 21-1)**. Presumably, APOBEC3G binds directly to the viral RNA and thereby impairs the movement of the RT along the RNA template. Most importantly, APOBEC3G can trigger massive deamination of dC to dU during synthesis of the viral minus-strand DNA **(Figure 21-2)**. During synthesis of the plus-strand DNA, adenosines are incorporated instead of the original guanines, resulting in G-to-A mutation. Finally, APOBEC3G also inhibits viral DNA integration and provirus formation, probably by inducing defects in tRNA cleavage during second strand transfer, leading to the formation of aberrant viral DNA ends **(Figure 21-3)**. APOBEC3G also interacts with the HIV-1 IN, which

might interfere with the integrity of the integration complex, resulting in diminished integration rates.
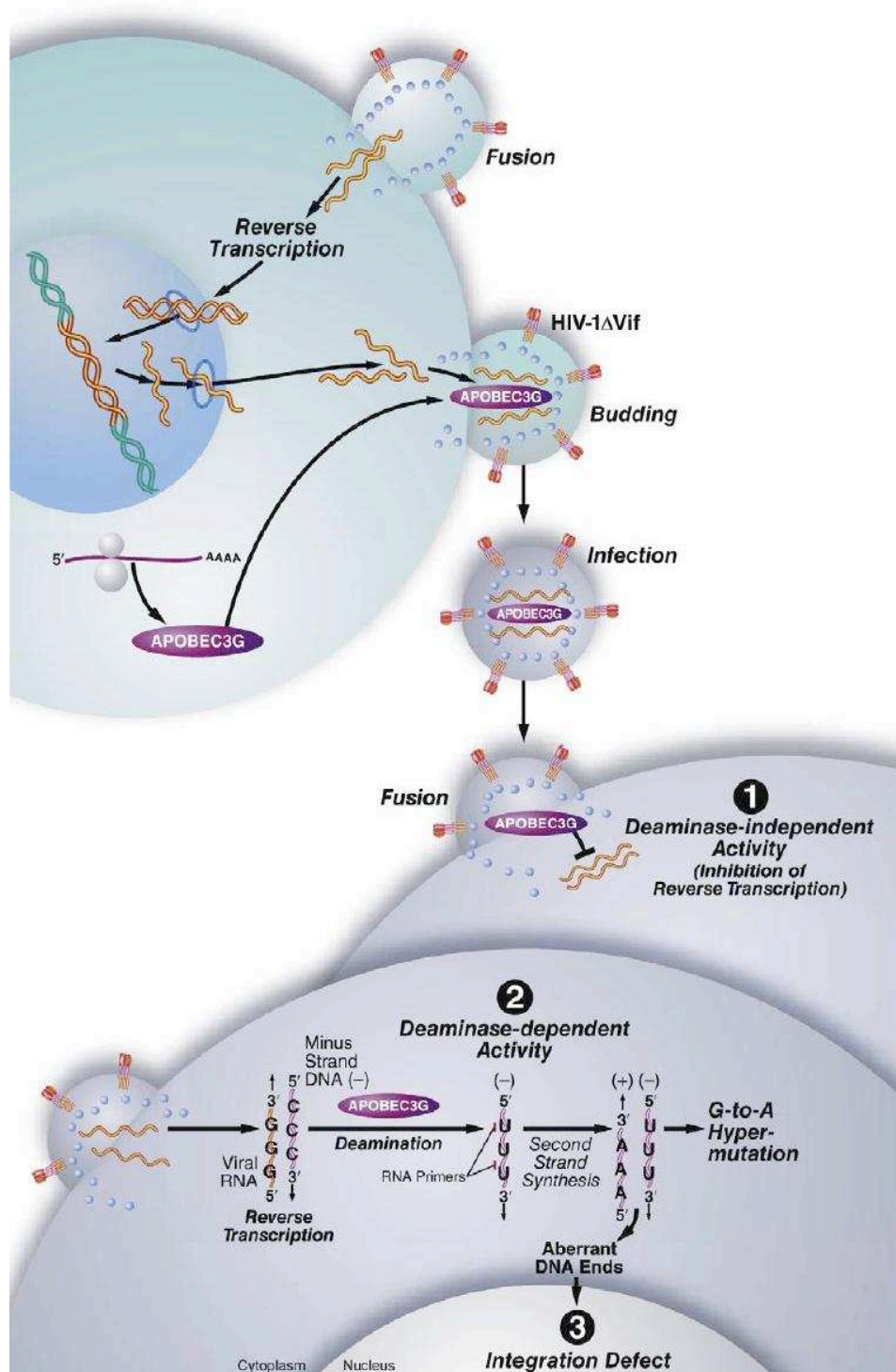
Figure 21. The impact of APOBEC3G on the life cycle of ΔVif HIV-1 (402).

2. HIV-1 proteins

The viral proteins **(Table 4)** are synthesized from more than 40 mRNAs, which are all derived from the same primary transcript **(Figure 12)**.

27

Table 4. HIV-1 proteins (source: http://www.hiv.lanl.gov/)

Name	Size	Function	Localization
Gag			
MA	p17	membrane anchoring; env 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 glycoproteins bind to CD4 and secondary receptors	plasma membrane, virion envelope
Tat	p16/p14	viral transcriptional transactivator	primarily in nucleolus/nucleus
Rev	p19	RNA transport, stability and utilization factor (phosphoprotein)	primarily in nuleolus/nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity	cytoplasm (cytosol, membranes), virion
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 protein
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

2.1 Viral envelope (Env)

The envelope (Env) spikes on HIV-1 define the viral tropism, mediate the fusion process and are the prime target of the humoral response. Three gp120 subunits comprise the 'head' of Env and three gp41 subunits comprise the 'stalk' and other membrane-associated elements. The mature gp120-gp41 proteins are bound by non-covalent interactions and are associated as a trimer on the cell surface. The gp120 moiety has five hypervariable regions, designated V1 through V5. V3 loop is not involved in CD4 binding, but is rather an important determinant of the preferential tropism of HIV-1 for either T lymphoid cell lines or primary macrophages (187). Sequences within the V3 loop interact with the HIV-1 co-receptors CXCR4 and CCR5, which partially determine the susceptibility of cell types to given viral strains (91). The V3 loop is also the principal target for neutralizing antibodies that block HIV-1 infectivity.

2.2 Structural proteins

2.2.1 *Matrix (MA)*

MA, which forms the N-terminal domain (NTD) of the Pr55^{Gag} precursor, has been implicated in the targeting of Gag to the plasma membrane, Env glycoprotein incorporation into virions, and early postentry events. The structure of the HIV-1 MA protein has been determined by both NMR spectroscopy and X-ray crystallography (69, 171, 256). The protein folds into a compact core domain, consisting largely of α-helices and a three-stranded β-sheet. The C-terminal ~30 residues of MA form an α-helix which may serve to connect MA and CA domains in Pr55^{Gag} **(Figure 18B)**. In the mature HIV-1 virus, MA is located under the virion envelope, which derives from the infected cell membrane. MA is myristoylated and contains basic amio acids within its N-terminus. Membrane binding is mediated by insertion of the myristoyl group into the lipid bilayer and by the basic patch, which binds acidic phospholipids, particularly PI(4,5)P$_2$, a phosphoinosotide that is concentrated in the plasma membrane **(Figure 22)** (273).
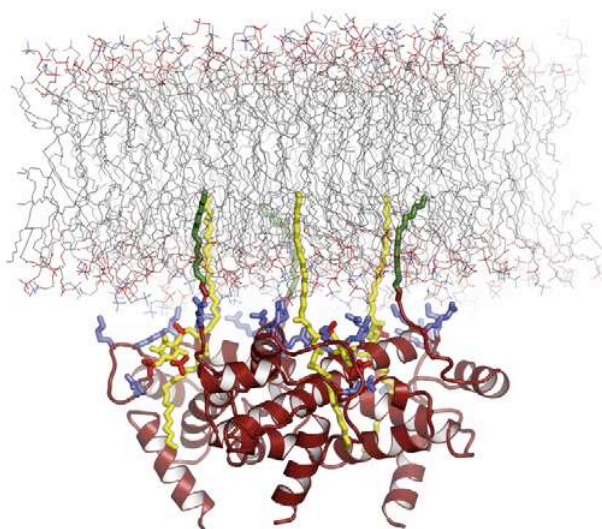
Figure 22. Models for membrane binding by the N-terminal MA domain showing a side view of the MA trimer bound to a lipid bilayer with the myristoyl chain (colored in green) inserted into the inner leaflet, and basic residues (blue) interacting with acidic phospholipid headgroups, including PI(4,5)P$_2$ (yellow) (132).

2.2.2 *Capsid (CA)*

In the mature virion, CA forms a shell surrounding the viral RNA genome and core-associated proteins. The CA domain of Pr55^{Gag} plays an important role in virus assembly, and the mature CA protein functions in virion maturation and also appears to be involved in early postentry steps. The CA protein is composed of two domains: an N-terminal region (the so-

called 'core' domain, composed of residues 1–145), which functions in virion maturation and incorporation of the cellular protein CypA, and a C-terminal 'dimerization' domain (residues 151–231), which contributes to Gag-Gag interactions **(Figure 23)**. The core domain is highly helical, being composed of seven α-helices, two β-hairpins, and an exposed loop. The exposed loop serves as the binding site for CypA (130).
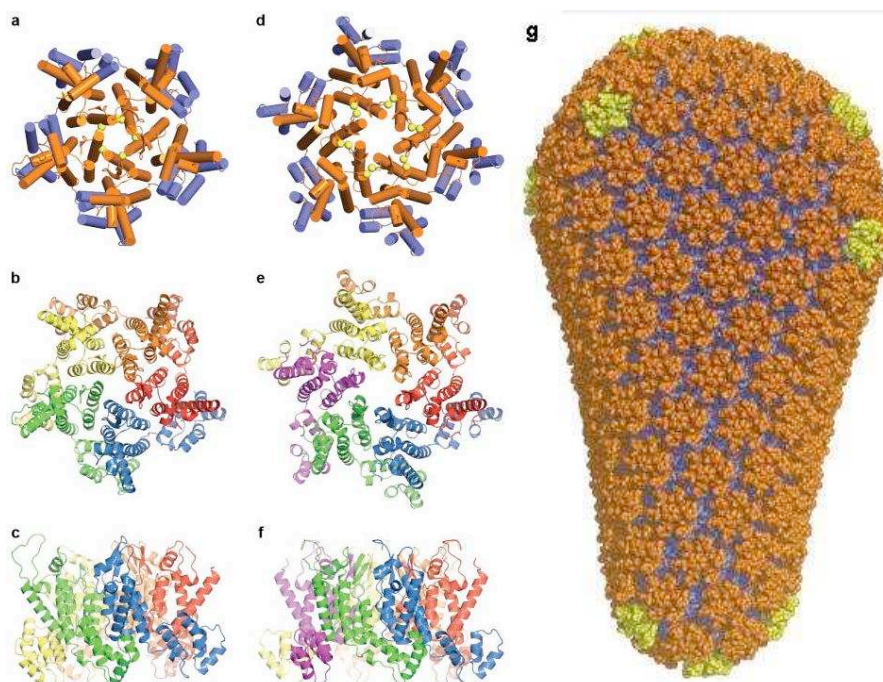
Figure 23. Model of capsid (Adapted from (319)). (a) Top view of the pentamer, with the NTD colored in orange and the CTD in blue. Helices are represented as cylinders. (b) and (c) Top view (b) and side view (c) of the pentamer, with the helices as ribbons. Each subunit is in a different color. (d), (e) and (f) Equivalent views of the hexamer. The yellow spheres in (a) and (d) indicate the positions of the pentamer-stabilizing (N21C/A22C) and hexamer-stabilizing (A14C/E45C) disulfide bonds, respectively. (g) Stereoview of a backbone-only fullerene cone model composed of 1,056 CA subunits. The hexamers, pentamers, and dimers are colored in orange, yellow, and blue, respectively.

2.2.3   *Nucleocapsid protein (NC)*

2.2.3.1   *Structure of NC*

The HIV-1 NC is a small basic nucleic acid-binding protein generated by the proteolytic cleavage of the Gag precursor. PR-directed cleavage of Gag **(Figure 18A)** first generates NCp15, next NCp9 (1-71) and ultimately NCp7 (currently named as NC) composed of 55 amino acid residues, containing two highly conserved CCHC zinc fingers that coordinate zinc ions with high affinity **(Figure 24)**.

2.2.3.2   *Properties of NC*

About 1000-1500 NC molecules coating the gRNA in a dimeric form (45, 56, 60) are

found within the infectious HIV-1 particles. The NC:gRNA ratio is therefore about 1 to 12-18 nucleotides.
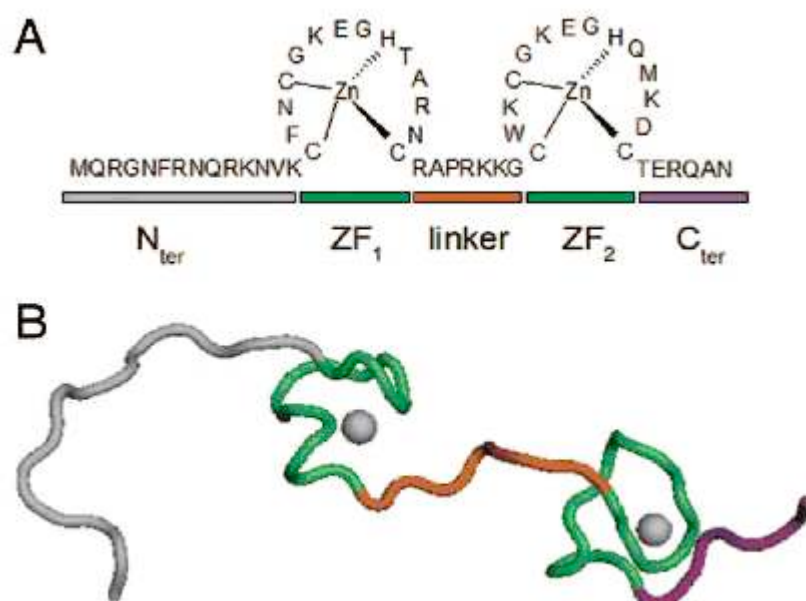
Figure 24. Sequence (A) and 3D structure (B) of NC (144). Two zinc finger domains (ZF_1 and ZF_2) are in green, separated by a highly basic linker (orange), and flanked by poorly folded N-terminal (grey) and C-terminal (deep purple) domains.

In vitro NC is endowed with nucleic acid (NA) binding, condensing, annealing, and strand transfer activities. NC is a NA chaperone (73, 336, 377), which is the key property and means that NC can remodel NA structures so that the most thermodynamically stable conformations that have the maximal number of base pairs (16, 73, 239) are gained **(Figure 25)**. Two main components account for the NC chaperone activities: 1) the transient destabilization of the nucleic acid secondary structure upon binding of NC molecules; 2) the NC-promoted annealing of complementary nucleic acid sequences (144).
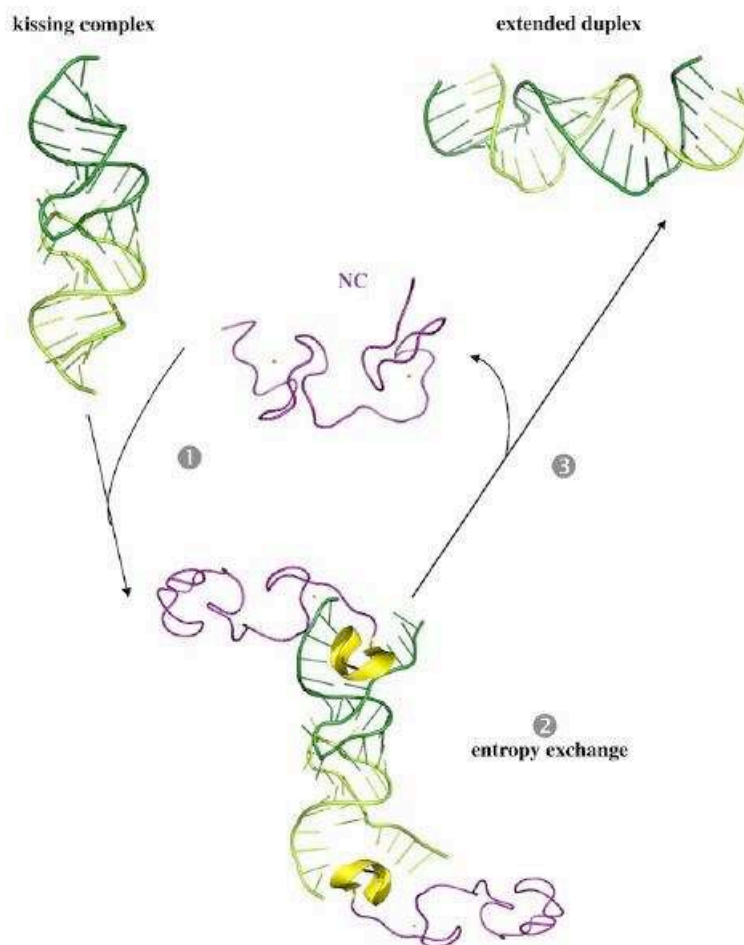
Figure 25. Viral NA structure remodeling by NC (82). Schematic model of the entropy exchange process between NC and the dimer initiation sequence of the HIV-1 gRNA. (1) Intrinsically unstructured NC molecules bind to the kissing complex form of the RNA dimer. (2) The binding is accompanied by mutual conformational rearrangements, leading to partial folding of NC (α helix in yellow) and partial melting of the RNA stem structures. (3) NC is released, allowing the RNA to conduct a conformational search, which - after successive cycles of binding and release (the on and off) - ultimately leads to the extended duplex conformation.

2.2.3.2.1 *Nucleic acid destabilization activity*

The capability of HIV-1 NC to destabilize NAs is mainly driven by its ZFs. The destabilization activity is dependent of the stability of the NA sequences. The motifs, such as bulges, internal loops, mismatches and terminal base pairs are likely important to ensure a selective destabilization by NC (144, 239). The NC-induced NA destabilization initiates preferentially at the single-stranded region. NC then opens the neighbouring base pairs in order to help the initial destabilization to propagate. As a consequence, this may allow the sliding of NC and/or the binding of additional NC molecule (144). Thus, the destabilization activity induces the secondary structure fluctuations and ultimately leads to the reactive

species required for the specific intra- or inter-molecular annealing processes occurring during the viral life cycle.

2.2.3.2.2 Nucleic acid aggregation activity

HIV-1 viral assembly and replication require multiple annealing reactions involving a large number of NA sequences. The ability of NC to non-specifically aggregate NAs (234)facilitates the nucleation step of the annealing reaction. Nucleation is a diffusion-limited association, which is slowed down by the electrostatic repulsion between the annealing strands and by the low probability of the correct positioning of nucleotides for annealing (315, 316). Multiple studies suggest that NC remains highly mobile when bound to NAs (371-373, 404). High mobility of multivalent cationic ligands in their NA-bound state is a key feature of efficient aggregating or condensing agents (166, 167, 217). NC is highly cationic with its N-terminal responsible for the strong NA aggregating properties. The NTD are unstructured in free NC, but form α-3_{10} helix upon NA binding (239). Together with the N-terminal tail, the basic residues in the linker region between the two NC ZFs contribute to robust and dynamic binding/dissociation interaction with nucleic acids, which ensures NAs aggregation (74, 355). Although NC preferentially binds to single-stranded G-rich sequences through stacking interactions via hydrophobic residues (29, 117, 239), NC's non-specific NA binding and aggregating ability are an essential part of its chaperone function.

2.2.3.2.3 Nucleic acid structure and NC binding properties

The binding constants of NC for NAs strongly depend on the nature, the sequence, and the folding of the interacting sequences and can thus vary by several orders of magnitude. This strong variation in the binding constants confers on the protein the ability to exert different functions, depending on the nature of the interacting NA sequences and the respective concentrations of the protein and the NA sequences (82). Studies based on NMR and isothermal titration calorimetry (ITC) showed that the RNA apical loops displaying the GNG sequences constitue high-affinity binding sites for NC (3, 86). An early surface plasmon resonance (SPR) experiment showed that in a moderate salt concentration (0.15 M NaCl) NC binds preferentially to the single-stranded DNA (117). The data obtained by a study based on single molecule stretching were consistent with the preferential binding of NC to the single-stranded DNA (401). However, in 10 mM sodium phosphate (pH 7.0), the affinity of NC for the double-stranded DNA is only slightly smaller in magnitude than the affinity for the single-

stranded DNA (378). NC also exhibits sequence-specific binding properties to a number of defined single-stranded sequences. These specific and strong binding properties play notably a critical role in the recognition, by the NC domain of Gag, of the encapsidation signal of the gRNA, enabling its specific recognition and selection among a large excess of cellular RNAs during virus assembly (2, 62, 238, 273, 274).

NC was also found to bind with high affinity to single-stranded TG, GNG, and TNG motifs (10, 11, 21, 29, 116, 117, 390) **(Table 5)**. These motifs are specifically recognized by NC in two viral (-) DNA stem-loops, namely, the (-)primer binding site (PBS) and the (14-39) complementary trans-activation response element (called mini-cTAR) (21, 40).

A model was proposed to describe the formation of specific complexes between NC and its RNA and DNA targets (21, 82) **(Figure 26)**. In this model, the interaction is initiated by the stacking between the W37 residue of the C-terminal ZF and a guanine residue **(Figure 26A-b and 26B-b)**. In the NC:DNA complexes, this primary contact positions the hydrophobic residues of NC (T24, F16, M46, and Q45) for hydrophobic interactions with the C2'atoms of the sugars **(Figure 26A-c)**, which in turn lead to the stretching and deformation of the phosphodiester backbone **(Figure 26A-c and 26D-b)**. Moreover, this modification of the phosphodiester backbone is accompanied by a weakening of the stacking interactions between the bases of the tetranucleotide sequence interacting with NC. In the NC:RNA complexes, the presence of the hydroxyl group at the C2'position (violet position in the sugars colored black) was found to prevent the interaction of the hydrophobic residues (F16 in the NC:SL2 complex) with the C2'atoms of the sugars at the expense of sugars at the O4' (red atoms in the sugar colored black) and O5' positions **(Figure 26B-b and 26B-c)**. In the NC:RNA complexes, it is mainly the O4'/O5'/C5'/C4'atoms that are involved in the contacts, while in the NC:DNA complexes, it is the C2'/C1'atoms. Note also that the number of protein residues involved in these contacts is larger in the NC:DNA complexes than in the NC:RNA complexes. As a consequence, even if the sequence preference rules and binding mode appear similar for DNA and RNA sequences, with notably an insertion of a G residue in the hydrophobic cleft formed on the C-terminal ZF, the RNA sequences were found to bind NC with an opposite polarity as compared to DNA sequences (21). This underlines the important role of the sugar-protein contacts in the determination of the NA binding polarity of NC. Moreover, in contrast to the DNA sequences, NC was found to only slightly destabilize and stretch the contacted RNA sequence and, thus, induce only limited changes in the stacking interactions between the contacted RNA bases **(Figure 26B)**.

Table 5. NC affinities for various target oligonucleotide sequences (82).

Name	Sequence	NA	NC	K_{obs} (nM)	Salt conditions	Method	References
SL3 (SL)	ggaCUAGC**GGAG**GCUAGucc	RNA-14nt RNA-20nt	NCp7	21–28	200 mM NaCl	TFQ	Paoletti et al.[91] Shubsda et al.[93] Athavale et al.[82]
	GGACUAGC**GGAG**GCUAGUCC	RNA-20nt	NCp7	100	25 mM CH₃COONa 25 mM NaCl	Native GE	De Guzman et al.[57]
	GGACTAGC**GGAG**GCTAGTCC	DNA-20nt	NC(1–72)	70	100 mM NaCl	TFQ	Vuilleumier et al.[94]
	GGACTAGC**GGAG**GCTAGTCC	DNA-20nt	NC(12–53)	1000	100 mM NaCl	TFQ	Vuilleumier et al.[94]
	AGGGCGCCUAGC**GGAG**GCUAGGCGCCC	RNA-27nt	NCp9	1450	100 mM NaCl 1 mM MgCl₂	Filter binding	Berglund et al.[98]
	751nt(−)CUAGC**GGAG**GCUAG(−)796nt	RNA 46nt	NCp7	~200	50 mM KCl	Native GE Filter binding	Clever et al.[162]
	GGACUAGC**GGAG**GCUAGUCC	RNA-20nt	NCp7	85	100 mM NaCl	TFQ	Xi et al.[163]
	GGACUAGC**GGAG**GCUAGUCC	RNA-20nt	NCp7	170	50 mM NaCl	ITC, NMR	Amarasinghe et al.[54]
	GGACUAGC**GGAG**GCUAGUCC		NCp7	178	150 mM CH₃COONH₄	ESI–FTMS	Hagan and Fabris[164]
	GGACUAGC**GGAG**GCUAGUCCdCdCdA	RNA-23nt	NCp7	551	150 mM NaCl 1 mM MgCl₂	FA	Takahashi et al.[165]
SL2 (SL)	GGCGACU**GGUG**AGUACGCC	RNA-19nt	NCp7	110	25 mM CH₃COONa 25 mM NaCl	ITC, NMR	Amarasinghe et al.[54]
	GGCGACU**GGUG**AGUACGCC	RNA-19nt	NCp7	23	200 mM NaCl	TFQ	Shubsda et al.[93]
	GGCGACU**GGUG**AGUACGCC	RNA-19nt	NCp7	62	150 mM CH₃COONH₄	ESI–FTMS	Hagan and Fabris[164]
	GGGGCGGCGACU**GGUG**AGUACGCCAAAAA	RNA-29nt	NCp7	~400	50 mM KCl	Native GE Filter binding	Clever et al.[162]
PBS (SL)	GTCCC**TGTT**CGGGC	DNA-14nt	NC(11–55)	80	30 mM NaCl 0.2 mM MgCl₂	TFQ, ITC	Bourbigot et al.[56]
cTAR (SL)	CCAGAGAGCTCCCGGGCTCGACC**TGG**	DNA-26nt	NC(11–55)	250	30 mM NaCl 0.2 mM MgCl₂	ITC	Bazzi et al.[97]
Linear single-stranded DNA or RNA	AA**TGCC**	DNA-6nt	NC(1–72)	260	100 mM NaCl	TFQ	Vuilleumier et al.[94]
	AA**UGCC**	RNA-6nt	NC(1–72)	285	100 mM NaCl	TFQ	Vuilleumier et al.[94]
	AA**TGCC**	DNA-6nt	NCp7	200	100 mM NaCl	TFQ	Avilov et al.[96]
	AA**TGCC**	DNA-6nt	NC(11–55)	400	30 mM NaCl	2fFCS	Didier et al.[166]
	AA**TGA**CTGAAAC	DNA-12nt	NCp7	1200	100 mM NaCl	2ApF	Avilov et al.[95]
	AATGAC**TG**AAAC	DNA-12nt	NCp7	560	100 mM NaCl	2ApF	Avilov et al.[95]
	A**TCGCC**	RNA-6nt	NC(1–72)	1100	100 mM NaCl	TFQ	Vuilleumier et al.[94]
	TGAC**CGTG**ACCG	DNA-12nt	NCp7	710	100 mM NaCl	2ApF	Avilov et al.[95]
	U**GUG**CCCGUCU	RNA-11nt	NCp7	120	100 mM NaCl	ITC	Spriggs et al.[101]
	TGTGTGTG	DNA-8nt	NCp7	3–5	150 mM NaCl	SPR, FA, TFQ	Fisher et al.[100]

Boldface underlined nucleotides correspond to the identified binding sites for NC. The reported binding sites affinities were obtained from 1:1 NC–NA complexes or from site-specific determination of the K_d. TFQ, tryptophan fluorescence quenching; GE, gel electrophoresis; 2fFCS, 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.
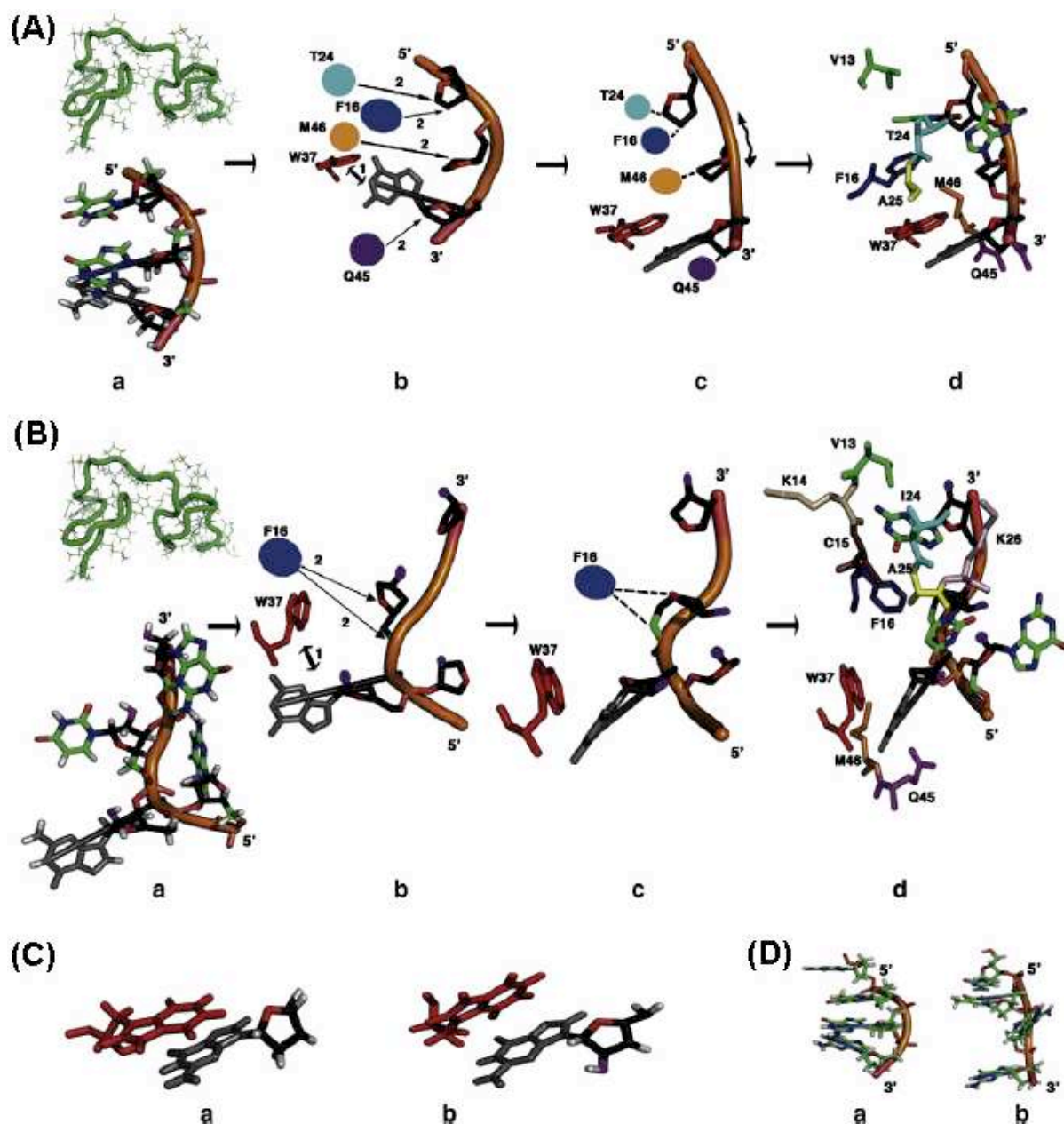
Figure 26. A simplified view of amino acid–sugar interactions directing the NA binding polarity of NC (82). (A) Hypothetical steps of the NC-DNA interaction illustrated by the binding of NC(11-55) to mini-cTAR. (B) Hypothetical steps of the NC-RNA interaction illustrated by the binding of NC to SL2. (C) View of the primary W37-guanine contact in the NC:mini-cTAR DNA complex (a) and in the NC:SL2 RNA complex (b). Note that the relative positioning of the two residues is very similar, including the relative sugar orientation. (D-a) the free CTGG sequence of mini-cTAR determined by NMR (413); (D-b) the same part of mini-cTAR complexed to NC (21).

2.2.3.3   *Roles of NC in the replication cycle*

Since NC can bind either specifically or nonspecifically to nearly any NA sequence, it plays multiple roles in virus replication cycle (82). HIV-1 NC has been extensively studied during the past 20 years and was shown to be implicated in the assembly of virus particles, genomic RNA dimerization and packaging, reverse transcription and integration into the host

genome (52, 151, 158). NC functions in both precursor and fully processed forms throughout the replication cycle **(Figure 27)**. Typically, in precursor forms, NC is involved in assembly processes, specifically recognizing and binding genomes as RNA dimers (56, 123). The precursor is also involved in placement of the tRNA primer onto genomes.
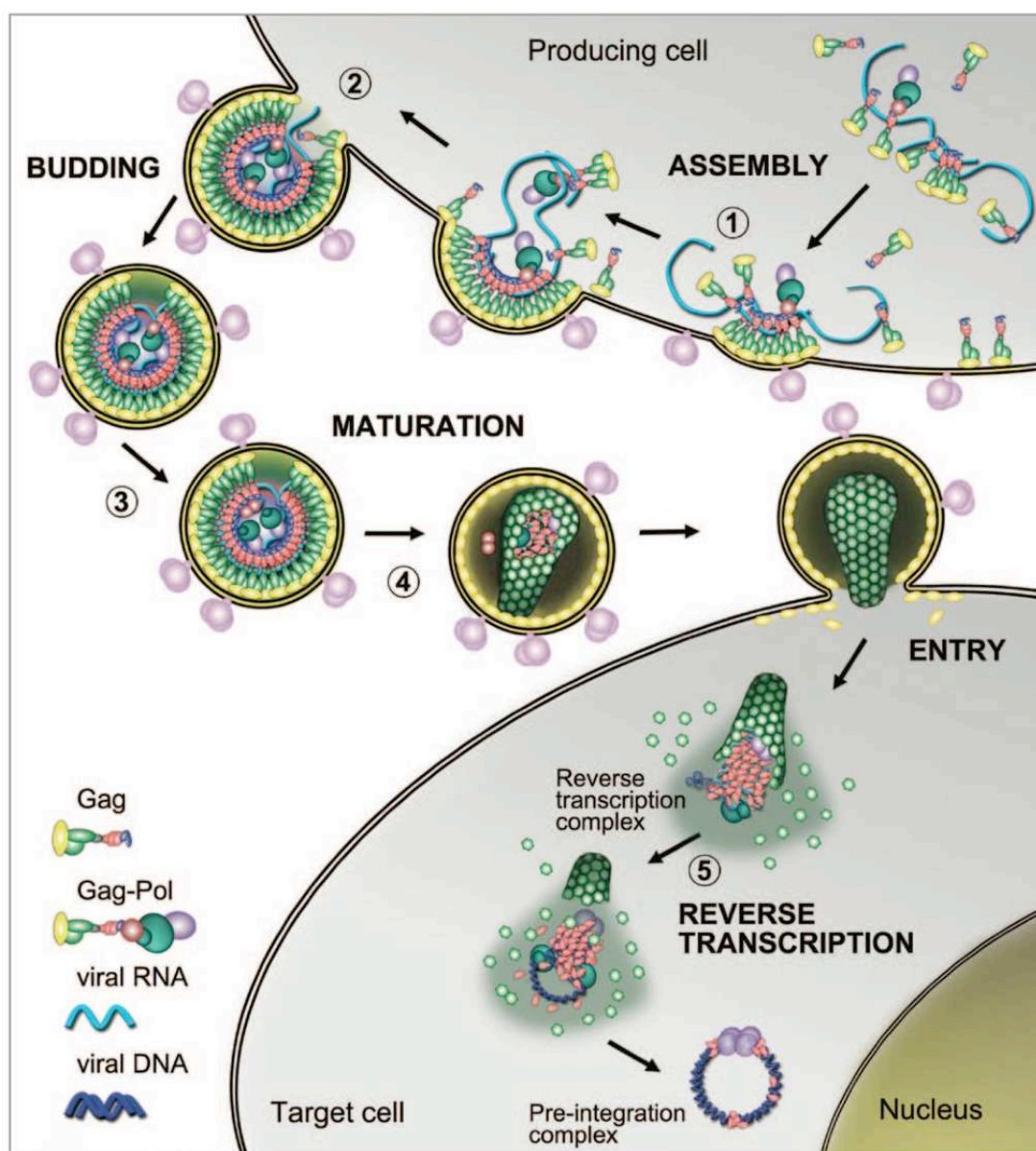
Figure 27. NC's involvement in HIV-1 replication cycle (267).

Assembly of gRNA and RNA packaging (Figure 27 step 1-2)

Gag binding through its NC domain initiates at the Psi RNA region ensuring viral RNA dimerization, which then acts as a nucleation point to aggregate other Gag and Gag-Pol units through Gag-Gag interactions and Gag-RNA interaction. Genetic analyses have demonstrated that the NC domain of Gag is critical for specific recognition and packaging of gRNA since

the ZF mutations (158) or multiple basic residues in the NC domain of Gag (63, 317) can significantly reduce the genome packaging. RNA packaging and Gag-Gag interactions mediate assembly of the particle that subsequently buds from the cell as an immature and non-infectious particle. Deletion of the NC domain of Gag caused an overall accumulation of Gag polyprotein at the plasma membran *in vivo* and essentially in diffuse pattern (154).

Virus budding and maturation (Figure 27 step 3-4)

A necessay step for virus particle maturation is the formation of a stable dimer complex between the two molecules of gRNA present in the virus particle. It has been demonstrated that NC is critical for the formation of this complex (124, 125). The mechanism for this stabilization is thought to begin with limited base pairing between the genomes, termed a "kiss loop structure", which, with the assistant of NC expands to form a more complex, extended structure (275, 312). This process, termed "dimer maturation", results in the formation of a more thermostable RNA dimer, and occurs during the maturation of the virus particle after budding (124, 125, 233, 351). Maturation results in NC:RNA complex condensation at the centre of the core where NC chaperone activity modulates RNA tertiary and quaternary structures. Mutations in ZFs have a negative impact on virus budding and Gag processing *in vivo* (154).

Reverse transcription (Figure 27 step 5)

After entry of the HIV-1 particle into the cytoplasm of a newly infected cell, RT catalyses viral DNA synthesis in the context of a ribonucleoproteic complex where NC accelerates the first and second strand transfers required for minus and plus strand DNA synthesis and induce a complete remodeling of the capsid and its contents, which are ultimately converted into a PIC, depleted of most of the NC initially bound to the gRNA. NC promotes specific cDNA synthesis and efficient strand transfer by preventing non-specific self-primed cDNA synthesis, and by increasing the rate and extent of annealing complementary sequences, respectively (83, 103, 157, 158, 230, 231, 410). More details will be given in Chapter 4.

Integration of viral DNA

The HIV-1 virion core that enters the cytoplasm is thought to consist of a CA shell which surrounds the nucleocapsid core (dimerized RNA genome, tRNA$^{(Lys,3)}$ primers, and ~1500 molecules of NC). In addition, purified cores have been shown to contain RT, IN, PR, Vpr and small amounts of Pr55^{Gag} and p41^{MA-CA} (263, 396). This nucleoprotein assembly is

termed as reverse transcription complex (RTC). When the RTC uncoats, reverse transcription proceeds and the complex is transported through the cytoplasm to the nucleus where it is converted into a PIC, which is actively transported into the nucleus (393). The presence of NC in PICs is expected to some extent because of its presence in RTCs and the experiments have shown nuclear localization of NC after infection (418). However, there has been a conspicuous inability to detect it in isolated HIV-1 PICs (287). The failure to detect NC may be due to weak NC-NA interaction in PICs. There have been proposed two models to explain the failure. The first model suggests that NC binds non-specifically dsDNA, forces IN to bind to areas of higher affinity, and/or stabilizes IN binding into LTR ends (313). Once IN is bound at the correct position, it may not be necessary for NC to remain present. The second model proposes that NC is bound at low concentrations to specific parts of the full-length viral DNA. TEM virualization showed that dsDNA contains a central DNA flap where may be bound NC (266) since it binds single-strand DNA better that dsDNA. In addition, the PR-digestion of NCp15 is unlikely to 100%, so that a small amount of NCp9 is expected, and NCp9 binds equally well to single-strand DNA and dsDNA (265). Interestingly, it has been demonstrated that NCp9 binding has a cooperative component (74), suggesting that perhaps a small amount of NCp9 could assist IN in forming complexes at LTR ends. It was also observed that NCp9 stimulated corrdinated integration to a greater degree than NC in *in vitro* integration assays (135).

2.3 Nonstructural proteins

2.3.1 Regulatory proteins and accessory proteins

HIV-1 encodes two regulatory proteins, the transcriptional transactivator (Tat) and the regulator of virion gene expression (Rev). The virus encodes four so-called accessory proteins: the ill-named 'negative effector' (Nef), viral infectivity factor (Vif), and the viral proteins r (Vpr) and u (Vpu).

Tat (86-101 amino acids) binds the TAR stem-loop structure (139). It is produced from incompletely and compleltely spliced mRNAs **(Figure 12)**. Tat is one of two essential viral regulatory factors of HIV-1 gene expression. It recruits several cellular proteins to make the chromatin structure accessible for the transcription machinery, to acquire the posttranslational modifications essential for its function, and to produce efficient viral replication (68). Low levels of both proteins are found in persistently infected cells. Tat has been localized primarily in the nucleolus/nucleus by immunofluorescence. The binding of Tat to TAR occurs

in conjunction with cellular proteins that contribute to the effects of Tat. The binding of Tat to TAR activates transcription from the HIV-1 LTR at least 1000 folds (114, 206).

Rev (116 amino acids) is a sequence-specific RNA binding protein. It is produced from completely spliced mRNAs **(Figure 12)**. The binding of Rev to the RRE facilitates the export of unspliced and incompletely spliced viral RNAs from the nucleus to the cytoplasm (115). Recent research showed that this binding was inhibited by Adenosine Deaminase Acting on RNA 1 (ADAR1) (39).

Nef (206 amino acids) is functionally complex. It is produced from completely spliced mRNAs **(Figure 12)**. It has a structured core, amino acids 62–147 and 179–200; flexible N- and C-termini, amino acids 2–61 and 201–206; and an internal flexible loop, amino acids 148–178 (standard NL4-3 numbering) (120). Nef is considered to be a pathogenic factor, but its role in the development of AIDS is not mechanistically understood. Although Nef effects are due to protein–protein interactions, only one host cell protein, the protein tyrosine kinase Hck, has been found to bind directly to HIV-1 Nef (K_D approximately 200 nM, (209, 236)). Nef has been shown to have multiple activities, including the down regulation of the cell surface expression of CD4, the perturbation of T cell activation, and the stimulation of HIV-1 infectivity.

Vpr (viral protein R, 96 amino acids) plays a role in the ability of HIV-1 to infect non-dividing cells by facilitating the nuclear localization of the PIC (169). It is produced from incompletely spliced mRNAs **(Figure 12)**. Other critical functions include transcriptional coactivation of viral and host genes, induction of cell-cycle arrest, both direct and indirect contributions to T-cell dysfunction, and regulation of NF-κB activity (218).

Vpu (viral protein U, 81 amino acids) is specific to HIV-1 and SIVcpz (the closest SIV relative of HIV-1). It is produced from completely spliced mRNAs **(Figure 12)**. There is no similar gene in HIV-2 or other SIVs. Two different biological functions have been attributed to Vpu: (a) HIV-1-induced CD4 receptor downregulation by mediating the proteasomal degradation of newly synthesized CD4 in the ER, and (b) enhancement of virion release from the plasma membrane of HIV-1-infected cells (361, 400). Recent discoveries suggest that these two functions may have impacted primate immunodeficiency virus cross-species transmission and this accessory protein helps HIV-1 to escape from host defenses (106).

Vif (viral infectivity factor, 220 amino acids) is Rev-dependent mRNA (137, 375). It is produced from completely spliced mRNAs **(Figure 12)**. Vif is essential for the replication of HIV-1 in peripheral blood lymphocytes, macrophages, and certain cell lines (358). Vif inhibits the antiviral activity of the cellular protein APOBEC3G from entering the virion

during budding from a host cell by targeting it for ubiquitination and proteasomal degradation (235, 340, 412).

2.3.2 Protease (PR)

HIV-1 PR activity is not required for virus production and release *per se*, but is essential for viral maturation leading to infectious viral particles. It generates mature infectious virus particles through cleavage of the viral Gag and Gag-Pol precursor proteins. HIV-1 PR is part of the family of aspartic proteases (with an aspartic acid in the active site at position 25) and is a symmetrically assembled homodimer consisting of two identical subunits of 99 amino acids (282, 403). The centre of the enzyme is formed by the substrate binding cleft, which interacts with the different substrate cleavage site sequences in the Gag and Gag-Pol proteins. Dimerization of Gag-Pol precursor proteins is required for the activation of viral protease that is embedded in the Gag-Pol protein.

2.3.3 Integrase (IN)

The HIV-1 IN is an essential enzyme for the integration of viral DNA into the host cell genome. The process of DNA incorporation occurs in two spatially and temporally distinct steps known as 3' processing and strand transfer **(Figure 11)**.

Mechanistically and structurally, IN belongs to a diverse family of polynucleotidyl transferases (107). In each newly formed virion, approximately 50 to 100 copies of the IN enzyme are packaged (258). IN is a 32 kD protein that comprises three structurally independent domains: the amino-terminal domain (NTD), the catalytic core domain (CCD) and the carboxy-terminal domain (CTD) **(Figure 28)**.
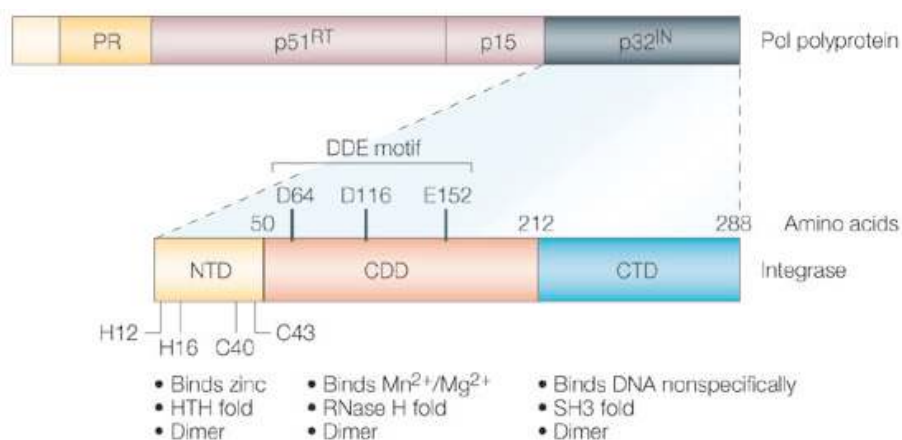
Figure 28. Structural domains of HIV-1 IN (314).

IN is supposed to be in equilibrium between its monomeric, dimeric, tetrameric and high order oligomeric states **(Figure 29)**. Only the dimer is able to bind the viral DNA, and thus it represents the active form of the enzyme (112, 155). The IN:DNA complex is then transported into the nucleus by several cofactors such as Transportin-SR2 (TR2) (381). LEDGF/p75 is another cellular cofactor of HIV-1 IN that interacts with IN through its IN binding domain (IBD) and tethers the viral PIC to the host cell chromatin (370). These two host cellular proteins have been recently reported to mediate PIC import.

A   **B**
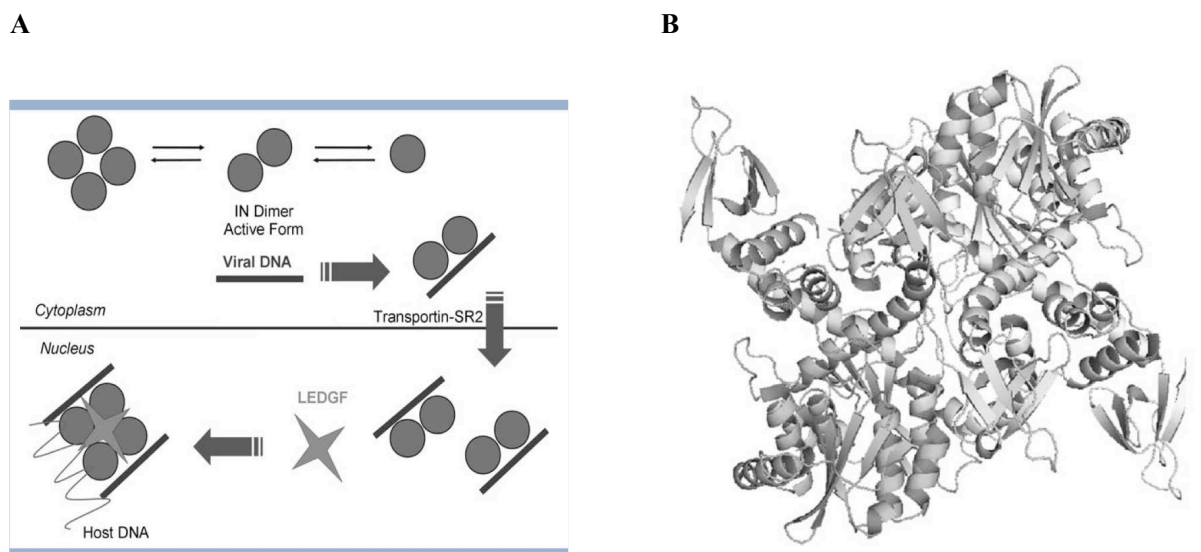
Figure 29. A) Hypothetical oligomerization IN equilibrium; B) Free IN tetramer. (370)

2.3.4   Reverse transcriptase (RT)

2.3.4.1   Structure of RT

The RT is an asymmetric heterodimer that is generated by cleavage of the Gag-Pol precursor and consists of two subunits, p66 (560 amino acids) and p51 (the first 440 amino acids of the p66 subunit) (219). The larger p66 subunit has two domains with the following functions: the NTD contains DNA- and RNA-dependent DNA polymerase activity, which is responsible for the critical step of conversion of the minus stranded RNA viral genome into double stranded DNA; the CTD contains RNase H activity, which degrades the RNA strand in an RNA-DNA hybrid. Both subunits are encoded by the same sequence in the viral genome. RNase H consists of the last 120 amino acids in the CTD of the p66 subunit, which correspond to the p15 fragment cleaved from the p66 subunit by the viral protease to generate the p51 subunit (195, 219). The p51 subunit lacks catalytic activity and has a structural role. Like other polynucleotide polymerases, the overall three-dimensional structure of the p66 subunit is often compared to a right hand **(Figure 30)**, with fingers (amino acids 1-85 and

118-155), a palm (amino acids 86-117 and 156-237) and a thumb (amino acids 238-318) domain (353). The palm domain contains the polymerase active site with its three aspartic acids (D110, D185 and D186) characteristic motif. The incoming dNTP binds between the palm and the finger subdomains, and the ribose and base make important contacts with residues including L74, Y115, M184 and Q151. The catalytic pocket is formed by the fingers folding down into the palm domain (182). Next to the catalytic domain, the p66 subunit also contains the RNase H domain (amino acids 427-560), linked by the connection domain (amino acids 319-426). The connection domain is also involved in interactions with the NA and the p51 subunit. Despite their sequence homology, the p66 subunit assumes a flexible and open structure, whereas the p51 subunit is rather compact, and seems to play a structural role, devoid of catalytic activity, with the three aspartic acids buried inside.
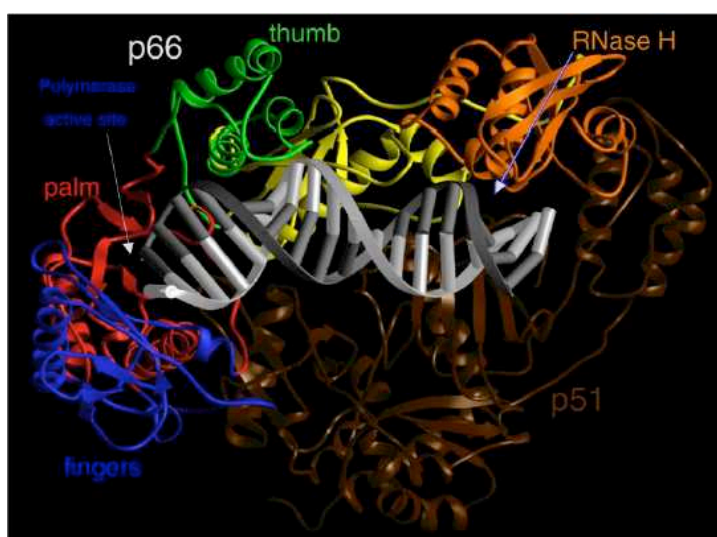
Figure 30. Crystal structure of HIV-1 RT with NA (333). The fingers, palm, thumb, connection, and RNase H subdomains of the p66 subunit are shown in blue, red, green, yellow, and orange, respectively. The p51 subunit is shown in dark brown. The template and primer DNA strands are shown in light gray and dark gray, respectively.

2.3.4.2   *Properties of RT*

RT has two enzymatic activities: a DNA polymerase activity that can copy either a DNA or an RNA template, and an RNase H activity that cleaves RNA only if the RNA is part of an RNA-DNA duplex. The two-enzymatic functions of RT, polymerase and RNase H, cooperate to convert the RNA into a double stranded linear DNA. However, HIV-1 RT lacks exonucleolytic proofreading activity which confirms that the DNA transcript produced is an accurate copy of the RNA template, and confers a mutation rate of approximately 3.4×10^{-5} mutations per base pair (251) in the replication cycle. Since the HIV-1 genome is an estimated 10^4 base pairs in length, and HIV infection is characterized by a very high replication rate,

with the production of 1 to 10 billion new virus particles per day in an untreated infected individual, millions of viral variants are produced in any infected person in a single day, and a mutant at each nucleotide position in the viral genome is produced every day (85, 306). After DNA synthesis has been completed, the resulting linear double-stranded viral DNA is translocated to the nucleus where the viral DNA is inserted into the host genome by IN.

2.3.4.2.1 DNA polymerase activity

The mechanism of DNA polymerization by HIV RT is reasonably well understood; extensive biochemical and crystallographic data have helped to define the individual steps of the process. The reaction begins with the binding of RT to the NA substrate, which results in a conformational change in the position of the p66 thumb, from a closed to an open conformation. Like many other DNA polymerases, RT requires both a primer and a template. In most sequence contexts, RT preferentially binds to a double-stranded NA so that the 3' end of the primer strand is bound at the priming site (P site), adjacent to the polymerase active site (41, 195, 252, 253). The initial step in nucleotide incorporation is the binding of the incoming dNTP at the nucleotide-binding site (N site) to form a ternary complex (182). The rate-limiting step in the polymerization reaction is a conformational change in which a portion of the p66 fingers subdomain close down on the incoming dNTP, which helps to precisely align the 3'-OH of the primer, the α-phospate of the dNTP, and the polymerase active site (182, 211, 327). The chemical step that follows leads to the formation of a phosphodiester bond between the newly incorporated nucleoside and the primer with the concomitant generation of pyrophosphate. The fingers open to allow the pyrophosphate to leave the active site. In processive DNA synthesis, the NA substrate must translocate relative to RT to free the nucleotide-binding site so that RT can bind the next incoming dNTP. The dependence of the rate-limiting step on correct dNTP binding and base-pairing forms the basis for the fidelity of polymerization (203). The chemical step requires two divalent metal ions and there is good reason to believe that the normal *in vivo* metals are both Mg^{2+}. The metals coordinate the oxygens of all three phosphates of the incoming dNTP and the side chains of the three catalytic Asp residues (D110, D185, D186). Metal coordination at the polymerase site facilitates the attack of the 3'-OH on the α-phosphate of the incoming nucleotide by activating the hydroxyl group **(Figure 31)**. The two metal ions also stabilize the charge of the reaction intermediates (354).
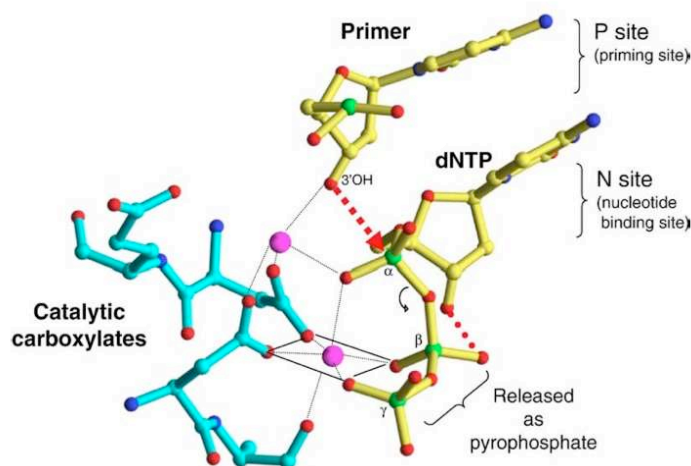
Figure 31. Metal chelation in the RT polymertase active site (333). The nucleotide binding site (N site or pre-translocation site) and the priming site (P site or post-translocation site) are shown.

2.3.4.2.2 RNase H activity

The RNase H active site in RT is located in the CTD and is separated from the polymerization site by 18 base pairs in DNA-RNA heteroduplex substrates (127, 150). This configuration allows RT to make cleavages in RNA during cDNA synthesis, which is known as polymerization-dependent RNase H activity (337) **(Figure 32)**. However, the polymerization rate of RT is greater than the rate of RNA hydrolysis, thus polymerization-independent cuts, made during revisits of RT molecules to remaining RNA-DNA hybrids, are necessary for complete removal of the genomic template (96, 337) **(Figure 32)**. HIV-1 RNase H was also shown to function in a two metal ion mechanism similar to other polymerase-associated nucleases. The crystal structure of the RNase H fragment shows two Mn^{2+} bound the active site. Calorimetry and solution NMR experiments demonstrated the presence of two Mn^{2+} or two Mg^{2+} bound to RNase H active site (71, 300). Biochemical assays suggested that the optimal RNase H activity is obtained in the presence of one Mg^{2+} and one Mn^{2+} (216). However, it is likely that, *in vivo*, both metals are Mg^{2+}. It has been suggested that the RNase H specificity for RNA-DNA hybrid relies on the width of the minor groove (13, 113). This proposal comes from the observation that RNA-DNA has an intermediate conformation between A- and B-form double-stranded NA named H-form, which has a minor groove width of ~9-10 Å (364). The crystal structure of RT bound to an RNA-DNA hybrid supports the idea that it is the width of the minor groove, in combination with the RNase H primer grip, that binds and positions the primer strand on the enzyme, which is responsible for cleavage specificity (364).
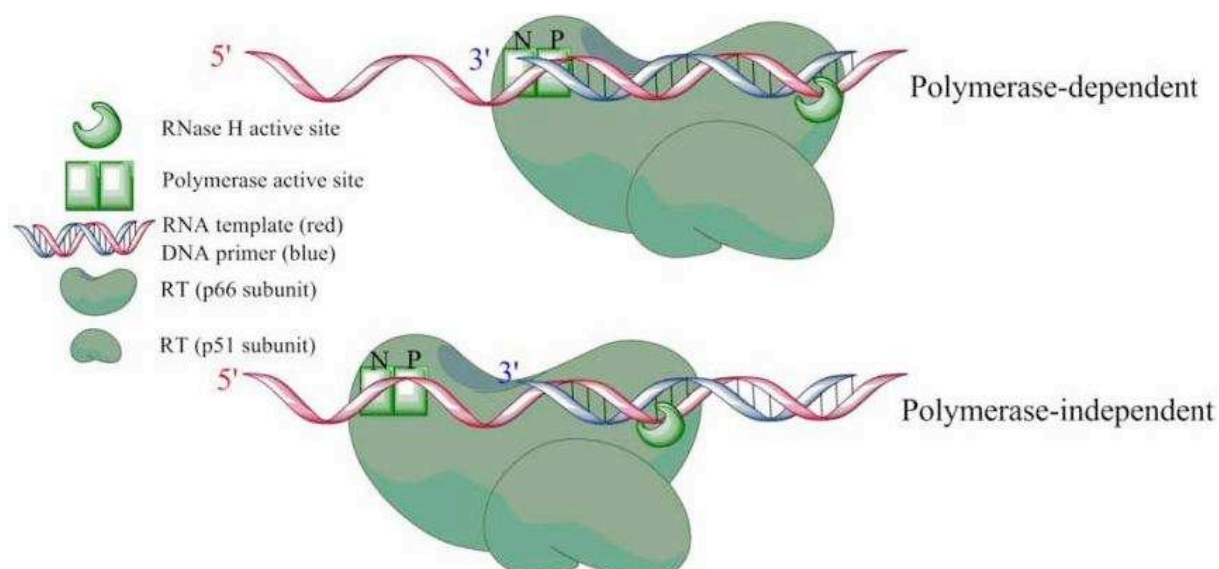
Figure 32. Two different cleavages by HIV-1 RT (24). RT can bind to a nucleic acid substrate in two distinct binding modes. The polymerase-dependent mode is characterized by the polymerase active site being in contact with the 3' primer terminus. All other possible conformations are considered polymerase-independent.

RNase H activity is responsible for the degradation of the RNA portion of the RNA-DNA hybrids that is formed during minus strand DNA synthesis. It is also responsible for the removal of the priming tRNA$^{(Lys,3)}$ (323) and the PPT (185). Viruses deficient in RNase H activity are non-infectious (335, 366). In general, degradation of the gRNA is not sequence specific. However, the RNase H of HIV-1 RT does make some specific cleavages. The removal of the tRNA$^{(Lys,3)}$ primer occurs one nucleotide from the RNA-DNA junction (323). This cleavage defines the right end of the unintegrated viral DNA. The cleavages that generate and remove the polypurine tract (PPT) primer used to initiate plus-strand DNA synthesis during the replication cycle are also specific. The PPT primer is completely removed by RNase H, which define the left end of the unintegrated viral DNA (397). These ends are the substrates for the integration of the DNA into the host genome.

Chapter 4. Reverse transcription of HIV-1 genomic RNA

Although some viral particles initiate reverse transcription before entering cells (245, 376, 419), this process generally begins when the virion enters the cytoplasm of the host cell. Recently, it has been shown that SAMHD1 restricts the replication of HIV-1 by depleting the intracellular pool of deoxynucleoside triphosphates (224). Recent studies showed that mutations in the NC domain affected the timing of reverse transcription (99, 177, 368). Indeed, reverse transcription occurred before viral particles were released from cells. These studies suggest that NC restricts the viral cDNA synthesis during viral formation and maturation. Many laboratoires have developed cell-free systems to determine the mechanisms involved in the key steps of reverse transcription. Thus, reverse transcription intermediates and products identical to those made in infected cells can be synthesized in the purified virions (referred to as natural endogenous reverse transcription or NERT) (415, 416, 419). Reverse transcription inside the purified virions is increased in the presence of deoxyribonucleotides and mild detergents (referred to as endogenous reverse transcription or ERT) (141, 176, 394). However, the yield of full-length products in ERT and NERT assays is extremely low (178, 224).

Reverse transcription consists of a complex series of events that culminate in the synthesis of a linear dsDNA copy from the single-stranded gRNA, which is ultimately integrated into the host chromosome **(Figure 33)** (170, 333). This process, catalyzed by RT, is an early postentry event that occurs in the cytoplasm of infected cells. The chaperone activity of NC is critical for reverse transcription (239, 240, 368).

1. <u>The primer binding site (PBS)</u>

Localized within the 5' UTR, the PBS sequence, complementary to the last 18 nucleotides of tRNA$^{(Lys,3)}$, is a requirement for the initiation of reverse transcription (254). From SHAPE (selective 2'-hydroxyl acylation analysed by primer extension) experiments performed with gRNA extracted from virions and cells, Weeks' group (395) proposed a secondary structure model for the 5' UTR **(Figure 34)**. SHAPE, a high-throughput RNA analysis techonology, was used to chemically interrogate local nucleotide flexibility at 99.4% of the 9,173 nucleotides in the NL4-3 HIV-1 RNA genome. 1-methyl-7-nitroisatoic anhydride (1M7), used as a chemical probe, preferentially acylates conformationally flexible nucleotides at the ribose 2'-OH position (261, 270). The resulting 2'-O-adducts are detected as stops to primer extension by RT using fluorescently labeled primers and capillary electrophoresis (270,

399) and are quantified by whole-trace Gaussian integration (384). SHAPE measurements are reproducible between independent biological replicates (395). SHAPE reactivities are highly sensitive to local nucleotide flexibility and disorder, but are insensitive to solvent accessibility (140, 261). SHAPE reactivities therefore provide direct model-free information about the overall level of structure, or architecture, for any RNA. The median SHAPE reactivity varies markedly across the HIV-1 genome (395). Regions with median reactivities below 0.25 indicate domains with substantial base-paired secondary RNA structure, whereas median SHAPE reactivities of 0.5 and greater indicate regions of largely unstructured nucleotides. Studies suggest that in the absence of tRNA, the PBS region of the NL4-3 isolate is almost not paired (297, 399).
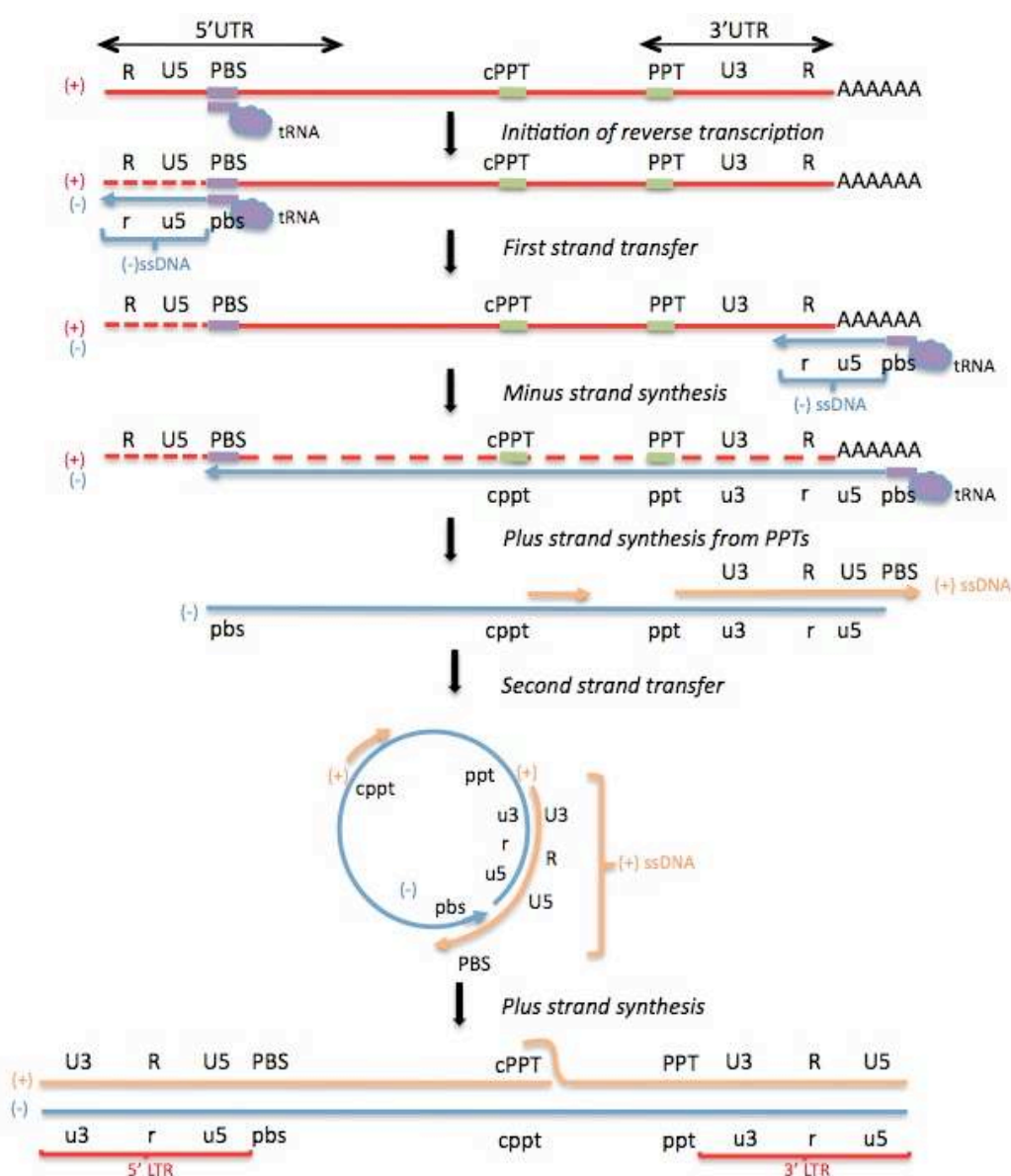
Figure 33. Schematic diagram of events in reverse transcription of HIV-1 genomic RNA. The tRNA primer is represented by a short violet line attached to a "cloud" (remaining tRNA bases); gRNA by a red solid line; (-)DNA by a blue solid line while (+)DNA by an orange solid line.
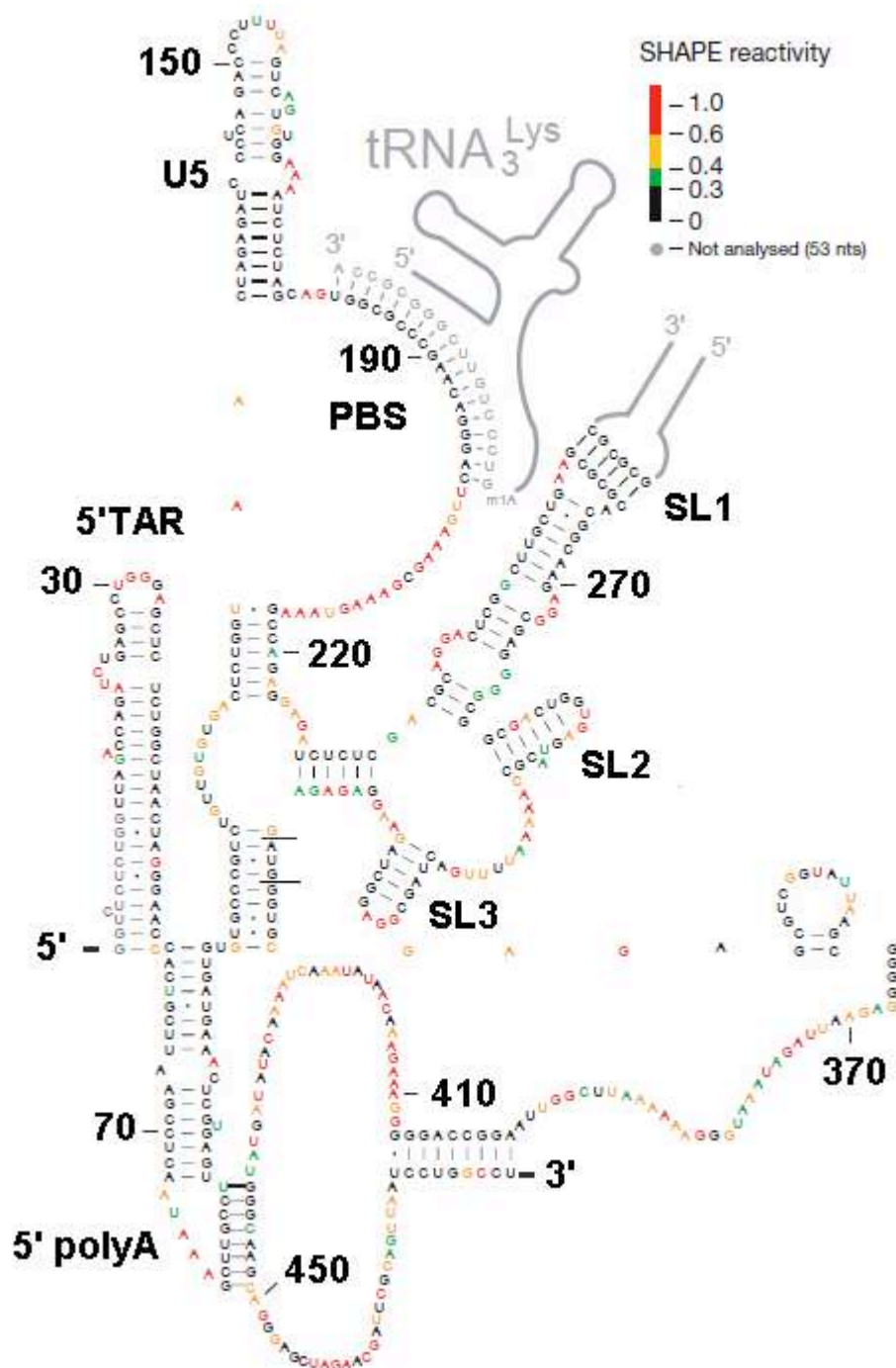
Figure 34. Secondary structure model of 5' UTR of HIV-1 (NL4-3 isolate) (Adapted from (395)). Nucleotides are coloured by their absolute SHAPE reactivities (see the scale).

2. Binding of primer tRNA^(Lys,3) to the HIV-1 genome

2.1 Annealing of tRNA^(Lys,3) to the PBS sequence

The 18 nt of the 3' end of the primer tRNA$^{(Lys,3)}$ is perfectly complementary to the 18-nt PBS in the 5' region of gRNA **(Figure 34)**. Even there are three cellular tRNA^{Lys} isoacceptor populations selectively incorporated into virions during assembly (198, 250, 303) along with

the tRNA^{Lys}-binding protein, lysyl-tRNA synthetase (LysRS) (53), only tRNA$^{(Lys,3)}$ serves as the primer for reverse transcription in HIV-1 (250, 254). This reaction is termed "tRNA primer placement". NC and RT are the main proteic factors involved in this process. Earlier studies found that tRNA placement onto the PBS occurs even in virions lacking HIV-1 PR activity, suggesting that the NC domain of Gag chaperones the annealing process *in vivo* (183). More recent studies showed that Gag-annealed tRNA is not as stable as NC-annealed tRNA (156, 331), suggesting that there is a two-step annealing process: initial tRNA annealing is promoted by the NC domain of Gag with final primer-template remodeling promoted by mature NC. Interestingly, Roldan *et al.* (330) found that NC annealing was more efficient when NC was combined with the C-terminal domain of CA. This suggests that other protein domains in the context of Gag or Gag-Pol may have an effect on gRNA-tRNA interactions.

RNA remodeling by NC likely facilitates the reverse transcription process. For example, NC was found to destabilize hairpins and other secondary structures in the 5' UTR to prevent pausing or stalling of RT *in vitro* (331). A strategy to rapidly measure RNA backbone flexibility using SHAPE provided a secondary structure model of the first 10% (976 nt) of gRNA of the NL4-3 isolate inside native virions (399). This strategy was also used to locate gRNA regions affected by NC's chaperone activity. Surprisingly, the helix destabilizing activity of NC was not observed uniformly throughout the gRNA, but was detected in the first 185 nt at six compact sites between the 5' end of gRNA and the PBS sequence (399). This region includes the site of tRNA annealing as well as the initial sequence copied by RT during ssDNA synthesis. Therefore, this localized destabilization activity is consistent with NC's role in facilitating both tRNA primer annealing and the reverse transcription process.

The interactions between NC and gRNA during packaging of the RNA genome place the NC domain of Gag at a key position for its role in the formation of the initiation complex prior to reverse transcription. Whether annealing of tRNA$^{(Lys,3)}$ to the gRNA occurs before or after packaging is unclear, but it is known that annealing is facilitated by the NA chaperone activity of Gag (16, 81, 239, 368). The basic amino acids of NC bring the complementary tRNA and PBS sequences together, while the two NC zinc finger domains are responsible for destabilizing structured RNA regions (239). It is unknown when tRNA$^{(Lys,3)}$ is displaced from PBS, but displacement is necessary to allow copying of the PBS and 18 nt of 3' end of the tRNA into DNA **(Figure 33)**. Therefore, the displacement of the tRNA primer could occur either before or after the first strand transfer.

2.2 Additional interactions between tRNA$^{(Lys,3)}$ and 5' UTR

Additional interactions between 5' UTR and tRNA$^{(Lys,3)}$ were identified *in vitro* (**Figure 35**). The first interaction is between the anticodon loop of tRNA$^{(Lys,3)}$ and a conserved A-rich loop (nucleotides 168-171 in **Figure 34**) located upstream of the HIV-1 PBS (188, 190). This template-primer loop-loop interaction was observed with the natural tRNA$^{(Lys,3)}$, but not with a syntheic tRNA$^{(Lys,3)}$ lacking the natural modified nucleotides (190). The second interaction is between the TΨC loop in tRNA$^{(Lys,3)}$ and a conserved 8-nt sequence downstream of the PBS (termed the "primer activation signal" or PAS, nucleotides 123-130 in **Figure 34**) (22, 146, 295, 331). Mutational analysis of gRNA template in regions upstream of the PBS supported the possibility that NC promotes an interaction between tRNA$^{(Lys,3)}$ (in particular, the 3' arm of the anticodon stem and a part of the variable loop) and the 142-148 sequence of the NL4-3 RNA (192).
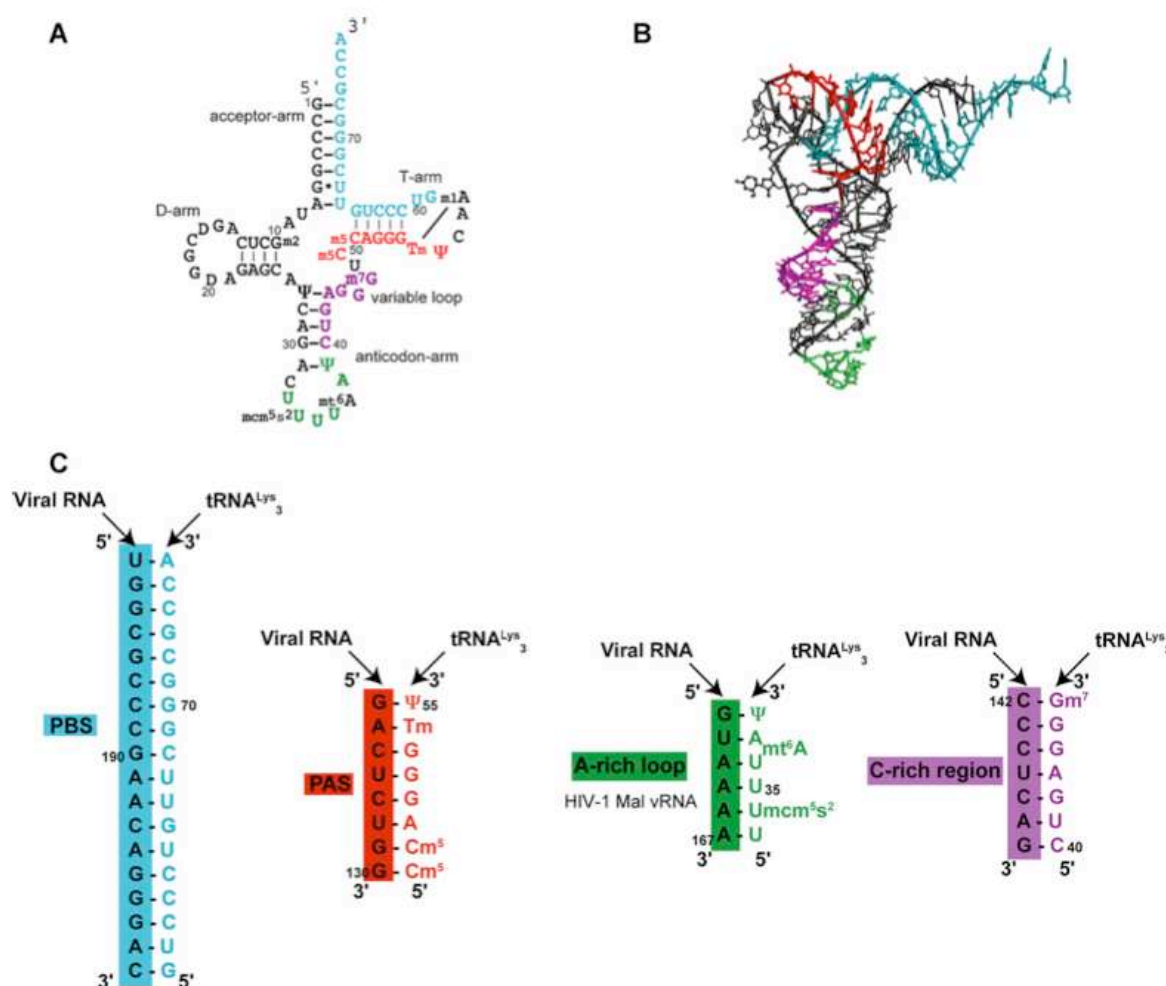
Figure 35. Interactions between 5' UTR and tRNA$^{(Lys,3)}$ (Adapted from (345)). Secondary structure (A) and three-dimensional structure (B) of tRNA$^{(Lys,3)}$. The motifs complementary to the viral RNA are indicated as follows: the sequence complementary to the PBS in blue, to the PAS in red, to the A-rich loop in green and to the nucleotides 142-148 in purple. (C) Base-pairing interactions between viral RNA and tRNA$^{(Lys,3)}$.

51

3. <u>Initiation of reverse transcription</u>

Nuclease mapping studies showed that the A-rich loop interaction was more important in the HIV-1 MAL isolate than in the HXB2 genome for efficient initiation of reverse transcription *in vitro* (147, 148, 189, 190). In addition, the interaction between the TΨC loop of tRNA$^{(Lys,3)}$ and PAS promotes efficient initiation of ssDNA synthesis *in vitro* in the HXB2 and LAI isolates (22, 146, 295, 331).

The initiation process is sensitive to the helical conformation of the NA duplexes (364) that react with RT. For example, when an 18-nt DNA complementary to the PBS (D18) was used instead of tRNA$^{(Lys,3)}$, synthesis of ssDNA bypassed the initiation mode entirely and proceeded exclusively in the elongation mode (189, 225, 226). In the absence of NC, efficient ssDNA synthesis *in vitro* required the presence of at least 24 nt downstream of the PBS in template RNA, when the primer was tRNA$^{(Lys,3)}$ or an 18-nt RNA complementary to the PBS (R18), but not D18 (192). Interestingly, NC abrogated the requirement for the 24-nt element only in tRNA$^{(Lys,3)}$-primed ssDNA synthesis but not in reactions primed by R18 (192). This suggested that NC might promote extended interactions between the tRNA primer and the gRNA template that are not possible with an 18-nt oligonucleotide primer.

In the absence of NC, early works proposed that additional interactions would facilitate RT binding to the substrate by preventing steric clashes between RT and the NA duplex (188, 191). However, gel-shift experiments showed that in the absence of dNTPs, NC did not affect RT binding to complexes constituted with either wild-type RNA or mutant templates having changes in the 143-149 sequence (192). In a similar assay in which there was also a +1 extension of the tRNA primer, NC stimulated incorporation with the wild-type, but not with the mutant templates (239). Thus, it would appear that the NC stimulation was dependent on RT binding to the substrate and extension of the primer by at least 1 nt.

4. <u>Minus-strand strong-stop DNA synthesis</u>

Extension of the tRNA primer by RT leads to synthesis of strong-stop DNA (ssDNA), a short DNA consisting of sequences complementary to the R-U5 region of gRNA **(Figure 33)**. The ssDNA was first discovered in the synthesis of Rous sarcoma virus (RSV) DNA *in vitro* by the virion endogenous DNA polymerase activity (67). A major product of this synthesis is a piece of DNA 101 nucleotides long but not full-length viral DNA, so named "strong-stop" DNA (ssDNA). There is evidence of limited DNA synthesis in HIV-1 virions prior to the entry of the virion into the target cell (245, 376, 419). Indeed, the ratio of ssDNA to genomic RNA is approximately $1:10^3$ (419). In HIV-1 virions, the number of ssDNA molecules is 10-

100 times greater than the number of DNA molecules synthesized after the first strand transfer (376, 419). When purified HIV-1 virions are mixed with dNTPs, approximately one copy of ssDNA is synthesized for one genomic RNA molecule (419).

During ssDNA synthesis, RT encounters RNA-DNA hybrid substrates and uses three distinct cleavage modes for degradation of the gRNA sequences (fragments represented by red broken lines in **Figure 33**): (1) 3' end directed cleavage, in which the 3'-OH of the DNA primer is recessed over a long RNA template and cleavage is coupled to DNA synthesis (polymerase-dependent cleavage, **Figure 32**); (2) RNA 5' end directed cleavage, in which the 5' end of an RNA is recessed over a long minus-strand DNA; and (3) internal cleavage. Both (2) and (3) represent polymerase-independent cleavage events **(Figure 32)** (337). Mechanistic studies of RNase H cleavage have revealed that the RNA-DNA substrate contacts both the polymerase and RNase H sites at the same time (25). There is evidence that during the course of polymerase-independent cleavages, NC helps to remove gRNA fragments resulting from primary cleavage events (194, 320). Thus, as RNase H cleavage proceeds during minus-strand ssDNA synthesis, the remaining hybrids become shorter and their thermodynamic stability is reduces. When this occurs, the helix destabilizing activity of NC displaces the RNA strand in the duplex and the RNA is released (240, 320). It has also been suggested that NC collaborates with RNase H by forming an NC-RT complex (50, 104, 237, 304), possibly through a zinc finger-dependent mechanism (104).

5. <u>First strand transfer</u>

Upon reaching the 5' end of the genome, the nascent ssDNA is generated. In order to complete synthesis of the minus strand DNA, ssDNA transfers to the 3' end of the same or the other co-packaged genome (19). This process is named as the first strand transfer and is highly efficient *in vivo*, since significant amounts of ssDNA did not accumulate in infected cells (383) **(Figure 33)**. More details on the first strand transfer will be given in the next chapter.

6. <u>Synthesis of full-length minus-strand DNA</u>

Following the first strand transfer, RT catalyzes elongation of ssDNA, which is accompanied by further degradation of gRNA by RNase H **(Figure 33)**. Very low amounts of full-length minus-strand DNA is detected in ERT assays (394). Synthesis of minus-strand DNA can also be carried out *in vitro* in totally reconstituted reactions including only a primer, an RNA template, RT and NC. However, a full-length DNA copy of a 7 kb gRNA template

cannot be detected in these reconstituted reactions (7). In the presence of NC, minus-strand DNAs of approximately 4 kb can be synthsized *in vitro*, but the yield is low (7). Using an 874-nt gRNA template *in vitro*, Drummond *et al.* (105) showed also that in the presence of high concentrations of NC, almost all the synthesized DNA products were full-length but with a low yield. The relative positions of the NC zinc finger domains were not important for this effect. Since the RNA template is highly structured (399), when RT encounters a secondary structure, it pauses at that site or in some cases, prematurely terminates polymerization. The duplex destabilization activity of NC alleviates pausing, thereby allowing RT to continue to traverse the template unimpeded (7, 227). Interestingly, it was shown that large aggregates are formed in the presence of NC, RT and nucleic acids (234, 237, 265, 356, 365). The aggregation activity of NC was more important than its duplex-destabilization activity for generation of long DNA products (7).

7. <u>Plus-strand DNA synthesis from PPTs</u>

Although almost the entire RNA genome is ultimately degraded by RNase H during the course of minus-strand DNA synthesis, there are two 15-nt regions containing only purine bases, which function as the primers for plus-strand DNA synthesis: (1) the 3' PPT, which abuts the 5' boundary of U3; and (2) the cPPT, which is found in the center of the IN coding region **(Figure 33)**. The sequences of the two PPTs (5'-AAA AGA AAA GGG GGG-3') are identical (111, 326, 337). These two PPTs promote the plus-strand DNA synthesis. The PPT duplexes are not like all other RNA-DNA hybrids, since they are resistant to RNase H cleavage. Several factors contribute to this unusual property. Especially the stretch of six Gs at the 3' end is required for proper RNase H cleavage and primer extension by RT (111, 262, 322, 326, 337). This unusual structure may play a role in the proper recognition of the PPT by HIV-1 RT. Studies utilizing different biochemical and biophysical techniques including X-ray crystallography, chemical probing alone or combined with mass spectrometry, and NMR revealed that the structure of the PPT is also crucial (111, 239, 326, 332). The 5' end of the PPT-DNA duplex consists of conventional, albeit weakened, Watson-Crick base pairs and assumes an A-form-like helical conformation, although certain structural anomalies were detected (409). Results from the NMR study and earlier work (223) are consistent with an inherent flexibility in the HIV-1 PPT-DNA duplex and with limited long-range structural effects (223). It was suggested that upon binding to RT, structural perturbations throughout the PPT sequence provide access to residues present in different sub-domains of RT, thereby achieving optimal binding and precise cleavage at the PPT-U3 junction (111).
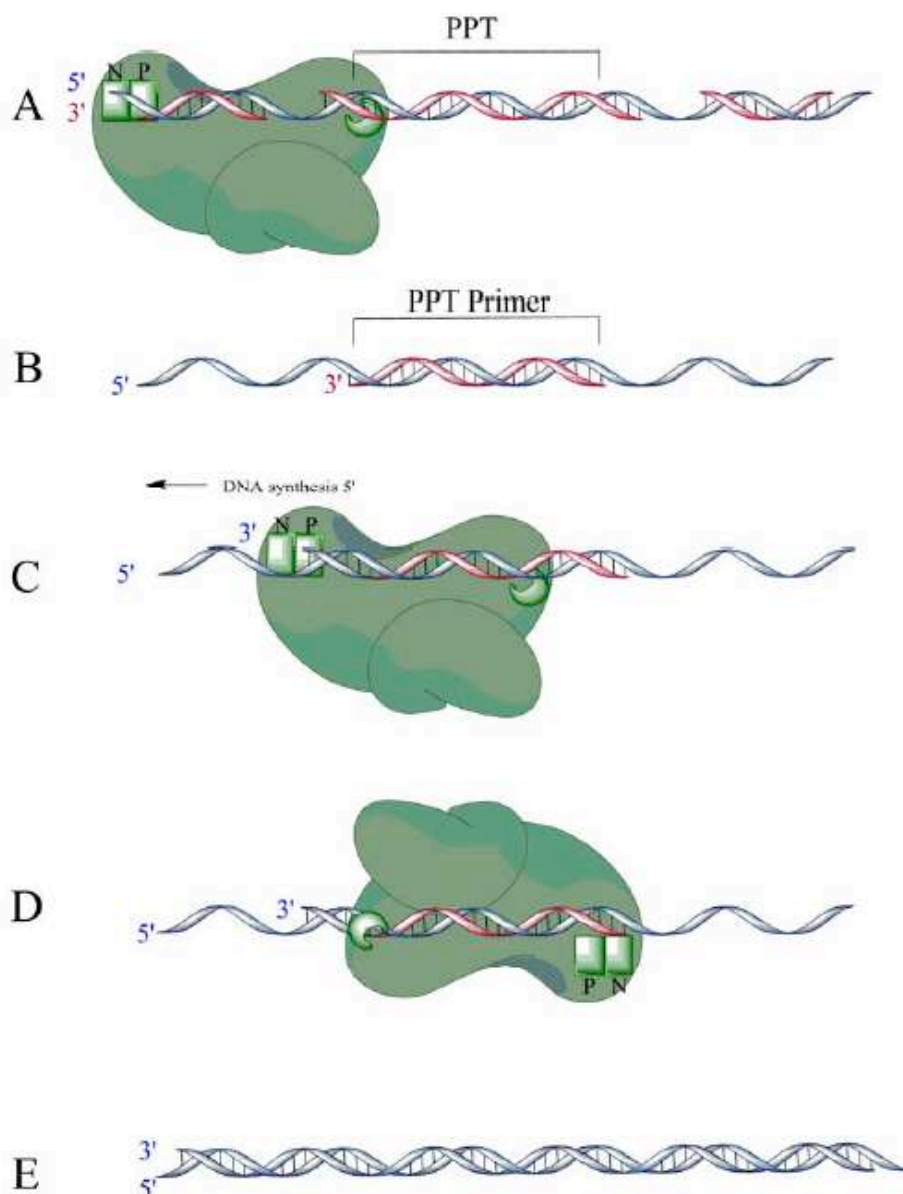
Figure 36. Generation and degradation of PPT (24). A) During minus-strand DNA synthesis, the PPT region of the RNA genome is resistant to RNase H cleavage, while a portion of the RNA genome is degraded concomitantly with minus-strand DNA synthesis. Here, a specific cleavage is made to create the PPT primer. B) The RNase H-resistant PPT sequence forms the primer for plus-strand DNA synthesis when the rest of the genome is completely degraded by RNase H. C) The RNA primer is extended by 12 nucleotides, D) then RT pauses and changes its orientation to a polymerase-independent binding mode in order to cleave at the DNA-RNA junction, and remove the PPT primer. E) The 12-nt fragment is extended toward the 5' end of the minus-strand DNA. After the second strand transfer event, the plus-strand DNA is fully synthesized resulting in a fully double-stranded provirus.

The 3' PPT is generated by a precise RNase H cleavage at the PPT-U3 junction **(Figure 36B)**. The 3' end of the newly formed RNA primer is extended by RT to position +12 downstream from the PPT RNA **(Figure 36C)**. RT then pauses, re-orients itself and makes a specific cleavage at the PPT-U3 junction **(Figure 36D)**. This cleavage is important as it

defines the end of the 3' LTR, which is used as a substrate by the viral IN. PPT primers with aberrant cleavages or that improperly begin DNA synthesis is not used efficiently and reverse transcription does not proceed (204). In order for RT to successfully initiate plus-strand DNA synthesis, RT must bind as a polymerase to extend the primer and as an RNase H to cleave at the RNA-DNA junction. As the active sites of RT are simultaneously positioned on opposite strands of the NA, RT must recognize its position on the genome and dynamically change its binding orientation appropriately for the task at hand. Thus, the primer removal reaction requires polymerase-independent RNase H activity. Cleavage at the 5' end to remove the PPT, once it is utilized as a primer, is less precise (337).

As minus-strand DNA synthesis proceeds, the 3' PPT primer is extended by RT, using a template consisting of minus-strand DNA still attached to the 5' end of the tRNA$^{(Lys,3)}$ primer **(Figure 33)**. The resulting DNA extension product is termed plus-strand strong-stop DNA ((+)ssDNA). Elongation is expected to terminate when RT has copied the base preceding the methyl A at position 58 at the 3' end of the tRNA primer, which reconstitutes the PBS sequence in (+)ssDNA.

8. <u>Second strand transfer</u>

The second strand transfer is required for synthesis of a full-length copy of minus-strand viral DNA. The (+) ssDNA transfer to the 3' end of the minus strand to complete viral replication. The second strand transfer is only between an intramolecular process (411). This process relies on the PBS region and RNase H cleavage. This event is mediated by base pairing of the 18-nt PBS region at the 3' end of (+)ssDNA, termed (+) PBS and the complementary region, termed (-) pbs, at the 3' end of minus strand DNA **(Figure 33)**, resulting in the formation of a circular intermediate. NC stimulates overall the second strand transfer and its NA chaperone activity has been shown to be crucial for efficient removal of the tRNA primer and for facilitating the annealing reaction (158, 240). Removal of the primer involves primary and secondary RNase H cleavages. Primary cleavage does not occur at the tRNA-DNA junction, but rather between the 3' end rA of the tRNA and the penultimate rC. This results in the attachment of an rA residue to the 5' end of minus-strand DNA and formation of a 17 nt hybrid consisting of sequences from the 3' end of the tRNA and the (+) PBS in plus ssDNA (239).

Using *in vitro* reconstituted systems that model the steps in the second strand transfer, it was shown that the primary RNase H cleavage was not sufficient to completely remove the tRNA, presumably because the 17-nt hybrid is relatively stable (346, 404). Thus, second

RNase H cleavages are also required. Interestingly, the NA chaperone activity of NC could mediate tRNA removal by dissociation of the 17-nt hybrid, but not as efficiently as RNase H activity (404). Removal was actually most efficient when both RNase H and NC were present. Subsequent studies demonstrated that the native ZFs are required for NC function in primer removal, indicating that NC's duplex destabilization activity is critical for this step (158, 159).

NC plays also a role in the annealing responsible for the second strand transfer. One NMR study suggested that NC-induced conformational changes in the (-) pbs that facilitate annealing to the (+) PBS (201). Another NMR study performed with NC(12-55) or full-length NC with a (-) pbs construct lacking the 3' overhang showed that binding is mediated by hydrophobic residues of the ZFs (40). Fluorescence studies demonstrated that NC only weakly destabilizes the PBS hairpins (108) and the NA aggregation also plays a role in the PBS annealing reaction, but its contribution is less than for the TAR hairpin annealing step (386, 387). Biochemical studies with longer NA constructs that model the donor and acceptor DNA substrates in the second strand transfer and contain the complementary PBS sequences showed that NC facilitates the annealing (404).

9. dsDNA synthezised with LTR

Due to the second strand transfer, the DNA replication intermediate is circularized **(Figure 33)**. Priming from both PPTs involves discontinuous DNA synthesis, which generates the ~99 nt central DNA flap at the center of the unintegrated dsDNA product (55). The flap is part of a triple-stranded structure containing two overlapping plus-strand DNA segments and a complementary minus-strand (55). The downstream segment is initiated from the cPPT, whereas the upstream segment is initiated from the 3' PPT. Flap removal and repair of the gap in dsDNA have not been demonstrated *in vivo*, although it is assumed that cellular enzymes are involved (326). More recently, two independent studies have presented evidence indicating that the cPPT protects HIV-1 from the activity of the human APOBEC3 restriction factors (179, 407).

Formation of a circular intermediate during the second strand transfer allows RT to extend both the plus- and minus-strands of DNA **(Figure 33)**. Following the circularization, RT displaces 636 nt to generate the LTR sequence (U3-R-U5) at the ends of both DNA strands. Completion of reverse transcription requires strand-displacement synthesis, which is strand synthesis through a region of dsDNA. RT by itself is capable of such synthesis, but again, the presence of NC greatly enhances this process (118, 161, 213, 302, 379). The failure

to complete strand-displacement synthesis would result in incomplete LTR ends. This, in turn, would create viral DNA that cannot integrate.

10. DNA strand transfers and recombination

Retroviruses have evolved a complex replication process that expands the gRNA into a DNA transcription unit that includes sequences that control transcription and proviral DNA integration. Although complicated, strand transfer has become an essential part of reverse transcription because it is crucial to carrying out this expansion and creations of the long terminal repeat sequences for integration. Since retrovirus particles contain two identical copies of gRNA, strand transfer can occur in an intra- or intermolecular manner (181, 202, 299, 382). In instance where the virus particles contain two non-identical RNA genomes, intermolecular strand transfer leads to recombination, which in turn generates viral diversity and provides a mechanism to resist antiviral host factors and the action of antiviral agents (6, 128, 180, 284, 293).

In addition to the requisite strand transfer events associated with replication at the ends of the genome, the RT has been shown to switch templates within internal regions during replication. Many of basic features of first and second strand transfer are also preserved during these internal transfer events. When the two co-packaged RNAs are genetically distinct, internal transfer can result in the production of a recombinant virus **(Figure 37)**. The low fidelity of the RT creates mutations that can be dispersed by the efficient strand transfers. This creates a favourable environment for rapid viral evolution by means of recombination.

10.1 Models of strand transfer

Models of strand transfer during minus strand synthesis include "forced copy choice", "copy choice" and "dynamic copy choice" (66, 286, 362) **(Figure 38)**. The forced copy choice and copy choice models describe the fundamental configurations of the templates that promote transfer, whereas dynamic copy choice describes the enzymatic requirements of efficient transfer by forced copy or copy choice. In the "forced copy choice" model proposed by Coffin (66), breaks in the RNA genome induce strand transfer during minus strand synthesis. In the model, primer elongation initiates on one RNA genome and continues until a break in the template is reached. At this point, on the basis of template homology, the primer switches to the other co-packaged RNA genome and continues synthesis. The "copy choice" model of strand transfer expands the possibility of template switching during minus strand synthesis to include regions of the genome containing intact templates (286). As with the

forced copy choice model, template homology is a prerequisite for copy choice transfer. In studies of non-homologous recombination, the average recombination rate was approximately 0.1-1.0% of that of homologous recombination (419). The evolution of the strand transfer model led to a new hypothesis, the "dynamic copy choice" model proposed by Pathak and co-workers (186, 289, 362). This model is descriptive of some factors that influence copy choice strand transfer. The model proposes that the ratio between the polymerase activity and RNase H activity of RT dictates the efficiency of strand transfer (362). In this model, a slower polymerization rate allows more time for RNase H cleavage of the template, and promotes transfer.
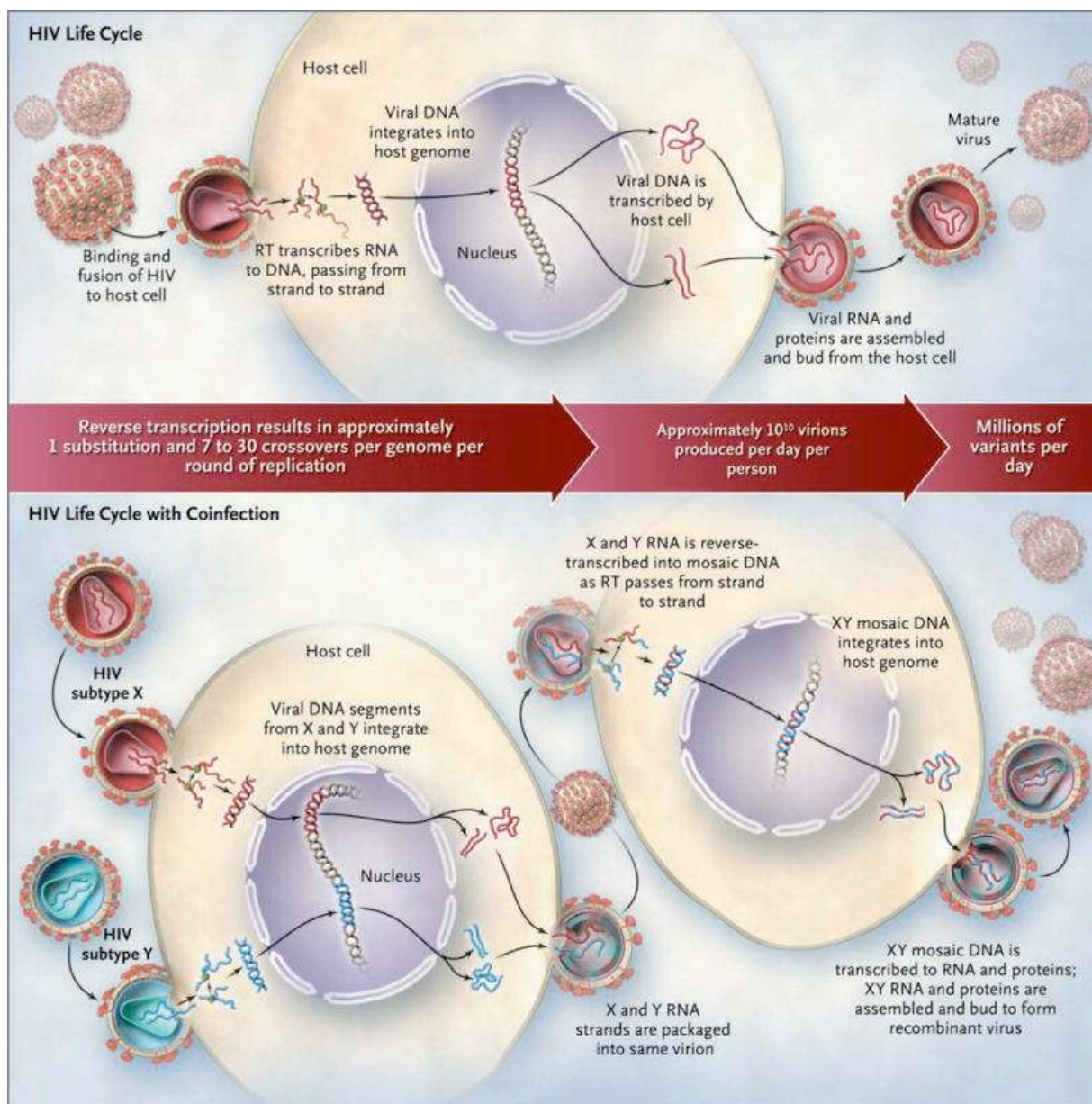
Figure 37. Evolution of diversity in HIV-1 during the viral life cycle and creation of unique recombinant forms in the context of coinfection with two subtypes (367).
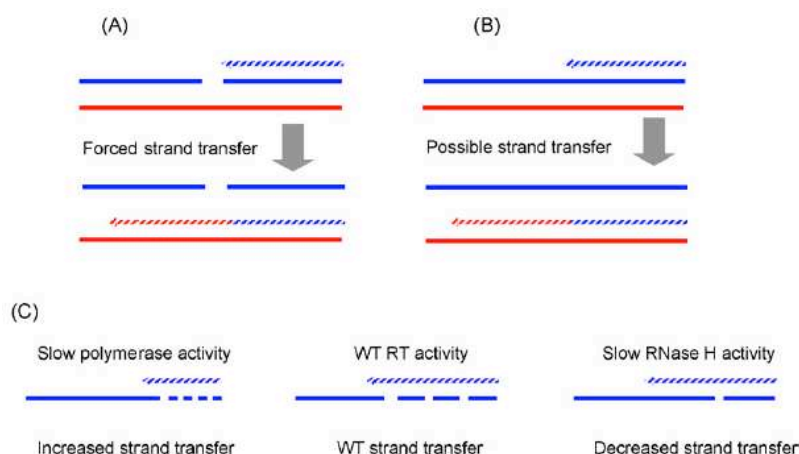
Figure 38. Models for strand transfer during minus strand synthesis (19). (A) Forced copy choice strand transfer. cDNA (patterned blue) extends along the donor RNA (solid blue) until it reaches a break in the donor RNA. To complete reverse transcription of the RNA template the cDNA must transfer to the acceptor RNA (solid red) and resume synthesis using the acceptor as the template (patterned red). (B) Copy choice strand transfer. cDNA (patterned blue) extends along the donor RNA (solid blue) until it transfers to the acceptor RNA (solid red). There is no donor RNA template configuration, such as a break in the template, which forces strand transfer. (C) Dynamic copy choice strand transfer. RT enzymatic activities (polymerase and RNase H) influence the efficiency of strand transfer. Compared to WT RT, slow polymerase activity results in more frequent RNase H cleavage and increased strand transfer. In contrast, slow RNase H activity results in reduced strand transfer.

10.2 Mechanisms of strand transfers

The mechanisms of strand transfers have been investigated *in vitro* using donor-acceptor RNA template systems. Typically, synthesis is initiated from a primer annealed to the 3' end of the donor template. As the primer is extended through a region of the donor RNA that is homologous to a corresponding 3'-terminal region of the acceptor template, the primer terminus switches to the acceptor and synthesis of DNA continues on the acceptor (19).

10.2.1 Invasion-driven transfer in vitro

Invasion-driven transfer is a mechanism in which the first interaction of the cDNA with the acceptor RNA occurs at a site upstream of the DNA primer terminus. The process is termed "docking" or "invasion" and the site is called the "invasion site" **(Figure 39)** (283).

The site is a region cleared by RT RNase H activity so that acceptor can interact with the cDNA. An early model proposed by Negroni and Buc (283) indicated that as synthesis proceeds, the chaperone activity of NC facilitates annealing of the acceptor RNA to the cDNA via complementary sequences of NA. As synthesis progresses and the acceptor-cDNA interaction expands, the pre-existing hairpins in the acceptor are disrupted while new secondary structures are transiently formed. Ultimately, transfer is accomplished by displacement of the cDNA primer from the donor and completion of synthesis on the acceptor

RNA template (283).

To determine the position of the invasion site, a series of DNA oligonucleotides were designed to interfere with strand transfer (58). These DNA oligomers were homologous to different regions of the acceptor RNA and therefore complementary to the cDNA in those regions. When both acceptor and these DNA oligomers were added into a reaction, the DNA oligomer would compete with the acceptor for binding to the cDNA at the invasion site, resulting in lower transfer efficiency due to inefficient invasion. Alternatively, the interaction of the acceptor RNA and the cDNA has been examined by measuring the susceptibility of the acceptor to RT RNase H cleavage in a time-dependent manner. The early cleavages define the point of invasion. This procedure has been termed "acceptor mapping". To address how and where the primer terminus switches templates, base substitutions were introduced throughout the acceptor. The point of strand transfer could be determined then by sequencing the transfer product and identifying the mutations that were incorporated. Using these techniques, studies of several different template systems have revealed that invasion occurs well upstream of the point of terminus transfer, thereby supporting the invasion-driven transfer mechanism (136, 329, 348).
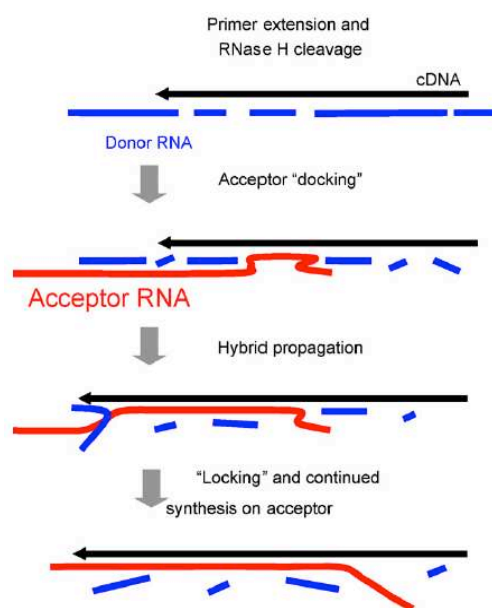
Figure 39. Invasion-driven transfer (19). When the RT extends the DNA primer (black), it periodically cleaves the donor template (blue). A series of closely spaced RNase H cuts creates a region of single stranded cDNA and allows the acceptor (red) to anneal to the cDNA. This step is termed 'docking' or 'invasion'. The acceptor–cDNA hybrid propagates by branch migration, displacing the cleaved pieces of donor template. Eventually, the primer terminus is released from the donor and annealed or 'locked' onto the acceptor, allowing synthesis to the 5' end of the acceptor.

10.2.2 Pause-dependent transfer

Secondary structures in the RNA genome, such as hairpins and G-quartets can pause the RT during DNA synthesis promoting RNase H cleavages in the RNA template (329, 341). *In vitro* experiments showed that RNase H activity is important for strand transfer, and stalling of RT on the donor template could facilitate more efficient transfer. In the pause-dependent transfer mechanism, there is a positive correlation between pausing of RT and strand transfer (95, 97, 163, 285). It was proposed that pausing allows the RT to make adjacent cuts on the donor, facilitating the interaction between cDNA and acceptor (95-97, 163, 285).

10.2.3 Pause-independent transfer

Efficient transfer can also be facilitated through a favourable interaction between the acceptor and the cDNA in the absence of strong pausing on the donor (14, 92, 93). Studies carried out by the DeStefano group showed that in the absence of NC, transfer efficiency was significantly greater with weakly structured sequences from the *pol-vif* region than with highly structured sequences from the *gag-pol* region (93). Another transfer mechanism has also been proposed based on experimental data using substrates harbouring the DIS that is responsible for linking the diploid genome in a dimeric state (232, 276, 344). Deletion of DIS from substrates tested *in vitro* resulted in about a 4-fold decrease in transfer efficiency (14, 15). These data suggest that the dimerization of the donor and acceptor RNAs improves the efficiency of template switching by increasing the local concentration of the acceptor and thus facilitating its initial interaction with the cDNA.

Chapter 5. The first strand transfer

The first strand transfer is mediated by base pairing of the complementary R and r sequences at the 3' end of the gRNA and the ssDNA **(Figure 33)**. This step is obligatory for reverse transcription and viral replication. Since HIV-1 viral particles contain two identical copies of the gRNA, the first strand transfer can occur in an intra- or intermolecular manner (181, 202, 299, 382). The vast majority of first strand transfers appear to occur after the completion of ssDNA (292). The first strand transfer reaction is inefficient in intact virions (176, 376, 414, 417). In contrast, most of the viral DNA detected in infected cells is full-length or nearly so, suggesting that the first strand transfer occurs rapidly *in vivo* (176). Most of studies dealing with the first strand transfer rely on *in vitro* systems.

1. Roles of RNase H and NC activities in the first strand transfer

The early studies on the first strand transfer were performed *in vitro* with systems containing relatively unstructured viral donor and acceptor RNA templates (5' and 3' ends of the genome, respectively) and no NC (122, 249, 305). These works demonstrated that during ssDNA synthesis, the 5' end of the template must be degraded by the RNase H activity of RT to allow subsequent strand transfer (122, 249, 305). This RNase H activity frees the ssDNA, enabling its annealing to the R sequence in the 3' end of gRNA and continuation of minus-strand synthesis. The excess RT molecules may be important in carrying out RNase H cleavage in order to clear the minus strand DNA to allow efficient strand transfer. Taking into account the RNase H activity, the team of Bambara proposed that first strand transfer is an invasion-driven transfer **(Figure 39)** (58, 59).

The first strand transfer is greatly enhanced by NC. This enhancement is due at least in part to the ability of NC to increase the annealing of the 5' r region of the ssDNA to the 3' R region of the genome (158, 304). Since highly structured RNA and DNA molecules must be annealed, the reaction is dependent on the ability of NC to destabilize these secondary structures (158, 159). In fact, unfolding of these structures was thought to be a rate-limiting step in the annealing reactions with DNA and RNA molecules containing most of R (410). The destabilizing activity of NC makes it possible for NC to perform another function in the first strand transfer, i.e. inhibition of a competing, non-specific reaction. It was previously reported that a secondary structure in the RNA template can lead to RT pausing and a decrease in the efficiency of full-length DNA synthesis during elongation of minus-strand DNA (406).

Unexpectedly, Guo *et al.* (157) demonstrated *in vitro* that DNAs, much longer than ssDNA, were synthesized in the absence of acceptor RNA, which is later on definited as "self-priming" (SP) product. Since ssDNA contains the cTAR sequence (sequence complementary to TAR RNA) that can form a stem-loop, it could induce self-annealing at the 3' end of ssDNA, thereby allowing ssDNA to act as both primer and template for synthesis of plus-strand DNA in the absence of acceptor RNA. It is clear that if SP occurs *in vivo*, virus replication would be adversely affected. However, no SP DNAs were detected in endogenous reactions with detergent-treated HIV-1 virions, which demonstrates that HIV-1 has overcome this potential problem. NC was supposed to be the important protein for inhibiting SP induced by the cTAR structure (157). It was reported that SP in the absence of acceptor could be blocked by NC alone (230), possibly because under the conditions used, increasing amount of NC inhibited overall reverse transcription. Other groups found that NC had little effect on SP in the absence of the acceptor (102, 149, 168, 175). However, if acceptor RNA and NC were both present, SP was dramatically reduced (46, 102, 103, 149, 157-159, 168, 175, 230) and a concomitant increase in strand transfer resulted.

2. <u>NA structures involved in the first strand transfer</u>

In vitro studies showed that pre-incubation of NC with the acceptor RNA rather than the donor RNA, significantly enhanced the efficiency of minus strand transfer (283). This suggests that destabilization by NC of structures within the 3' UTR is beneficial for the annealing reaction which is necessary for the first strand transfer.
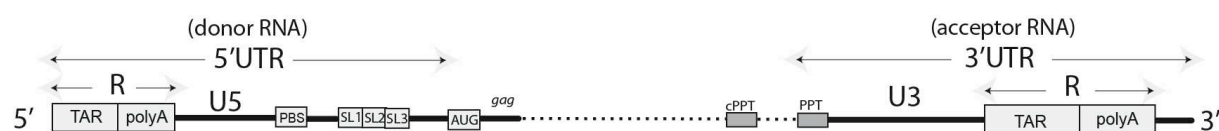
Figure 40. The 5' and 3' UTRs in HIV-1 genomic RNA (MAL isolate, GenBank: X04415.1).

The HIV-1 NL4-3 isolate is widely used by several teams in the world. Our team and other teams in France use the HIV-1 MAL isolate which is a complex recombinant of HIV-1 subtypes A, D, and I, and it contains a 23-nucleotide duplication, including parts of the PBS and the downstream regions (20). However, as shown by the aligment of the 5' and 3' UTR sequences of the two isolates **(Figure 41)**, there are no large differences in the R sequences, motifs SL1 to SL3, segment 1 and '9nt', while the PBS and PPT sequences are highly conserved.

5′ UTRs of HIV-1 MAL and NL4-3 isolates

```
                                              R
MAL    1  GGTCTCTCTTGTTAGACCAGGTC-GAGCCCGGGGAGCTCTCTGGCTAGC-AAGGAACCCAC   58
          |||||||||||||||||||||||| ||||| ||||||||||||||||||| |||||||||||
NL4-3  1  GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCAC     60

                           R
MAL   59  TGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTC-AAGCAGTGTGTGCCCATCTGTTG   117
          ||||||||||||||||||||||||||||||||| || |||||||||||||||||| |||||
NL4-3 61  TGCTTAAGCCTCAATAAAGCTTGCCTTGAGTG-CTCAAAGTAGTGTGTGCCCGTCTGTTG   119

MAL  118  TGTGACTCTGGTAACTAGAGATCCCTCAGACCACTCTAGACGGTGT-AAAAATCTCTAGC   176
          ||||||||||||||||||||||||||||||||| |||| ||| |||| |||||||||||||
NL4-3 120 TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGC   179

               PBS                            23-nt insertion in MAL isolate
MAL  177  AGTGGCGCCCGAACAGGGACTTTAAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCC   236
          ||||||||||||||||||||||||| |||||| |||||  |   |  |     | |  |||
NL4-3 180 AGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAA-A--G-C-C---A---G-AG----   223

                                                     SL1
MAL  237  AGAGAAGTTCTCTCGACGCAGGACTCGGCTTGCTGAGGTGCACACAGCAAGAGGCGA-GA   295
          || ||| |  |||||||||||||||||||||||||| |||| | ||||||||||||| ||
NL4-3 224 -GAG-A--TCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGG   279

               SL2                           SL3                  AUG
MAL  296  GCGGCGACTGGTGAGTACGCC--AATTTTTGACTAGCGGAGGCTAGAAGGAGAGAGATG   352
          ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
NL4-3 280 GCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATG   338
```

3′ UTRs of HIV-1 MAL and NL4-3 isolates

```
                         PPT
MAL  8651 GCCACTTTTTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTGGTCCCCAAAAAGACAAG  8710
          |||||||||||||||||||||||||||||| ||| ||||    |||| ||| ||||||||
NL4-3 8595 GCCACTTTTTAAAAGAAAAGGGGGGACTGGAAGGGCTAATTCACTCCCAAAGAAGACAAG  8654

MAL  8711 AAATCCTTGATCTGTGGGTCTACCACACACAAGGCTACTTCCCTGATTGGCAGAATTACA  8770
          | |||||||||||||||||| ||||||||||||||||||||||||||||||||||| ||||
NL4-3 8655 ATATCCTTGATCTGTGGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACA  8714

MAL  8771 CACCAGGGCCAGGGGATTAGATTCCCACTGACCTTCGGATGGTGCTTTAAGTTAGTACCAA  8830
          |||||||||||||||| |||| | ||||||||| ||||||||||| ||| ||||||||| |
NL4-3 8715 CACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTACCAG  8774

MAL  8831 T-GAGTCCAGAGGAAGTAGAGGAGGCCAATGAAGGAGAGAACAACTGTCTGTTACACCCT  8889
          | |||  |||| |||||||||| ||||||||| ||||||||||||  || |||||||||||
NL4-3 8775 TTGAG-CCAGATAAGGTAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCT  8833

MAL  8890 ATTAGCCAACATGGAATGGAGGACGCAGAAAGAGAAGTGCTAAAATGGAAGTTTGACAGC  8949
          | || ||| |||||||||||| | |  || || ||||||||| || |||||||||||||||
NL4-3 8834 GTGAGCCTGCATGGAATGGATGACCCTGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGC  8893

                                                               Seg-
MAL  8950 AGCCTAGCACTAAGACACAGAGCCAGAGAACAACATCCGGAGTACTACAAAGACTGCTGA  9009
          |||||||| || ||| |||||| |||||||||||||  |||||||| |||  |||||||||
NL4-3 8894 CGCCTAGCATTTCATCACGTGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGA  8953

          ment 1                     '9nt'
MAL  9010 CACAGAAGTTGCTGACAGGGGACTTTCCGCTGGGGACTTTCCAGGGGAGGCGTAACTTGG  9069
          ||  |||  ||||| || |||||||||||||||||||||||||||| | ||||| | |||
NL4-3 8954 CATCGAGCTTGCT-ACAAGGGACTTTCCGCTGGGGACTTTCCAGGG-AGGCGTGGCCTGG  9011

MAL  9070 GCGGGACCGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTTTTCGCCTGT  9129
          |||||| | |||||||||||| | |||||||||||||||||||||||||||||||||||||
NL4-3 9012 GCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCTGCTTTTTGCCTGT  9071

                                              R
MAL  9130 ACTGGGTCTCTCTTGTTAGACCAGGTC-GAGCCCGGGGAGCTCTCTGGCTAGCAAGG-AAC  9187
          ||||||||||||| |||||||||||||| ||||| ||||||||||||||||| |||| |||
NL4-3 9072 ACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAAC  9131

               R
MAL  9188 CCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAA  9229
          ||||||||||||||||||||||||||||||||||||| ||||
NL4-3 9132 CCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCAA  9173
```

Figure 41. Alignment of 5' and 3' UTRs of HIV-1 MAL and NL4-3 isolates. R regions and AUG codon are indicated in orange; PBS in red; PPT in blue; motifs SL1-3, segment 1 and '9nt' in green; the 23-nt insertion in MAL isolate is indicated in violet.

2.1 RNA structures formed by the UTRs

2.1.1 Folding of the R sequence

The R sequence (96 nt in the MAL isolate) present at both ends of the genome **(Figure 40)** is required for the first strand transfer. Deletion and mutational analysis of the R region shows that not all the R region is required for transfer in HIV-1. HIV-1 can still effectively complete replication with an R homology much shorter than 97 nt (32). Note that the sizes of the R elements in different retroviruses vary significantly from 15 nt in Mouse mammary tumor virus (MMTV) to 247 bases in HLTV-2 (79, 308, 347). Extensive *in vitro* studies strongly suggest that the structure adopted by the R region play an important role in the first strand transfer (30, 34, 222, 271).

The 5' R sequence folds into secondary structures corresponding to the TAR and the poly(A) hairpins **(Figure 34)**. In an RNA corresponding to the 3' UTR, 3' R sequence folds *in vitro* into secondary structures **(Figure 42)** corresponding to the 3' TAR stem-loop and the upper part of the poly(A) hairpin (205). The stem-loop structure of 3' TAR is important for reverse transcription (164). Mutations in TAR that either alter the upper stem-loop orientation or the apical loop sequence greatly reduced the synthesis of minus-strand "strong-stop" DNA (ssDNA) (164, 165). The 3' poly(A) sequence can form a hairpin (31, 205), but there is no data showing that this secondary structure is present in the entire HIV-1 RNA genome. *In vitro* studies suggest that the 3' poly(A) hairpin structure is not of critical importance for the strand transfer (205, 271).

Figure 42. Secondary structure model proposed for the 3'R region of the MAL isolate (Adapted from (205)). The palindrome sequence is boxed.

66

2.1.2 Circularization of the genome by the TAR-TAR interaction

A palindromic sequence of 10 nucleotides in the apical stem-loop of the TAR hairpin could be involved in genome dimerization, in addition to the primary site of dimerization that is named DIS (5). Since the TAR is present at both ends of the viral RNA, the palindromic sequences could interact and circularize the HIV-1 genome **(Figure 43)**. Stem-loops and loop-loop kissing structures formed by sequences at both ends of the RNA genome were demonstrated to participate in genome circularization in many plant viruses (264).
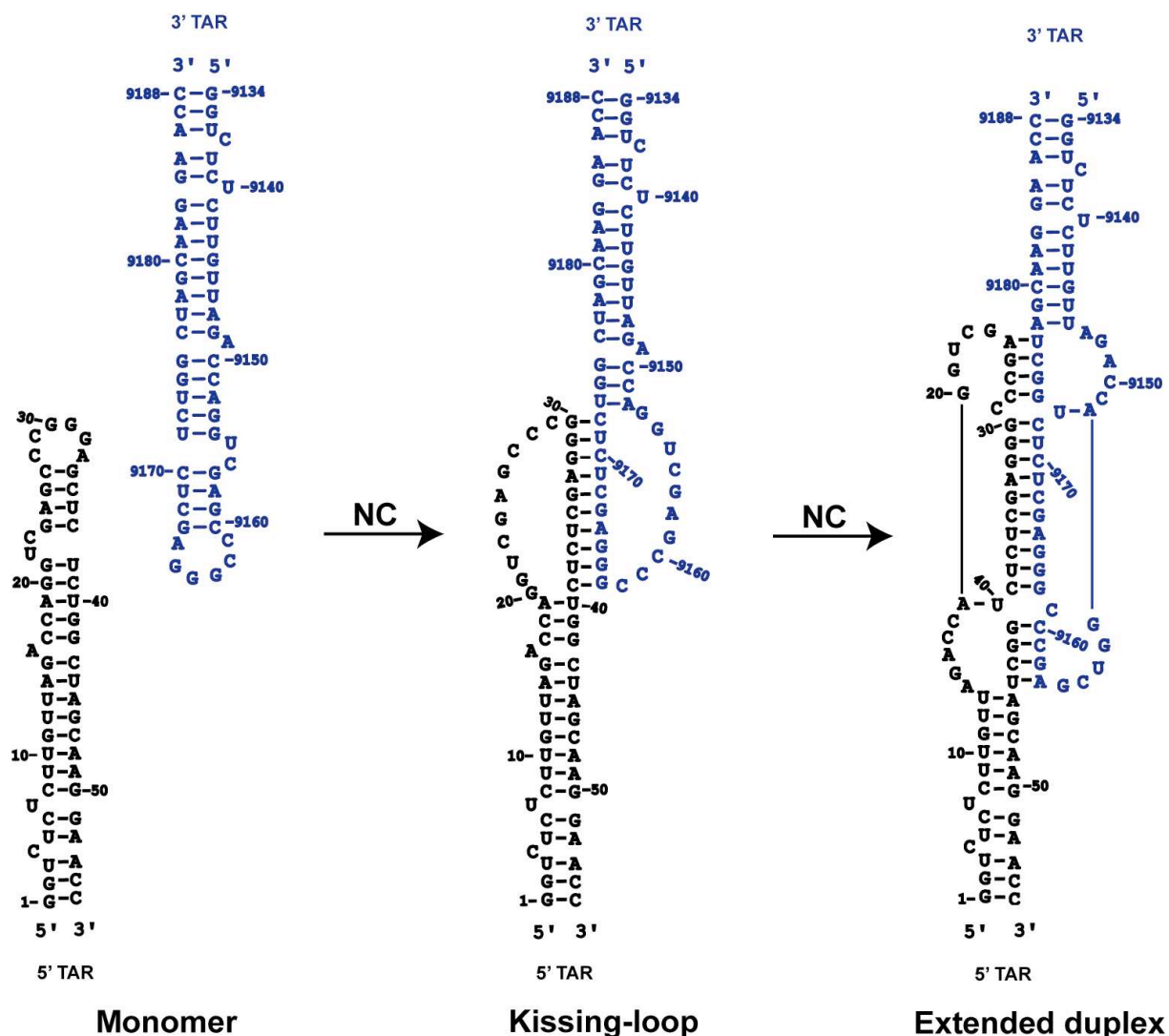
Figure 43. Model of the structural rearrangements in the TAR hairpins of HIV-1 (MAL isolate). TAR-TAR dimerization induced by NC gives rise to a kissing-loop complex and an extended duplex.

2.1.3 Circularization of the genome by the U3-tRNA$^{(Lys,3)}$ interaction

Sequence U3 (457 nt in the MAL isolate) is unique and located at the 3' end of the genomic RNA. Early studies have identified a conserved nonanucleotide sequence in the U3 region that is complementary to the anticodon stem of tRNA$^{(Lys,3)}$ (46). The primer tRNA was

proposed to serve as a bridging factor holding RNA genome ends together. The primer tRNA[(Lys,3)] may interact with the 3' end of the viral gRNA by a base pairing interaction involving nine nucleotides of U3 and nine nucleotides of the anticodon stem of tRNA[(Lys,3)] (46). Recent studies revealed the striking result that the 9 nt sequence is a part of a much larger sequence with strong homology to the entire tRNA[(Lys,3)] **(Figure 44)** (310). This sequence stretches over regions surrounding the U3 and R border. Segment 1 is located just upstream of motif 9nt and resembles the PBS, which interacts with the 3'end of tRNA[(Lys,3)] **(Figure 44a)**.
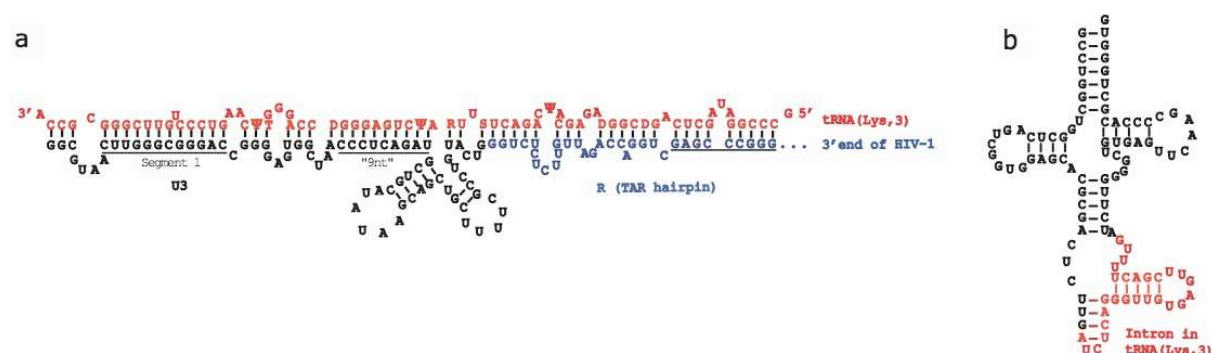
Figure 44. Sequences in HIV-1 U3R complementary to tRNA[(Lys,3)] (MAL isolate, adapted from (310)). (a) The 9nt segment in U3 proposed to interact with tRNA[(Lys,3)] is a part of a longer sequence with additional complementarities to tRNA[(Lys,3)] and could derive from an ancient tRNA gene incorporated at an early stage of HIV-1 evolution. The U3 and R sequences are indicated. The black underlines indicate probable regions of interaction with tRNA[(Lys,3)]. (b) An example of a tRNA with an intron present in its gene. The sequence shown is that of tRNA[Lys] (gene scaffold_208396) of the domestic cat (*Felis catus*).

2.1.4   *Circularization of the genome by the gag-U3,poly(A) interaction*

Detailed structure probing of RNA heterodimers indicated that sequences in the poly(A) and U3 region base pair with *gag* sequences *in vitro* (294). This *gag*-U3,poly(A) interaction can be proposed for all HIV-1 subtypes (294), suggesting that this base-pairing interaction involving the 5' and 3' ends of the genome has been conserved during HIV-1 evolution. This interaction involves nucleotides of the *gag* ORF and 16 nt of U3 and 28 nt of the poly(A) hairpin of 3' R in the NL4-3 isolate (23, 294). This interaction has been proposed to circularize the HIV-1 genome *in vivo* **(Figure 45)** (23).

2.2   **Structure of the strong-stop DNA**

The r region of ssDNA **(Figure 33)** is totally complementary to the R region of gRNA and contains the sequences complementary to TAR and poly(A), which are named cTAR and cpoly(A). Since TAR and poly(A) fold into stem-loop structures, it was thought that the

complementary sequences could also fold into stem-loop structures. Using NMR and probing methods, Zargarian *et al.* (413) investigated the structural and dynamical properties of the top half of the cTAR DNA (mini-cTAR). It was shown that the upper stem located between the apical and the internal loops is stable, but that the lower stem of mini-cTAR is unstable. The results are consistent with the functions of cTAR DNA in the first strand transfer and help to define the structure of cTAR DNA. The longest ssDNA fragment studied was of 128-nt length (157, 168) and was investigated for self-priming (SP) which can occur during the first strand transfer (157, 168). When I started my thesis, the structures adopted by the cTAR and cpoly(A) sequences in the full-length ssDNA, i.e. in the natural sequence context, were not characterized.
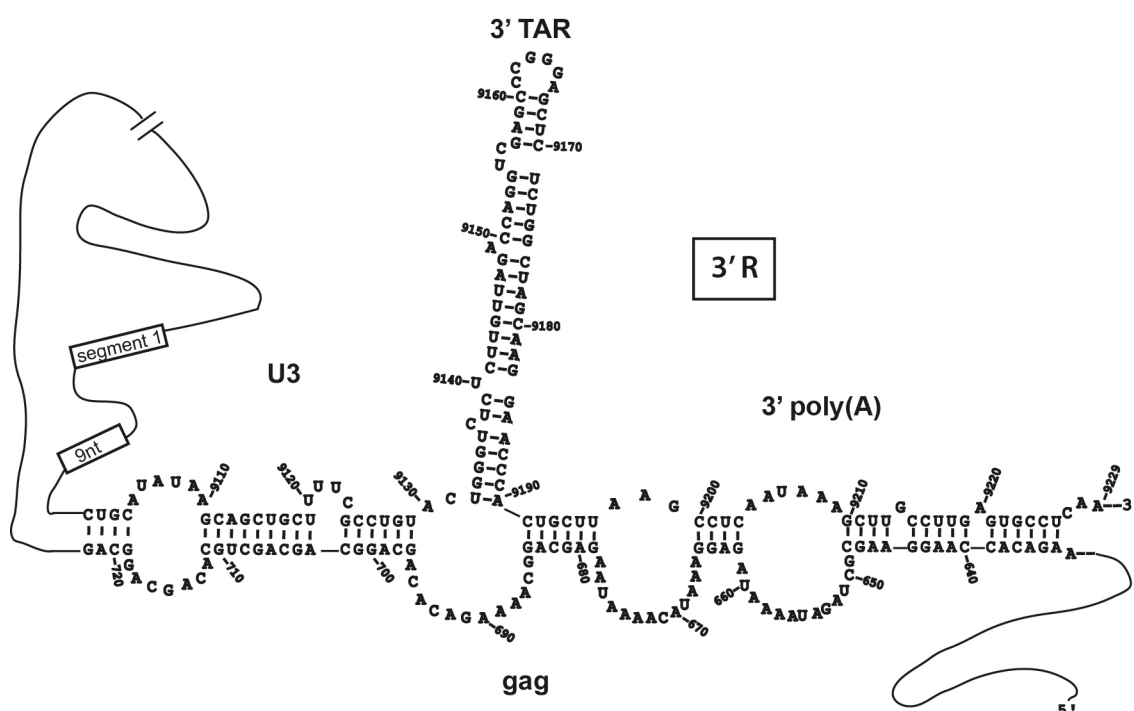
Figure 45. Proposed RNA structure of the *gag*–U3poly(A) interaction in the MAL isolate. The motif '9nt' and 'segment 1' are indicated by boxes.

3. <u>Mechanisms involved in the first strand transfer</u>

3.1 Circularization of the HIV-1 genome would facilitate the first strand transfer

Circularization of the HIV-1 genome in a way that juxtaposes the R elements should facilitate the first strand transfer. As mentioned previously, several RNA-RNA interactions **(Figure 46)** were described *in vitro* that may mediate circularization of the HIV-1 genome. The first strand transfer *in vitro* was not significantly affected (23) by mutations that prevented TAR RNA dimerization. Early studies showed that the '9nt' sequence (5'-

CCCUCAGAU, in **Figure 44**) promotes the first strand transfer *in vitro* (46). Recent studies showed that the 9 nt segment "motif 9nt" and a nearby sequence 'segment 1' (5'-CUUGGGCGGGAC, in **Figure 44**), both increased transfer *in vitro*, and most efficiently when present together (310). The team of Bambara also reported that mutations in the U3 region within 'motif 9nt' and 'segment 1' decreased from 9% to 26% the efficiency of ssDNA in infected cells (311). They showed that mutations in 'segment 1' had the greatest effect on the strand transfer, whereas the alteration of 'motif 9nt' or both sequences caused a smaller reduction than the 'segment 1' alone (311). The interaction with 'segment 1' could facilitate the displacement of tRNA$^{(Lys,3)}$ and promote the second strand transfer. Using RNA heterodimers and mutational analysis, it has been shown that the gag-U3, poly(A) interaction stimulates the first strand transfer *in vitro* (23). Circularization of the HIV-1 genome would be exclusive for unspliced gRNA because all spliced subgenomic mRNAs lack the *gag* interaction domain. To date, circularization of the HIV-1 RNA genome in the viral particle or in the infected cells remains to be demonstrated.
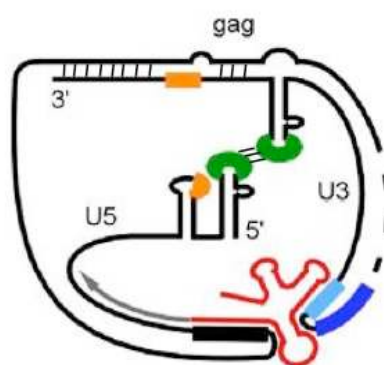
Figure 46. Proposed interactions of 5' end and 3' end of the HIV-1 RNA genome (Adapted from (309)). The model shows the interactions between the *gag* and 3'U3poly(A) sequences, and between tRNA$^{(Lys,3)}$ (red), bound to the PBS region (black thick line), and motif 9 nt (light blue) in U3. Sequences of loops of the hairpins TAR (green) and poly(A) (orange) are indicated. Sequences of segment 1 (dark blue) and the synthesized minus strand DNA (gray) are also shown. The rest of the HIV-1 genome is designated by a dotted line.

3.2   Roles of the poly(A) and cpoly(A) hairpins

In the 5' UTR, the R sequence forms the TAR and poly(A) haipins (395). Note that the lower stem of the poly(A) hairpin is extended by base-pairing with a complementary portion of the U5 region. In the 3' UTR, the TAR hairpin is formed, whereas the poly(A) hairpin is predicted but not demonstrated (395). A model has been proposed whereby the first strand transfer is facilitated by two loop-loop interactions involving the TAR and poly(A) hairpins of the acceptor RNA and the complementary cTAR and cpoly(A) hairpins of ssDNA (34).

However, an *in vitro* mutational analysis based on the insertion of the R sequence into different RNA contexts suggested that the poly(A) hairpin is not important for strand transfer (271). Consistent with this finding, R-mediated strand transfer remained efficient from an artificial internal template position, in which the TAR hairpin was formed but not the poly(A) hairpin (205). However, using progressively elongated DNA acceptors in an *in vitro* competitive approach, the bottom of TAR, i.e. the poly(A) sequence, was identified as the invasion site for the annealing reaction (349). Indeed, the DNA acceptors containing the poly(A) sequence greatly inhibit the transfer of ssDNA to the RNA acceptor when compared to other DNA acceptors containing only the TAR sequence. This annealing mechanism relies on the hypothesis that the strand transfer process occurs through a two-step mechanism: docking of the acceptor RNA onto the nascent DNA and displacement of the donor RNA by the acceptor RNA (283, 329).

3.3   Roles of the TAR and cTAR hairpins

An *ex vivo* study based on mutational analysis of HIV-1 gRNA showed that the great majority of first strand transfers occur after the completion of ssDNA synthesis (292), i.e. the entire cTAR sequence would be required for efficient strand transfer. Our group found that the upper part of the TAR hairpin structure (nucleotides 17-44 in **Figure 43**) was conserved within three different acceptor RNAs containing the R sequence that were able to promote strand transfer *in vitro* (205). The *in vitro* strand stransfer efficiency did not decrease when both donor and acceptor RNAs contained the upper part of the TAR hairpin but not the lower part of the TAR hairpin (34). In contrast, the *in vitro* strand transfer efficiency was strongly reduced when the 3' RNA could not form the upper part of the TAR hairpin because R was limited to the 5' first 29 nucleotides (83). These results are consistent with the notion that the upper part of the TAR stem-loop in the acceptor RNA is important for the strand transfer.

Since mutations in the TAR apical loop strongly decreased the strand transfer *in vitro*, Berkhout *et al.* (34) suggested that the first strand transfer involves a 'kissing' complex formed by the apical loops of TAR and cTAR hairpins **(Figure 47)**. This loop-loop interaction may be a docking step when strand transfer occurs in R in the absence of NC (271). This hypothesis was consistent with the findings of our group showing that mutations preventing the putative loop-loop interaction almost completely abolished the annealing of the cTAR hairpin to 3' UTR in the absence of NC (205). Furthermore, in the presence of NC the annealing rate of loop mutant was 2.8-fold reduced compared to the wild-type (205).

Most of *in vitro* studies on the annealing mechanism were performed with the short TAR and cTAR sequences. The TAR RNA sequence forms a stable hairpin that is barely destabilized by NC (one base pair per molecule) (36). Our group showed that the NC zinc fingers interact with the guanine residues of the TAR apical loop, while the basic residues interact with the adjacent stem. FRET-based assays performed with the cTAR sequence support a dynamic structure of the cTAR hairpin, involving equilibrium between both the closed conformation and the partially open 'Y' conformation (12, 36, 175, 243). In the 'Y' conformation, the lower stems are open and the upper stems are closed. FRET studies suggest that NC largely enhances the fraying of the cTAR hairpin ends, shifting the distribution of hairpin conformations toward the more open structures (26, 36, 175). Studies using TAR RNA and cTAR DNA hairpins derived from the HIV-1 NL4-3 isolate, suggest that both the apical loops and the 3'/5' termini of complementary hairpins are the initiation sites for the annealing reaction (242, 386). Indeed, Vo *et al.* (386) reported that cTAR DNA-TAR RNA annealing in the absence of NC depends on nucleation through the apical loops, whereas cTAR DNA-TAR RNA annealing in the presence of saturated NC depends on nucleation through the 3'/5' termini, resulting in the formation of a 'zipper' intermediate **(Figure 47)**.
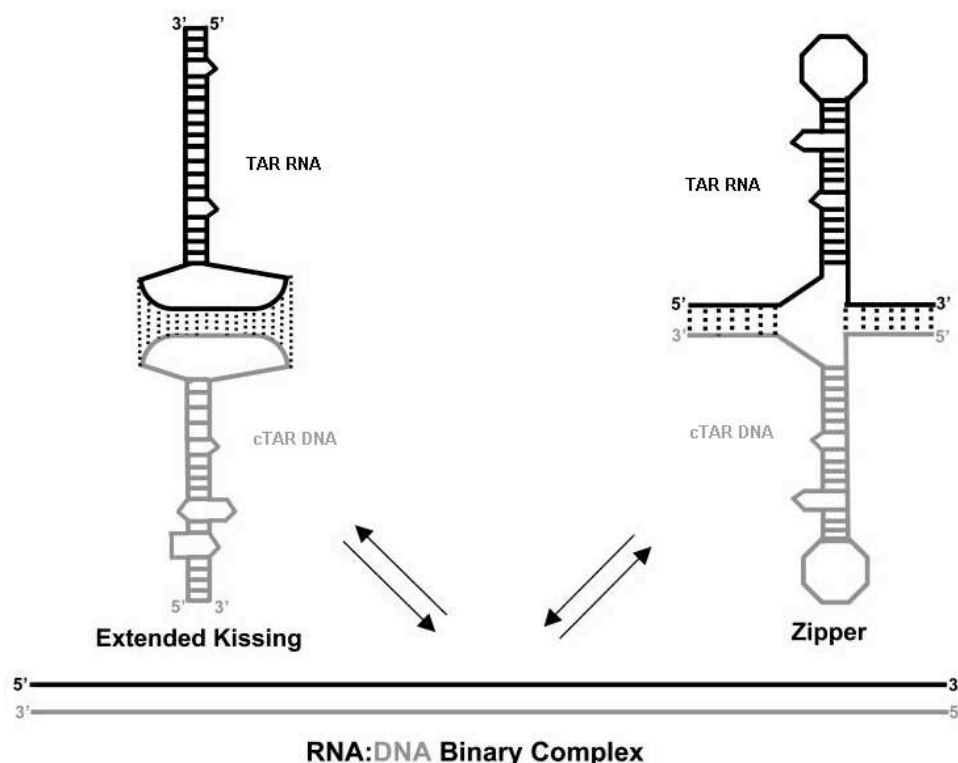
Figure 47. Two pathways of full-length TAR RNA-cTAR DNA annealing (Adapted from (386)). Left: loop-loop 'kissing' pathway, which involves initial formation of an extended kissing complex followed by subsequent strand exchange to form the fully annealed duplex. Right: 'zipper' pathway involving nucleation through the 3'-5' termini resulting in formation of a zipper intermediate followed by conversion to the fully annealed duplex.

Moreover, cTAR DNA and TAR RNA anneal *via* both 'kissing' and 'zipper' pathways under subsaturating concentrations of NC (386). The annealing process has also been investigated using sequences derived from the HIV-1 MAL isolate (143, 205). Thus, Godet *et al.* (143) reported that under low-salt concentrations, the TAR hairpins anneal almost exclusively *via* the 'zipper' pathway in the presence of NC(12-55), a truncated form of NC which lacks the N-terminal domain and is therefore a poor aggregating agent. 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 gRNA.

74

<u>OBJECTIVES</u>

Characterization of mechanisms involved in the first strand transfer contributes to gain insights into the reverse transcription process and the associated genetic recombination. Indeed, the strand transfer process is responsible for the recombination events that produce resistance to antiretroviral drugs that constitute a major problem in the anti-HIV-1 therapies. In the laboratory, we investigate the NC-mediated annealing process that is in a large part responsible for the first strand transfer in HIV-1. The main purpose of the laboratory is to understand how NC interacts with ssDNA and the 3' end of gRNA (3' UTR) and mediates annealing of these molecules that are folded.

As mentioned in the introduction, the cTAR sequence, a part of ssDNA, plays an important role in the annealing process. The first aim of my PhD project was to determine the cTAR secondary structure and the NC binding sites within cTAR.

Although the full-length ssDNA is the natural actor in the NC-mediated annealing process, its annealing to 3' UTR has not been studied. Thus, the annealing pathways are unknown for the full-length ssDNA. The second aim of my project was to study the annealing of the full-lengh ssDNA to the 3' UTR under conditions mimicking the natural conditions of the first strand transfer in order to identify the annealing pathway(s).

The ssDNA is probably folded when it interacts with NC and the 3' UTR. It is likely that ssDNA's folding has an effect on the efficiency of the NC-mediated annealing process. To date, the ssDNA structure and the NC binding sites within this structure have not been determined. In addition, we do not know if double-stranded regions of ssDNA are preferentially destabilized by NC. So the third aim of my PhD project was to determine the secondary structure of the full-lengh ssDNA in the absence or in the presence of NC using chemical and enzymatic probes. This study should allow to charaterize within ssDNA the double-stranded regions destabilized by NC and the regions protected by NC.

My study should provide new data on the molecular mechanisms that are responsible for annealing of ssDNA to the 3' end of the genomic RNA.

MATERIALS and METHODS

1. <u>NC preparation</u>

Full-length NC (55 amino acids) was synthesized by the Fmoc/opfp chemical method and purified to homogeneity by HPLC (87). NC was dissolved at a concentration of 2 mg/ml in a buffer containing 25 mM HEPES (pH 6.5), 50 mM NaCl and 2.2 mol of $ZnCl_2$ per mole of peptide.

2. <u>Oligonucleotides</u>

DNA oligonucleotides were purchased from Eurogentec. In the following oligonucleotide sequences, the upper-case 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.

2.1 Oligonucleotides for construction of plasmids

O1 5'-gcagaattctggagataatacgactcactataGGTCTCTCTTGTTAGAC-3' (1-17)
O1DIS 5'-CCAGCGCTGAGGaaaACACAGCAAGAGGCG-3' (263-292)
O2 5'-gcagcggtcgacGGTATCACTTCTGGGC-3' (876-861)
O3 5'-gcaatacGGTCTCTCTTGTTAGACCAGGTCGAGCaatttgGCTCTCTGGC
   TAGC-3' (1-47)
O4 5'-gcaatacGGTCTCTTGTTAGACCAGGTCGAGCCCttGAGCTCTCTGGC
   TAG-3' (1-46)
O5 5'-ccgccgtgactaccgcgagacccac-3'
O6 5'-gctcgaaCCCGGGCTCGACCTGGTCTAACAAGttcttCCtatagtgagtcgtattc-3'
   (27-7)
O7 5'-catccaCCCGGGAGCTCTCTGGCTAGCAAGtctctggCACTGCTTAAGCC
   TC-3' (32-70)
O8 5'-gaagaattctggagtaatacgactcactataGGAGGCGTAACTTGGG-3' (9055-9070)
O9 5'-atcgataccgtcgacc-3'
O10 5'-gcagaattctggagtaatacgactcactataGGAGCTTTTGATCTCAG-3' (8635-9229)
O11 5'-gcagaattctggagtaatacgactcactataGGCGCCCGAACAGG-3' (180-193)
O12 5'-gcagccttctggagtaatacgactcactataGGGACTCGAAAGCGG-3' (215-229)

2.2 Oligonucleotides for labeling and synthesis of cTAR, ssDNAs and DNA size markers

cTAR 5'-GGTTCCTTGCTAGCCAGAGAGCTCCCGGGCTCGACCTGGTCTAA
   CAAGAGAGACC-3' (55-1)
cPBS 5'-GTCCCUGTTCGGGCGCCA-3' (196-179)
P158 5'-CTGCTAGAGATTTTTACACCG-3' (158-178)
CompSS 5'-ctGGTCTCTCTTGTTAGACCAG-3' (1-20)
PU3b1M 5'-TCCGGATGTTGTTCTC-3' (8990-8975)

3. <u>Construction of plasmids</u>

Standard procedures were used for restriction enzyme digestion and plasmid construction (312). 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 (Eurofins MWG Operon).

Plasmids pHIVCG-4 and pHIVCG8.6 contain DNA fragments of the HIV-1 genome derived from the MAL isolate (80, 229). 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. Plasmid pYC5'DIS was generated by PCR amplification of linearized pYC5' with EcoRI using oligonucleotides O1DIS and O2. The resulting PCR product was digested with Bpu10 I and SalI and ligated into pYC5' digested with the same enzymes. Plasmids pBCSL1 and pBCSL2 were generated by PCR amplification of linearized pYC5' with EcoRI using oligonucleotide O5 with oligonucleotides O3 and O4, respectively. The resulting PCR products were digested with BsaI and ligated into pYC5' digested with the same enzyme. Plasmid pCCinv was generated by PCR amplification of linearized pYC5' with AvaI using oligonucleotides O6 and O7. 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 O8 and O9. The resulting PCR product was digested with EcoRI and XhoI and ligated into pCG44 (119) digested with the same enzymes. Plasmid pFC3'UTR was generated by PCR amplification of linearized pHIVCG8.6 with EcoRI using oligonucleotides O9 and O10. The resulting PCR product was digested with EcoRI and XhoI and ligated into pCG44 digested with the same enzymes. Plasmids pBC180wt and pBC215wt were generated by PCR amplification of linearized pYC5' with EcoRI using oligonucleotide O2 with oligonucleotides O11 and O12, respectively. The resulting PCR products were digested with EcoRI and SalI and ligated into pYC5' digested with the same enzyme. Plasmids pBC180DIS and pBC215DIS were generated by PCR amplification of linearized pYC5'DIS with EcoRI using oligonucleotide O2 with oligonucleotides O11 and O12, respectively. The resulting PCR products were digested with EcoRI and SalI and ligated into pYC5' digested with the same enzyme.

4. *In vitro* RNA synthesis and purification

Plasmids pYC5', pYC5'DIS, pBCSL1, pBCSL2 and pCCinv were digested with HaeIII to generate templates for *in vitro* synthesis of RNAs 1-415wt, 1-415DIS, 1-415SL1, 1-

415SL2 and 1-415inv, respectively. Plasmids pBC180wt, pBC180DIS, pBC215wt and pBC215DIS were digested with AvrII to generate templates for *in vitro* synthesis of RNAs 180-816wt, 180-816DIS, 215-816wt and 215-816DIS, 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 3'-2 (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 (615 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). Plasmid pHIVCG 8.6 was digested with XhoI to generate the template for *in vitro* synthesis of RNA 3'HIV-1 (662 nucleotides). Five μg of the cleaved plasmid was transcribed with T3 RNA polymerase under the conditions stipulated by the MEGAscript® T3 high yield transcription kit (Ambion). RNAs were purified as described (312). RNAs were loaded onto a denaturing polyacrylamide gel in 89 mM Tris-borate, 2 mM EDTA, electrophoresed, visualized by UV shadowing. The gel bands containing RNAs were excised from the gel, and RNAs were extracted from the gel bands by passive elution that relies on (diffusion processes). Then, RNAs were precipitated by ethanol and the RNA precipitates were dissolved in distilled water and dialyzed (Millipore filters type V6, 0.025 mm) for 40 min against distilled water. The purity and integrity of the RNA transcripts were checked on denaturing polyacrylamide gels and the concentration was determined by UV spectroscopic measurement at 260 nm.

5. 5' end labeling of RNA 3'-2

Purified RNA 3'-2 was treated with alkaline phosphatase from calf intestine (Roche Molecular Biochemicals) and 5' end labeled using phage T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (Perkin Elmer). The 5' end-labeled RNA 3'-2 was then purified by electrophoresis on a denaturing 8% polyacrylamide gel. The gel band containing RNA 3'-2 was excised from the gel, and RNA 3'-2 was extracted from the gel band by passive elution that relies on diffusion processes. Then, the purified 5' end-labeled RNA 3'-2 was precipitated by ethanol and the RNA precipitate was dissolved in distilled water and conserved at -20°C.

6. <u>Assays of labeling DNAs at their 3' end</u>

One pmol of cPBS and 1.5 pmol of RNA 1-415 in 7.5 µl of water were heated at 90 °C for 2 min and frozen for 5 min on dry ice-ethanol bath. Then, 6.9 µl of reaction buffer (final concentrations: 78 mM KCl, 1 mM DTT and 52 mM Tris-HCl, pH 7.8) containing 0.5 units of HIV-1 reverse transcriptase (Worthington) was added and the sample was pre-incubated at 37 °C for 10 min before the primer extension reaction was initiated by 0.3 µl of MgCl$_2$ and 0.3 µl of dNTPs. Final reaction contained 50 mM Tris-HCl (pH 7.8), 75 mM KCl, 2 mM MgCl$_2$, 1 mM DTT and 300 µM dNTPs. Reaction was incubated at 37°C for 1h and terminated by addition of 10 µl of 0.3 M NaOH-0.05 M EDTA and heating at 90°C for 15 min in order to destroy the RNA template. Then the sample was extracted by phenol-chloroform followed by ethanol precipitation and the dried pellet was resuspended in 10 µl of Tris-HCl (10 mM, pH 7.8). The sample underwent another ethanol-precipitation and the dried pellet was dissolved in distilled water. Since we estimated that the recovery rate of full-length ssDNAs was approxomately 50%, the pellet contained about 0.5 pmol of full-lenth ssDNA; 0.75 pmol CompSS (ssDNA:CompSS, v:v = 1:1.5) was then added and the mixture was heat-denatured like mentioned above.

For the assays using Klenow fragment of DNA polymerase I (New England Biolabs), 1 µl of Klenow buffer (final concentration: 50 mM NaCl, 10 mM MgCl$_2$ and 10 mM Tris-HCl, 1 mM DTT, pH7.9) was added. The annealed ssDNA was then incubated at 25 or 37 °C for 15 min with 0.8 µM [α-^{32}P] dATP (Perkin Elmer) and 0.05 unit of Klenow fragment. For the assays using Taq DNA polymerase (New England Biolabs), 1 µl of Taq buffer (final concentration: 50 mM KCl, 0.5 or 1.5 mM MgCl$_2$ and 10 mM Tris-HCl, pH8.3) was added. The annealed ssDNA was then incubated at 75°C for 15 min with 0.8 µM [α-^{32}P] dATP (Perkin Elmer) and 0.5 unit of Taq DNA polymerase. For the assays using AMV RT (Invitrogen), 1 µl of RT buffer (final concentrations: 30 mM KCl, 8 mM MgCl$_2$, 1 mM DTT and 50 mM Tris-HCl, pH8.5) was added. The annealed ssDNA was then incubated at 42°C for 15 min with 0.8 µM [α-^{32}P] dATP and 2 units of AMV RT. All the reactions were stopped by adding 10 µl EDTA (75 mM) and followed by phenol-chloroform extraction and two ethanol precipitations. The dried pellet resuspended in 7 µl of loading buffer A (7 M urea, 0.03 % w/v bromophenol blue and 0.03 % w/v xylene cyanol). G and T+C sequence markers of the labeled ssDNAs were produced by the Maxam-Gilbert method (257).

7. <u>Labeling and purification of cTAR</u>

Full-length cTAR was first purified by electrophoresis as described (312). The cTAR DNA was loaded onto a 10% denaturing polyacrylamide gel, electrophoresed and visualized by UV shadowing. The gel band containing cTAR was excised from the gel, and cTAR was extracted from the gel band by passive elution that relies on diffusion processes. Then, the purified cTAR was precipitated by ethanol and the DNA precipitate was dissolved in distilled water. The purity and integrity of the cTAR DNA was checked on a denaturing 11% polyacrylamide gel and the concentration was determined by UV spectroscopic measurement at 260 nm. The purified cTAR DNA was then 5' end labeled using phage T4 polynucleotide kinase (New England Biolabs) and $[\gamma\text{-}^{32}P]$ ATP (Perkin Elmer). The reaction was carried out in a final volume of 10 µl. 50 pmol of cTAR was dissolved in 3 µl of distilled water. Then 1 µl of labeling buffer was added (final concentrations: 70 mM Tris-HCl, 10 mM $MgCl_2$ and 5 mM dithiothreitol (DTT), pH 7.6). T4 polynucleotide kinase (10 U) and $[\gamma\text{-}^{32}P]$ ATP (50 µCi, 16.67 pmol) were also added into the mixture. This mixture was incubated 45 min for 37°C. 10 µl of loading buffer A was added to the mixture. The 5' end labeled cTAR was then heat-denatured and ready for purification. Full-length cTAR was also labeled at its 3' end as follows. Twenty-five pmol of the cTAR DNA and 37.5 pmol of the oligonucleotide CompSS were mixed in a 1:1.5 ratio and incubated at 75°C for 10 min in the Taq polymerase buffer (final concentrations: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.5 mM $MgCl_2$ and 0.4 µM $[\alpha\text{-}^{32}P]$ dATP (Perkin Elmer)). Then, five units of Taq DNA polymerase (New England Biolabs) were added to the mixture. The reaction was carried out in a final volume of 50 µl and incubated at 75°C for another 15 min. The labeling reaction was then stopped by adding 10 µl EDTA (75 mM) and followed by two phenol-chloroform extractions and two ethanol precipitations. The dried pellet resuspended in 40 µl of loading buffer A and heat denatured was ready for purification. The 5' and 3' end-labeled cTAR DNAs were loaded onto a 10 % denaturing polyacrylamide gel. At the end of electrophoresis, the gel band containing cTAR was excised from the gel and cTAR was extracted from the gel band by passive elution that relies on diffusion processes; Then, the purified labeled cTAR was precipitated by ethanol and the DNA precipitate was dissolved in distilled water and conserved at -20°C.

8. <u>Synthesis, labeling and purification of ssDNAs and DNA size markers</u>

First, the DNA oligonucleotides cPBS, P158 and PU3b1M were 5'-end labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (Perkin Elmer), and purified as described above for cTAR. Second, these labeled oligonucleotides were used as primers for synthesis of 5' end-labeled ssDNAs and DNA size markers (M1 and M2). More precisely, 90 pmol of 5'-end-labeled primer (3×10^5 cpm/pmol) and 60 pmol of RNAs 1-415, 3'UTR or 3'HIV-1 in 150 µl of water were heated at 90 °C for 3 min and frozen for 5 min on dry ice-ethanol bath. Then, 138 µl 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 pre-incubated at 37 °C for 10 min before the primer extension reaction was initiated by 6 µl of MgCl$_2$ and 6 µl of dNTPs. Final reaction contained 50 mM Tris-HCl (pH 7.8), 75 mM KCl, 2 mM MgCl$_2$, 1 mM DTT and 300 µM dNTPs. Reaction was incubated at 37°C for 1 h and terminated by addition of 200 µl of 0.3 M NaOH-0.05 M EDTA and heating at 90 °C for 15 min in order to destroy the RNA template. Then, the sample was extracted by phenol-chloroform followed by ethanol precipitation and the dried pellet was resuspended in 40 µl of loading buffer A.

The ssDNAs were labeled at their 3'-end by adding one labeled adenine residue as follows. The primer extension reaction with HIV-1 RT was the same as described above except that oligonucleotide cPBS or P158 was not labeled and the dried pellet containing 30 pmol of ssDNA was resuspended in 50 µl of Taq polymerase buffer containing 30 units of Taq DNA polymerase (New England Biolabs), 0.8 µM [α-^{32}P] dATP (Perkin Elmer) and 45 pmol of oligonucleotide CompSS. Following incubation at 75 °C for 15 min, the reaction was terminated by addition of 10 µl of 75 mM EDTA. The 3'-end-labeled ssDNA was extracted by phenol-chloroform followed by ethanol precipitation and the dried pellet was resuspended in 40 µl of loading buffer A.

The 5' and 3' end-labeled DNAs were loaded onto a 6% denaturing polyacrylamide gel. At the end of electrophoresis, the gel band containing ssDNA or a DNA size marker was excised from the gel and ssDNA was extracted from the gel band by passive elution that relies on diffusion processes. Then, the purified labeled DNAs were precipitated by ethanol and the DNA precipitates were dissolved in distilled water and conserved at -20°C.

9. Gel-shift annealing assay

The annealing assay was carried out in a final volume of 10 µl. The 5' end-labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) in 4 µl of water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 1 µl of renaturation buffer was added (final concentrations: 75 mM KCl, 0.2 or 2 mM MgCl$_2$ and 50 mM Tris-HCl, pH 7.8) and the sample was incubated at 37 °C for 15 min. Unlabeled RNA 3'-2 (0.45 pmol) underwent the same renaturation treatment and was then added to refolded ssDNA. The reaction mixture was then incubated at 37 °C for 15 min in presence of NC at various concentrations or incubated at 37 °C for various times in the absence of NC. At the end of incubations, 4 µl of loading buffer B was added to assays without protein, and the assays with NC were phenol-chloroform extracted and each aqueous phase was mixed with 4 µl of loading buffer B. The heat-denatured control of 5' end-labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol in 10 µl of water) was performed by heating at 90 °C for 2 min and chilling for 2 min on ice, and mixing with 4 µl of loading buffer B. The samples were analyzed by electrophoresis on a 2 % agarose (QA-AgaroseTM, Qbiogene) gel at 25 °C in 0.5 X TBE (45 mM Tris-borate (pH 8.3), 1 mM EDTA). After electrophoresis, the gel was fixed, dried and autoradiographed as described (119). Unannealed and annealed ssDNAs were quantified using a Typhoon^{TM} TRIO (GE Healthcare) and ImageQuant software. The percent of annealed ssDNA-RNA was determined as $100 \times$ (annealed/(annealed + unannealed)).

The 5' end-labeled RNA 3'-2 (0.45 pmol at 7×10^4 cpm/pmol) in 4 µl of water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 1 µl of renaturation buffer was added (final concentrations: 75 mM KCl, 0.2 or 2 mM MgCl$_2$ and 50 mM Tris-HCl, pH 7.8) and the sample was incubated at 37 °C for 15 min. Five µl of annealing buffer (final concentration as same as the renaturation buffer) was added. The samples was then incubated at 37 °C for various times in the absence of NC. At the end of incubations, 4 µl of loading buffer B was added. The heat-denatured control of 5' end-labeled RNA (0.45 pmol at 7×10^4 cpm/pmol in 10 µl of water) was performed by heating at 90 °C for 2 min and chilling for 2 min on ice, and mixing with 4 µl of loading buffer B. The samples were analyzed by electrophoresis on a 2 % agarose (QA-AgaroseTM, Qbiogene) gel at 25 °C in 0.5 X TBE (45 mM Tris-borate (pH 8.3), 1 mM EDTA). After electrophoresis, the gel was fixed, dried and autoradiographed as described (119).

10. Gel-shift analysis of ssDNAs folding

The assay was carried out in a final volume of 10 µl. The 5'-end-labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) was dissolved in 7.2 µl of water, heated at 90 °C for 2 min and chilled for 2 min on ice. Then 0.8 µl of the renaturation buffer was added (final concentrations: 75 mM KCl, 0.2 or 2 mM $MgCl_2$ 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 presence of NC at various concentrations. The incubations were stopped by extraction with phenol-chloroform and each aqueous phase was mixed with 4 µl 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 µl of water) was performed by heating at 90 °C for 2 min and chilling for 2 min on ice, and mixing with 4 µl of loading buffer B. Formation of homoduplexes was analyzed by electrophoresis on 2% agarose (QA-AgaroseTM, Qbiogene) gels at 4 °C in 0.5 X TBM (45 mM Tris-borate (pH 8.3), 0.1 mM $MgCl_2$). After electrophoresis, the gel was fixed, dried and autoradiographed as described (119). Monomeric conformations were analyzed by electrophoresis on 6 % polyacrylamide gels (37.5:1 (w/w), acrylamide/bisacrylamide) at 25 °C in 1 X TBE (90 mM Tris-borate (pH 8.3), 2 mM EDTA) and at 20 °C in 0.5 X TBM-0.2 (45 mM Tris-borate (pH 8.3), 0.2 mM $MgCl_2$) or 0.5 X TBM-2 (45 mM Tris-borate (pH 8.3), 2 mM $MgCl_2$). After electrophoresis, the gels were fixed, dried and autoradiographed. Conformers of ssDNAs were quantified using a Typhoon[TM] TRIO (GE Healthcare) and ImageQuant software.

11. Structural probing of cTAR and ssDNAs

Potassium permanganate ($KMnO_4$) and piperidine were purchased from Sigma-Aldrich. Mung bean nuclease (MB) and DNase I were purchased from New England Biolabs and Promega, respectively. Structural probing of cTAR or ssDNAs was carried out in a final volume of 10 µl.

11.1 Probing of purified cTAR

The 5' or 3' end-labeled cTAR (1 pmol at 5×10^4 cpm/pmol) in 5.5 µl of distilled water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 2.5 µl of renaturation buffer (final concentrations: 75 mM KCl, 0.2 or 7 mM $MgCl_2$ and 50 mM Tris-HCl, pH 7.8 for probing with DNase I; 75 mM KCl, 0.2 or 7 mM $MgCl_2$ 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 30 min. The reaction mixtures were then incubated at 37 °C for 15 min in the absence or in the presence of NC at various concentrations. The samples were then incubated with an enzyme as follows: 1.5, 3 or 6 U of MB for 15 min at 37 °C; or 0.05, 0.1 or 0.2 U 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 µl of loading buffer A. The cleavages sites in the labeled cTAR were identified by running in parallel the sequencing of DNA. G and T+C sequence markers of the labeled cTAR were produced by the Maxam-Gilbert method (257). All samples were analyzed by different migration times on denaturing 14 % polyacrylamide gels.

11.2 Probing of purified ssDNAs

The 5'- or 3'-end-labeled ssDNA (0.15 pmol at 4×10^5 cpm/pmol) in 7.2 µl of water was heated at 90 °C for 2 min and chilled for 2 min on ice. Then, 0.8 µl of renaturation buffer (final concentrations: 75 mM KCl, 0.2 or 2 mM $MgCl_2$ and 50 mM Tris-HCl, pH 7.8 for probing with $KMnO_4$ or DNase I; 75 mM KCl, 0.2 or 2 mM $MgCl_2$ and 50 mM sodium cacodylate, pH 6.5 for probing with MB) 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 in the presence of NC at various concentrations. The samples were then incubated with 0.75, 1.25 or 1.5 U of mung bean nuclease for 15 min at 37 °C or with 0.2, 0.3 or 0.4 U 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 µl of loading buffer A. For potassium permanganate probing, ssDNA was treated with 0.5, 1 or 2 mM of $KMnO_4$ for 1 min at 37 °C. The treatment was stopped by adding 40 µl of the termination buffer (0.7 M β-mercaptoethanol, 0.4 M NaOAc (pH 7.0), 10 mM EDTA, 25 µg/ml 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 µl of freshly diluted 1 M piperidine and heating at 90 °C for 30 min. The samples were lyophilized, resuspended in 20 µl of water, and lyophilized again. After a second lyophilization from 15 µl of water, the samples were resuspended in 7 µl of loading buffer A. G, G+A and T+C sequence markers of the labeled ssDNAs were produced by the Maxam-Gilbert method (257). In some cases, the cleavage sites in the 5'-end-labeled ssDNAs were identified by running in parallel the sequencing of RNAs 1-415 performed with the avian myeloblastosis virus reverse transcriptase (Invitrogen) following the Sanger method (205). All samples were analyzed by

short and long migration times on denaturing 8-14% and 6% polyacrylamide gels, respectively.

11.3 Probing of ssDNAs in the reverse transcription mixture

The DNA oligonucleotide cPBS was 5'-end labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (Perkin Elmer), and purified as described above for cTAR. The labeled oligonucleotide was used as primer for synthesis of 5' end-labeled ssDNAs. More precisely, 1.5 pmol of 5'-end-labeled primers (1×10^5 cpm/pmol) and 1 pmol of RNA 1-415 in 7.5 µl of water were heated at 90 °C for 2 min and frozen for 3 min on dry ice-ethanol bath. Then, 6.9 µl of reaction buffer (final concentrations: 78 mM KCl, 1 mM DTT and 52 mM Tris-HCl, pH 7.8) containing 0.5 units of HIV-1 reverse transcriptase (Worthington) was added and the sample was pre-incubated at 37 °C for 10 min before the primer extension reaction was initiated by 0.3 µl of MgCl$_2$ and 0.3 µl of dNTPs. Final reaction contained 50 mM Tris-HCl (pH 7.8), 75 mM KCl, 2 mM MgCl$_2$, 1 mM DTT and 100 µM dNTPs. Reaction was incubated at 37°C for 1h. The samples were then incubated with different concentrations of mung bean nuclease for 15 min at 37 °C or with differeny concentration of DNase I for 7 min at 37 °C. Reaction was then terminated by addition of 10 µl of 0.3 M NaOH-0.05 M EDTA and heating at 90 °C for 5 min. Then, the sample was extracted by phenol-chloroform followed by ethanol precipitation and the dried pellet was resuspended in 7 µl of loading buffer A. For potassium permanganate probing, the sample was treated with different concentrations of KMnO$_4$ for 1 min at 37 °C. The treatment was stopped by adding 60 µl of the termination buffer (0.7 M β-mercaptoethanol, 0.4 M NaOAc (pH 7.0), 10 mM EDTA, 25 µg/ml 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 µl of freshly diluted 1 M piperidine and heating at 90 °C for 30 min. The samples were lyophilized, resuspended in 20 µl of water, and lyophilized again. After a second lyophilization from 15 µl of water, the samples were resuspended in 7 µl of loading buffer A. The cleavage sites in the 5'-end-labeled ssDNAs were identified by running in parallel the sequencing of RNAs 1-415 performed with the avian myeloblastosis virus reverse transcriptase (Invitrogen) following the Sanger method (205). All samples were analyzed by short and long migration times on denaturing 8 % and 6% polyacrylamide gels, respectively.

12. Basis of structural analysis

A large part of my thesis was devoted to determination of the folding of HIV-1 single-stranded DNAs. For this purpose, we used the Mfold program (423), two enzymatic probes and one chemical probe providing informations on single-stranded and double-stranded regions. DNase I is a double-stranded-specific endonuclease that produces single-strand nicks (162). Mung bean nuclease (MB) is highly selective for single-stranded nucleic acids and single-stranded regions in double-stranded nucleic acids (94). Note that single-base mismatches in double-stranded DNA are poor substrates for MB cleavage at 37 °C (94, 369). Potassium permanganate ($KMnO_4$) can be used to detect regions of DNA that are unpaired or distorted (72, 174): 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 DNA with piperidine, the DNA backbone was cleaved at the sites of the modified thymines. The probes were used under statistical conditions where less than one cleavage or modification occurs per molecule, i.e. less than 50% of the DNA should be cleaved or modified (47, 109, 215). These conditions prevent additional cuts (named secondary cuts) in the same DNA molecule that could not reflect the original structure. Indeed, one cut in a DNA molecule can induce conformational rearrangements that potentially provide new targets to the probe **(Figure 48)**. To identify the cleavage products by electrophoresis on a denaturing polyacrylamide gel, the DNA molecule should be labeled at either the 5' or 3' end. One advantage of using 5' and 3' end-labeled DNA in parallel is that it allows distinction between primary and secondary cuts **(Figure 48)**, since the secondary cuts cannot be detected with the two types of labeling.

13. Determination of cleavage sites at the nucleotide level

The cleavage sites induced by the probes in DNAs are identified by running in parallel the sequence markers produced by the Maxam-Gilbert method (413). The Maxam-Gilbert method (257), a series of chemical reactions which cleave at one or two of the four bases, involves three consecutive steps: modification of a base, removal of the modified base from its sugar and DNA strand scission at that sugar.

Here **(Figure 49)**, a part of cTAR (nucleotides 7-13) **(Figure 49A)** is taken as an example to compare the cleavages induced by the Maxam-Gilbert method and those induced by the chemical and enzymatic probes that were used in our study.
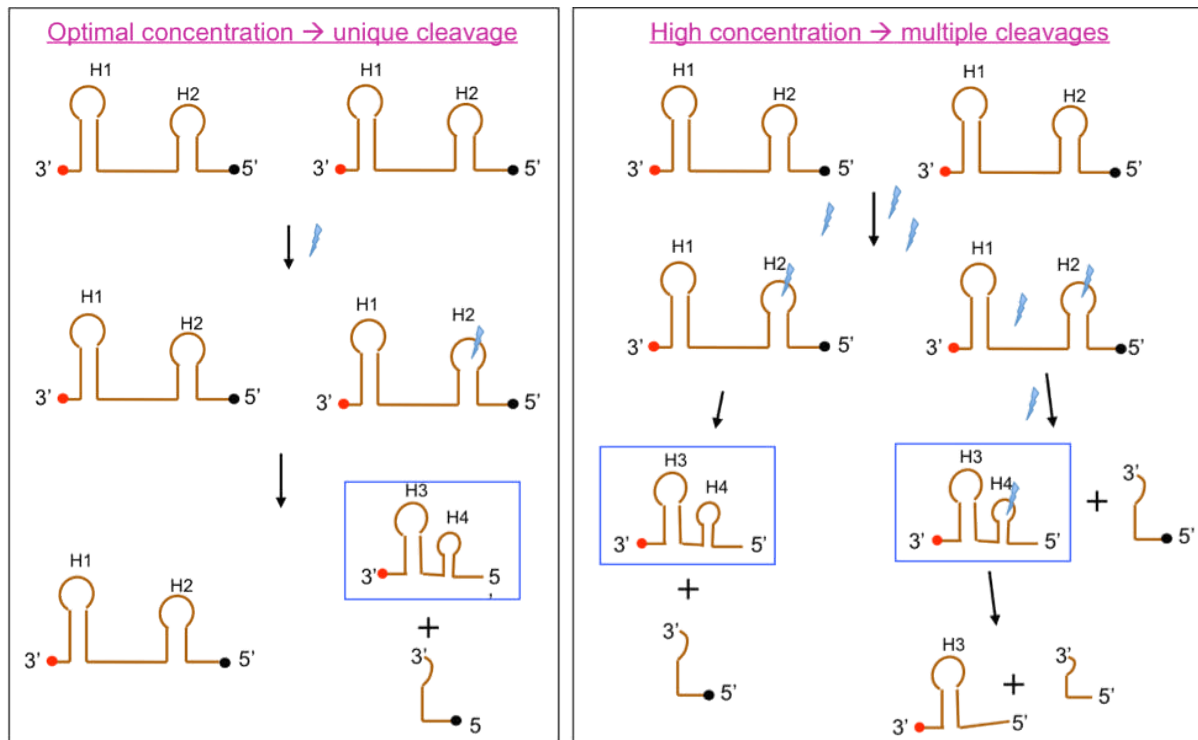
Figure 48. Examples of cleavages induced by optimal or high concentrations of structural probe. DNA molecule folded into two hairpins (H1 and H2) is an example. This molecule can be labeled at either 5' (black dot) or 3' (red dot) end. In the presence of optimal concentration of the probe (blue lightning), less than one cleavage occurs in the loop of H2 region per DNA molecule. In the presence of high concentration of the probe, primary cut induces a conformational rearrangement in the DNA molecule (indicated by the blue rectangle) that may fold into an other two-hairpin structure (H3 and H4). The H4 region could be a new target for the probe that would induce a secondary cut in this region, which distorts the structural analysis of DNA molecule.

DNA fragments generated by DNase I cleavage between G_9-C_{10} **(Figure 49B)** are compared to the sequence markers (C and G) produced by the Maxam-Gilbert method **(Figure 49C)**. For the 5' end-labeled DNA, DNase I probing product **(Figure 49B-a)** consists of T_7-T_8-G_9 with a 3'-hydroxyl terminus, whereas the sequence marker C **(Figure 49C-a)** is composed of three bases (T_7-T_8-G_9) with a 3'-phosphorylated terminus. Since the 3'-phosphate end displays an additional charge than the 3'-hydroxyl end, DNA marker C labeled at its 5' end migrates slightly faster in a denaturing polyacrylamide gel than the DNase I DNA fragment labeled at its 5' end (413). In general, the difference in electrophoretic mobility is barely detected for DNA fragments higher than 30 nucleotides. In contrast, the difference in mobility varies between 0.5 to 2 bands for short DNA fragment (121). Interestingly, if the DNA molecule is labeled at its 3' end, there is no difference in mobility between the labeled products derived from sequencing and enzymatic probing **(Figures 49 C-b and B-b)**. For chemical probing (for instance, KMnO$_4$), there is also no

difference in mobility between the probing products and the sequence markers because they have the same 5' and 3' end at the level of the cleavage site **(Figures 49 D and E)**.

For the purpose of well-defined secondary structure of studied DNAs, it is necessary to label these DNAs at their 3' end to confirm some enzymatic cleavage sites determined from the 5' end-labeled DNAs.
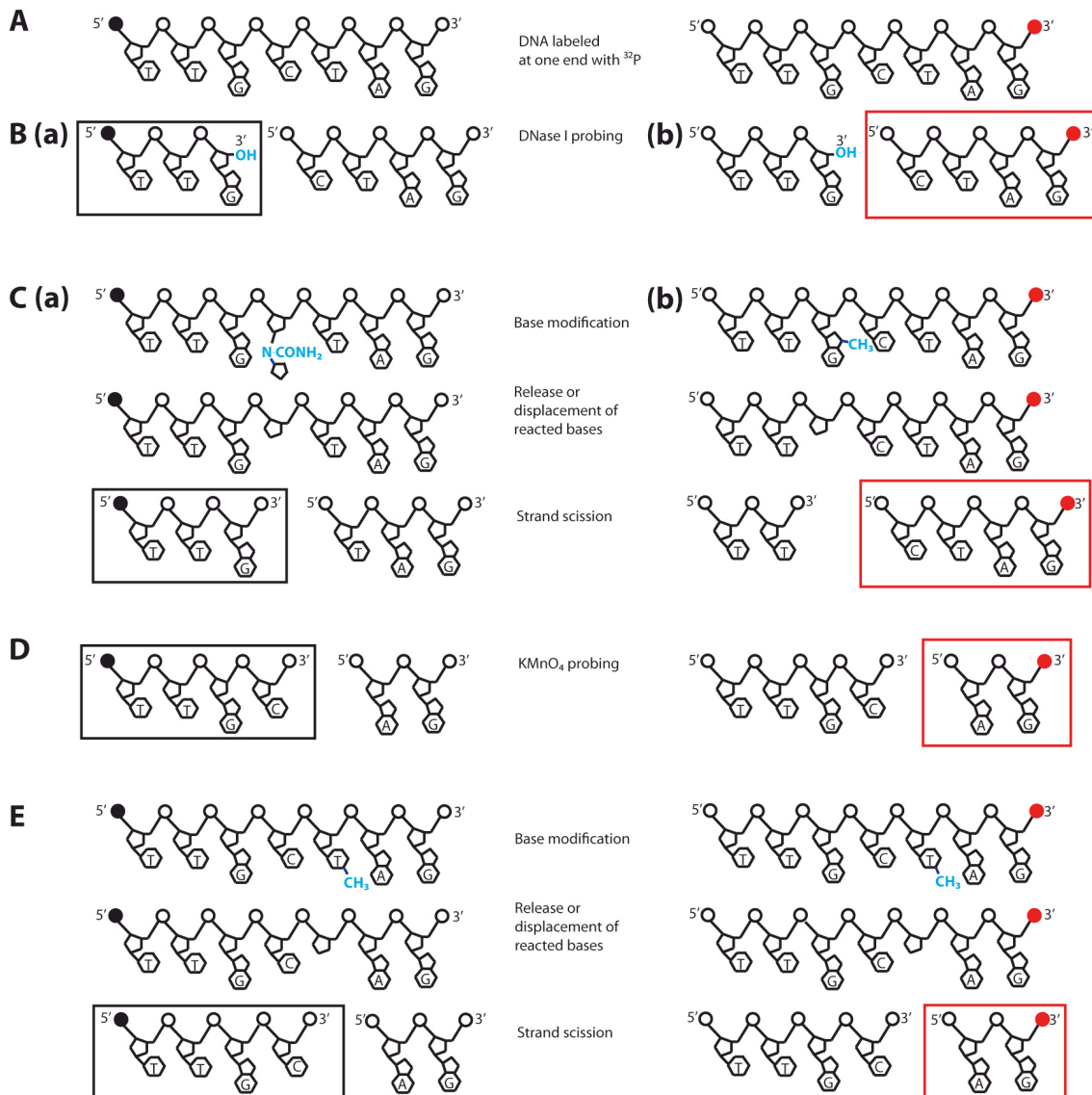
Figure 49. Determination of cleavage sites induced by enzymatic and chemical probes. (A) A part of cTAR (nucleotides 7-13) is taken as an example. DNA fragment is labeled with ^{32}P at its 5' (black dot) or 3' end (red dot). **(B)** DNase I probing. There is a strong cleavage by DNase I at G_9 in cTAR DNA (data shown in the paper in the end of this part). The black and red rectangles show the 5' end- and 3' end-labeled fragments that can be detected on the denaturing gel, respectively. **(C)** Sequence markers C (a) and G (b) by the Maxam-Gilbert method. Maxam-Gilbert sequencing requires a DNA fragment labeled at one end. First, in a limited reaction, some reagent specifically modifies one or two of the four DNA bases, generally by substituting a puring ring or a pyrimidine ring. Then, a second reaction removes this modified base from its sugar, and finally a third reaction cleaves the phosphodiaster bond of DNA by eliminating the sugar moiety and generating 3'- and 5'-phosphate ends (circles). **(D)** KMnO$_4$ probing. There is a weak cleavage by KMnO$_4$ at T_{11} in cTAR DNA (data shown in the paper in the end of this part). **(E)** Sequence marker T by the Maxam-Gilbert method.

RESULTS and DISCUSSION

Part 1. Structural analysis of cTAR DNA

1. Introduction

Mutational analysis of HIV-1 gRNA suggests that the great majority of first strand transfers occur after completion of ssDNA synthesis, i.e. with the entire cTAR sequence (292). FRET-based assays performed with the cTAR sequence suggest that the majority (~70-80%) of the cTAR DNA molecules adopt the hairpin conformation (closed form) and the remaining (20-30%) are partially or totally melted (open forms) (36, 175). Furthermore, single-molecule techniques (12, 243) were used to show that the NC-coated cTAR hairpin exhibits a dynamic equilibrium between a "Y"-shaped and a closed conformation. In the "Y" conformation, the lower part of the stem is open and the upper part of the stem is closed. Note that the secondary structure of the full-length cTAR sequence has not been determined until our study.

Unfolding of the complementary hairpins is thought to be rate-limiting in the annealing process leading to the first strand transfer (410). *In vitro* studies show that NC **(Figure 50)** stimulates the first strand transfer (58, 83, 157, 304), due, at least in part, to the NC stimulatory effect on the rate of annealing of the complementary hairpins (239, 410). NC induces a limited melting of the TAR RNA secondary structure (one base pair per molecule) (36), but largely enhances the fraying of the cTAR hairpin ends, shifting the distribution of hairpin conformations toward the more open structures (26, 36, 175). The cTAR sequence binds eight NC molecules at saturating protein concentrations under low ionic strength conditions (30 mM NaCl, 0.2 mM $MgCl_2$) (27), through both weak and strong binding sites, that were not identified (342).
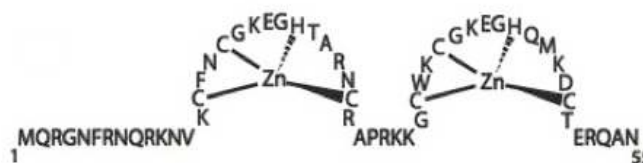
Figure 50. Sequence of NC used in our study.

The *in vitro* reverse transcription and strand transfer assays are usually performed under high-salt concentrations, i.e. at KCl concentrations ranging from 50 to 100 mM and at $MgCl_2$ concentrations ranging from 5 to 7 mM (46, 157, 227, 271, 310). Before my arrival to the laboratory, our group had begun the investigation of the annealing process using DNA and RNA sequences derived from the HIV-1 MAL isolate under high-salt concentrations (75 mM KCl, 7 mM $MgCl_2$). Thus, our group has shown that efficient annealing of cTAR DNA to the

3' end of gRNA relies on sequence complementarities between TAR and cTAR apical loops under high-salt concentrations allowing cDNA synthesis by RT and strand transfer (205). Recently, using cTAR labeled at its 5' end and chemical ($KMnO_4$) probe, and gel retardation assays, our group has obtained data on the cTAR secondary structure and the NC binding sites. Since my presence in the group, we used cTAR DNA labeled at either the 5' or 3' end and enzymatic probes (DNase I and mung bean nuclease (MB)) to provide supplementary informations on the secondary structure of the cTAR hairpin and the NC binding sites within this DNA molecule.

2. Purification and labeling of cTAR DNA

The oligonucleotide cTAR DNA (55 nt) was first purified by electrophoresis on a denaturing polyacrylamide gel to eliminate the abortive DNA fragments generated during the chemical synthesis **(Figure 51A)**. Usually, the cTAR DNA is 5' end-labeled **(Figure 51B)** for structural analysis because this type of labeling is efficient and easy to perform. However, as described in the Part 13 of the 'Materials and Methods' section, DNA should be labeled at its 3' end **(Figure 51B)** to determine the precise location of cleavage at some sites. We used the 3' end-labeling method that we developed for the structural analysis of full-length ssDNA (see the Part 2 of the 'Results and Discussion' section). The cTAR DNA was then checked for purity and integrity on a 10% denaturing polyacrylamide gel **(Figure 51C)**. Purified cTAR DNA was therefore usable for structural and functional analysis.
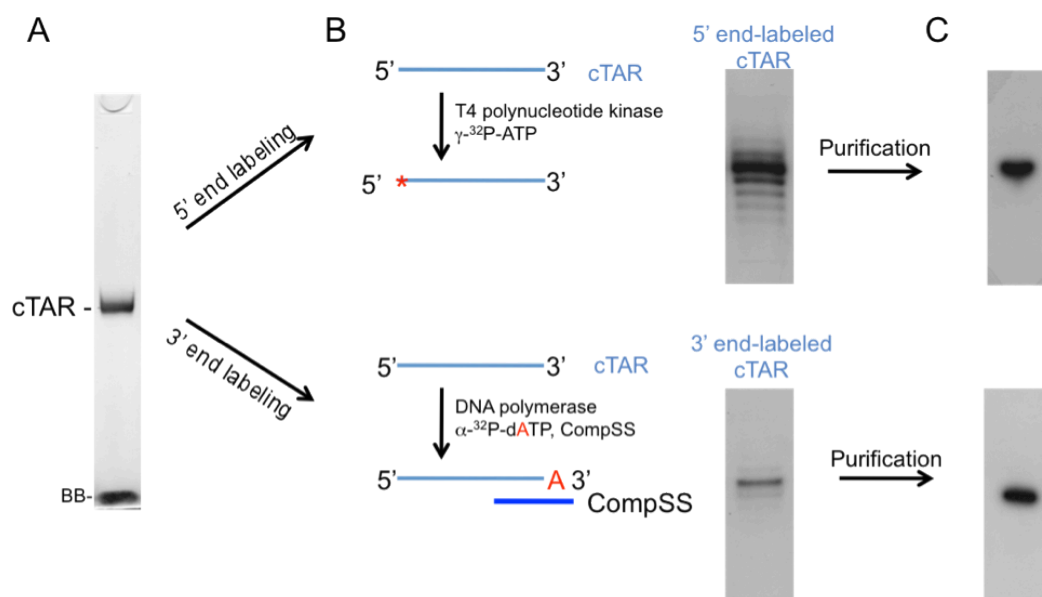
Figure 51. Purification and labeling of cTAR DNA. (A) Analysis of purified cTAR DNA on a 11% denaturing gel. BB: Bromophenol blue. **(B)** 5' and 3' end-labeling of cTAR DNA and second purification on a 10% denaturing polyacrylamide gel. **(C)** Analysis of 5' and 3' end-labeled cTAR DNA after second purification on a 10% denaturing polyacrylamide gel.

3. Structural analysis of cTAR under high-magnesium concentrations

This part of work has already been published in *Nucleic Acid Research* (see the paper at the end of this Part), and all my contributions are presented in Figure 5 of this paper.

3.1 Analysis of cTAR secondary structure in the absence of NC

The secondary structure of cTAR DNA was investigated at 37°C in the strand transfer buffer (75 mM KCl, 7 mM $MgCl_2$ and 50 mM Tris-HCl, pH7.8), i.e. under conditions allowing strand transfers and annealing of cTAR DNA to the 3' end of the gRNA (157, 205, 271). The cTAR was monomeric after incubation in the absence of NC as assessed by native agarose gel electrophoresis in the TBM and TBE buffers (data not shown). Therefore, the structural analysis was not complicated by the presence of homoduplexes. The secondary structure of cTAR DNA was probed with MB and DNase I. MB is highly selective for single-stranded nucleic acids and single-stranded regions in double-stranded nucleic acids (94). Note that single-base mismatches in double-stranded DNA are poor substrates for MB cleavage at 37 °C (94, 369). DNase I is a double-strand-specific endonuclease that produces single-strand nicks (162). Running Maxam-Gilbert sequence markers of cTAR on the same gels in parallel identified the cleavage sites. Representative examples of probing experiments are shown in **Figure 52 (A-D)** and the results of a series of independent experiments are summarized in **Figure 52E**. As mentioned in the Part 13 of 'Materials and Methods' section, the nucleases cleave the phosphodiester bond and generate a 3'-hydroxyl terminus on 5' end-labeled DNA, whereas, the Maxam-Gilbert reactions generate a 3'-phosphorylated terminus (257). The electrophoretic mobility of Maxam-Gilbert sequence markers is therefore slightly greater than that of fragments produced by nucleases. To determine the precise position of cleavage at some sites, the 3' end-labeled cTAR was used (examples in **Figure 52 A and B**).

The partially melted 'Y' conformation is the most stable cTAR structure predicted by the Mfold program (423). It has been proposed that the cTAR sequence forms the closed conformation (143, 243, 386). Therefore, the probing data were suprimposed on these two conformations **(Figure 52E)**. In both conformations, the 6-48 sequence forms the same stem-loop structure. There were strong and moderate DNase I cleavages between G_9-C_{10} and C_{10}-T_{11}, all of which are predicted to lie at the ends of stems. There were moderate and strong MB cleavages between C_{25}-C_{26}, C_{26}-C_{27}, A_{34}-C_{35} and C_{35}-C_{36} that are predicted to be in the apical and internal loops. The moderate MB cleavage between C_5 and C_6 is consistent with the partially melted conformation of cTAR.

Figure 52. Enzymatic probing of cTAR under high-magnesium concentrations. Enzymatic probing experiments were performed as described in the 'Materials and Methods' section. **(A and B)** In the absence of NC, the 3' end-labeled cTAR DNA was incubated with DNase I (0.05, 0.1 and 0.2 U) or mung bean nuclease (MB) (1.5, 3 and 6 U). **(C)** In the absence of NC, the 5' end-labeled cTAR DNA was incubated with MB (1.5, 3 and 6 U). Lanes C are controls without enzyme. G and T+C refer to Maxam-Gilbert sequence markers. Arrows indicate the cleavage sites. **(D)** 5' end-labeled cTAR DNA was incubated with MB (3 U) in the absence (lane 1) or in the presence of NC (lanes 2-5). The protein to nucleotide molar ratios were 1:8 (lane 2), 1:4 (lane 3), 1:2 (lane 4) and 1:1 (lane 5). The strong protections induced by NC at the level of MB cleavage sites are indicated by asterisks. **(E)** Secondary structures for the partially melted and closed forms of cTAR. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium, and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for MB).

94

However, the 3'/5' terminal ends may be involved in transient interactions to form the closed conformation since there was moderate DNase I cleavage between T_4 and C_5. Taken together, the enzymatic probing data are consistent with a structural heterogeneity for the 3'/5' terminal ends of cTAR in the absence of NC. This heterogeneity is supported by $KMnO_4$ probing of cTAR (see the paper at the end of this Part).

3.2 Analysis of cTAR secondary structure in the presence of NC

The cTAR was also monomeric after incubation with NC as assessed by native polyacrylamide gel electrophoresis in the TBE buffer (data not shown). Therefore, the structural analysis was not complicated by the presence of homoduplexes. To identify destabilized regions and protections induced by NC in the cTAR hairpin, enzymatic probing patterns of cTAR were compared in the absence or presence of increasing concentrations of NC. In the presence of increasing concentrations of NC, DNase I cleavage at the level of all sites decreased at the same rate, i.e. NC did not induce specific protection against DNase I (data not shown). These results indicate that the double-stranded regions of cTAR do not contain preferential binding sites for NC. Interestingly, NC induced a strong decrease in MB sensitivity for A_{34} and C_{35} **(Figure 52D)** that are located in the internal loop **(Figure 52E)**.

PhosphorImager quantification indicated that the cleavage rate for MB in the internal loop at a protein to nucleotide molar ratio of 1:4 corresponded to 52% of that observed in the absence of NC. In contrast, at the same ratio, the cleavage rate for MB in the apical loop (residues C_{25} and C_{26}) corresponded to 96% of that observed in the absence of NC. These data indicate that the internal loop is a strong binding site for NC, and that, if NC binds the apical loop, its binding is too weak to be detected at a protein to nucleotide molar ratio of 1:4. However, PhosphorImager quantification indicated that the cleavage rate for MB in the apical loop at a protein to nucleotide molar ratio of 1:2 corresponded to 55% of that observed in the absence of NC, i.e. NC binds the apical loop at protein to nucleotide molar ratio equal to or greater than 1:2 **(Figure 52D, lanes 4 and 5)**. Taken together the data indicate that the apical and internal loops are weak and strong binding sites for NC, respectively.

4. Structural analysis of cTAR under low-magnesium concentrations

Although the *in vitro* reverse transcription and strand transfer assays are usually performed at $MgCl_2$ concentrations ranging from 5 to 7 mM (46, 157, 227, 271, 310), the NC-mediated annealing process is significantly inhibited at $MgCl_2$ concentrations equal or greater than 2 mM, presumably because Mg^{2+} competes with NC for non-specific binding to the

negatively charged phosphate groups of nucleic acids (385, 405). In addition, a DNA hairpin can be greatly stabilized by increasing the $MgCl_2$ concentration from 0.2 to 2.5 mM (70). The Mg^{2+} concentration in virions and in infected cells is not known, but it could be in the range of 0.2-0.25 mM, as in the case for uninfected cells ((145) and references therein). To determine the role of magnesium in the folding of cTAR and the interactions between cTAR and NC, probing experiments with MB were performed in the presence of 0.2 mM $MgCl_2$ (see the "Materials and Methods" section).
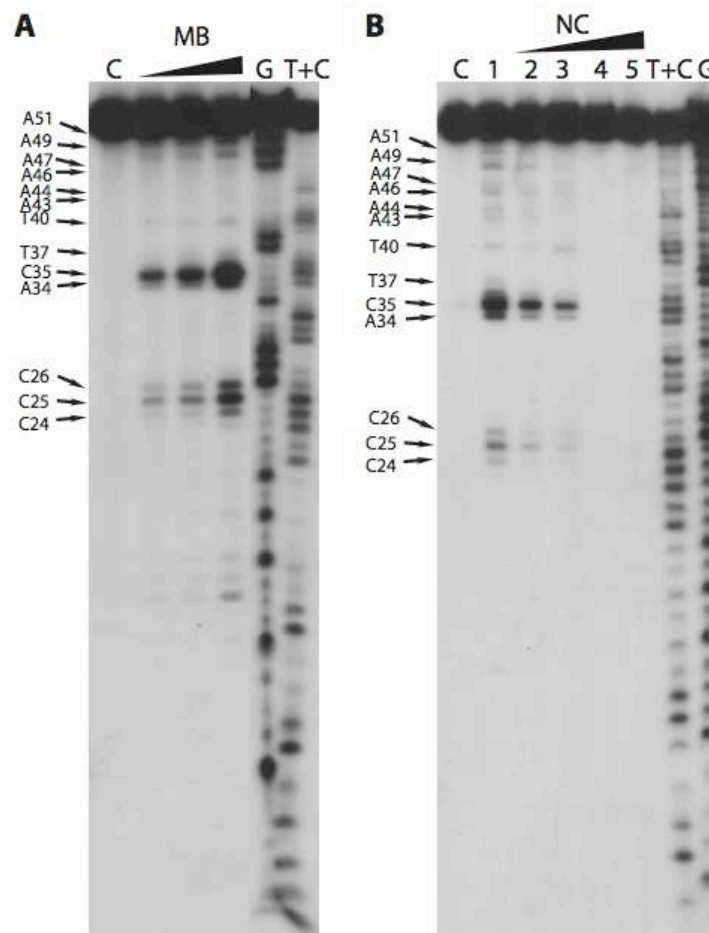
Figure 53. Enzymatic probing of cTAR under low-magnesium concentrations. MB probing was performed as described in the 'Materials and Methods' section. (**A**) In the absence of NC, the 5' end-labeled cTAR DNA was incubated with MB (1.5, 3 and 6 U). Lanes C are controls without enzyme. G and T+C refer to Maxam-Gilbert sequence markers. Arrows indicate the cleavage sites. (**B**) 5' end-labeled cTAR DNA was incubated with MB (1.5 U) in the absence (lane 1) or in the presence of NC (lanes 2-5). The protein to nucleotide molar ratios were 1:8 (lane 2), 1:4 (lane 3), 1:2 (lane 4) and 1:1 (lane 5).

In the absence of NC, the MB cleavage patterns in 0.2 mM $MgCl_2$ (**Figure 53A**) and 7 mM $MgCl_2$ (**Figure 53C**) were similar. Therefore, the folding of cTAR does not seem to depend on magnesium concentration. In the presence of NC (**Figure 53B**), the MB sensitivity decreased dramatically for both internal (residues A_{34} and C_{35}) and apical (residues C_{25} and

C_{26}) loops. PhosphorImager quantification indicated that the cleavage rate for MB in both loops at a protein to nucleotide molar ratio of 1:8 corresponded to 43-44% of that observed in the absence of NC. Moreover, at a protein to nucleotide molar ratio of 1:4, the cleavage rate for MB corresponded to 22-23% of that observed in the absence of NC. Taken together these data indicate that the degree of protection by NC was the same in both loops under low-magnesium concentrations. Therefore, in contrast to the NC interactions occuring under high-magnesium concentrations, NC does not preferentially bind the internal loop of the cTAR hairpin under low-magnesium concentrations. Finally, MB probing data are consistent with: 1) the folding of cTAR into a stem-loop structure under low-magnesium concentrations; 2) binding of NC to the apical and internal loops under low-magnesium concentrations.

5. <u>Discussion</u>

5.1 cTAR adopts two alternative conformations in the absence of NC

FRET-based assays have been developed to study the folding of the cTAR DNA molecule derived from the MAL (36) or NL4-3 isolate (175). These studies suggest that the majority (~70-80%) of the cTAR DNA molecules adopt the hairpin conformation (closed form) and the remaining (20-30%) are partially or totally melted (open forms) in the absence of NC. Using single-molecule spectroscopy (SMS) techniques and cTAR DNA molecules derived from the NL4-3 isolate, it has been reported that the partially open 'Y' conformation is the dominant form for the cTAR hairpin in the NC-meditaed annealing process (70, 243). These studies suggest that the 5' and 3' unpaired termini of the 'Y' conformation are accessible for TAR RNA-DNA annealing. Note that the double-stranded and single-stranded regions of the 'Y' conformation were deduced from the analysis of hairpin mutants. Here, nucleases were used to determine the secondary structure of cTAR derived from the wild-type MAL isolate. The probing experiments were performed under both high-salt concentrations allowing strand transfer and annealing (157, 205, 271) and low-magnesium concentrations that promote the NC-mediated annealing process (385, 405) and that are more likely in virions (70, 145). Our MB probing data suggest that the folding of cTAR is the same under low- and high- magnesium concentrations. The 6-48 sequence forms a stem-loop structure containing a C-A mismatch, a T bulge in the looped-out conformation and an internal loop **(Figure 52E)**. Consistent with the stem involving nucleotides 6-9/45-48, there was a strong DNase I cleavage between G_9 and C_{10}. Nucleotides 1-5 and 50-55 are paired in the closed conformation and are unpaired in the 'Y' conformation. Our probing data are consistent with the equilibrium between the closed and 'Y' conformations of the cTAR hairpin in the absence

of NC.

5.2   NC' effects on the structure of the cTAR DNA hairpin

FRET-based assays suggest that the populations of the partially or totally melted forms of the cTAR hairpin are moderately increased in the presence of NC (36, 175). The regions opened by NC are expected not to become accessible to MB because they probably bind the NC. Consistent with this notion, MB cleavage did not increase in the presence of NC **(Figure 52D and 53B)**. Probing data obtained with a small probe ($KMnO_4$) do not support the hypothesis (70) that a population of cTAR DNA molecules is totally melted in the presence of NC (see the paper in end of this part). Taken together, chemical and enzymatical probing data suggest that NC slightly destabilizes the lower stem that is adjacent to the internal loop and shifts the equilibrium toward the 'Y' conformation **(Figure 54)**.
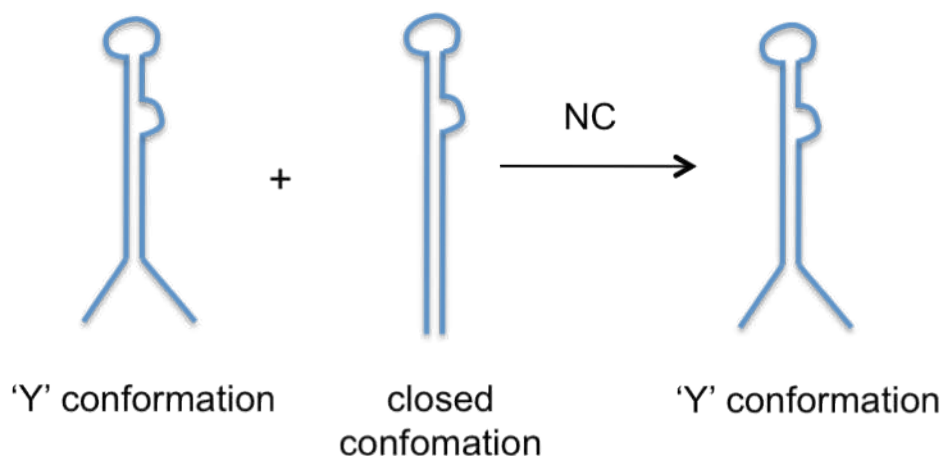
Figure 54. NC's effects on the equilibrium of 'Y' and closed conformations of cTAR DNA.

5.3   NC binding sites in the cTAR hairpin

A recent study suggests that the cTAR sequence contains weak and strong binding sites for NC in the absence of magnesium and in the presence of 100 mM NaCl (342). Identification of preferred binding sites in the cTAR molecule is necessary to propose models for the annealing mechanism mediated by NC. Our study is the first to identify the NC binding sites using NC and cTAR DNA under both high- and low-magnesium concentrations **(Figure 55)**. The DNase I probing data indicate that the double-stranded regions of cTAR bind NC but do not contain preferential binding sites for this protein. In a previous report, our group showed that the N-terminal basic domain of NC is required for providing protection against RNase V1 that cleaves double-stranded RNA (205). These results are consistent with

the notion that NC binds the double-stranded regions of nucleic acids non-specifically through electrostatic interactions of the basic residues with the phosphodiester backbone (229, 391). At high- but not low-magnesium concentrations, NC preferentially binds the internal loop than the apical loop. This may be due to stabilization of the T.G mismatch base pair by magnesium. This stablization effect would decrease the flexibility of the apical loop and therefore its ability to interact with NC.
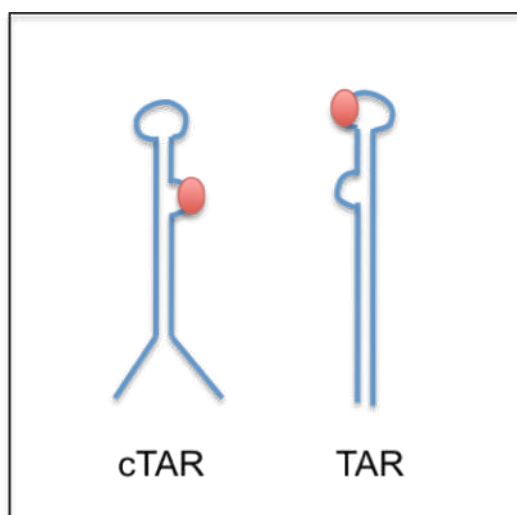
Figure 55. **Preferential binding sites for NC in cTAR DNA and TAR RNA under high-magnesium concentrations.** In cTAR DNA, NC (red oval) binds the internal loop, whereas it binds the apical loop in TAR RNA (205).

6. **Publication**

Structural determinants of TAR RNA-DNA annealing in the absence and presence of HIV-1 nucleocapsid protein

Igor Kanevsky[1], Françoise Chaminade[1], Yingying Chen[1], Julien Godet[2], Brigitte René[1], Jean-Luc Darlix[3], Yves Mély[2], Olivier Mauffret[1] and Philippe Fossé[1,*]

[1]LBPA, ENS de Cachan, CNRS, 61 avenue du Président Wilson, 94235 Cachan, France, [2]Laboratoire de Biophotonique et Pharmacologie, CNRS UMR 7213, Faculté de Pharmacie, Université de Strasbourg, 74 Route du Rhin, 67401 Illkirch, France and [3]LaboRetro INSERM #758, Ecole Normale Supérieure de Lyon, IFR 128 Biosciences Lyon-Gerland, 69364 Lyon Cedex 07, France

ABSTRACT

Annealing of the TAR RNA hairpin to the cTAR DNA hairpin is required for the minus-strand transfer step of HIV-1 reverse transcription. HIV-1 nucleocapsid protein (NC) plays a crucial role by facilitating annealing of the complementary hairpins. To gain insight into the mechanism of NC-mediated TAR RNA–DNA annealing, we used structural probes (nucleases and potassium permanganate), gel retardation assays, fluorescence anisotropy and cTAR mutants under conditions allowing strand transfer. In the absence of NC, cTAR DNA-TAR RNA annealing depends on nucleation through the apical loops. We show that the annealing intermediate of the kissing pathway is a loop–loop kissing complex involving six base-pairs and that the apical stems are not destabilized by this loop–loop interaction. Our data support a dynamic structure of the cTAR hairpin in the absence of NC, involving equilibrium between both the closed conformation and the partially open 'Y' conformation. This study is the first to show that the apical and internal loops of cTAR are weak and strong binding sites for NC, respectively. 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.

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1), like all retroviruses, replicates through reverse transcription, a complex process leading to the synthesis of double-stranded DNA from a single-stranded genomic RNA and requiring the HIV-1 nucleocapsid protein (NC) as well as at least two strand transfer reactions (1,2). The first strand transfer requires the repeat (R) sequences at both ends of the genome. During this reaction, 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 (ss-cDNA) is transferred to the 3′-end of the genomic RNA (acceptor RNA). This transfer is mediated by base pairing of the complementary R sequences at the 3′-ends of the RNA and DNA molecules. The 3′-R sequence of the genomic RNA folds into secondary structures corresponding to the transactivator response element (TAR) hairpin and the upper part of the poly(A) hairpin (3–6). The R sequence of ss-cDNA 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 (7,8). Mutational analysis of HIV-1 genomic RNA suggests that the great majority of first-strand transfers occur after completion of ss-cDNA synthesis, i.e. with the entire cTAR sequence (9). FRET-based assays performed with the cTAR sequence suggest that the majority (∼70–80%) of the cTAR DNA molecules adopt the hairpin conformation (closed form) and the remaining (20–30%) are partially or totally melted (open forms) (8,10). Furthermore, single-molecule techniques (11,12) were used to show that the NC-coated cTAR hairpin exhibits a dynamic equilibrium between a 'Y'-shaped and a closed conformation. In the 'Y' conformation, the lower stems are open and the upper stems are closed. Recently, using NMR and probing methods, we determined the structural and dynamic properties of the top half of the cTAR DNA (mini-cTAR, Figure 1A) (13). We showed that the upper stem located between the apical and the internal loops is stable, but that the lower stem of mini-cTAR is unstable. Note that the secondary
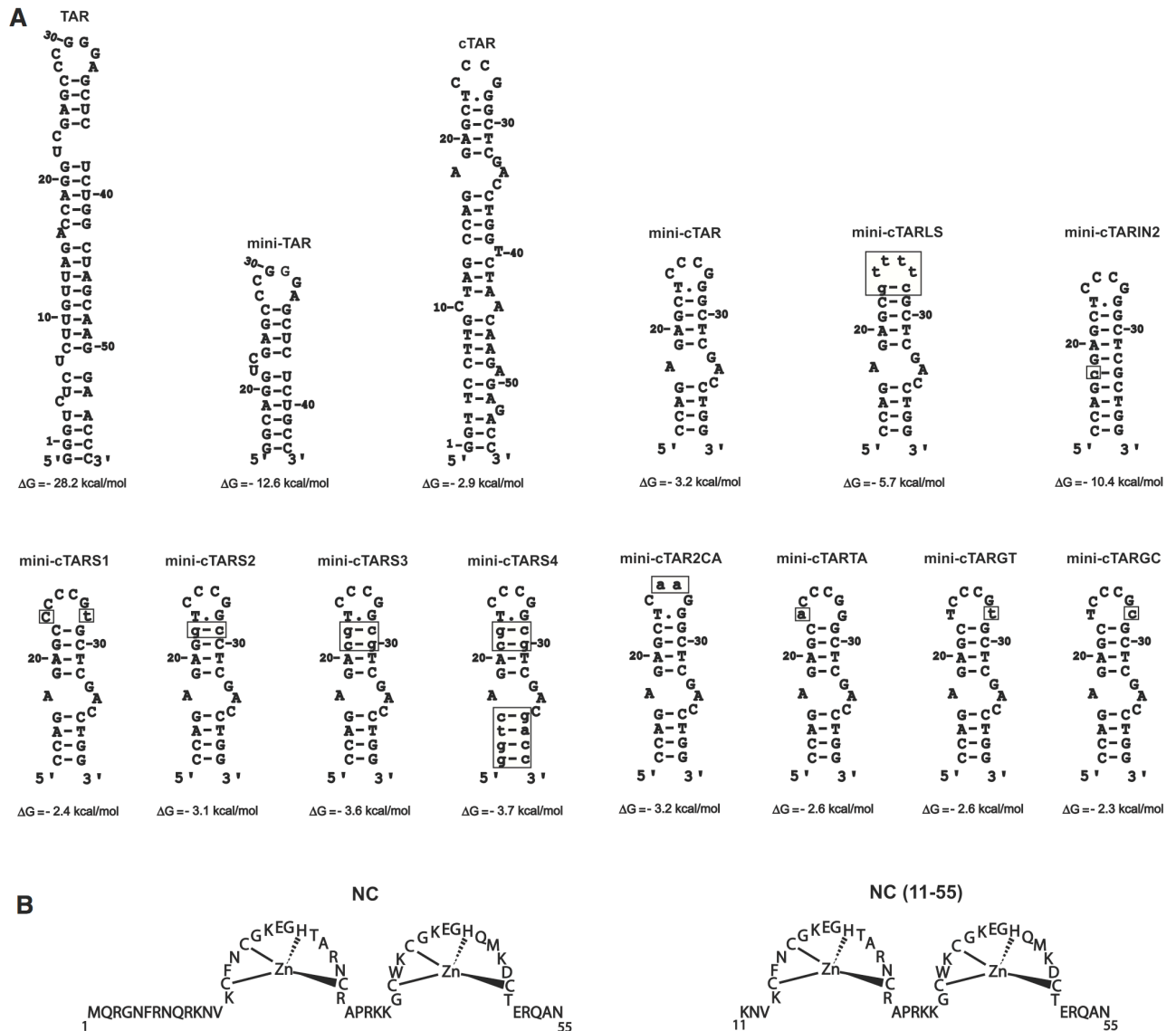
Figure 1. (**A**) Predicted secondary structures for the TAR and cTAR sequences. Zuker's folding programs (48) were used to predict the most stable secondary structures for RNAs and DNAs. The TAR RNA and cTAR DNA sequences are derived from the HIV-1 MAL isolate. Numbering is relative to the TAR and cTAR sequences. Mutations are shown as lower case letters in boxes. (**B**) Sequences of proteins used in this study.

structure of the full-length cTAR sequence has not been determined to date.

Unfolding of the complementary hairpins is thought to be rate-limiting in the annealing process leading to the first-strand transfer (14). *In vitro* studies show that NC stimulates the first-strand transfer (15–18), due, at least in part, to the NC stimulatory effect on the rate of annealing of the complementary hairpins (14,19). NC is a nucleic acid chaperone that facilitates the rearrangement of nucleic acids into the most thermodynamically stable structures containing the maximum number of base pairs (20). This chaperone function of NC relies on two main components: duplex destabilization by the zinc fingers and nucleic acid aggregation, primarily by the

basic N-terminal domain (Figure 1B) (19). NC induces a limited melting of the TAR RNA secondary structure (one base pair per molecule) (8), but largely enhances the fraying of the cTAR hairpin ends, shifting the distribution of hairpin conformations toward the more open structures (8,10,21). The cTAR sequence binds eight NC molecules at saturating protein concentrations under low ionic strength conditions (22), through both weak and strong binding sites (23), not yet identified.

Since mutations in the TAR apical loop decrease the first strand transfer *in vitro*, Berkhout *et al.* (7) suggested that this process involves a 'kissing complex' formed by the apical loops of TAR and cTAR. Studies using TAR RNA and cTAR DNA hairpins derived from the HIV-1

NL4-3 isolate, suggest that both the apical loops and the 3′/5′ termini of complementary hairpins are the initiation sites for the annealing reaction (24,25). Indeed, Vo *et al.* (24) reported that cTAR DNA–TAR RNA annealing in the absence of NC depends on nucleation through the apical loops, whereas cTAR DNA–TAR RNA annealing in the presence of saturated NC depends on nucleation through the 3′/5′ termini, resulting in the formation of a 'zipper' intermediate. Moreover, cTAR DNA and TAR RNA anneal *via* both 'kissing' and 'zipper' pathways under subsaturating concentrations of NC (24). Single-molecule FRET studies support the notion that the annealing process relies on two pathways (25). The annealing process has also been investigated using sequences derived from the HIV-1 MAL isolate (5,26). Thus, Godet *et al.* (26) reported that under low-salt concentrations, the TAR hairpins anneal almost exclusively via the 'zipper' pathway in the presence of NC(12–55), a truncated form of NC which lacks the N-terminal domain and is therefore a poor aggregating agent. Finally, we showed that efficient annealing of cTAR DNA to the 3′-end of the genomic RNA relies on sequence complementarities between TAR and cTAR apical loops under high-salt concentrations allowing cDNA synthesis by RT and strand transfer (5,17,27).

In this study, we investigated the annealing process using DNA and RNA sequences derived from the HIV-1 MAL isolate (Figure 1A) under high-salt concentrations. The 'kissing complex' was trapped using site-directed mutagenesis and investigated by fluorescence anisotropy (FA). Using chemical and enzymatic probes, and gel retardation assays, we determined the cTAR secondary structure and the NC binding sites within cTAR.

MATERIALS AND METHODS

Protein, DNA and RNA preparation

Full-length NC [NC(1–55)] and the truncated version of NC [NC(11–55)] were synthesized by the Fmoc/opfp chemical method and purified to homogeneity by HPLC (28). Proteins were 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 $ZnCl_2$ per mole of peptide. DNA oligonucleotides were purchased from Eurogentec and were 5′-end labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (Perkin Elmer). Full-length cTAR DNA was labeled at its 3′-end as follows. An oligonucleotide (5′-CTGGTCTCTCTTGTT AGACCAG-3′) was annealed to the 3′-end of cTAR. This oligonucleotide carries at its 5′-end a non-hybridizing tail of two nucleotides (5′-CT-3′). To add one labeled adenine residue to the 3′-end of cTAR, the annealed cTAR was incubated at 75°C with [α-^{32}P] dATP and Taq DNA polymerase (New England Biolabs). The 5′- and 3′-end-labeled DNA oligonucleotides were purified by electrophoresis on 12% or 15 % denaturing polyacrylamide gels and isolated by elution followed by ethanol precipitation. DNA template for transcription of TAR RNA was prepared by PCR amplification of pHIVCG8.6 (29) using the expand^{TM} high-fidelity PCR

system (Roche Molecular Biochemicals) and oligonucleotides from Eurogentec. The upstream primer included the T7 promoter and its sequence was: 5′-GAGTAATACGA CTCACTATAGGGTCTCTCTTGTTAG-3′. The sequence for the downstream primer was: 5′-GGGTTCC TTGCTAGCC-3′. The PCR product was purified as described previously (30). DNA template for transcription of mini-TAR RNA was prepared by annealing of two DNA oligonucleotides in the transcription buffer (31). The oligonucleotides contained the T7 promoter and the sequences were: 5′-TAATACGACTCACTATAG-3′ and 5′-GGCAGAGAGCTCCCGGGCTCGACCTGCCTAT AGTGAGTCGTATTA-3′. RNAs were transcribed *in vitro* using the T7-MEGAshortscript^{TM} high yield transcription kit (Ambion). RNAs were purified by electrophoresis on 12% or 15% denaturing polyacrylamide gels as described previously (30). RNAs were treated with alkaline phosphatase from calf intestine (Roche Molecular Biochemicals) and 5′-end-labeled using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P] ATP (Perkin Elmer). The 5′-end-labeled RNAs were purified by electrophoresis on 12% or 15% denaturing polyacrylamide gels and isolated by elution followed by ethanol precipitation. The mini-TAR RNA labeled at its 5′-terminus by the 5/ 6-Rh6G was purchased from IBA GmbH Nucleic Acids Product Supply.

Gel-shift annealing assay

For preparation of loose heteroduplexes, 100 pmol of mini-TAR ^{32}P-RNA (2×10^2 cpm/pmol) or 120 pmol of mini-cTAR ^{32}P-DNA (2×10^2 cpm/pmol) in 4 μl of water was heated at 90°C for 2 min and chilled for 2 min on ice. Then, 1 μl of strand transfer buffer was added (final concentrations: 75 mM KCl, 7 mM $MgCl_2$ and 50 mM Tris–HCl, pH 7.8) and the sample was incubated at 37°C for 30 min. Unlabeled mini-cTAR DNA (120 pmol) or unlabeled mini-TAR RNA (100 pmol) underwent the same renaturation treatment and was then added to the labeled hairpins. The reaction mixture was then incubated at 4°C for 60 min. Then, 3.5 μl of loading buffer was added to the sample. The samples were loaded on 4% agarose (QA-AgaroseTM, Qbiogene) gels. Electrophoresis was carried out at 4°C in TBM buffer [45 mM Tris–borate (pH 8.3), 1 mM $MgCl_2$] or at 25°C in TBE buffer [45 mM Tris–borate (pH 8.3), 1 mM EDTA]. After electrophoresis, the gel was fixed, dried and autoradiographed as described (32).

Gel retardation assays

Assays were carried out in a final volume of 10 μl. Mini-cTAR ^{32}P-DNA (10 pmol) at 2×10^3 cpm/pmol was dissolved in 6 μl of water, heated at 90°C for 2 min and chilled for 2 min on ice. Then, 2 μl of strand transfer buffer was added (final concentrations: 75 mM KCl, 7 mM $MgCl_2$ and 50 mM Tris–HCl, pH 7.8) and the sample was incubated for 15 min at 37°C in the absence or presence of protein at various concentrations. Gel loading buffer (final concentrations: 10 % w/v glycerol, 0.01 % w/v bromophenol blue, 0.01 % w/v xylene cyanol) was added and the samples were analyzed by electrophoresis

on a 14% polyacrylamide gel [29:1 (w/w), acrylamide/bisacrylamide] at 4°C in 0.5× TBE buffer [45 mM Tris–borate (pH 8.3), 1 mM EDTA]. After electrophoresis, the gel was fixed, dried and autoradiographed. Free DNA and protein/DNA complexes were quantified using a PhosphorImager (Molecular Dynamics) and ImageQuant software. The fraction of bound mini-cTAR DNA (FR) was determined using the formula FR = 1−(IF/IB+IF), where IF and IB are the band intensities of free and bound mini-cTAR DNAs, respectively.

Structural probing of cTAR DNA and loose heteroduplexes

Potassium permanganate and piperidine were purchased from Sigma-Aldrich. Mung bean nuclease, DNase I and RNase T1, were purchased from New England Biolabs, Promega and Life Technologies, respectively. Structural probing of cTAR DNA was carried out in a final volume of 10 µl. cTAR ^{32}P-DNA (1 pmol at 5×10^4 cpm/pmol) in 5.5 µl of water was heated at 90°C for 2 min and chilled for 2 min on ice. Then, 2.5 µl of renaturation buffer (final concentrations: 75 mM KCl, 7 mM MgCl$_2$ and 50 mM Tris–HCl, pH 7.8 for probing with KMnO$_4$ or DNase I; 75 mM KCl, 7 mM MgCl$_2$ and 50 mM sodium cacodylate, pH 6.5 for probing with mung bean nuclease) was added and the sample was incubated for 30 min at 37°C. At the end of this incubation, the samples were incubated for 15 min at 37°C in the absence or presence of NC at various concentrations. The samples were then incubated with KMnO$_4$ or an enzyme as follows: 1.5, 3 or 6 U of mung bean nuclease for 15 min at 37°C; or 0.05, 0.1 or 0.2 U 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 10 µl of loading buffer (7 M urea, 0.03 w/v% bromophenol blue and 0.03 w/v% xylene cyanol). For potassium permanganate probing, cTAR DNA was treated with 0.25, 0.5 or 1 mM of KMnO$_4$ for 1 min at 37°C. The treatment was stopped by adding 40 µl of the termination buffer [0.7 M β-mercaptoethanol, 0.4 M NaOAc (pH 7.0), 10 mM EDTA, 25 µg/ml 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 µl of freshly diluted 1 M piperidine and heating at 90°C for 30 min. The samples were then lyophilized, resuspended in 20 µl of water and lyophilized again. After a second lyophilization from 15 µl of water, the samples were resuspended in 10 µl of loading buffer. G and T+C sequence markers of the labeled cTAR were produced by the Maxam–Gilbert method (33). All samples were analyzed by short and long migration times on denaturing 14% polyacrylamide gels.

Structural probing of loose heteroduplexes was carried out in a final volume of 10 µl. For RNase T1 probing, 100 pmol of mini-TAR ^{32}P-RNA (3 × 10^2 cpm/pmol) in 4 µl of water was heated at 90°C for 2 min and chilled for 2 min on ice. Then, 1 µl of strand transfer buffer was added (final concentrations: 75 mM KCl, 7 mM MgCl$_2$ and 50 mM Tris–HCl, pH 7.8) and the sample was

incubated at 37°C for 30 min. Unlabeled mini-cTARS3 DNA (50, 100 or 200 pmol) underwent the same renaturation treatment and was then added to refolded mini-TAR ^{32}P-RNA. The reaction mixture was incubated at 4°C for 60 min. Then, 0.2 U of RNase T1 was added and the sample was incubated at 4°C for 10 min. This cleavage reaction was stopped by phenol–chloroform extraction followed by ethanol precipitation. The dried pellet was resuspended in 10 µl of loading buffer. For mung bean nuclease probing, 100 pmol of mini-cTARS3 ^{32}P-DNA (3 × 10^2 cpm/pmol) in 4 µl of water was heated at 90°C for 2 min and chilled for 2 min on ice. Then, 1 µl of renaturation buffer was added (final concentrations: 75 mM KCl, 7 mM MgCl$_2$ and 50 mM sodium cacodylate pH 6.5) and the sample was incubated at 37°C for 30 min. Unlabeled mini-TAR RNA (50, 100 or 200 pmol) underwent the same renaturation treatment and was then added to refolded mini-cTARS3 ^{32}P-DNA. The reaction mixture was incubated at 4°C for 60 min. Then, 30 U of mung bean nuclease were added and the sample was incubated at 4°C for 20 min. This cleavage reaction was stopped by phenol-chloroform extraction followed by ethanol precipitation. The dried pellet was resuspended in 10 µl of loading buffer. All samples were analyzed by short and long migration times on denaturing 17% polyacrylamide gels.

Fluorescence anisotropy

Binding of the mini-cTAR DNA sequences to 5′ 5/6-Rh6G mini-TAR RNA was investigated using FA. Measurements were performed in 50 µl quartz cuvettes using a Fluorolog spectrophotometer (Jobin Yvon) equipped with a Peltier thermostated cell holding the sample at 4°C. FA values were obtained by recording the I_{VV}, I_{VH}, I_{HV} and I_{HH} intensities (V and H characterizing the vertical and horizontal light polarisation for the excitation and the emission paths, respectively) of each solution. The steady-state anisotropy r was calculated according to: $r = (I_{VV}−G.I_{VH})/(I_{VV}+2\,G.I_{VH})$ in which $G = I_{HV}/I_{HH}$. Excitation and emission wavelengths were set at 520 nm and 560 nm, respectively. FA data are the result of two independent experiments performed in duplicate. Apparent equilibrium dissociation constants were determined from the FA values plotted as a function of DNA concentrations. The binding curves were fitted according to:

$$r = \{r_0+\nu(r_f.R − r_0)\}/\{\nu(R − 1)+1\} \qquad (1)$$

where r_0 and r_f are the anisotropy of mini-TAR RNA in the free and in fully bound form, respectively, and R corresponds to the fluorescence quantum yield ratio of bound mini-TAR relative to free mini-TAR. Assuming a 1:1 stoichiometry, the molar fraction ν of bound mini-cTAR DNA can be expressed as:

$$\nu = \{N+D+K_d − [(N+D+K_d)^2 − 4.D.N]^{0.5}\}/\{2.N\}$$

in which N and D stand for the concentrations of mini-TAR RNA and mini-cTAR DNA, respectively.

RESULTS

Analysis of the loop–loop interaction in the absence of NC

The mini-TAR and mini-cTAR hairpins (Figure 1) are appropriate substrates to investigate the kissing pathway in the absence of NC (24,34). To better understand the nucleation through the apical loops, we investigated the mini-TAR RNA–DNA annealing process in the strand transfer buffer, i.e. under high-salt conditions previously shown to promote annealing of cTAR DNA to the 3′-end of the genomic RNA (5). The apical loop of mini-cTAR is probably constituted of six but not four residues, since the T.G base-pair predicted by mfold (Figure 1) was not detected by NMR methods (13). An intermolecular loop–loop interaction can be characterized by the formation of loose duplexes that are stable upon gel migration at 4°C in Tris–borate magnesium (TBM) but not in TBE at 25°C (35–37). In contrast, the tight duplexes can be seen under both electrophoretic conditions (36–38). Despite numerous attempts, we could not detect loose duplexes using the wild-type and mutant sequences at low concentrations (data not shown). Therefore, annealing assays were performed at 4°C and at high concentrations of complementary hairpin molecules, i.e. under conditions

allowing observation of RNA–DNA kissing complexes (39,40). To identify the annealing products, either mini-TAR ^{32}P RNA (Figure 2A, lanes 1–6) or mini-cTAR ^{32}P DNAs (Figure 2A, lanes 7–12) were incubated with or without the unlabeled complementary sequences. After renaturation and incubation at 4°C in the strand transfer buffer, mini-TAR RNA and mini-cTAR DNAs were mostly monomeric (Figure 2A, lanes 2 and 10–12). Addition of wild-type mini-cTAR to the wild-type mini-TAR strongly increased the band intensity that migrated at a rate expected for the heteroduplex formed by mini-TAR and mini-cTAR (Figure 2A, lanes 3). Since the band intensity did not decrease in the TBE gel, the heteroduplex was stable. These results suggest that the loop–loop interaction was converted into an extended duplex under our experimental conditions. Consistent with an annealing mechanism involving nucleation through a loop–loop interaction, mini-cTAR2CA and mini-cTARS1 (Figure 1A) did not form tight heteroduplexes with mini-TAR (Figure 2B). In addition, the results obtained with mini-cTARS1 show that a loose heteroduplex cannot be formed through a loop-loop interaction involving four base-pairs. It has been reported that the apical loop of mini-TAR RNA and the terminal loop of a DNA
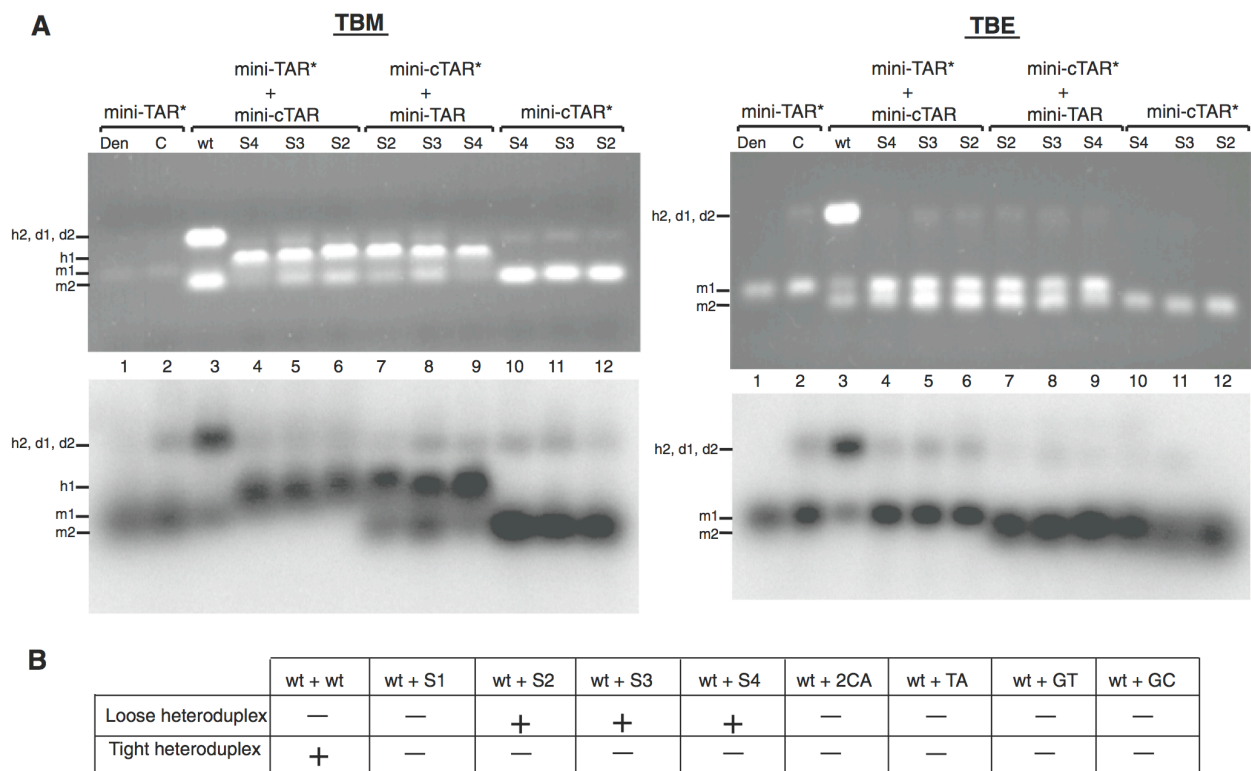
Figure 2. Mutational analysis of mini-TAR RNA–DNA annealing in the absence of NC. Annealing was performed as described in 'Materials and Methods' section. (**A**) Each sample was divided in two aliquots which were analyzed by 4% agarose gel electrophoresis at 4°C in TBM buffer and at 25°C in TBE buffer. Nucleic acids were visualized by ethidium bromide staining (top) and autoradiography (bottom). Lane 1, heat-denatured mini-TAR ^{32}P-RNA. Mini-TAR ^{32}P-RNA was incubated in the absence (lane 2) or presence of unlabeled mini-cTAR DNAs (lanes 3–6). Mini-cTAR ^{32}P-DNAs were incubated in the absence (lanes 10–12) or in the presence of unlabeled TAR RNA (lanes 7–9). Monomeric and dimeric forms of mini-TAR RNA are indicated by m1 and d1, respectively. Monomeric and dimeric forms of mini-cTAR DNAs are indicated by m2 and d2, respectively. Loose and tight heteroduplexes are indicated by h1 and h2, respectively. (**B**) The table summarizes the results of the mutational analysis.

aptamer form a DNA–RNA loop–loop complex involving 5 bp (39,40). Interestingly, this complex and the RNA–RNA loop–loop complexes formed by the dimerization sequences of HIV-1 and avian leukosis virus are flanked by at least one unpaired nucleotide (30,41). The mini-cTARTA, mini-cTARGT and mini-cTARGC mutants were designed so that they have the potential to form a loop–loop complex involving 5 bp and one unpaired nucleotide (Figure 1). These mutants did not form loose heteroduplexes (Figure 2B), suggesting that the loop–loop complex was not formed or was very unstable. To trap the loop–loop interaction involving 6 bp, the mini-cTARS2 and mini-cTARS3 mutants were designed (Figure 1). Both mutants formed loose heteroduplexes that were observed on the TBM gel but not on the TBE gel (Figure 2A, lanes 5–8). Since the mini-cTARS4 mutant formed loose heteroduplexes (Figure 2A, lanes 4 and 9), we can exclude the possibility that the mini-cTAR mutants formed loose heteroduplexes through the 5′-/3′-terminal ends.

To confirm that formation of loose heteroduplexes relies on an intermolecular loop–loop interaction, probing experiments were performed with mini-TAR RNA and mini-cTARS3 (Figure 3). First, we probed the structure of mini-TAR RNA with RNase T1 (specific for unpaired guanines) in the absence and presence of mini-cTARS3 (Figure 3A). In the absence of mini-cTARS3, there were strong and moderate T1 cleavages in the apical loop (Figure 3A, lane 1). As expected, the guanine residues in the predicted stems were not cleaved by RNase T1. The guanine residues in the apical loop became markedly unreactive to RNase T1 in the presence of increasing concentrations of mini-cTARS3 (Figure 3A, lanes 2–4). Second, we probed the structure of mini-cTARS3 DNA with mung bean nuclease (MB) in the absence and presence of mini-TAR RNA (Figure 3B). MB is highly selective for single-stranded nucleic acids and single-stranded regions in double-stranded nucleic acids (42). Note that single-base mismatches in double-stranded DNA are poor substrates for MB cleavage at 37°C (42,43). In the absence of mini-TAR RNA, there were moderate and strong MB cleavages in the apical loop and within the sequence between positions 34 and 38 (Figure 3B, lane 1). These results are consistent with our study showing that the lower stem of mini-cTAR does not form a stable double-stranded structure (13). The residues in the apical loop became markedly unreactive to MB in the presence of increasing concentrations of mini-TAR RNA (Figure 3B, lanes 2–4). In contrast, the cleavage rate for MB in the lower part of mini-cTARS3 was not significantly reduced in the presence of increasing concentrations of mini-TAR RNA. Taken together, our results show that the linkage structure of the loose heteroduplex is the loop–loop complex. To determine the apparent dissociation constants (Kd) of loop–loop complexes, 5′ 5/6-Rh6G mini-TAR RNA was mixed with increasing amounts of mini-cTARS2 or mini-cTARS3. Heteroduplexes were allowed to form for one hour at 4°C in the strand transfer buffer, before the fluorescence anisotropy of the mixture was measured. Adding the mini-cTAR mutants to 5′ 5/6-Rh6G mini-TAR RNA

induced a slight but significant increase of the anisotropy. Since no change in the steady-state fluorescence spectra were observed when DNA sequences were added, the anisotropy increase indicated that mini-cTAR DNAs interacted with mini-TAR RNA. Using Equation (1) to fit the binding curves shown in Figure 4, K_d values of 3.5 (± 2) $\times 10^6 M^{-1}$ were obtained with both mini-cTARS2/mini-TAR and mini-cTARS3/mini-TAR.

Analysis of cTAR secondary structure in the absence of NC

The secondary structure of cTAR DNA was investigated at 37°C in the strand transfer buffer, i.e. under conditions previously shown to promote strand transfer and annealing of cTAR DNA to the 3′-end of the genomic RNA (5,17,27). Using gel-shift annealing assays, we checked that the cTAR DNA hairpin anneals to the TAR RNA hairpin under our high-salt conditions (Supplementary Figure S1). Consistent with previous reports (24,26), the TAR RNA–DNA annealing reaction is extremely slow in the absence of NC (Supplementary Figure S1). In addition, the full-length cTAR was monomeric after incubation as assessed by native agarose gel electrophoresis in the TBM and TBE buffers (data not shown). Therefore, the structural analysis was not complicated by the presence of homoduplexes. The secondary structure of cTAR DNA was probed with potassium permanganate ($KMnO_4$), MB and DNase I. DNase I is a double-strand-specific endonuclease that produces single-strand nicks (44). $KMnO_4$ can be used to detect regions of DNA that are unpaired or distorted (45,46): 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 cTAR DNA 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 14% polyacrylamide gels. Running Maxam–Gilbert sequence markers of cTAR on the same gels in parallel identified the cleavage sites. Representative examples of probing experiments are shown in Figures 5 and 6 and the results of a series of independent experiments are summarized in Figure 7. 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 (33,47). The electrophoretic mobility of Maxam–Gilbert sequence markers is therefore slightly greater than that of fragments produced by nucleases. To determine the precise position of cleavage at some sites, the 3′-end-labeled cTAR was used (examples in Figure 5).

The partially melted 'Y' conformation is the most stable cTAR structure predicted by the Mfold program (48). It has been proposed that the cTAR sequence forms the closed conformation (12,24,26). Therefore, the probing data were superimposed on these two conformations (Figure 7). In both conformations, the 6–48 sequence
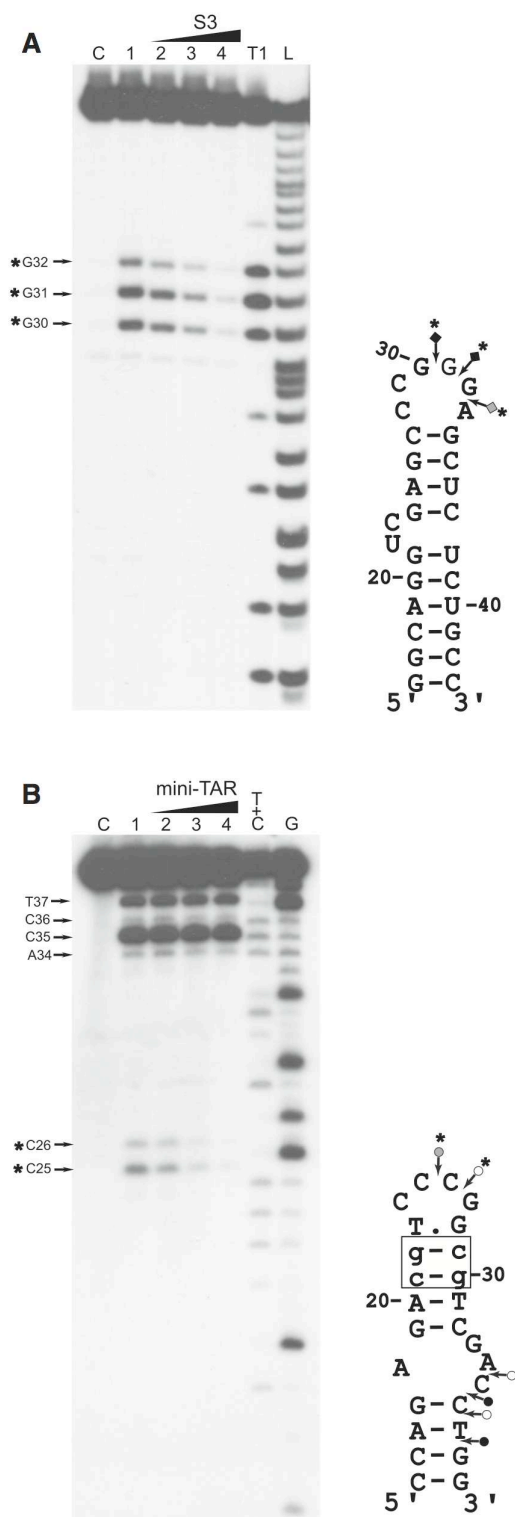
Figure 3. Probing of mini-TAR:mini-cTARS3 duplex. (**A**) Enzymatic probing with RNase T1. Mini-TAR ^{32}P-RNA alone (lanes C and 1) or mixed with mini-cTARS3 (lanes 2–4) was incubated in the absence (lane C) or in the presence of RNase T1 (lanes 1–4). Lanes 2–4, the RNA:DNA ratios were 1:0.5, 1:1 and 1:2, respectively. Lanes T1 and L refer to RNAse T1 sequencing and to alkaline ladder, respectively. (**B**) Enzymatic probing with mung bean nuclease. Mini-cTARS3 ^{32}P-DNA

forms the same stem–loop structure. There were strong and moderate DNase I cleavages between G_9–C_{10} and C_{10}–T_{11}, all of which are predicted to lie at the ends of stems. There were moderate and strong MB cleavages between C_{25}–C_{26}, C_{26}–G_{27}, A_{34}–C_{35} and C_{35}–C_{36} that are predicted to be in the apical and internal loops. The moderate and high sensitivities of T_{23} and T_{40} to $KMnO_4$ (Figure 6A), respectively, indicate that at least one side of the plane of the heterocyclic ring of these thymine residues is exposed. These data suggest that the predicted T.G mismatched base pair is not involved in strong stacking interactions with the residues of the apical loop and T_{40} adopts the looped-out bulge conformation. Consistent with the formation of stems, T_{31} and T_{42} were unreactive to $KMnO_4$ and the sensitivity of T_8, T_{11} and T_{37} to $KMnO_4$ was very low. It is likely that the C–A mismatch slightly destabilizes the adjacent base-pairs and the internal loop destabilizes the G_{17}–C_{36} base pair (13), since T_8, T_{11} and T_{37} were not totally protected by the surrounding base pairs. Fraying of the C_6–G_{48} base pair is probably responsible for the low sensitivity of T_7. The high and moderate sensitivities of T_3 and T_4 to $KMnO_4$, respectively and the moderate MB cleavage between C_5 and C_6 are consistent with the partially melted conformation of cTAR. However, the 3′-/5′-terminal ends may be involved in transient interactions to form the closed conformation since there was moderate DNase I cleavage between T_4 and C_5. Taken together, our probing data support a structural heterogeneity for the 3′-/5′-terminal ends of cTAR.

Analysis of cTAR secondary structure in the presence of NC

We checked that the cTAR DNA hairpin anneals to the TAR RNA hairpin in the presence of NC under structural analysis conditions (Supplementary Figure S1). Consistent with previous reports (24,26), the TAR RNA–DNA annealing reaction is dramatically accelerated by NC (Supplementary Figure S1). The full-length cTAR was monomeric after incubation with NC as assessed by native polyacrylamide gel electrophoresis in the TBE buffer (data not shown). Therefore, the structural analysis was not complicated by the presence of homoduplexes. To identify destabilized regions and protections induced by NC in the cTAR hairpin, we compared the enzymatic and $KMnO_4$ probing patterns of cTAR in the absence or presence of increasing concentrations of NC (Figures 5 and 6). Consistent with the formation of stable stems, T_{11}, T_{31} and T_{42} did not become more reactive to $KMnO_4$ in the presence of NC (Figure 6B).

Figure 3. Continued
alone (lanes C and 1) or mixed with mini-TAR (lanes 2–4) was incubated in the absence (lane C) or in the presence of mung bean nuclease (lanes 1–4). Lanes 2–4, the DNA:RNA ratios were 1:0.5, 1:1 and 1:2, respectively. Lanes G and T+C refer to Maxam–Gilbert sequence markers. Samples were analyzed by electrophoresis on 17% denaturing polyacrylamide gels. Arrows indicate the cleavage sites. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively. The strong protections induced by mini-cTARS3 (A) or mini-TAR (B) are indicated by asterisks.
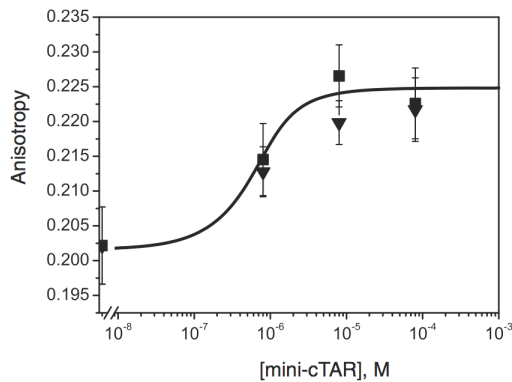
Figure 4. Characterization of loose heteroduplexes by fluorescence anisotropy. $5'$ 5/6-Rh6G mini-TAR RNA was mixed with increased amounts of mini-cTARS2 (triangles) or mini-cTARS3 (squares) in the absence of NC and incubated at 4°C for 60 min as described in 'Materials and Methods' section. The fluorescence anisotropy of the mixture was measured at the end of incubation.

The sensitivity of T_3, T_4, T_7 and T_8 to $KMnO_4$ was slightly increased with NC at protein to nucleotide molar ratios of 1:8, 1:4 and 1:2 (Figure 6B, lanes 2–4), suggesting that the population of cTAR in the 'Y' conformation increases in the presence of NC. It is likely that the stem containing T_{37} is slightly destabilized by NC, since the sensitivity of this thymine residue was slightly increased with NC at protein to nucleotide molar ratios of 1:8, 1:4 and 1:2 (Figure 6B, lanes 2–4). In the presence of NC at protein to nucleotide molar ratios higher than 1:2, the efficiency of DNA recovery after phenol–chloroform extraction was reduced for $KMnO_4$ assays. Therefore, the sensitivity of thymine residues to $KMnO_4$ seems to decrease with NC at a protein to nucleotide molar ratio of 1:1 (Figure 6B, lane 5). The double-stranded region containing the T bulge is probably a binding site for NC since the protein induced a significant decrease in $KMnO_4$ sensitivity for T_{40}. In the presence of increasing concentrations of NC, DNase I cleavage at the level of all sites decreased at the same rate, i.e. NC did not induce specific protection against DNase I (data not shown). These results indicate that the double-stranded regions of cTAR do not contain preferential binding sites for NC. Interestingly, NC induced a strong decrease in MB sensitivity for A_{34} and C_{35} (Figure 5D) that are located in the internal loop (Figure 7). PhosphorImager quantification indicated that the cleavage rate for MB in the internal loop at a protein to nucleotide molar ratio of 1:4 corresponded to 52% of that observed in the absence of NC. In contrast, at the same ratio, the cleavage rate for MB in the apical loop (residues C_{25} and C_{26}) corresponded to 96% of that observed in the absence of NC. These data indicate that the internal loop is a strong binding site for NC, and that, if NC binds the apical loop, its binding is too weak to be detected at a protein to nucleotide molar ratio of 1:4. However, phosphorImager quantification indicated that the cleavage rate for MB in the apical loop at a protein to nucleotide molar ratio of 1:2 corresponded to 55% of that observed in the absence of NC, i.e. NC binds the

apical loop at protein to nucleotide molar ratios equal to or greater than 1:2 (Figure 5D, lanes 4 and 5). Taken together the data indicate that the apical and internal loops are weak and strong binding sites for NC, respectively. Finally, our results show that NC preferentially binds the internal loop, slightly destabilizes the lower stem that is adjacent to this loop, and shifts the equilibrium toward the 'Y' conformation.

Binding of NC to mini-cTAR

To confirm that the apical loop is not a strong binding site for NC, we carried out gel retardation assays with wild-type and mutant mini-cTAR DNAs (Figure 1A). Since NC binds nucleic acids with a preference for sequences containing unpaired guanine residues (49–55), we altered the apical and internal loops containing guanine residues. The LS mutant was designed so that four T bases replaced the CCCG sequence in the apical loop. The internal loop sequence was deleted in the IN2 mutant. Protein:DNA complexes were formed under the same salt conditions and at protein to nucleotide molar ratios as those used in probing experiments (see 'Materials and Methods' section). Heat-denatured mini-cTAR DNAs (Figure 8A and B, lanes 2) were used to locate the position of monomeric mini-cTAR DNAs. To identify the positions of bands corresponding to dimeric mini-cTAR DNAs, the proteins were removed before gel analysis (Figure 8A and B, lanes 1). After renaturation and incubation at a protein to nucleotide molar ratio of 1:1, the majority of mini-cTAR DNAs remained monomeric (Figure 8A and B, lanes 1). Addition of increasing amounts of NC resulted in the appearance of two bands (Figure 8A): band CI migrating at a rate expected for the NC:mini-cTAR complex and band CII corresponding to high-molecular-weight protein:mini-cTAR complexes (aggregates) that stuck in the wells. The LS mutation did not lead to reduction in NC binding (Figure 8C). Interestingly, deletion of the internal loop significantly reduced the binding of NC to the IN2 mutant (Figure 8C).

The nucleic acid aggregating activity of NC resides primarily in the basic N-terminal domain (29,56–58), and the duplex destabilizing activity of NC has been mapped to its zinc finger structures (22,59–62). NC exhibits sequence-specific binding to single-stranded regions through interactions that involve the zinc fingers (51,53,63). To characterize specific interactions between the zinc-finger structures and the mini-cTAR hairpin, we used NC(11–55) which lacks the basic N-terminal domain (Figure 1B). As expected, NC(11–55) did not form aggregates with mini-cTAR DNAs (Figure 8B). Addition of increasing amounts of NC(11–55) resulted in the appearance of band CI migrating at a rate expected for the NC(11–55):mini-cTAR complex. Binding of NC(11–55) to the LS mutant was not impaired by mutations in the apical loop (Figure 8B and D). Interestingly, NC(11–55) did not bind to the IN2 mutant. Taken together, these results show that the internal loop sequence is important for tight binding of NC(11–55) to cTAR and that the NC zinc fingers do not bind the apical loop under conditions allowing strand transfer.

107

Figure 5. Enzymatic probing of cTAR. Enzymatic probing experiments were performed as described in 'Materials and Methods' section. (**A** and **B**) In the absence of NC, the 3′-end-labeled cTAR DNA was incubated with DNase I (0.05, 0.1 and 0.2 U) or mung bean nuclease (MB) (1.5, 3 and 6 U). (**C**) In the absence of NC, the 5′-end-labeled cTAR DNA was incubated with MB (1.5, 3 and 6 U). Lanes C are controls without enzyme. G and T+C refer to Maxam–Gilbert sequence markers. Arrows indicate the cleavage sites. (**D**) 5′-end-labeled cTAR DNA was incubated with MB (3 U) in the absence (lane 1) or in the presence of NC (lanes 2–5). The protein to nucleotide molar ratios were 1:8 (lane 2), 1:4 (lane 3), 1:2 (lane 4) and 1:1 (lane 5). The strong protections induced by NC at the level of MB cleavage sites are indicated by asterisks. (**E**) Secondary structures for the partially melted and closed forms of cTAR. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease).
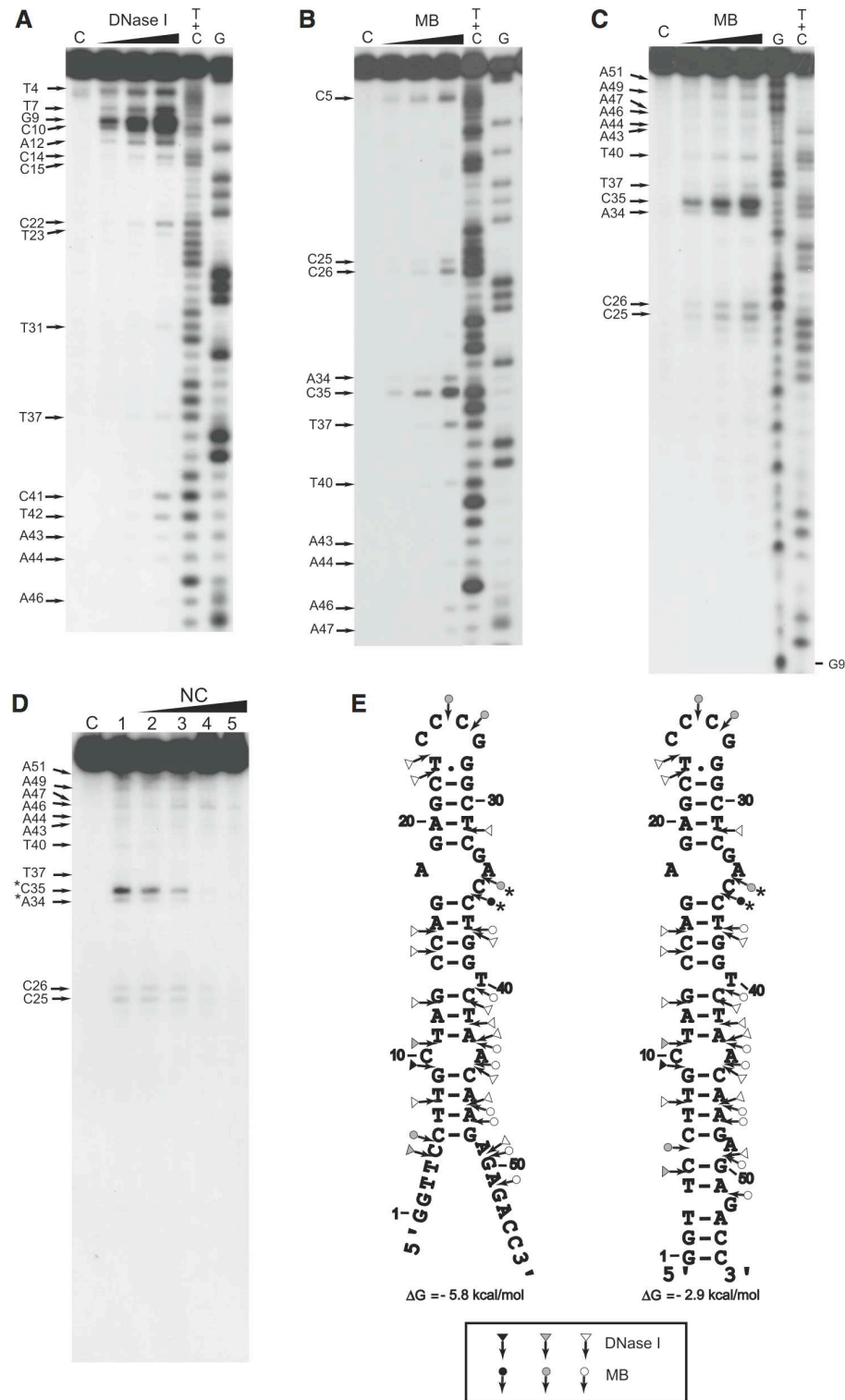
Figure 6. KMnO4 probing of cTAR. KMnO4 probing experiments were performed as described in 'Materials and Methods' section. (**A**) In the absence of NC, the 5′-end-labeled cTAR DNA was incubated with $KMnO_4$ (0.25, 0.5 and 1 mM). (**B**) In the absence (lane 1) or in the presence of NC (lanes 2–5), the 5′-end-labeled cTAR DNA was incubated with $KMnO_4$ (0.25 mM). Lanes C are controls without NC and any chemical treatment. Lane C_{pip} is the control without NC and $KMnO_4$ treatment but with piperidine treatment. The protein to nucleotide molar ratios were 1:8 (lane 2), 1:4 (lane 3), 1:2 (lane 4) and 1:1 (lane 5). Arrows indicate the reactive thymine residues. The strong protection induced by NC at the level of T_{40} is indicated by an asterisk. The thymine residues where the reactivity to $KMnO_4$ is increased by NC are indicated by the stars. G and T+C refer to Maxam-Gilbert sequence markers. (**C**) Secondary structures for the partially melted and closed forms of cTAR. Delta G values were predicted by mfold. The color codes used for the reactivity of thymine residues are indicated in the inset.

DISCUSSION

The full-length cTAR adopts two alternative conformations in the absence of NC

FRET-based assays have been developed to study the folding of the cTAR DNA molecule derived from the MAL (8) or NL4-3 isolate (10). These studies suggest that the majority (~70–80%) of the cTAR DNA molecules adopt the hairpin conformation (closed form) and the remaining (20–30%) are partially or totally melted (open forms) in the absence of NC. Using single-molecule spectroscopy (SMS) techniques and cTAR DNA molecules derived from the NL4-3 isolate, it has been reported that the partially open 'Y' conformation is the dominant form for the cTAR hairpin in the NC-mediated annealing process (12,64). These studies suggest that the 5′ and 3′ unpaired termini of the 'Y' conformation are accessible for TAR RNA–DNA annealing. Note that the double-stranded and single-stranded regions of the 'Y'

conformation were deduced from the analysis of hairpin mutants. Here, we used structural probes (nucleases and potassium permanganate) to directly determine the secondary structure of the full-length cTAR derived from the MAL isolate. The probing experiments were performed under high-salt conditions previously shown to promote strand transfer and annealing (5,17,27). It is unlikely that a population of cTAR DNA molecules is totally melted in the absence of NC under our conditions, since T_{31} and T_{42} were unreactive to potassium permanganate and the sensitivity of T_8, T_{11} and T_{37} to this reagent was very low (Figures 6 and 7). The 6–48 sequence forms a stem–loop structure containing a C–A mismatch, a T bulge in the looped-out conformation and an internal loop (Figure 7). Consistent with the stem involving nucleotides 6–9/45–48, the sensitivity of T_8 to potassium permanganate was very low and there was a strong DNase I cleavage between G_9 and C_{10}. Nucleotides 1–5 and 50–55 are paired in the closed conformation and are unpaired in

Figure 7. Secondary structures for the partially melted and closed forms of cTAR. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (tri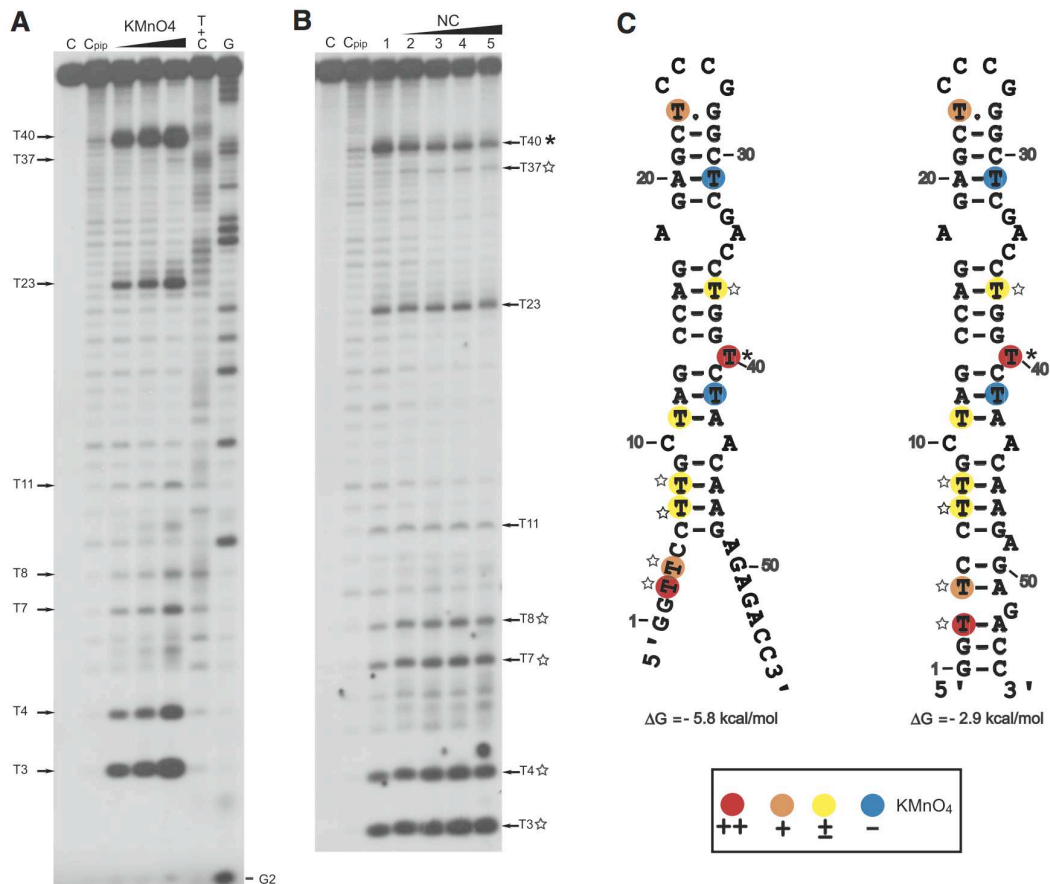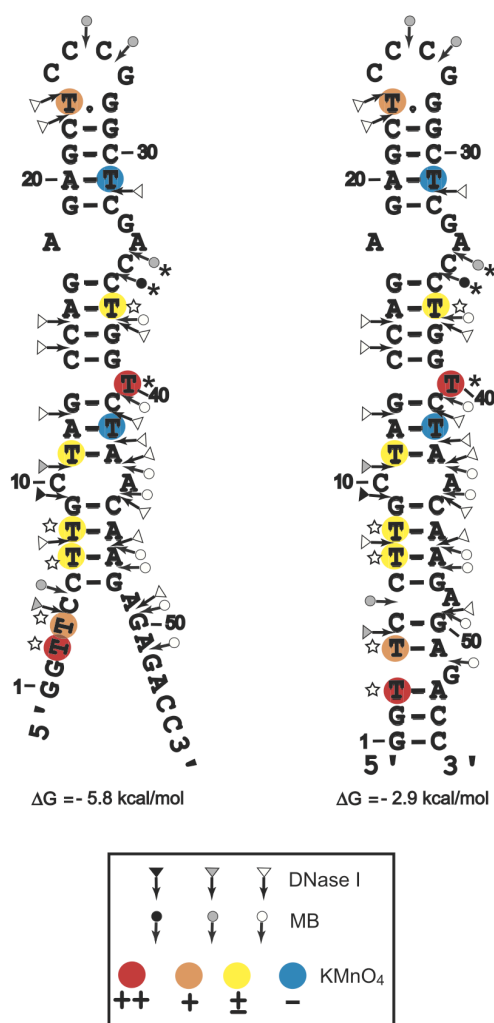angle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The strong protections induced by NC at the level of MB cleavage sites and thymine residues are indicated by asterisks. The thymine residues where the reactivity to KMnO$_4$ is increased by NC are indicated by the stars.

the 'Y' conformation. In summary, our probing data are consistent with the equilibrium between the closed and 'Y' conformations of the cTAR hairpin in the absence of NC.

NC' effects on the structure of the cTAR DNA hairpin

FRET-based assays suggest that the populations of the partially or totally melted forms of the cTAR hairpin are moderately increased in the presence of NC or NC(12–55) (8,10). The regions opened by NC are expected not to become accessible to MB because they probably bind the nucleocapsid protein. Consistent with this notion, MB cleavage did not increase in the presence of NC (Figure 5D). Studies performed with DNA

oligonucleotides showed that NC binds to the single-stranded regions with the following preference: TG > TNG, GNG > G, where N corresponds to either A or C (50,52,65). Recently, NMR studies performed with DNA oligonucleotides containing TG and TNG sequences showed that the nucleobase of the thymine residue is partly inserted in the N-terminal zinc finger of NC and is involved in stacking interactions with the Phe16 side chain (63,66). Residues T_3, T_4, T_{23}, T_{40} and T_{42} are predicted not to be involved in stacking interactions with the Phe16 side chain of NC, since they are not present in TG and TNG sequences. In addition, a close examination of the structures of the NC:DNA complexes (63,66) shows that the 5,6 double bond of the nucleobase of the thymine residue is accessible to potassium permanganate (KMnO$_4$) that is a small probe. Therefore, KMnO$_4$ is expected to interact with the nucleobases of unpaired thymine residues that interact with the N-terminal zinc finger of NC. Since T_{31} and T_{42} did not become sensitive to KMnO$_4$ in the presence of NC (Figure 6B), it is likely that the $_{12}A–T_{42}$ and $_{20}A–T_{31}$ base pairs are stable. Therefore, there is no evidence that a population of cTAR DNA molecules is totally melted in the presence of NC under our experimental conditions. Our probing data are consistent with the SMS study that does not report the observation of totally melted forms (64). NC induced a slight increase in KMnO$_4$ sensitivity for residues T_3, T_4, T_7 and T_8 (Figure 6B), suggesting that the population of cTAR in the 'Y' conformation (Figure 7) increases in the presence of NC. Our results are in agreement with the hypothesis that NC shifts the equilibrium toward the O1 and O2 partially open forms (8). In the lower part of the cTAR hairpin, the O1 and O2 forms exhibit 12 and 22 unpaired nucleotides, respectively. The sensitivity of T_{37} to KMnO$_4$ was slightly increased with NC (Figure 6B), suggesting that the stem involving nucleotides $_{14}CCAG_{17}/_{36}CTGG_{39}$ is slightly destabilized by NC. For the full-length cTAR hairpin in the absence of NC, this stem is stable since the sensitivity of T_{37} to KMnO$_4$ was barely detectable (Figure 6A). However, NMR investigation of mini-cTAR in the absence of NC suggests that this stem may be destabilized by alternative possibilities of base pairing involving the residues of the internal loop and the $_{17}G–C_{36}$ base pair (13). An attractive hypothesis is that binding of NC to the internal loop facilitates these conformational exchange phenomena. Finally, our data suggest that NC destabilizes the lower stem that is adjacent to the internal loop and shifts the equilibrium toward the 'Y' conformation.

NC binding sites in the cTAR hairpin

A recent study suggests that the cTAR sequence contains weak and strong binding sites for NC (23). Identification of preferred binding sites in the cTAR molecule is necessary to propose models for the annealing mechanism mediated by NC. Our study is the first to identify the NC binding sites using full-length NC and full-length cTAR DNA under salt conditions previously shown to promote strand transfer and annealing (5,17,27). The DNase I probing data indicate that the double-stranded
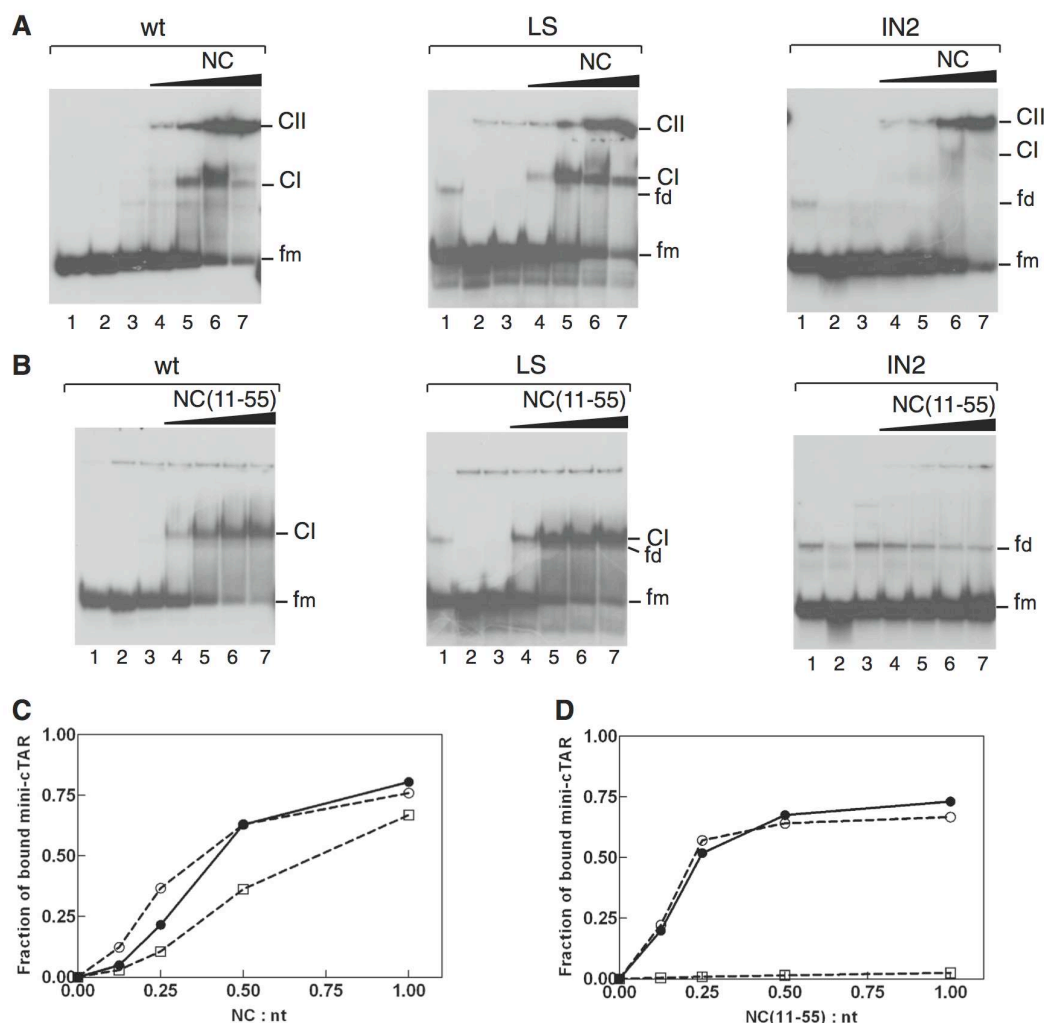
Figure 8. Gel retardation assays of protein:mini-cTAR DNA complexes formed *in vitro*. Mini-cTAR ^{32}P-DNAs were incubated in presence of NC (**A**) or NC(11-55) (**B**) and analyzed by electrophoresis on a 14% polyacrylamide gel as described in 'Materials and Methods' section. Lanes 1, controls mini-cTAR dimerization induced by NC or NC(11-55) at a protein to nucleotide molar ratio of 1:1 [NC and NC(11-55) were removed by phenol/chloroform before gel electrophoresis]. Lanes 2, heat-denatured mini-cTAR DNAs. Lanes 3, controls without protein; lanes 4–7, protein to nucleotide molar ratios were 1:8, 1:4, 1:2 and 1:1. Monomeric and dimeric forms of free mini-cTAR DNAs are indicated by fm and fd, respectively. CI indicates the protein:mini-cTAR complexes. CII indicates the high molecular mass protein:mini-cTAR complexes (aggregates). (**C**) The graph was derived from the experiments shown in (A). (**D**) The graph was derived from the experiments shown in (B). Symbols: filled circles, wt; open circles, LS; open squares, IN2.

regions of cTAR bind NC but do not contain preferential binding sites for this protein. In a previous report, we showed that the N-terminal basic domain of NC is required for providing protection against RNase V1 that cleaves double-stranded RNA (5). These results are consistent with the notion that NC binds the double-stranded regions of nucleic acids non-specifically through electrostatic interactions of the basic residues with the phosphodiester backbone (29,67).

Binding of NC to the double-stranded region containing the T bulge is probably responsible for the significant decrease in KMnO$_4$ sensitivity for T$_{40}$ (Figure 6B). A likely explanation is that this thymine residue is stacked between the flanking bases in the presence of NC while it

was in looped out conformation in the absence of NC. Interestingly, Kalnik *et al.* (68) reported an equilibrium between the stacked and looped-out conformations of a T bulge. Since NC can bend short segments of double-stranded nucleic acids (67), an attractive hypothesis is that the NC-induced bending of cTAR shifts the equilibrium towards the stacked conformation of the T bulge. Using mung bean nuclease (MB), we showed that the apical and internal loops are weak and strong binding sites for NC, respectively. Moreover, since mutations in the apical loop had no influence on the binding of the mini-cTAR to NC and NC(11–55) (Figure 8) and since NC(11–55) did not bind the mini-cTARIN2 mutant displaying the apical loop but not the internal loop, it is likely

that binding of NC to the apical loop relies mainly on non-specific interactions involving the N-terminal basic domain. Moreover, NMR experiments performed with mini-cTAR in low ionic strength buffer showed that there is no significant NC(11–55) binding to the apical loop (66). It is likely that the NC zinc fingers do not interact with the two unpaired guanine residues of the apical loop because these residues are involved in stacking or transient hydrogen bonding interactions with neighbor nucleotides (13).

Mechanisms of annealing and roles of NC

Our probing data support the equilibrium between the closed and 'Y' conformations of the cTAR hairpin, i.e. the existence of two sites (the single-stranded ends and the apical loop) for initiation of the annealing reaction. The secondary structure of the 'Y' conformation (Figure 7) suggests that nucleation through the $3'/5'$ termini results in formation of a zipper intermediate involving at least 12 bp which is subsequently converted into the final extended duplex. Since the apical loop of the cTAR DNA hairpin is a weak binding site for NC, it can probably interact with the apical loop of the TAR RNA hairpin in the presence of NC. Using mini-cTAR mutants, gel-shift annealing assays, probing experiments and fluorescence anisotropy we showed that the annealing intermediate of the kissing pathway is the loop–loop interaction involving 6 intermolecular base pairs (Figures 2–4). The enzymatic and chemical probing data suggest that the apical stems are not destabilized by the loop–loop interaction (Figure 3 and data not shown). In their preceding work, Vo *et al.* (34) used a mini-cTAR mutant named Inv-1 that contains inverted sequences for the internal loop and the apical stem. This mutant designed to trap the 6 bp kissing complex, did not form heteroduplex with mini-TAR RNA (34). From these results, Vo *et al.* (34) suggested that the annealing intermediate involves 17 intermolecular base pairs, including the nucleotides from the apical stem–loop and from the internal loop of cTAR. The apparent discrepancy between our findings and this previous work can be explained by the fact that the secondary structure of our mutants is similar to that of the wild-type mini-cTAR, whereas this is not the case for the Inv-1 mutant. Indeed, the configuration of the internal loop in our mutants is similar to that of the wild-type mini-cTAR hairpin (the internal loop is located near the $3'$-end), whereas it is not the case for the Inv-1 mutant (the internal loop is located near the $5'$-end). In addition, we used experimental conditions (low temperature, high concentrations of nucleic acids) that were different from those used in the previous study. Our findings in conjunction with previous works (24,26) are consistent with annealing mechanisms that involve the fast formation of unstable kissing and zipper intermediates, followed by a rate-limiting strand exchange between the stems of hairpins. TAR RNA–DNA annealing via the zipper intermediate is favored in the presence of saturated NC (24). This finding is consistent with our probing data showing that the stem end is less stable than the apical stem which is not destabilized by NC. In addition, NC

shifts the equilibrium toward the 'Y' conformation and slightly destabilizes the lower stem that is adjacent to the internal loop. Preferential binding of NC zinc fingers to the internal loop is probably responsible for destabilization of the adjacent stem.

CONCLUSIONS

Here, we investigated the TAR RNA–DNA annealing mechanism in the absence or presence of NC under high-salt conditions previously shown to promote strand transfer (17,27). Our study is the first to show that the annealing intermediate of the kissing pathway is the loop–loop complex involving 6 intermolecular base pairs. Our results support a model for the NC–cTAR interactions in which one molecule of NC binds the internal loop tightly via its zinc fingers, whereas other NC molecules bind the apical loop and the double-stranded regions weakly via the basic residues scattered along the protein. The timing of HIV-1 core disassembly is essential for viral DNA synthesis, but there is still no consensus on exactly when uncoating occurs (69). Dissociation of the capsid shell surrounding the nucleoprotein complex is believed to take place during reverse transcription. The NC molecules are probably diluted by the disassembly of viral core. Dilution of NC is expected to induce a significant dissociation of NC molecules that are weakly bound to the minus-strand strong stop DNA. An attractive hypothesis, based on our results, is that the initiation sites for the annealing reaction (the apical loop and the $3'/5'$ termini of cTAR) are not bound to NC and that NC binds only its strong binding site (the cTAR internal loop) when uncoating and the first strand occur simultaneously. Binding of NC to the internal loop of the cTAR DNA hairpin would be essential for destabilization of this hairpin, and therefore for the annealing step of first strand transfer.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. 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.
2. Mougel,M., Houzet,L. and Darlix,J.L. (2009) When is it time for reverse transcription to start and go? *Retrovirology*, **6**, 24.

3. Feng,S. and Holland,E.C. (1988) HIV-1 tat trans-activation requires the loop sequence within tar. *Nature*, **334**, 165–167.
4. 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.
5. Kanevsky,I., Chaminade,F., Ficheux,D., Moumen,A., Gorelick,R., Negroni,M., Darlix,J.L. and Fosse,P. (2005) Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J. Mol. Biol.*, **348**, 1059–1077.
6. 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.
7. Berkhout,B., Vastenhouw,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.
8. Bernacchi,S., Stoylov,S., Piemont,E., Ficheux,D., Roques,B.P., Darlix,J.L. and Mely,Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J. Mol. Biol.*, **317**, 385–399.
9. 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.
10. 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.
11. Azoulay,J., Clamme,J.P., Darlix,J.L., Roques,B.P. and Mely,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.
12. 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. USA*, **104**, 5261–5267.
13. Zargarian,L., Kanevsky,I., Bazzi,A., Boynard,J., Chaminade,F., Fosse,P. and Mauffret,O. (2009) Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. *Nucleic Acids Res.*, **37**, 4043–4054.
14. 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.
15. Darlix,J.L., Vincent,A., Gabus,C., de,R.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.
16. 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.
17. 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.
18. 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.
19. 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.
20. Thomas,J.A. and Gorelick,R.J. (2008) Nucleocapsid protein function in early infection processes. *Virus Res.*, **134**, 39–63.
21. Beltz,H., Azoulay,J., Bernacchi,S., Clamme,J.P., Ficheux,D., Roques,B., Darlix,J.L. and Mely,Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J. Mol. Biol.*, **328**, 95–108.

22. Beltz,H., Clauss,C., Piemont,E., Ficheux,D., Gorelick,R.J., Roques,B., Gabus,C., Darlix,J.L., de Rocquigny,H. and Mely,Y. (2005) Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J. Mol. Biol.*, **348**, 1113–1126.
23. Shvadchak,V.V., Klymchenko,A.S., de,R.H. and Mely,Y. (2009) Sensing peptide-oligonucleotide interactions by a two-color fluorescence label: application to the HIV-1 nucleocapsid protein. *Nucleic Acids Res.*, **37**, e25.
24. 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.
25. 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.
26. Godet,J., de,R.H., Raja,C., Glasser,N., Ficheux,D., Darlix,J.L. and Mely,Y. (2006) During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J. Mol. Biol.*, **356**, 1180–1192.
27. Moumen,A., Polomack,L., Roques,B., Buc,H. and Negroni,M. (2001) The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res.*, **29**, 3814–3821.
28. de Rocquigny,H., Ficheux,D., Gabus,C., Fournie-Zaluski,M.C., Darlix,J.L. and Roques,B.P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.*, **180**, 1010–1018.
29. Lapadat-Tapolsky,M., de,R.H., Van,G.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.
30. Polge,E., Darlix,J.L., Paoletti,J. and Fosse,P. (2000) Characterization of loose and tight dimer forms of avian leukosis virus RNA. *J. Mol. Biol.*, **300**, 41–56.
31. Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
32. Fosse,P., Motte,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.
33. Maxam,A.M. and Gilbert,W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.*, **65**, 499–560.
34. 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.
35. 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. USA*, **91**, 4945–4949.
36. Laughrea,M. and Jette,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.
37. Ben Ali,M., Chaminade,F., Kanevsky,I., Ennifar,E., Josset,L., Ficheux,D., Darlix,J.L. and Fosse,P. (2007) Structural requirements for nucleocapsid protein-mediated dimerization of avian leukosis virus RNA. *J. Mol. Biol.*, **372**, 1082–1096.
38. Lanchy,J.M., Rentz,C.A., Ivanovitch,J.D. and Lodmell,J.S. (2003) Elements located upstream and downstream of the major splice donor site influence the ability of HIV-2 leader RNA to dimerize in vitro. *Biochemistry*, **42**, 2634–2642.
39. Boiziau,C., Dausse,E., Yurchenko,L. and Toulme,J.J. (1999) DNA aptamers selected against the HIV-1 trans-activation-responsive RNA element form RNA-DNA kissing complexes. *J. Biol. Chem.*, **274**, 12730–12737.

40. Collin,D., van,H.C., Boiziau,C., Toulme,J.J. and Guittet,E. (2000) NMR characterization of a kissing complex formed between the TAR RNA element of HIV-1 and a DNA aptamer. *Nucleic Acids Res.*, **28**, 3386–3391.

41. Paillart,J.C., Shehu-Xhilaga,M., Marquet,R. and Mak,J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.*, **2**, 461–472.

42. Desai,N.A. and Shankar,V. (2003) Single-strand-specific nucleases. *FEMS Microbiol. Rev.*, **26**, 457–491.

43. Till,B.J., Burtner,C., Comai,L. and Henikoff,S. (2004) Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.*, **32**, 2632–2641.

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. 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.

46. 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.

47. Rubin,C.M. and Schmid,C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.*, **8**, 4613–4619.

48. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

49. Morellet,N., Demene,H., Teilleux,V., Huynh-Dinh,T., de,R.H., Fournie-Zaluski,M.C. and Roques,B.P. (1998) Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J. Mol. Biol.*, **283**, 419–434.

50. 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.

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

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

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

54. Shubsda,M.F., Paoletti,A.C., Hudson,B.S. and Borer,P.N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry*, **41**, 5276–5282.

55. Paoletti,A.C., Shubsda,M.F., Hudson,B.S. and Borer,P.N. (2002) Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. *Biochemistry*, **41**, 15423–15428.

56. Lapadat-Tapolsky,M., Pernelle,C., Borie,C. and Darlix,J.L. (1995) Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res.*, **23**, 2434–2441.

57. Stoylov,S.P., Vuilleumier,C., Stoylova,E., de,R.H., Roques,B.P., Gerard,D. and Mely,Y. (1997) Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. *Biopolymers*, **41**, 301–312.

58. Cruceanu,M., Gorelick,R.J., Musier-Forsyth,K., Rouzina,I. and Williams,M.C. (2006) Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J. Mol. Biol.*, **363**, 867–877.

59. 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.

60. Williams,M.C., Gorelick,R.J. and Musier-Forsyth,K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl Acad. Sci. USA*, **99**, 8614–8619.

61. Heath,M.J., Derebail,S.S., Gorelick,R.J. and DeStefano,J.J. (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.

62. Narayanan,N., Gorelick,R.J. and DeStefano,J.J. (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.

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

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

65. Avilov,S.V., Godet,J., Piemont,E. and Mely,Y. (2009) Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. *Biochemistry*, **48**, 2422–2430.

66. Bazzi,A., Zargarian,L., Chaminade,F., Boudier,C., de,R.H., Rene,B., Mely,Y., Fosse,P. and 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.

67. 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.

68. Kalnik,M.W., Norman,D.G., Li,B.F., Swann,P.F. and Patel,D.J. (1990) Conformational transitions in thymidine bulge-containing deoxytridecanucleotide duplexes. Role of flanking sequence and temperature in modulating the equilibrium between looped out and stacked thymidine bulge states. *J. Biol. Chem.*, **265**, 636–647.

69. 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.

114

Part 2. Structure-function relationship of the minus "strong-stop DNA"

1. Introduction

The main object of my thesis is minus-strand "strong-stop" DNA (ssDNA), synthesized during the first step of reverse transcription **(Figure 33)**. The first requirement for synthesis of ssDNA is the annealing of the human primer tRNA$^{(Lys,3)}$ onto the primer binding site (PBS) of the genomic RNA. It is unknown when tRNA$^{(Lys,3)}$ is displaced from PBS, but displacement is necessary to allow copying of the 3' end of the tRNA into DNA (311). Therefore, the displacement of the tRNA primer could occur either before **(Figure 56A, hypothesis a)** or after the first strand transfer **(Figure 56A, hypothesis b)**. The first strand transfer is mediated by base pairing of the R region at the 3' end of the genomic RNA and the complementary r region at the 3' end of ssDNA **(Figure 56A)**. The R sequence contains the transactivator response element (TAR) and a portion of the poly(A) signal (31, 114). The 3' TAR sequence folds into a hairpin in the entire HIV-1 RNA genome extracted from virions (395). The 3' poly(A) sequence can form a hairpin (31, 205), but there is no data showing that this secondary structure is present in the entire HIV-1 RNA genome. Moreover, *in vitro* studies suggest that the 3' U3-poly(A) and *gag* sequences are involved in base pairing interactions to circularize the HIV-1 genome (23, 294). These interactions stimulate the first strand transfer during reverse transcription *in vitro* (23). 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 (34, 36). 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.

Mutations within the first 10 nucleotides of the 5' R sequence produce virions that are markedly defective for reverse transcription (292). These results suggest that the great majority of first strand transfers occur after completion of ssDNA synthesis, i.e. the entire cTAR sequence would be required for efficient strand transfer *in vivo*. 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 (12, 36, 175, 243). As described in the Part 1 of the 'Results and Discussion' section, 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

115

twelve unpaired nucleotides in its lower part. We also showed that the apical and internal loops of cTAR are binding sites for NC. To date, the NC binding sites within the full-length ssDNA have not been identified.
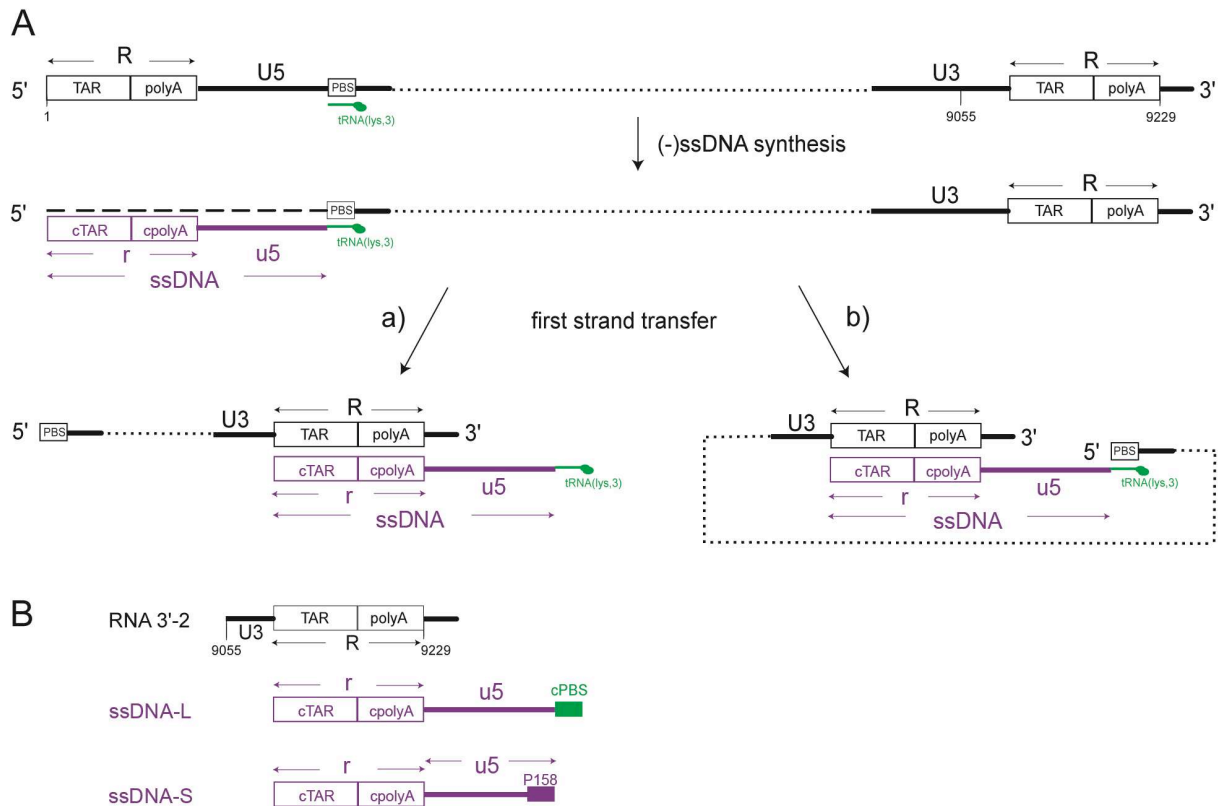
Figure 56. (A) Schematic diagram of events leading to the first strand transfer. The 3' 18 nucleotides of the tRNA primer are annealed to the PBS, a complementary sequence in the viral genome (black line); numbering is relative to the genomic RNA cap site (+1). Once annealed, RT catalyzes extension of the primer tRNA (green) to form ssDNA (purple line). The 5' end of the gRNA is degraded (broken line) by the action of the RNase H activity of RT, and ssDNA is transferred to the 3' end of the gRNA. This transfer is mediated by base pairing of the R and r regions that are complementary. The displacement of the tRNA primer could occur either before (pathway a) or after the first strand transfer (pathway b). **(B)** RNA and ssDNAs used in this study. RNA 3'-2 (200 nucleotides) contains the 3' end of the U3 sequence and the full-length R sequence; ssDNA-L (196 nucleotides) and ssDNA-S (178 nucleotides) are the extension products of primers cPBS and P158, respectively. Primer cPBS represents the 3'-terminal 18 nucleotides of tRNA^{Lys,3} that are complementary to the primer binding site (PBS). Primer P158 is complementary to the 3' end of the U5 sequence and corresponds to the first 21 nt of ssDNA.

Since mutations in the TAR apical loop decrease the first strand transfer *in vitro*, Berkhout *et al.* (34) 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 under high-salt concentrations allowing cDNA synthesis by RT and strand transfer (157, 205, 271). 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 (242, 385). 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 (143, 385). 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.

Here, we used two full-lenth ssDNAs **(Figure 56B)** to investigate the relationship between the structure and function of ssDNA in 0.2 mM (intracellular concentration) and 2 mM $MgCl_2$ (concentration required for reverse transcription and strand transfer *in vitro*). ssDNA-L represents the full-length ssDNA that is not annealed to the PBS region **(Figure 56A, hypothesis a)**, i.e. it contains the complementary sequence to PBS (cPBS) that might influence the folding of ssDNA. If the full-length ssDNA is annealed to the PBS region **(Figure 56A, hypothesis b)**, then its folding is not influenced by the cPBS sequence; this possibility is represented by ssDNA-S that is deleted of the cPBS sequence.

The number of Gag molecules in immature HIV-1 particles was estimated to be 1400-2400 (51, 421). About 1000-1500 NC molecules coating the gRNA in a dimeric form were found within the infectious HIV-1 particles (45, 60). Therefore, the NC to nucleotide molar ratios were probably between 1:18 and 1:7 within the infectious HIV-1 particles. To take into account these observations, the NC assays were performed at NC to nucleotide molar ratios of 1:24, 1:18, 1:12 and 1:7.

2. <u>Synthesis, labeling and purification of ssDNAs</u>

2.1 Purification and labeling of ssDNA-S generated by chemical synthesis

We purchased from IBA (Germany) and Eurogentec (Belgium) oligonucleotides that were obtained by chemical synthesis and should correspond to ssDNA-S (178 nt). Unfortunately, even the best oligonucleotide solution contained a large amount of oligonucleotides that were shorter than the full-length ssDNA-S **(Figure 57A, lane 3)**. Even after purification by electrophoresis on a denaturing polyacrylamide gel, the oligonucleotide solution contained products that formed a smear above ssDNA-S **(Figure 57B, lane 3)**.

As mentioned for 5' end labeling of cTAR (see in the 'Materials and Methods' section), the purified oligonucleotide ssDNA-S could also be labeled at its 5' end. As shown in **Figure 57C**, the oligonucleotide solution was not quite pure. Three slices of gel (a, b and c) were eluted, and the oligonucleotides were purified and analyzed in parallel with ssDNA-L and

ssDNA-S generated by reverse transcription. Lanes 3, 4 and 5 in **Figure 57D** correspond to oligonucleotides purified from slices a, b and c in **Figure 57C**, respectively. Band a corresponds to ssDNA-S, whereas band b and c probably correspond to the +1 and -1 products of ssDNA-S, respectively **(Figure 57D)**. Despite the purification of labeled ssDNA-S, a smear was again observed **(Figure 57D, lane 3)**. It is likely that this smear correspond to partially protected oligonucleotides, i.e. the deprotection step in oligonucleotide synthesis was not complete. In addition, the recovery rate of ssDNA-S was low (about 10%). Finally, we chose to synthesize ssDNAs by reverse transcription because the purified ssDNA-S produced by chemical synthesis contained protected nucleotides and the yield of purification was low.
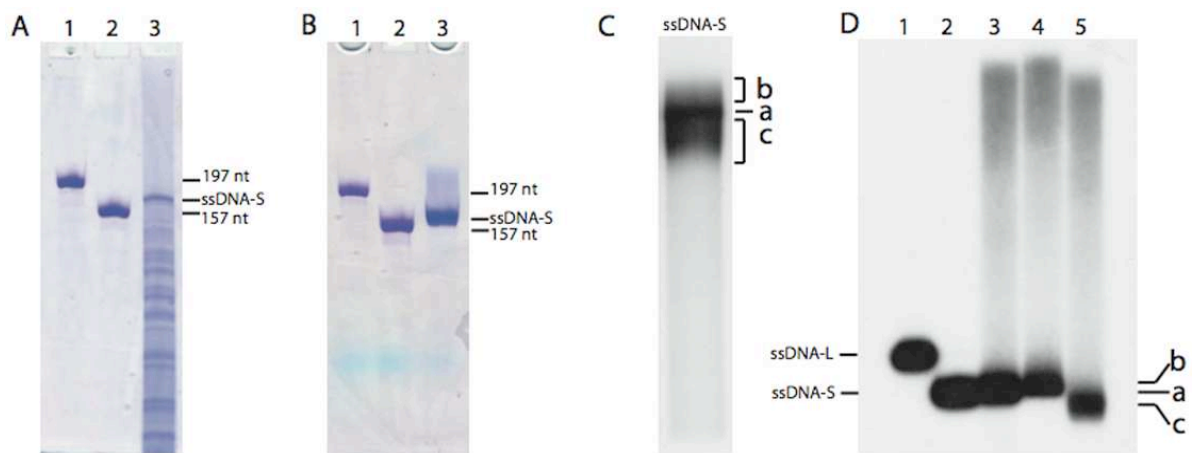
Figure 57. Analysis of oligonucleotide ssDNA-S. (A) Five µg of oligonucleotide ssDNA-S was analyzed on a 6% denaturing polyacrylamide gel. **(B)** Seven hundred ng of the first purification of ssDNA-S was analyzed on a 6% denaturing polyacrylamide gel. Lanes 1 and 2 in **A and B** indicate the size markers (197 nt and 157 nt). The 5' end-labeled oligonucleotide ssDNA-S was analyzed before **(C)** and after **(D)** the second purification. Lanes 1 and 2 in **D** correspond to ssDNA-L and ssDNA-S synthesized by reverse transcription, respectively. Lanes 3, 4 and 5 in **D** correspond to purification of slices a, b and c indicated in **C**, respectively.

2.2 Synthesis, labeling and purification of ssDNAs by reverse transcription

Previous studies in our group found that the yield of reverse transcription products was higher using the RNA template (RNA 1-415) than other smaller or longer RNA template. Both ssDNA-L and ssDNA-S were then synthesized by reverse transcription of RNA 1-415 from the cPBS and P158 primers, respectively **(Figure 58)**.
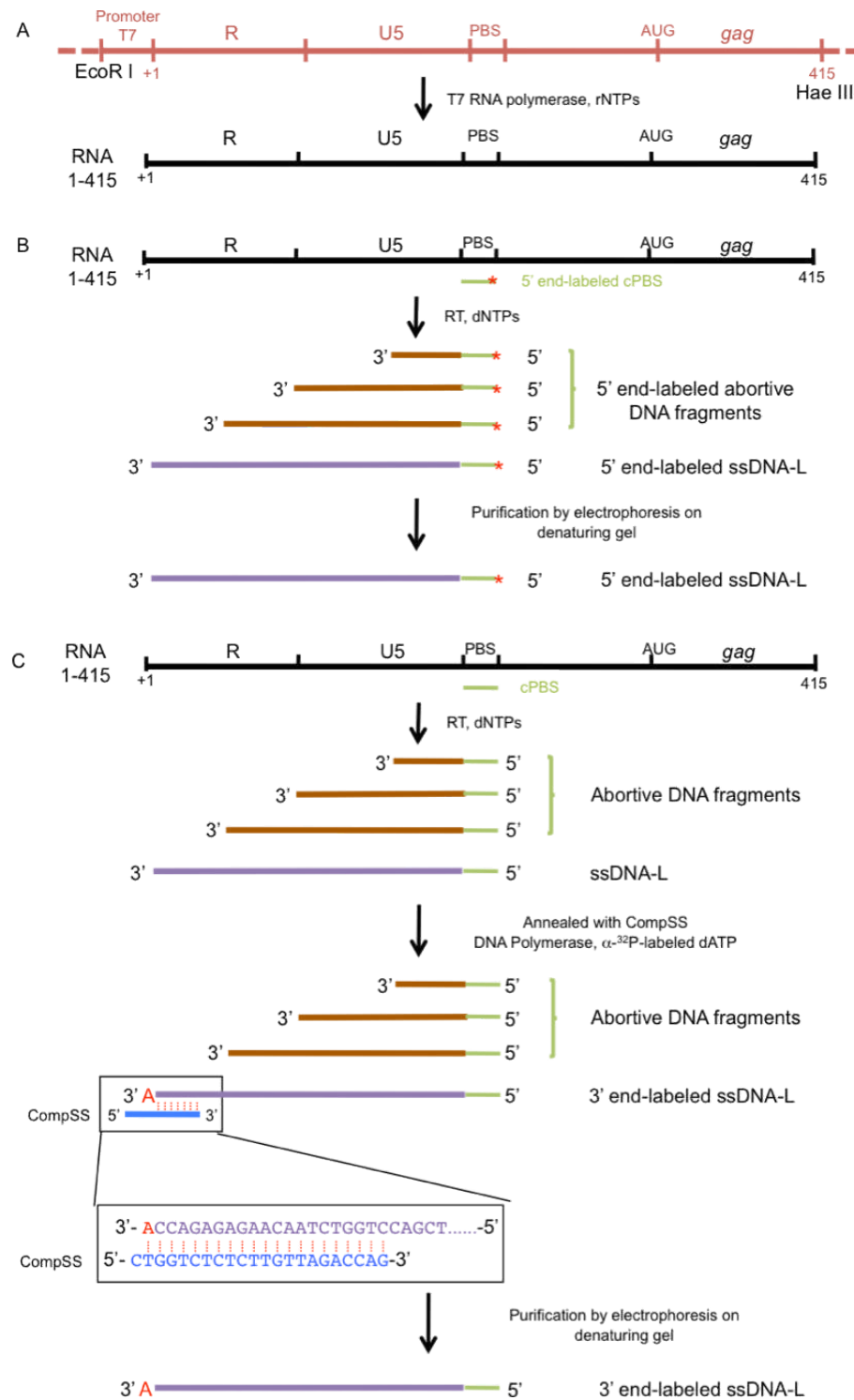
Figure 58. Synthesis, labeling and purification of ssDNA-L generated by reverse transcription. **(A)** *In vitro* transcription of RNA 1-415 using the template generated from plasmid pYC5' digested by Hae III. **(B)** *In vitro* reverse transcription of RNA 1-415 initiated by 5' end-labeled cPBS primer. All the reverse transcription products, including ssDNA-L and abortive DNA fragments were labeled at their 5' end. **(C)** *In vitro* reverse transcription of RNA 1-415 initiated by the cPBS primer. In the presence of oligonucleotide CompSS, a DNA polymerase and α-^{32}P-labeled dATP, ssDNA-L was labeled at its 3' end. Both 5' and 3' end-labeled ssDNA-L were purified by electrophoresis on a denaturing polyacrylamide gel. The plasmid is indicated in red, RNA 1-415 in black, cPBS primer in green, abortive DNA fragment in brown, ssDNA-L in purple and CompSS in blue. The red asterisk and the red A indicate the 5' and 3' end-labeling, respectively.

119

2.2.1   Synthesis and purification of RNA 1-415

To optimize full-length ssDNA synthesis, it is important that RNA 1-415 is not partially degraded or contaminated by RNA fragments that are produced by abortive transcription. As shown in **Figure 59**, the preparation of RNA 1-415 contains a significant amount of an abortive RNA fragment at the end of transcription by T7 RNA polymerase. Therefore, RNA 1-415 was purified using a denaturing polyacrylamide gel (see in the 'Materials and Methods' section). After purification, RNA 1-415 was full-length and ready to serve as template for ssDNA synthesis **(Figure 58A)**.
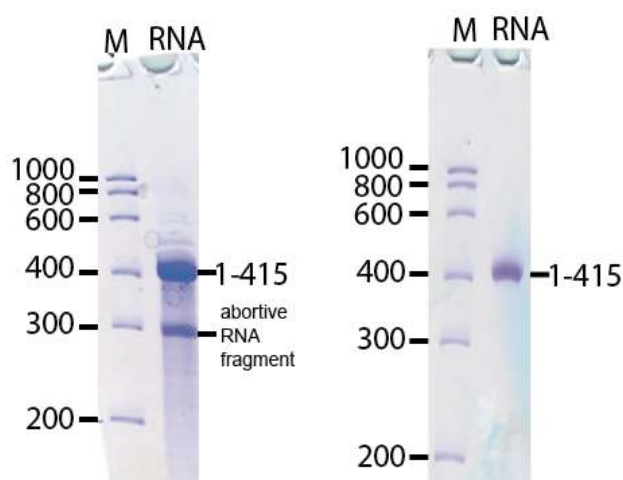
Figure 59. Analysis of transcription products before and after purification on a 5% denaturing polyacrymide gel. On the left, the mixture of products (RNA1-415 and an abortive RNA fragment) obtained at the end of transcription. On the right, RNA1-415 after purification.

2.2.2   Synthesis and 5' end labeling of ssDNAs

The cPBS or P158 primer was firstly 5' end-labeled using T4 polynucleotide kinase and γ-^{32}P-labeled ATP. The reverse transcription of the RNA template (RNA 1-415) initiated by the 5' end-labeled primer generates ssDNA-L or ssDNA-S and other DNA fragments that were labeled at their 5' end **(Figure 58B)**. After purification, the 5' end-labeled ssDNAs were then checked for purity and integrity on a 6% denaturing polyacrylamide gel. As shown in **Figure 60**, ssDNAs were pure and therefore usable for structural and functional analysis. All the dimerization, conformation and annealing assays in this Part were carried out with the 5' end-labeled ssDNAs.
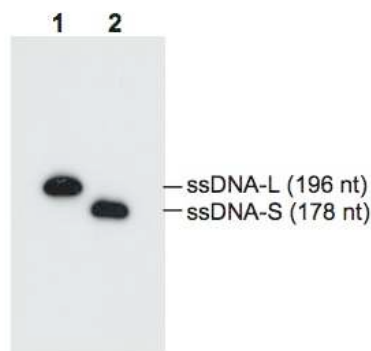
Figure 60. Analysis of purified 5' end-labeled ssDNAs. One hundred fifty fmol (6 x 10^4 cpm) of the 5' end-labeled ssDNA-L (lane 1) and ssDNA-S (lane 2) in 4 μl of loading buffer A (7 M urea, 0.03 w/V% bromophenol blue and 0.03 w/v% xylene cyanol) were heated at 90 °C for 2 min, chilled for 2 min on ice and analysed by electrophoresis on a 6% denaturing polyacrylamide gel. After electrophoresis, the gel was fixed, dried and autoradiographed.

2.2.3 Synthesis and 3' end labeling of ssDNAs

2.2.3.1 Experimental strategy

Since the synthesis of ssDNA generates a lot of abortive DNA fragments during the reverse transcription of HIV-1 RNA (23, 33, 103, 145, 157, 348), we developed a new method to label only the full-length ssDNA-L and ssDNA-S at their 3' end among the mixture of reverse transcription products **(Figure 58C)**. Our labeling method is based on the study of Huang and Szostak (184) that used an oligonucleotide to label the 3' end of an RNA. The short DNA template with a two nucleotide 5' overhang of 3'-TA-5', 3'-TG-5' or 3'-TC-5' is annealed to the 3' end of an RNA and the Klenow fragment of DNA polymerase I can then cleanly and efficiently extend the 3' end of the RNA by the incorporation of a single a-^{32}P-labeled dATP residue. Thus, we used the CompSS oligonucleotide that is complementary to the 3' end of ssDNA and cTAR DNA with a non-hybridizing tail of two nucleotides (5'-CT-3').

2.2.3.2 Labeling assays

Purified RNA 1-415 was the template and the cPBS DNA oligonucleotide was used as primer for the synthesis of ssDNA-L as described in the 'Materials and Methods' section. Reverse transcription initiated from the 5' end-labeled cPBS primer was taken as control showing the bands corresponding to ssDNA-L and abortive DNA fragments **(Figure 61, lane 5)**. Note that the length of 5' end-labeled DNAs was not greater than that of ssDNA-L. After reverse transcription of RNA 1-415, the 3' end-labeling assay was performed at 25 °C or 37 °C for 15 min using Klenow fragment of DNA polymerase I, [α-^{32}P] dATP and in the

absence or in the presence of oligonucleotide CompSS. Consistent with the optimal temperature for Klenow fragment activity, DNAs were more efficiently labeled at 37 °C than at 25 °C. Despite the absence of oligonucleotide CompSS, ssDNA-L (196 nucleotides) and DNA fragments (165-170 nucleotides), named DF, were labeled at their 3' end **(Figure 61, lanes 1 and 3)**. A possible hypothesis is that the 3' ends of these DNAs fold into stem-loop structures serving as self-primers that incorporate the labeled adenine in the presence of the Klenow fragment of DNA polymerase I **(Figure 62a)**. This hypothesis is supported by studies reporting self-priming from ssDNA (149, 157). As expected, the 3' end labeling of ssDNA-L was strongly increased in the presence of oligonucleotide CompSS **(Figure 61, lanes 2 and 4)**. The DFs (165-170 nt) did not contain the complementary sequence of oligonucleotide CompSS if they are synthesized from the cPBS primer **(Figure 62a)**. However, labeling of these DNAs was increased in the presence of oligonucleotide CompSS **(Figure 61, lanes 2 and 4)**. Therefore, we propose that these DNAs did not originate from cPBS extension by RT. Our hypothesis is that the DFs were generated from extension of RNA fragments produced by abortive reverse transcription events **(Figure 62b)**. These RNAs contained the template for the complementary sequence of oligonucleotide CompSS and were able to form hairpins serving as self-primers for reverse transcription. The full-length RNA 1-415 and large fragments of this RNA were probably also self-primers for DNA synthesis **(Figure 63)**, since DNAs longer than ssDNA-L (named LDF) were observed **(Figure 61)**.
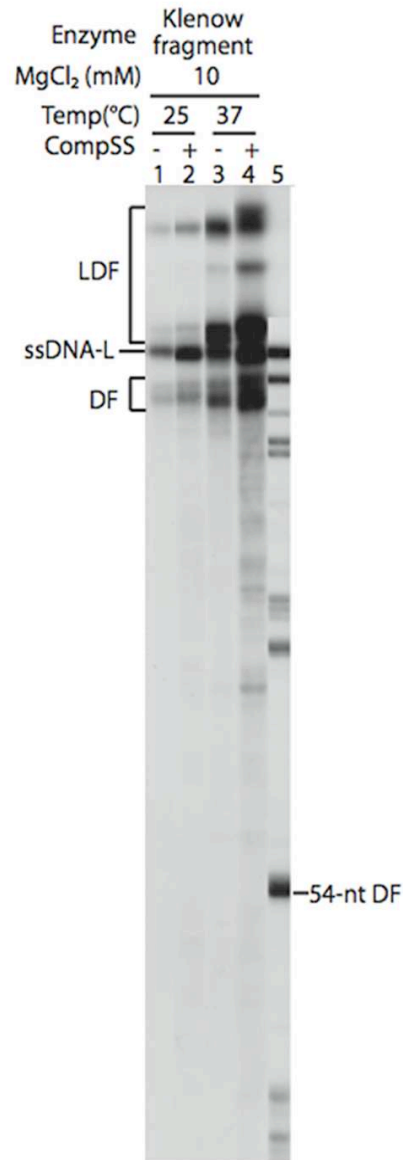
Figure 61. 3' end-labeling assays of full-length ssDNA-L performed by Klenow fragment of DNA polymerase I after reverse transcription. Lanes 1 and 2: 3' end-labeling performed at 25 °C for 15 min. Lanes 3 and 4: 3' end-labeling performed at 37 °C for 15 min. The Klenow buffer contains a final concentration of 10 mM $MgCl_2$. Lanes 1 and 3: controls without oligonucleotide CompSS. Lanes 2 and 4: assays with CompSS. Lane 5: ssDNA-L synthesized by reverse transcription with the 5' end-labeled cPBS primer and RNA 1-415. DF indicates 3' end-labeled abortive DNA fragments shorter than ssDNA-L. LDF indicates 3' end-labeled DNA fragments longer than ssDNA-L. 54-nt DF indicates the 54-nucleotide DNA fragment.
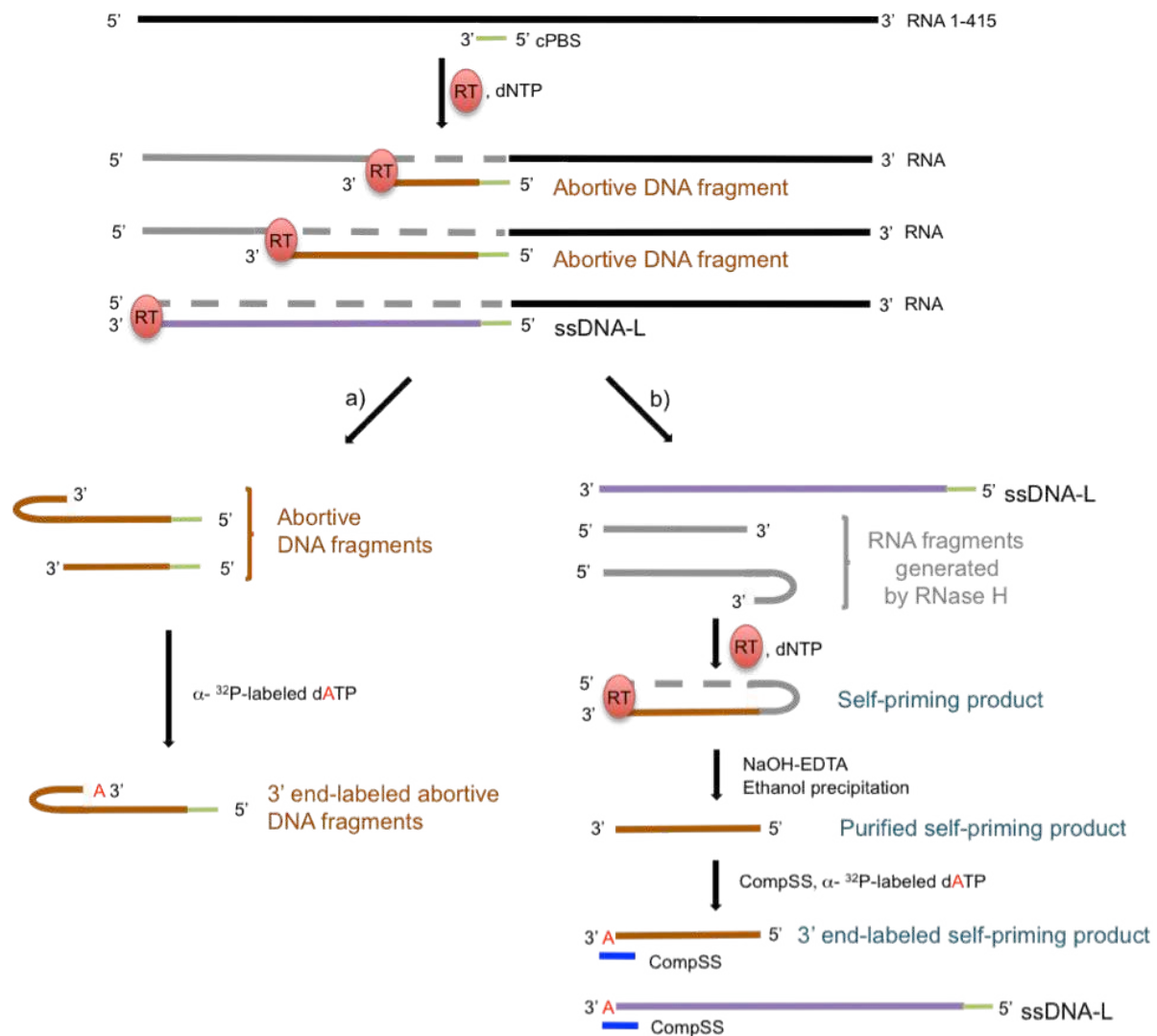
Figure 62. 3' end-labeling in the absence or presence of oligonucleotide CompSS. Reverse transcription of HIV-1 RNA 1-415 generates a mixture of DNAs, including ssDNA-L and abortive DNA fragments. **(a)** Abortive DNA fragments are generated during reverse transcription. Some of these abortive DNA fragments could fold into hairpin that serves as self-primer for labeling at its 3' end in the absence of oligonucleotide CompSS. **(b)** Due to abortive reverse transcription events, RNA 1-415 is partially degraded by the RNase H activity of RT. RNAs shorter than RNA 1-415 could fold into a primer-template that could serve as self-primer for synthesis of DNA fragments. Self-priming products contain also the complementary sequence to CompSS. Thus, in the presence of oligonucleotide CompSS, both ssDNA-L and self-priming products are labeled at their 3' ends. RNA fragments generated by RNase H are indicated in gray and abortive DNA fragments in brown.

Finally, the 3' end labeling by the Klenow fragment occurred under conditions (10 mM MgCl$_2$ and 37 °C) that probably favor formation of hairpins at the 3' end of DNAs. In addition, these conditions may facilitate nonspecific annealing of oligonucleotide CompSS to DNAs containing sequences that are partially complementary to this oligonucleotide. To improve the 3' end-labeling method, lower magnesium concentrations and higher temperatures than 37 °C were tested for the labeling step. AMV RT and Taq DNA polymerase were chosen because the optimal activity temperature for these enzymes is 42 and

75 °C, respectively. Note that the melting temperature (T_m) of oligonucleotide CompSS is about 56°C, i.e. between the optimal temperatures for AMV RT and Taq DNA polymerase.
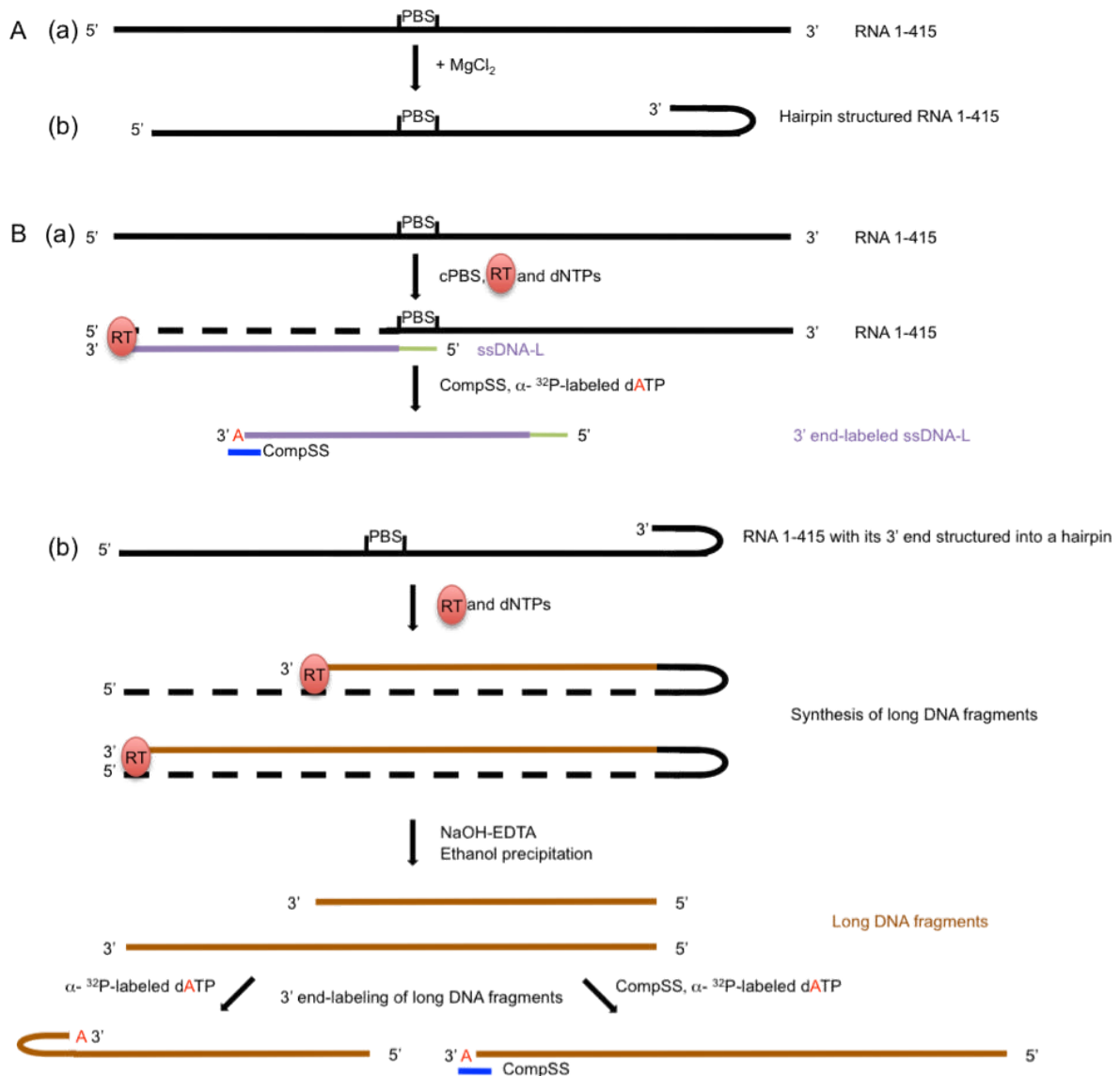
Figure 63. Hypothesis on folding of RNA 1-415 and synthesis of long DNA fragments. (A) In the presence of MgCl$_2$ (2 mM in reverse transcription assay), some of the 3' end of RNA 1-415 folds into a hairpin structure. **(B)** Reverse transcription of RNA 1-415. (a) ssDNA-L is synthesized from the RNA template in the presence of primer cPBS, RT and dNTPs. (b) In the presence of RT and dNTPs, long DNA fragment is synthesized from the hairpin located at the 3' end of RNA 1-415. After reverse transcription, the mixture of DNA and RNA is treated with NaOH-EDTA to destroy the RNA template (see Materials and Methods). RNA 1-415 is indicated in black, cPBS in green, ssDNA-L in purple and long DNA fragment in brown.

As expected, ssDNA-L was barely labeled by AMV RT in the absence of oligonucleotide CompSS, whereas it was strongly labeled in the presence of this oligonucleotide **(Figure 64, compare lanes 1 and 2)**. However, 3' end labeling by AMV RT was not chosen because an abortive DNA (54 nt long) was significantly labeled. Nonspecific

labeling of 54-nt DF is probably due to its folding into a stem-loop structure that is stable at 42 °C in the presence of 8 mM $MgCl_2$. Note that this abortive DNA fragment was also identified by reverse transcription from the 5' end-labeled cPBS primer **(Figure 61, lane 5)**.
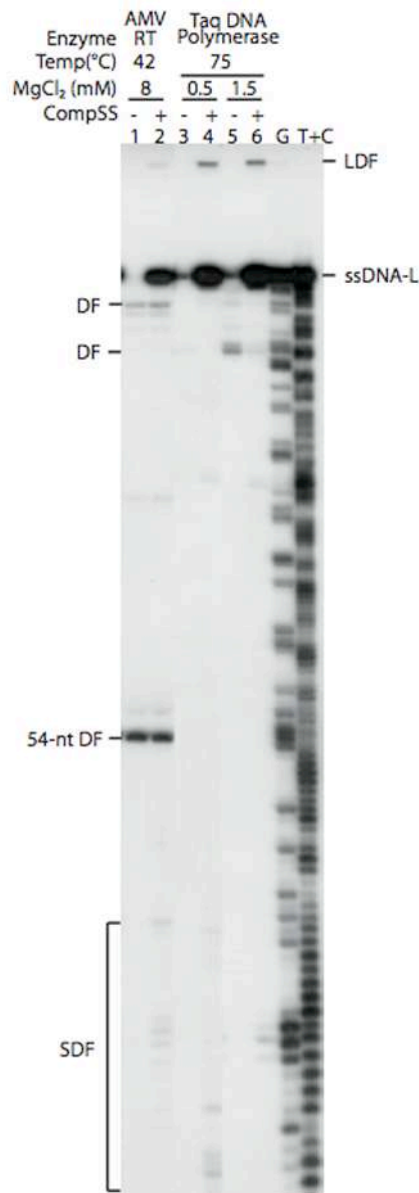
Figure 64. 3' end-labeling assays of ssDNA-L performed by AMV RT and Taq DNA polymerase after reverse transcription. Lanes 1 and 2: 3' end-labeling by AMV RT incubated at 42°C for 15 min. The RT buffer contains a final concentration of 8 mM $MgCl_2$. Lanes 3-6: 3' end-labeling by Taq DNA polymerase incubated at 75°C for 15 min. Lanes 3 and 4: assays were incubated in the presence of 0.5 mM $MgCl_2$. Lanes 5 and 6: assays were incubated in the presence of 1.5 mM $MgCl_2$. Lanes 1, 3 and 5: controls without CompSS. Lanes 2, 4 and 6: assays with CompSS. Lanes G and T+C refer to Maxam-Gilbert sequence markers. DF indicates 3' end-labeling abortive DNA fragments. LDF indicates 3' end-labeled DNA fragments longer than ssDNA-L. SDF indicates short DNA fragments. 54-nt DF indicates the 54-nucleotide DNA fragment.

To prevent the 3' end labeling of 54-nt DF, we used Taq DNA polymerase at 75 °C in the presence of low magnesium concentrations (0.5 or 1.5 mM $MgCl_2$) **(Figure 64, lanes 3-6)**.

Consistent with our hypothesis, 54-nt DF was not labeled **(Figure 64, lanes 3-6)**. Furthermore, nonspecific labeling was low in the absence of oligonucleotide CompSS **(Figure 64, lanes 3 and 5)**. Interestingly, the yield of 3' end-labeled ssDNA-L was higher with Taq DNA polymerase than with AMV RT **(Figure 64, compare lane 2 to lanes 4 and 6)**. In addition, the data show that 3' end-labeling of ssDNA-L by Taq DNA polymerase was slightly more efficient at 0.5 mM $MgCl_2$ than at 1.5 mM $MgCl_2$ **(Figure 64, lanes 4 and 6)**. Note that there were short DNA fragments (SDF) and only one LDF that were weakly labeled in the presence of oligonucleotide CompSS **(Figure 64, lanes 4 and 6)**. Since nonspecific labeling of DNA fragments by Taq DNA polymerase was low at 75 °C in the presence of 0.5 mM $MgCl_2$, we chose these conditions to specifically label ssDNA-L and ssDNA-S.

In summary, the ssDNAs were labeled at their 3' end in the presence of Taq DNA polymerase, oligonucleotide CompSS and α-^{32}P-labeled dATP. After purification, the 3' end-labeled ssDNAs were checked for purity and integrity on a 6% denaturing polyacrylamide gel. As shown in **Figure 65**, ssDNAs were pure and therefore usable for structural and functional analysis.
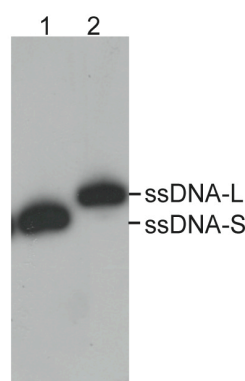
Figure 65. Analysis of purified 3' end-labeled ssDNAs. One hundred fifty fmol (6 x 10^4 cpm) of the 3' end-labeled ssDNA-S (lane 1) and ssDNA-L (lane 2) in 4 μl of loading buffer A were heated at 90 °C for 2 min, chilled for 2 min on ice and analysed by electrophoresis on a 6% denaturing polyacrylamide gel.

3. <u>Dimerization assays of ssDNAs</u>

Previous studies showed that both loose and tight duplexes can be characterized by native agarose gel electrophoresis at 4 °C in the TBM buffer (28, 232, 344). The labeled ssDNAs were monomeric after incubation with or without NC and analyzed under the electrophoretic conditions described above **(Figure 66)**. Therefore, the structural and functional analysis were not be complicated by the presence of homoduplexes.
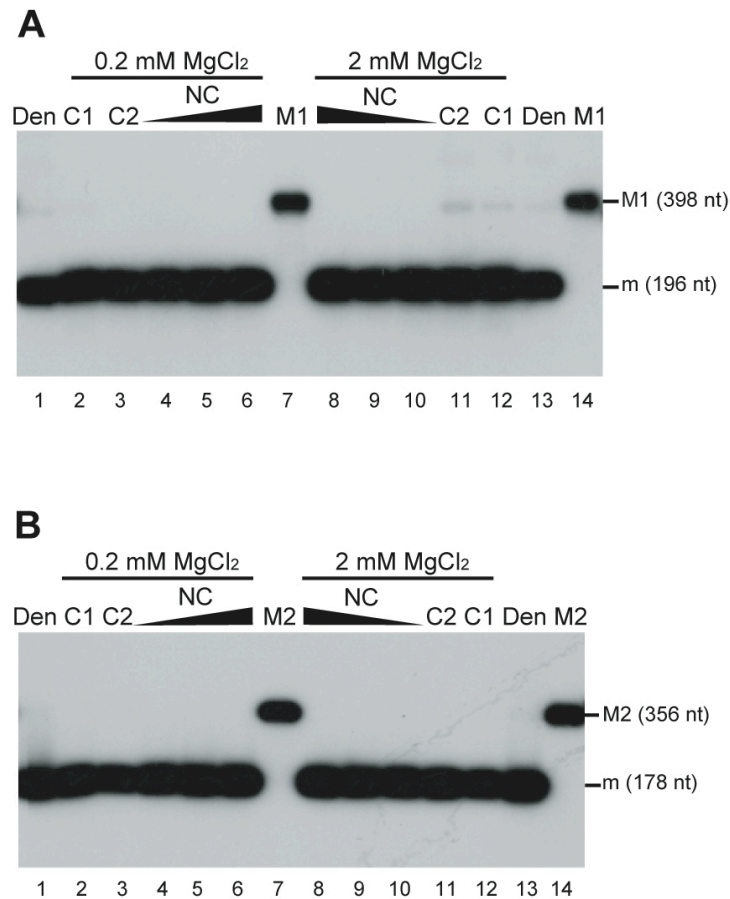
Figure 66. Analysis of ssDNAs dimerization. The samples were analyzed by 2% agarose gel electrophoresis at 4 °C in the TBM buffer and visualized by autoradiography. 5' end-labeled ssDNA-L **(A)** and ssDNA-S **(B)** in the absence (lances C1 and C2) or in the presence of NC (lanes 4-6 and 8-10) were incubated at 37°C in the presence of 0.2 or 2 mM $MgCl_2$ as described in the 'Materials and Methods' section. The protein to nucleotide molar ratios were 1:18 (lanes 4 and 10), 1:12 (lanes 5 and 9) and 1:7 (lanes 6 and 8). Lanes Den, heat-denatured ssDNAs. Lanes C1 are controls without phenol-chloroform extraction and ethanol precipitation. Lanes C2 are controls with phenol-chloroform extraction and ethanol precipitation. Monomeric forms of ssDNAs are indicated by m. M1 and M2 indicate the size markers.

4. <u>Analysis of ssDNA-3'UTR annealing</u>

As mentioned previously, the annealing reaction between the full-length ssDNA and the 3' end of the gRNA has not been studied. Here we investigated the annealing process using the two full-length ssDNAs and RNA 3'-2 representing the 3' end of gRNA **(Figure 56B)**. To determine whether the magnesium concentration has an effect on the annealing pathway, the annealing assays were performed in 0.2 and 2 mM $MgCl_2$ as described in the 'Materials and Methods' section. Since the 3' UTR and ssDNA are probably folded during the first strand transfer, RNA 3'-2 and ssDNA-L or ssDNA-S were denatured and renatured before to be mixed together.

4.1 Annealing in the absence of NC

Annealing of ssDNA-S to RNA 3'-2 was tested in the absence of NC in order to determine whether annealing of the full-length ssDNA to 3' UTR is possible without the chaperone activity of NC. After one-hour incubation in the presence of 0.2 mM or 2 mM $MgCl_2$, most of ssDNA-S molecules were in the unannealed form **(Figure 67)**.
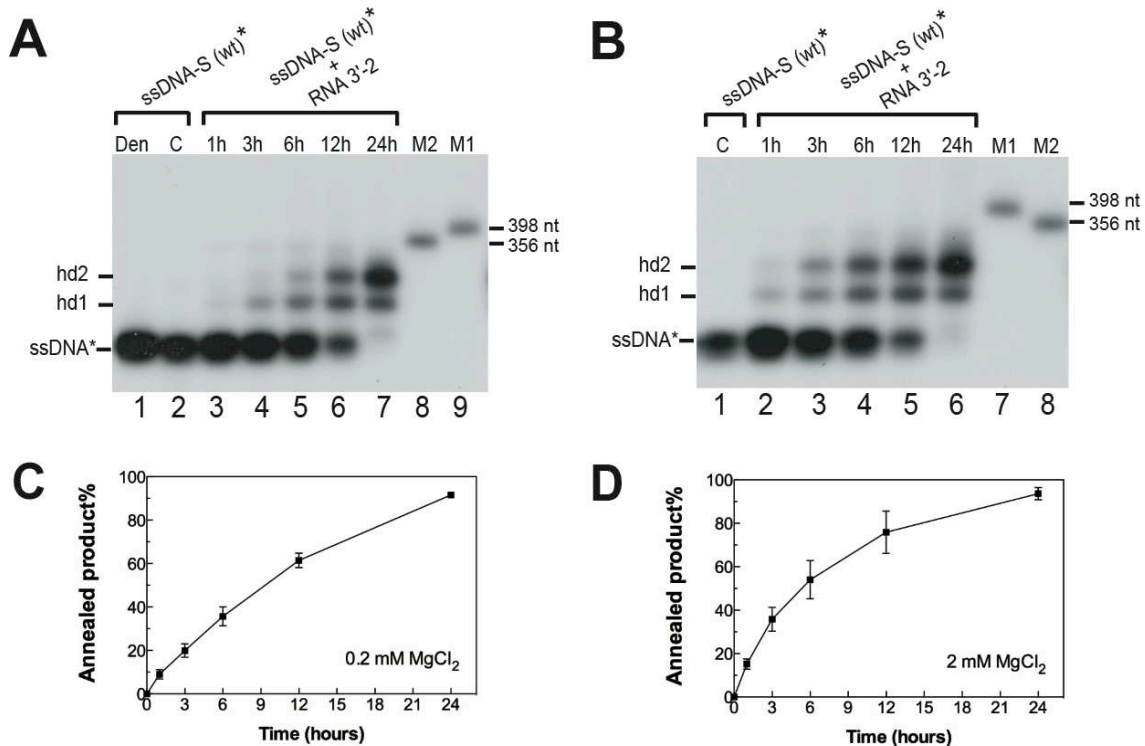
Figure 67. Time course of ssDNA-S annealing with RNA 3'-2 in the absence of NC and in the presence of 0.2 mM (A and C) or 2 mM $MgCl_2$ (B and D). (A and B) Annealing was performed as described in the 'Materials and Methods' section. Lane Den, heat-denatured 5' end-labeled ssDNA-S. Lanes C, 5' end-labeled ssDNA-S was incubated at 37 °C for 24 h in the absence of RNA 3'-2. Lanes 3-7 in A and lanes 2-6 in B, 5' end-labeled ssDNA-S and RNA 3'-2 were incubated together at 37 °C for various times. Lanes M1 and M2, size markers. **(C and D)** All experiments were repeated at least three times and error bars show standard deviations.

Surprisingly, two new bands appeared for incubation times equal to or greater than three hours. The positions of these bands suggest that they did not correspond to the full-length heteroduplex (378 nt) of RNA 3'-2 and ssDNA-S but to shorter heteroduplexes. It is well established that the magnesium ion accelerates the spontaneous degradation of RNA. Interestingly, the annealing rate was greater in 2 mM $MgCl_2$ than in 0.2 mM $MgCl_2$ **(Figure 67, compare C and D)**. Therefore, we propose that the two heteroduplexes resulted from annealing between ssDNA-S and two RNA fragments generated by spontaneous cleavages of RNA 3'-2 **(Figure 68)**. Note that formation of heteroduplexes was not efficient, since the annealing rate was not maximal before 24-hour incubation **(Figure 67)**.
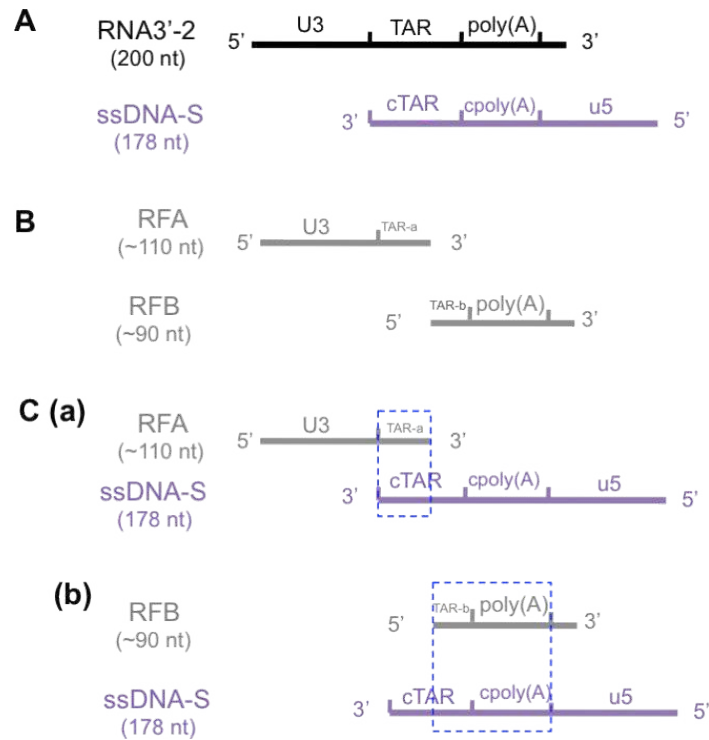
Figure 68. Hypothesis for the formation of heteroduplexes shorter than the full-length heteroduplex of ssDNA-S and RNA 3'-2. (A) RNA 3'-2 (200 nt) and ssDNA-S (178 nt) used in the annealing assays. **(B)** A putative spontaneous cleavage in the apical loop of TAR divides the TAR region into two parts (TAR-a and TAR-b). This cleavage generates RNA fragment A (RFA, ~110 nt) and RNA fragment B (RFB, ~90 nt). **(C)** (a) The heteroduplex (~288 nt) generated by the annealing of ssDNA-S to RFA that is mediated by base pairing of the complementary TAR and cTAR sequences. (b) Another heteroduplex (~268 nt) generated by the annealing of ssDNA-S to RFB that is mediated by base pairing between the (TAR-b)-polyA region and the cTAR-cpoly(A) region. RNA 3'-2 is indicated in black, ssDNA-S in purple, and RNA fragment in gray. The blue rectangle (broken line) indicates the base-pairing region in the heteroduplexes.

In the absence of ssDNA-S and after incubation at 37 °C for various times, most of RNA 3'-2 molecules were in the full-length form but not in the cleaved forms **(Figure 69)**. Nevertheless, there were weak spontaneous cleavages in RNA 3'-2 for incubation times equal to or greater than three hours. As expected, the spontaneous cleavage of RNA 3'-2 was slightly greater in the presence of 2 mM $MgCl_2$ than 0.2 mM $MgCl_2$ **(Figure 69)**. There are two possible explanations for our results. First, the sensitivity of RNA 3'-2 to spontaneous cleavage was increased strongly in the presence of ssDNA-S, i.e. the cleavage of RNA 3'-2 increased with the incubation time, thus facilitating formation of heteroduplexes shorter that the full-length heteroduplex of RNA 3'-2 and ssDNA-S. Note that the cleavage of RNA 3'-2 probably destroyed the stable TAR hairpin. Second, most of RNA 3'-2 molecules remained in the full-length form in the presence of ssDNA-S which annealed exclusively to RNA

fragments generated by the spontaneous cleavage of RNA 3'-2. In summary, formation of the full-length heteroduplex of RNA 3'-2 and ssDNA-S was not observed in the absence of NC.
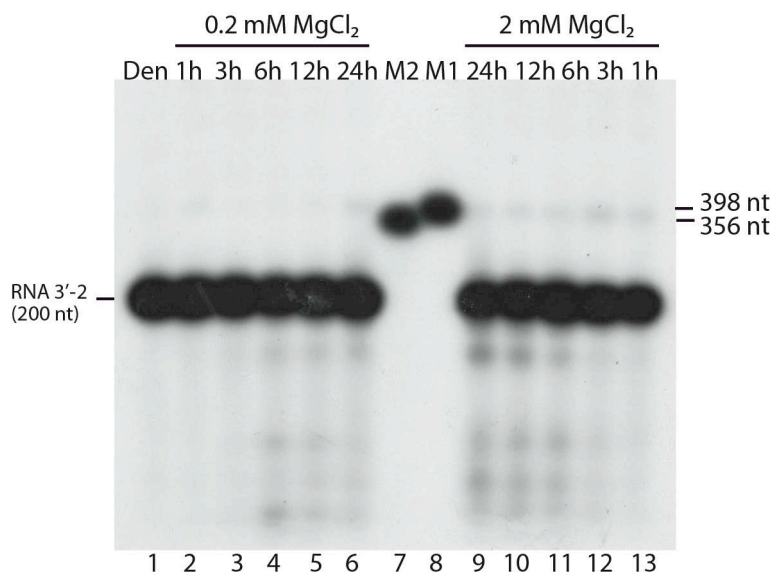
Figure 69. Time course of RNA 3'-2 cleavage in the absence of ssDNA-S. 5' end-labeled RNA 3'-2 was incubated in the annealing buffer (containing 0.2 mM or 2 mM MgCl$_2$) at 37 °C for various times: 1 h (lanes 2 and 13); 3 h (lanes 3 and 12); 6 h (lanes 4 and 11); 12 h (lanes 5 and 10) and 24 h (lanes 6 and 9). Lane Den, heat-denatured 5' end-labeled RNA 3'-2. Lanes M1 and M2, size markers.

4.2 Annealing assays of ssDNAs and RNA 3'-2 in the presence of NC

Consistent with our results mentioned above, annealing of RNA 3'-2 to ssDNAs was barely detected after 15 min incubation in the absence of NC **(Figure 70, lanes C3)**. In contrast, wild-type ssDNA-L and ssDNA-S were annealed to RNA 3'-2 and formed full-length heteroduplexes after 15 min incubation in the presence of NC **(Figure 70A, lanes 4-7; Figure 70B, lanes 3-6)**. Thus, the chaperone activity of NC stimulated strongly the annealing process between ssDNAs and RNA 3'-2. There were no large difference in annealing efficiency between the assays performed in 0.2 mM MgCl$_2$ and those performed in 2 mM MgCl$_2$ **(Figure 70C)**. Note that ssDNA-S was more efficient than ssDNA-L to anneal to RNA 3'-2 in the presence of 2 mM MgCl$_2$. The bands above the full-length heteroduplex may correspond to multimers induced by NC in which the dimeric form of RNA 3'-2 was annealed to one or two ssDNA molecules **(Figure 70 A and B)**. In agreement with this hypothesis, a tiny fraction of RNA 3'-2 molecules dimerized in the absence of NC **(Figure 69**, see the bands at the level of size markers).
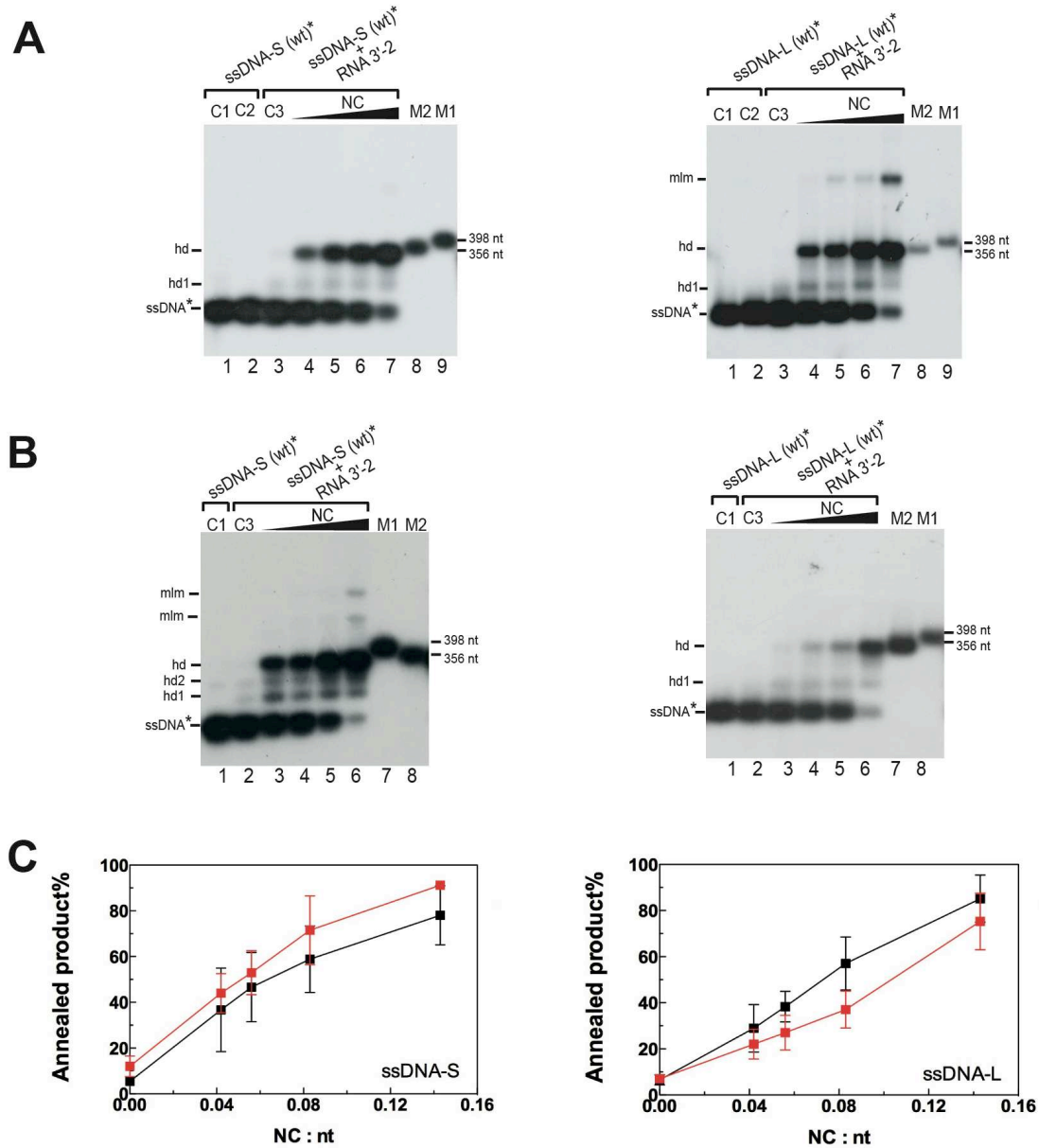
Figure 70. Annealing of wild-type ssDNAs to RNA 3'-2 is facilitated by NC. The NC-mediated annealing assays in the presence of 0.2 mM MgCl$_2$ **(A)** or 2 mM MgCl$_2$ **(B)** were performed as described in the 'Materials and Methods' section. The samples were analyzed by 2% agarose gel electrophoresis at 25 °C in the TBE buffer. Lanes C1, 5' end-labeled ssDNA-S or ssDNA-L was incubated with NC in the absence of RNA 3'-2. Lanes C2, 5' end-labeled ssDNA-S or ssDNA-L was heat-denatured. Lanes C3, 5' end-labeled ssDNA-S or ssDNA-L was incubated with RNA 3'-2 in the absence of NC. Lanes M1 and M2, size markers. mlm: multimer; hd: full-length heteroduplex; hd1 and hd2: heteroduplexes shorter than full-length heteroduplex; ssDNA: 5' end-labeled ssDNA-S or ssDNA-L. **(A)** Lanes 4-7, 5' end-labeled ssDNA-S or ssDNA-L was incubated with RNA 3'-2 in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 4), 1:18 (lanes 5), 1:12 (lanes 6) and 1:7 (lanes 7). **(B)** Lanes 3-6, 5' end-labeled ssDNA-S or ssDNA-L was incubated with RNA 3'-2 in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 3), 1:18 (lanes 4), 1:12 (lanes 5) and 1:7 (lanes 6). **(C)** The graphs show the percentage of 5' end-labeled ssDNA-S or ssDNA-L that was annealed to RNA 3'-2 in the presence of NC. All experiments were repeated at least three times and error bars show standard deviations. Black square: in the presence of 0.2 mM MgCl$_2$; red square: in the presence of 2 mM MgCl$_2$.

132
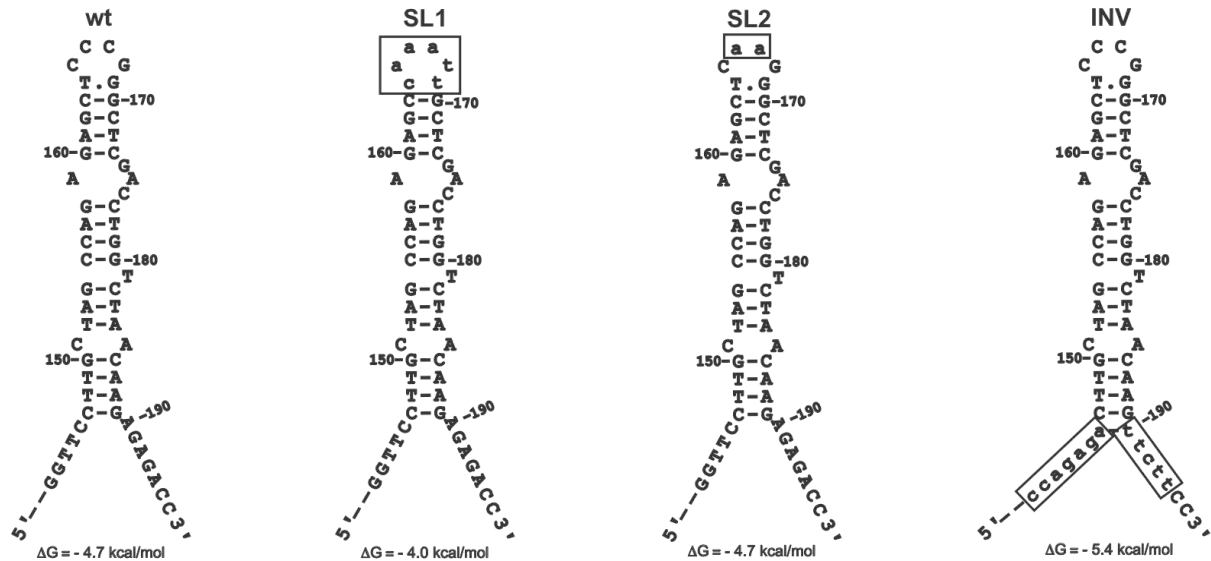
Figure 71. Predicted secondary structures for the cTAR sequence in the wild-type and mutant ssDNA-L and ssDNA-S. Zuker's DNA folding program (423) was used to predict the most stable secondary structure for each cTAR sequence. Numbering is relative to the first nucleotide of ssDNA-L. Mutations are shown as lower case letters in boxes.

To determine the role of the apical loop and the lower stem of the cTAR hairpin in the NC-mediated annealing process, three mutants were investigated **(Figure 71)**. The SL1 mutant was designed such that it forms the stem-loop structure in the cTAR hairpin of ssDNAs, but its apical loop cannot base-pair with the apical loop of the TAR hairpin in RNA 3'-2. The SL2 mutant was designed so that the loop-loop interaction should be prevented by single-base substitutions that introduce two base-pair mismatches. The INV mutant was designed so that the cTAR stem-loop can form only the "Y" conformation and its lower part cannot base-pair with the bottom of the TAR hairpin in RNA 3'-2. The rate of annealing in 0.2 mM $MgCl_2$ and in the presence of NC was not significantly reduced with the ssDNA mutants, compared to the wild-type ssDNAs **(Figures 72C and 73C)**. These results show that NC-mediated annealing of ssDNA to 3' UTR does not rely only on one pathway ('kissing' or 'zipper') in 0.2 mM $MgCl_2$. The results are also consistent with the notion that the NC-mediated annealing process could be initiated through sequences that do not involve the apical loop and the bottom of the TAR hairpin. In the presence of 2 mM $MgCl_2$, the NC-mediated annealing process was not significantly impaired by the INV and SL2 mutations **(Figures 72C and 73C)**. However, it seems that the annealing rate of ssDNA-S (SL2) to RNA 3'-2 was slightly reduced. Interestingly, the rate of annealing in 2 mM $MgCl_2$ and in the presence of NC was on average 1.8-fold reduced with the SL1 mutants, compared to the wild-type ssDNAs **(Figures 72C and 73C)**. These results suggest that about fifty percent of ssDNA molecules annealed to RNA 3'-2 through the 'kissing' pathway in 2 mM $MgCl_2$.

133
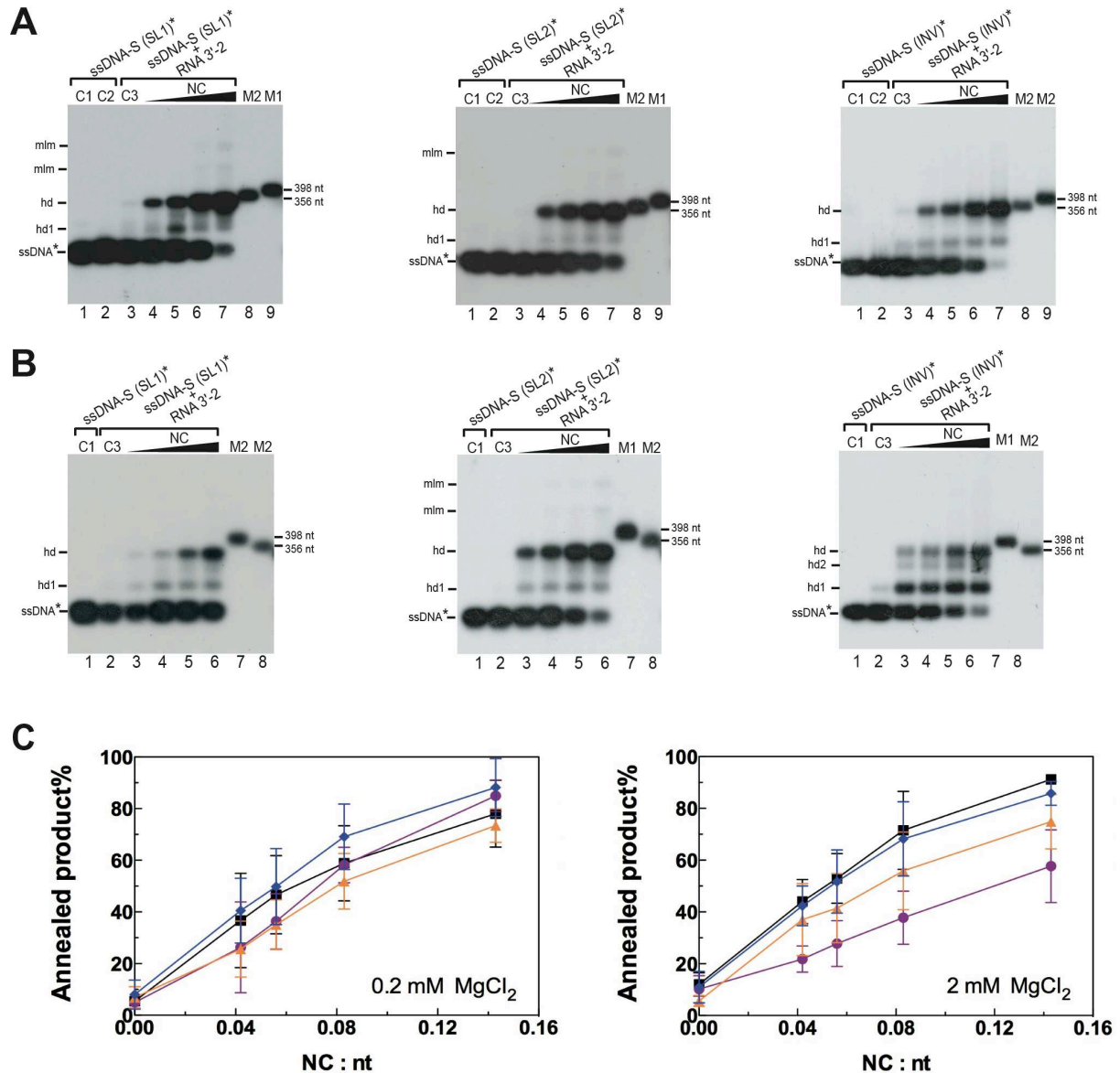
Figure 72. Mutational analysis of the annealing of RNA 3'-2 and ssDNA-S in the presence of NC. The NC-mediated annealing assays were performed in the presence of 0.2 mM MgCl$_2$ **(A)** or 2 mM MgCl$_2$ **(B)** as described in the 'Materials and Methods' section. The samples were analyzed by 2% agarose gel electrophoresis at 25 °C in the TBE buffer. Lanes C1, 5' end-labeled ssDNA-S incubated with NC in the absence of RNA 3'-2. Lanes C2, 5' end-labeled ssDNA-S was heat-denatured. Lanes C3, 5' end-labeled ssDNA-S was incubated with RNA 3'-2 in the absence of NC. Lanes M1 and M2, size markers. mlm: multimer; hd: full-length heteroduplex; hd1 and hd2: heteroduplexes shorter than the full-length heteroduplex; ssDNA: 5' end-labeled ssDNA-S. **(A)** Lanes 4-7, 5' end-labeled ssDNA-S and RNA 3'-2 were incubated in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 4), 1:18 (lanes 5), 1:12 (lanes 6) and 1:7 (lanes 7). **(B)** Lanes 3-6, 5' end-labeled ssDNA-S and RNA 3'-2 were incubated in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 3), 1:18 (lanes 4), 1:12 (lanes 5) and 1:7 (lanes 6). **(C)** The percentage of 5' end-labeled ssDNA-S that was annealed to RNA 3'-2 in the presence of NC. All experiments were repeated at least three times and error bars show standard deviations. Left graph: 0.2 mM MgCl$_2$; right graph: 2 mM MgCl$_2$. Black square: wild type; purple circle: SL1; yellow triangle: SL2; blue diamond: INV.

134
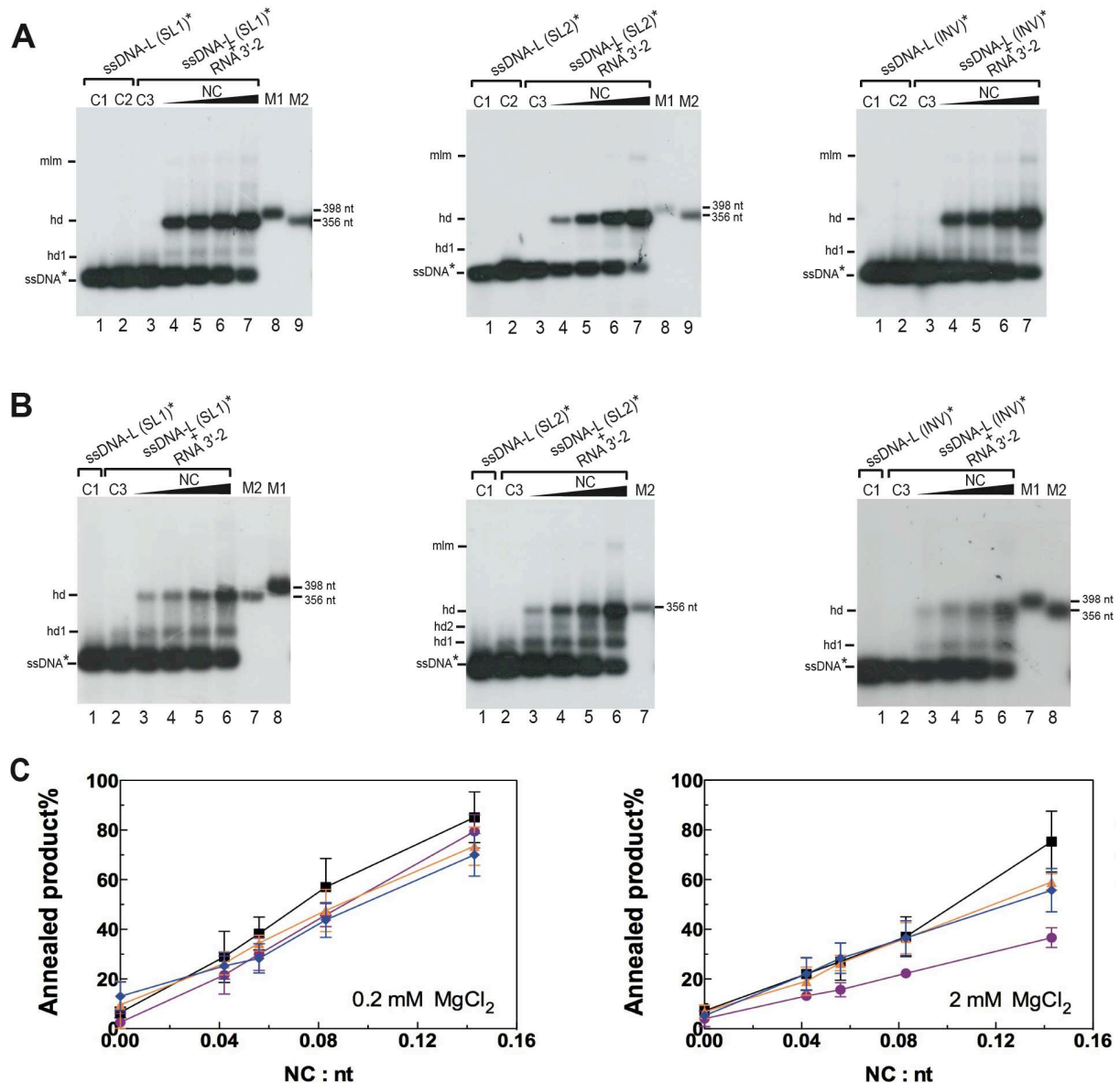
Figure 73. Mutational analysis of the annealing of RNA 3'-2 and ssDNA-L in the presence of NC.
The NC-mediated annealing assays were performed in the presence of 0.2 mM $MgCl_2$ **(A)** or 2 mM $MgCl_2$ **(B)** as described in the 'Materials and Methods' section. The samples were analyzed by 2% agarose gel electrophoresis at 25 °C in the TBE buffer. Lanes C1, 5' end-labeled ssDNA-L incubated with NC in the absence of RNA 3'-2. Lanes C2, 5' end-labeled ssDNA-L was heat-denatured. Lanes C3, 5' end-labeled ssDNA-L incubated with RNA 3'-2 in the absence of NC. mlm: multimer; hd: full-length heteroduplex; hd1 and hd2: heteroduplexes shorter than the full-length heteroduplex; ssDNA: 5' end-labeled ssDNA-L. Lanes M1 and M2, size markers. **(A)** Lanes 4-7, 5' end-labeled ssDNA-S and RNA 3'-2 were incubated in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 4), 1:18 (lanes 5), 1:12 (lanes 6) and 1:7 (lanes 7). **(B)** Lanes 3-6, 5' end-labeled ssDNA-L and RNA 3'-2 were incubated in the presence of NC. The protein to nucleotide molar ratios were 1:24 (lanes 3), 1:18 (lanes 4), 1:12 (lanes 5) and 1:7 (lanes 6). **(C)** The percentage of 5' end-labeled ssDNA-L that was annealed to RNA 3'-2 in the presence of NC. All experiments were repeated at least three times and error bars show standard deviations. Left graph: 0.2 mM $MgCl_2$; right graph: 2 mM $MgCl_2$. Black square: wild type; purple circle: SL1; yellow triangle: SL2; blue diamond: INV.

It seems suprising that the SL2 mutation did not efficiently inhibit heteroduplex formation by introducing two adjacent destabilizing A.G mismatches in the loop-loop

complex. A possible explanation is that these mismatches are stablized by the flanking G-C base-pairs, i.e. a transient loop-loop interaction could be recognized by NC and sufficient to initiate heteroduplex formation.

In summary, NC-mediated annealing of ssDNA to 3' UTR can initiate through different regions of the R sequence in the presence of 0.2 mM $MgCl_2$ (intracellular concentration). We do not know the Mg^{2+} concentration in the partially disassembled virus particles that are in the infected cells. Therefore, we cannot exclude the possibility that this concentration is locally greater than 0.2 mM. Our results suggest that the 'kissing' pathway plays an important role in the NC-mediated annealing of ssDNA to 3' UTR if this process occurs in the presence of 2 mM $MgCl_2$ (concentration required for reverse transcription and strand transfer *in vitro*). It is possible that ssDNA-S and ssDNA-L fold differently in 2 mM $MgCl_2$, since ssDNA-S was more efficient than ssDNA-L to anneal to RNA 3'-2 **(Figure 70C)**. Thus, it is important to investigate the conformations of ssDNA-S and ssDNA-L to gain insight into the mechanism of NC-mediated ssDNA-3' UTR annealing.

5. <u>Conformation analysis of ssDNAs</u>

Two conformers of an RNA molecule can be separated by non-denaturing polyacrylamide gel electrophoresis (214). Furthermore, single-strand conformation polymorphism (SSCP) showed that different conformations of a DNA molecule can be characterized by non-denaturing polyacrylamide gel electrophoresis (343, 352). Consistent with the analysis of ssDNAs by native agarose gel electrophoresis in a TBM buffer **(Figure 66)**, ssDNAs were monomeric after incubation with or without NC and analysed by non-denaturing polyacrylamide gel electrophoresis in the TBM buffers **(Figure 74)**.

Interestingly, in the presence of 0.2 mM $MgCl_2$ in the incubation and electrophoresis buffers, the 5' end-labeled ssDNAs separated into two distinct bands in the absence or presence of NC **(Figure 74 A and C)**. These observations suggest that ssDNAs adopted two different conformations in the presence of 0.2 mM $MgCl_2$. PhosphorImager quantification indicated that 60% of ssDNA molecules folded into the m1 conformation. In contrast, only one band was observed in the presence of 2 mM $MgCl_2$ in the incubation and electrophoresis buffers **(Figure 74 B and D)**. We also observed only one band when the assays were analyzed by non-denaturing polyacrylamide gel electrophoresis in the TBE buffer **(Figure 75)**. These observations suggest that the conformer identified by the TBM-2 gel required magnesium to be formed, whereas the conformer identified by the TBE gel did not require this divalent cation to be formed. It is likely that at low magnesium concentrations as 0.2 mM,

there were the two types of conformers. Determination of secondary structures of ssDNAs may allow characterizing the folding of conformers.

Figure 74. Analysis of conformers by non-denaturing PAGE in TBM buffers. 5' end-labeled ssDNA-L **(A and B)** and ssDNA-S **(C and D)** in the absence (lanes C1 and C2) or in the presence of NC (lanes 4-6 and 10-12) were incubated at 37 °C in the presence of 0.2 mM MgCl$_2$ **(A and C)** or 2 mM MgCl$_2$ **(B and D)** as described in the 'Materials and Methods' section. The samples were analyzed by 6% polyacrylamide gel electrophoresis at 20°C in the TBM-0.2 or TBM-2 buffer. The protein to nucleotide molar ratios were 1:18 (lanes 4 and 10), 1:12 (lanes 5 and 11) and 1:7 (lanes 6 and 12). Lanes Den, heat-denatured ssDNAs. Lanes C1 are controls without phenol-chloroform extraction and ethanol precipitation. Lanes C2 are controls with phenol-chloroform extraction and ethanol precipitation. Lanes 7-12 correspond to a short-exposure autoradiography of lanes 1-6. Monomeric forms of ssDNAs are indicated by m. Note that two monomeric conformations of ssDNAs (m1 and m2) can be separated by electrophoresis in the TBM-0.2 buffer.
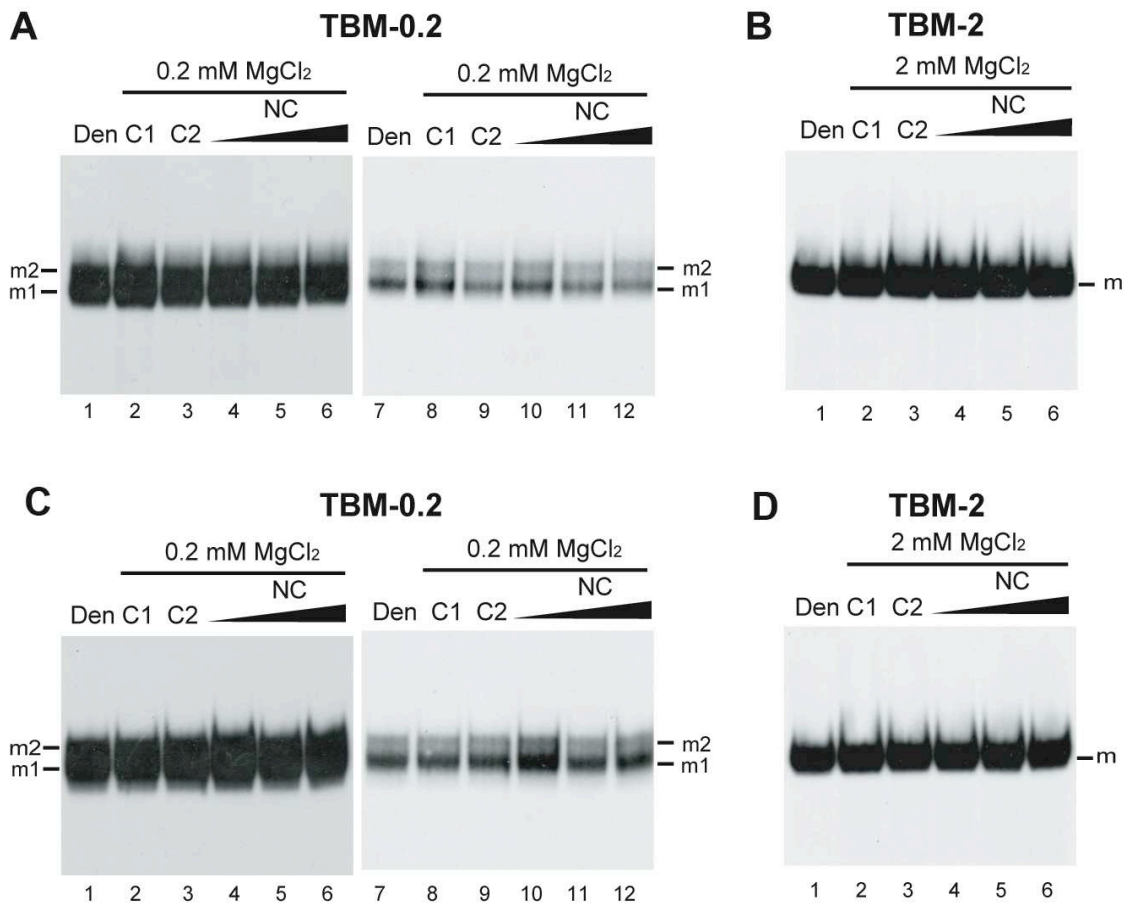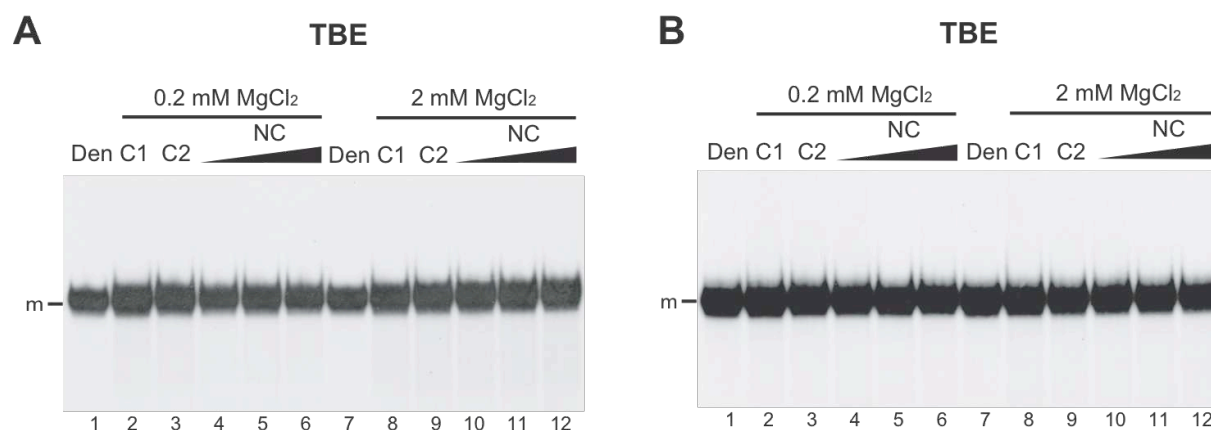
Figure 75. Analysis of conformers by non-denaturing PAGE in the TBE buffer. 5' end-labeled ssDNA-L **(A)** and ssDNA-S **(B)** in the absence (lanes C1 and C2) or in the presence of NC (lanes 4-6 and 10-12) were incubated at 37 °C in the presence of 0.2 or 2 mM $MgCl_2$ as described in the 'Materials and Methods' section. The samples were analyzed by 6% polyacrylamide gel electrophoresis at 25 °C in TBE. Monomeric forms of ssDNAs are indicated by m. The protein to nucleotide molar ratios were 1:18 (lanes 4 and 10), 1:12 (lanes 5 and 11) and 1:7 (lanes 6 and 12). Lanes Den, heat-denatured ssDNAs. Lanes C1 are controls without phenol-chloroform extraction and ethanol precipitation. Lanes C2 are controls with phenol-chloroform extraction and ethanol precipitation.

6. <u>Structural analysis of ssDNAs</u>

Using potassium permanganate ($KMnO_4$), mung bean nuclease (MB) and DNase I, we determined the secondary structures of the two ssDNAs and the NC binding sites within these DNAs. Note that due to the size of DNase I and MB, the absence of cuts by these enzymes may also be a result of steric hindrance. As mentioned previously, the nucleases cleave the phosphodiester bond and generate a 3'-hydroxyl terminus in the 5' end-labeled DNA. In contrast, the Maxam-Gilbert reactions generate a 3'-phosphorylated terminus in the 5' end-labeled DNA. The electrophoretic mobility of Maxam-Gilbert sequence markers is therefore slightly greater than that of fragments produced by nucleases. Short and long migrations were used to identify the cleavages sites (examples in **Figures 76 and 87**). To determine the precise position of cleavage, the 5' and 3' end-labeled ssDNA were used (examples in **Figures 81 and 87**). Only the cleavage sites that were observed with the two types of labeling (5' and 3' ends) were taken into account to build the secondary structure models of ssDNAs. To generate the secondary structure models, we selected the foldings predicted by the Mfold program (423), which were the most consistent with the experimental data. To facilitate comparisons between the ssDNA-S and ssDNA-L, we used the same numbering for nucleotides. Consistent with the probing data and the results of the conformation assays, two conformations of ssDNAs in the presence of 0.2 mM $MgCl_2$ **(Figures 79 and 83)** and one conformation in the presence of 2 mM $MgCl_2$ **(Figures 80 and 84)** are proposed.

6.1 Structural analysis of ssDNAs in the absence of NC

Representative examples of probing experiments are shown in **Figures 76-78** (ssDNA-S) and **81-82** (ssDNA-L) and the results of a series of independent experiments are summarized in **Figures 79-80** (ssDNA-S) **and 82-83** (ssDNA-L).

6.1.1 Probing of ssDNA-S
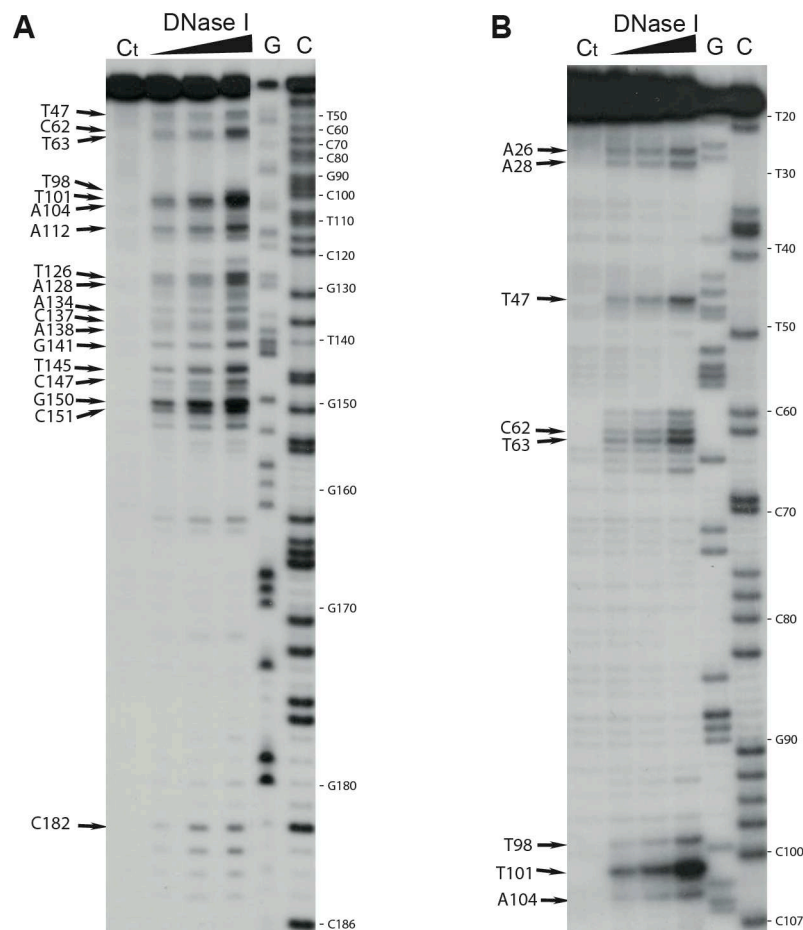
Figure 76. DNase I probing of 3' end-labeled ssDNA-S in the presence of 2 mM MgCl$_2$. DNase I probing experiments were performed as described in the 'Materials and Methods' section. **(A)** Short migration on an 8% denaturing polyacrylamide gel. **(B)** Long migration on a 6% denaturing polyacrylamide gel. Lanes C$_t$ are controls without DNase I. The 3' end-labeled ssDNA-S was incubated with DNase I (0.2, 0.3 and 0.4 U). G and C refer to Maxam-Gilbert sequence markers. Arrows indicate the medium and strong DNase I cleavage sites.

The DNase I and MB cleavage patterns **(Figures 76 and 77)** were consistent with the secondary structures proposed for ssDNA-S in 0.2 mM and 2 mM MgCl$_2$ **(Figures 79 and 80)**. Indeed, most of the moderate and strong DNase I cleavages occured within stems or at the ends of stems, whereas most of moderate and strong MB cleavages occured within unpaired regions. Consistent with the formation of stems, T$_{23}$, T$_{61}$, T$_{63}$, T$_{98}$, T$_{101}$, T$_{102}$, T$_{110}$, T$_{140}$, T$_{172}$ and T$_{178}$ were unreactive to KMnO$_4$ and the sensitivity of T$_{59}$, T$_{126}$, T$_{132}$, T$_{133}$ and

T_{183} to $KMnO_4$ were low **(Figure 78)**. In contrast, the thymine residues in the loops displayed high and moderate sensitivities to $KMnO_4$. The high sensitivity of T_{181} **(Figure 78)** indicates that at least one side of the plane of the heterocyclic ring of the thymine residue is exposed. This suggests that T_{181} adopts the looped-out bulge conformation **(Figures 79 and 80)**.
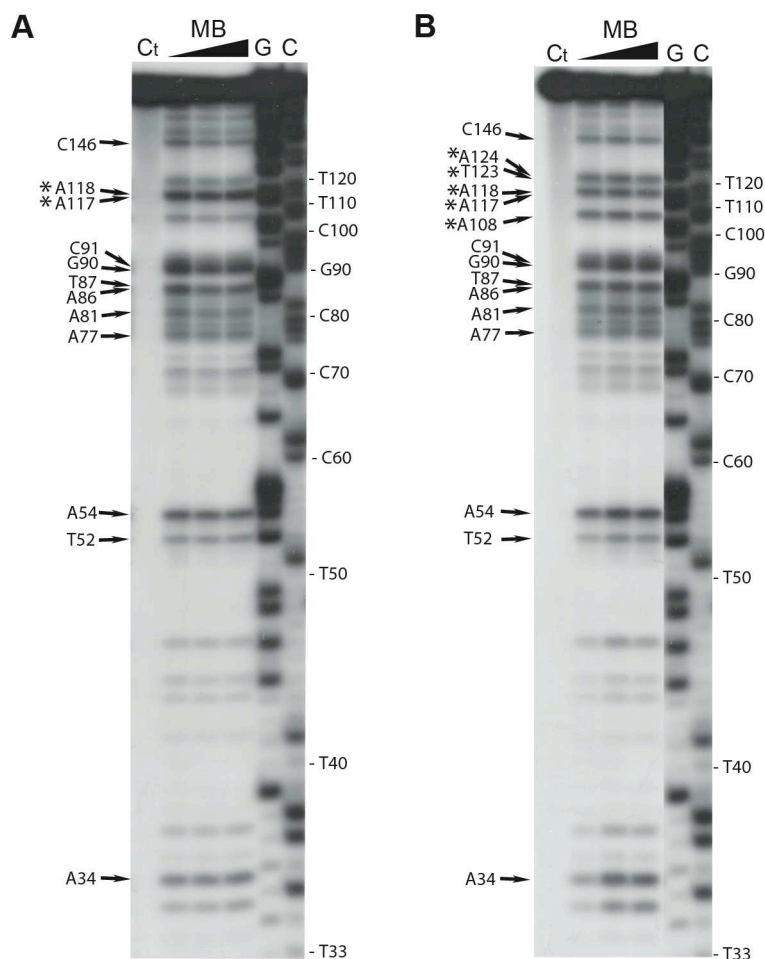
Figure 77. MB probing of 5' end-labeled ssDNA-S in the presence of 0.2 mM MgCl$_2$ (A) or 2 mM MgCl$_2$ (B). MB probing experiments were performed as described in the 'Materials and Methods' section. The 5' end-labeled ssDNA-S was incubated with MB (0.75, 1 and 1.25 U). Lanes C_t are controls without MB. G and C refer to Maxam-Gilbert sequence markers. Arrows indicate the medium and strong MB cleavage sites. The differences in cleavage rate between 0.2 mM and 2 mM MgCl$_2$ are indicated by asterisks.

We did not find significant differences between the cleavage patterns generated by DNase I in 0.2 mM and 2 mM MgCl$_2$ (data not shown). The cleavage patterns generated by MB in 0.2 mM and 2 mM MgCl$_2$ were similar but not identical **(Figure 77)**. Indeed, the rate of MB cleavage between A_{117}-A_{118} and A_{118}-G_{119} was greater in 0.2 mM MgCl$_2$ than in 2 mM MgCl$_2$. In contrast, the rate of MB cleavage between A_{108}-C_{109}, T_{123}-A_{124} and A_{124}-T_{125} was greater in 2 mM MgCl$_2$ than in 0.2 mM MgCl$_2$. These cleavage differences were restricted to the upper part of a long stem-loop structure **(Figures 79 and 80)** and probably resulted from

differences in the accessibility of this stem-loop to MB. It is likely that the three-dimensional folding of ssDNA-S is modulated by the magnesium ions. The $KMnO_4$ probing patterns of ssDNA-S in 0.2 mM and 2 mM $MgCl_2$ exhibited differences for T_{148}, T_{149} and T_{152} **(Figures 78-80)**. Interestingly, the differences in the reactivity of T_{148}, T_{149} and T_{152} to $KMnO_4$ between 0.2 and 2 mM $MgCl_2$ **(Figure 78)** are consistent with a dynamic structure of the cTAR hairpin in ssDNA-S. Our $KMnO_4$ probing data are consistent with the equilibrium between two conformations of the cTAR hairpin in 0.2 mM $MgCl_2$ **(Figure 78)**. Our data also suggest that the cTAR sequence folds into the partially closed conformation in the majority of ssDNA-S molecules that are in the presence of 2 mM $MgCl_2$.
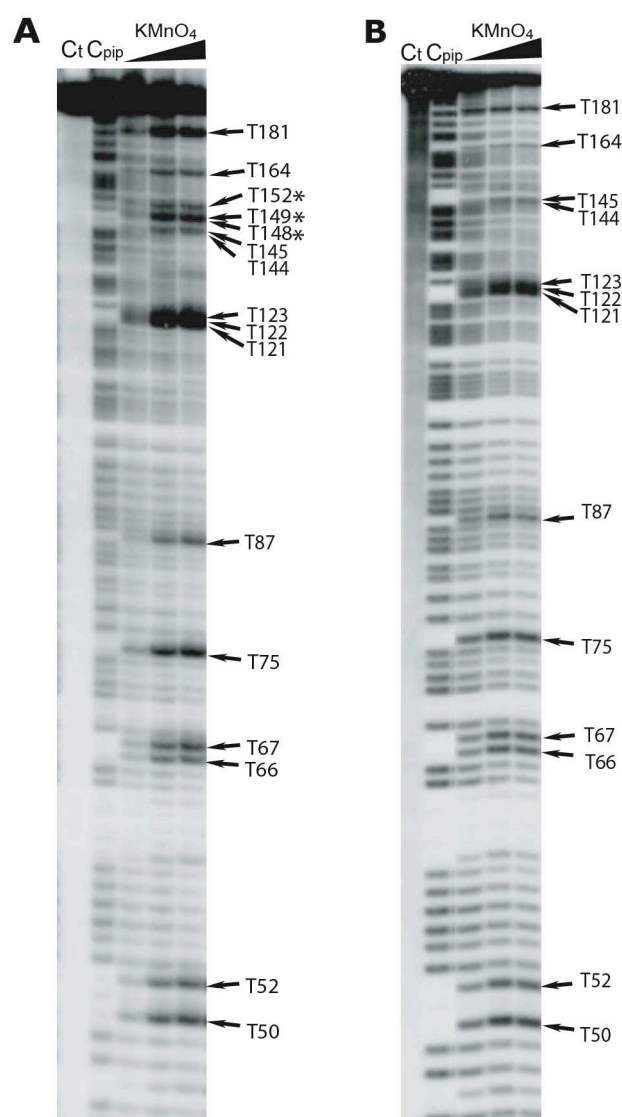
Figure 78. $KMnO_4$ probing of 5' end-labeled ssDNA-S. $KMnO_4$ probing experiments were performed in the presence of 0.2 mM $MgCl_2$ **(A)** or in the presence of 2 mM $MgCl_2$ **(B)** as described in the 'Materials and Methods' section. Lanes C_t are controls without any chemical treatment. Lanes C_{pip} are controls without $KMnO_4$ treatment but with piperidine treatment. The 5' end-labeled ssDNA-S was incubated with $KMnO_4$ (0.5, 1 and 2 mM). Arrows indicate the medium and strong $KMnO_4$ reactive thymine residues. The difference in $KMnO_4$ sensitivity between 0.2 mM and 2 mM $MgCl_2$ are indicated by asterisks.
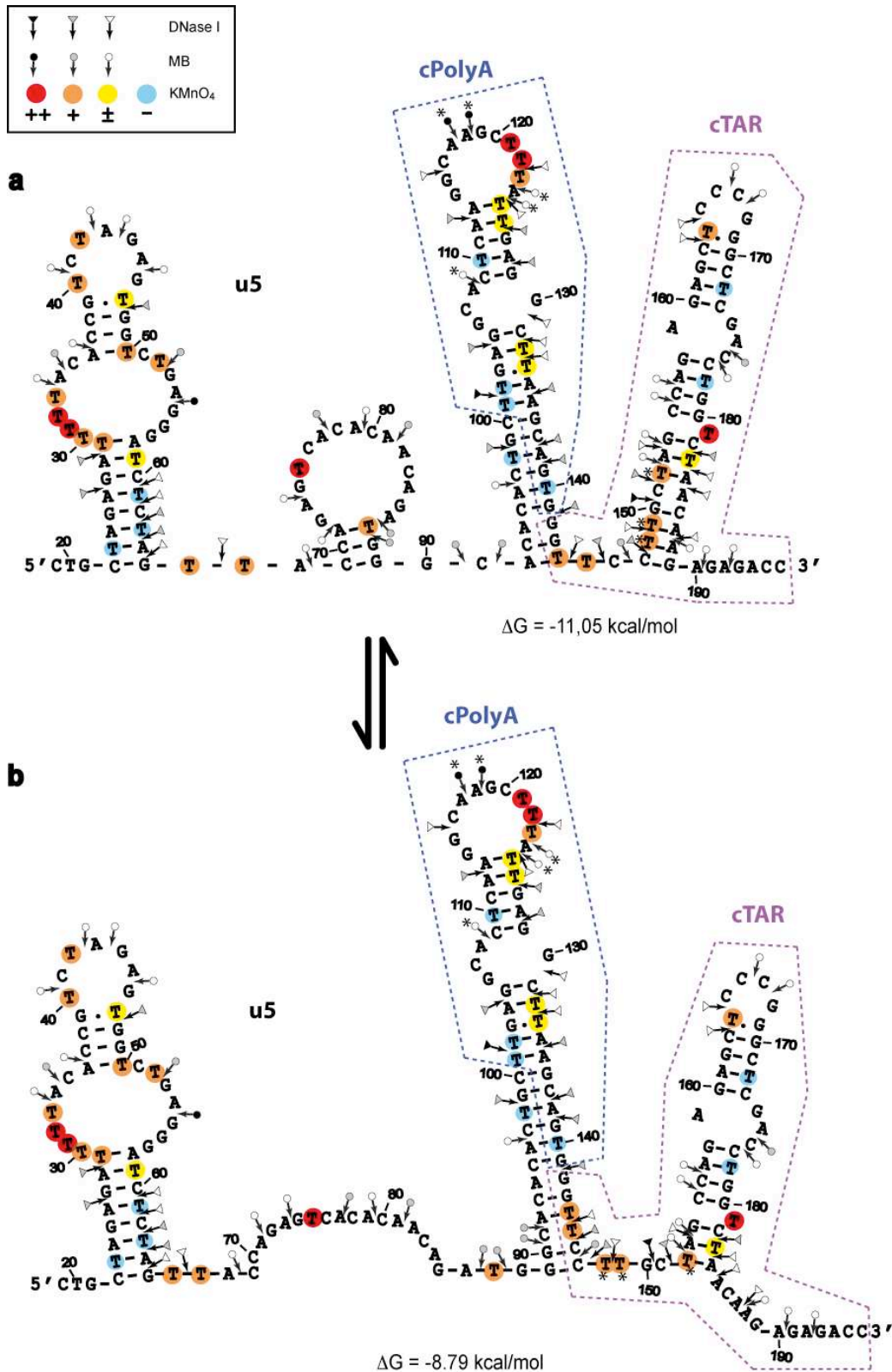
Figure 79. Secondary structure models of ssDNA-S in the presence of 0.2 mM MgCl₂. ssDNA-S may adopt two conformations **(a and b)**. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for MB). The color codes used for the reactivity of thymine residues are indicated in the inset. The differences in KMnO₄ sensitivity and in the cleavage rate between 0.2 mM and 2 mM MgCl₂ are indicated by asterisks. Numbering of ssDNA-S is relative to the first nucleotide of ssDNA-L.

142
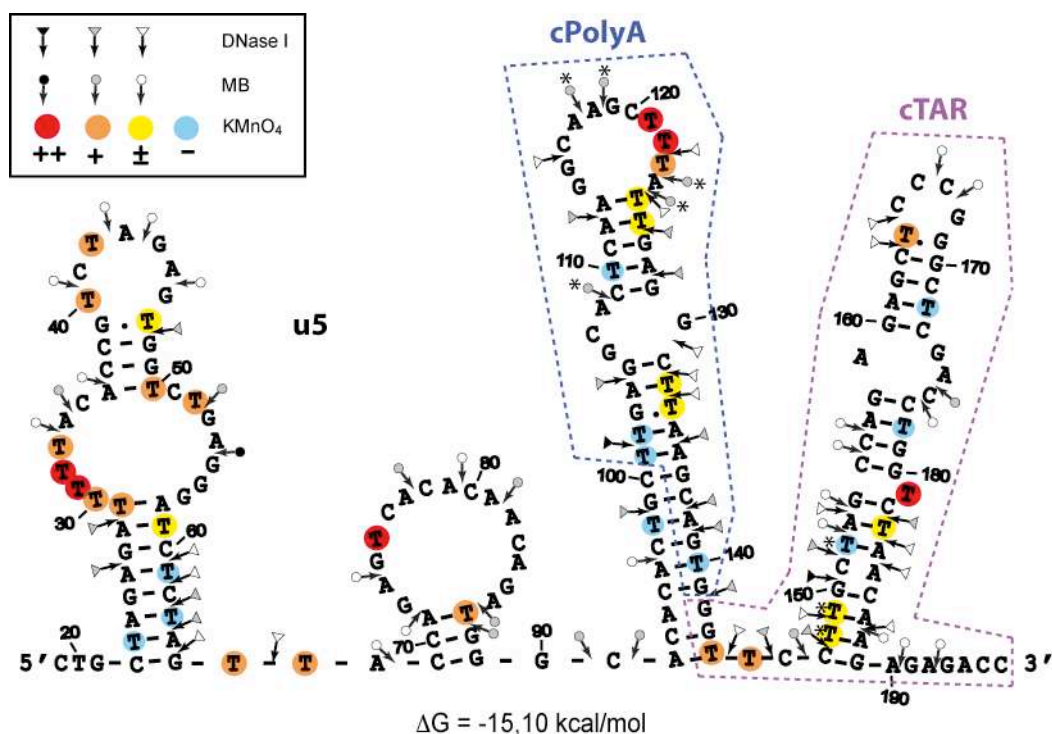
Figure 80. Secondary structure model of ssDNA-S in the presence of 2 mM MgCl$_2$. ssDNA-S forms a stable conformation. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The differences in KMnO$_4$ sensitivity and in the cleavage rate between 0.2 mM and 2 mM MgCl$_2$ are indicated by asterisks. Numbering of ssDNA-S is relative to the first nucleotide of ssDNA-L.

6.1.2 Probing of ssDNA-L

The DNase I and MB cleavage patterns **(Figures 81 and 87)** were consistent with the secondary structures proposed for ssDNA-L in 0.2 mM and 2 mM MgCl$_2$ **(Figures 83 and 84)**. Indeed, most of the moderate and strong DNase I cleavages occured within stems or at the ends of stems, whereas most of moderate and strong MB cleavages occured within unpaired regions. Consistent with the formation of stems, T$_{47}$, T$_{61}$, T$_{63}$, T$_{98}$, T$_{101}$, T$_{102}$, T$_{110}$, T$_{126}$, T$_{140}$, T$_{172}$ and T$_{178}$ were unreactive to KMnO$_4$ and the sensitivity of T$_{23}$, T$_{132}$, T$_{133}$ and T$_{183}$ to KMnO$_4$ were low **(Figure 82)**. In contrast, the thymine residues in the loops displayed high and moderate sensitivities to KMnO$_4$. The high sensitivity of T$_{181}$ **(Figure 82)** indicates that at least one side of the plane of the heterocyclic ring of the thymine residue is exposed. This suggests that T$_{181}$ adopts the looped-out bulge conformation **(Figures 83 and 84)**. Surprisingly, T$_{42}$ displayed a moderate sensitivity to KMnO$_4$, although it is predicted to lie in a stem. T$_{42}$ could be in a distorted DNA helix, since KMnO$_4$ also allows to detect this type of structure (72, 174).

143
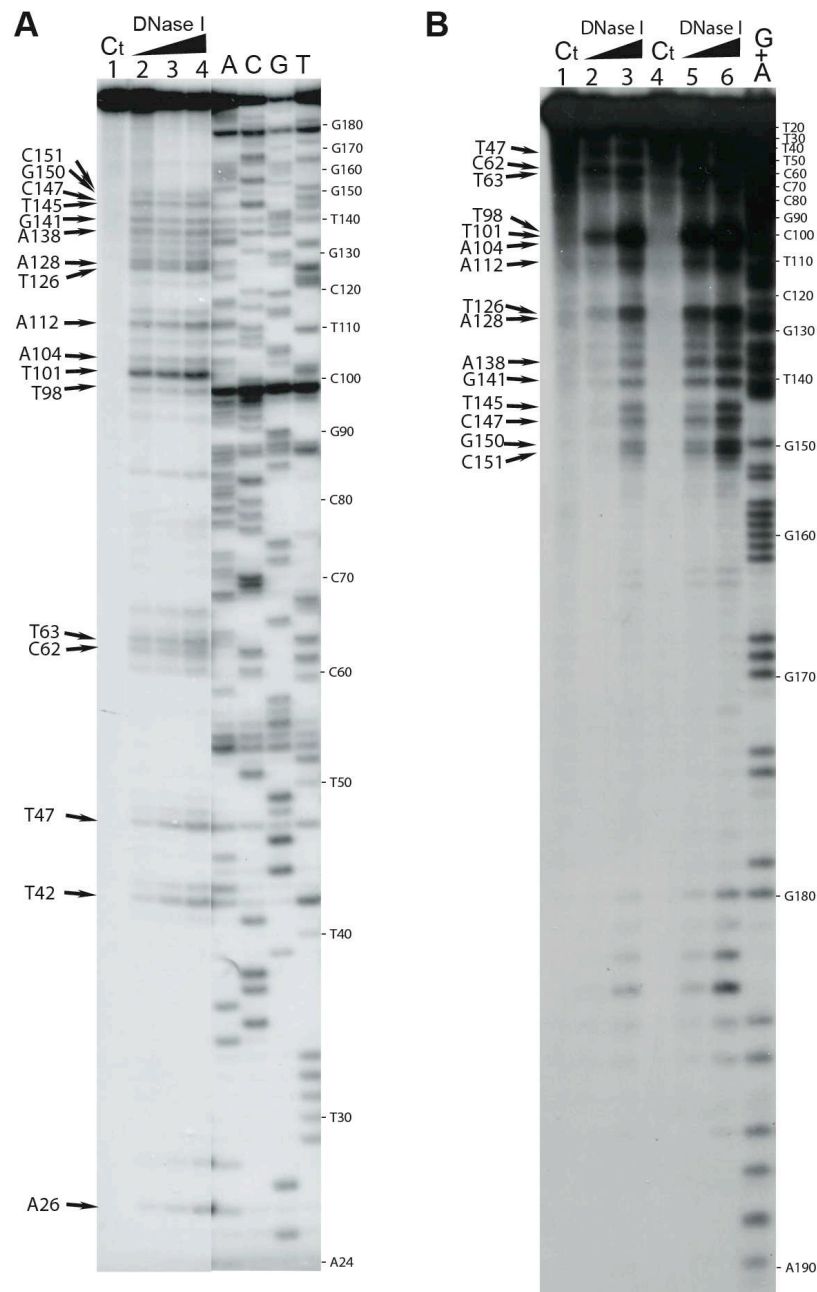
Figure 81. Enzymatic probing of ssDNA-L. Enzymatic probing experiments were performed as described in the 'Materials and Methods' section. **(A)** In the presence of 0.2 mM MgCl$_2$, the 5' end-labeled ssDNA-L was incubated with DNase I (0.2, 0.3 and 0.35 U). Lane C$_t$ is the control without DNase I. Sequence lanes (A, C, G and T) were run in parallel. **(B)** The 3' end-labeled ssDNA-L was incubated with DNase I (0.2 mM: 0.12 and 0.15 U; 2 mM: 0.05 and 0.08 U). Lanes Ct are controls without enzymes. Lanes 1 to 3, assays performed in the presence of 0.2 mM MgCl$_2$. Lanes 4-6, assays performed in the presence of 2 mM MgCl$_2$. G+A refers to Maxam-Gilbert sequence markers. Arrows indicate the medium and strong cleavage sites.

No significant difference was found between the cleavage patterns generated by DNase I in 0.2 mM and 2 mM MgCl$_2$ **(Figure 81)**. The cleavage patterns generated by MB in 0.2 mM and 2 mM MgCl$_2$ were similar but not identical (data not shown). Indeed, the rate of MB cleavage between C$_{146}$-C$_{147}$ and A$_{166}$-C$_{167}$ was greater in 2 mM MgCl$_2$ than in 0.2 mM MgCl$_2$.

144

In contrast, the rate of MB cleavage between C_{175}-C_{176} was greater in 0.2 mM $MgCl_2$ than in 2 mM $MgCl_2$. These cleavage differences were restricted to the cTAR stem-loop structure **(Figures 83 and 84)** and probably resulted from differences in the accessibility of this stem-loop to MB. It is likely that the three-dimensional folding of ssDNA-L is also modulated by the magnesium ions.
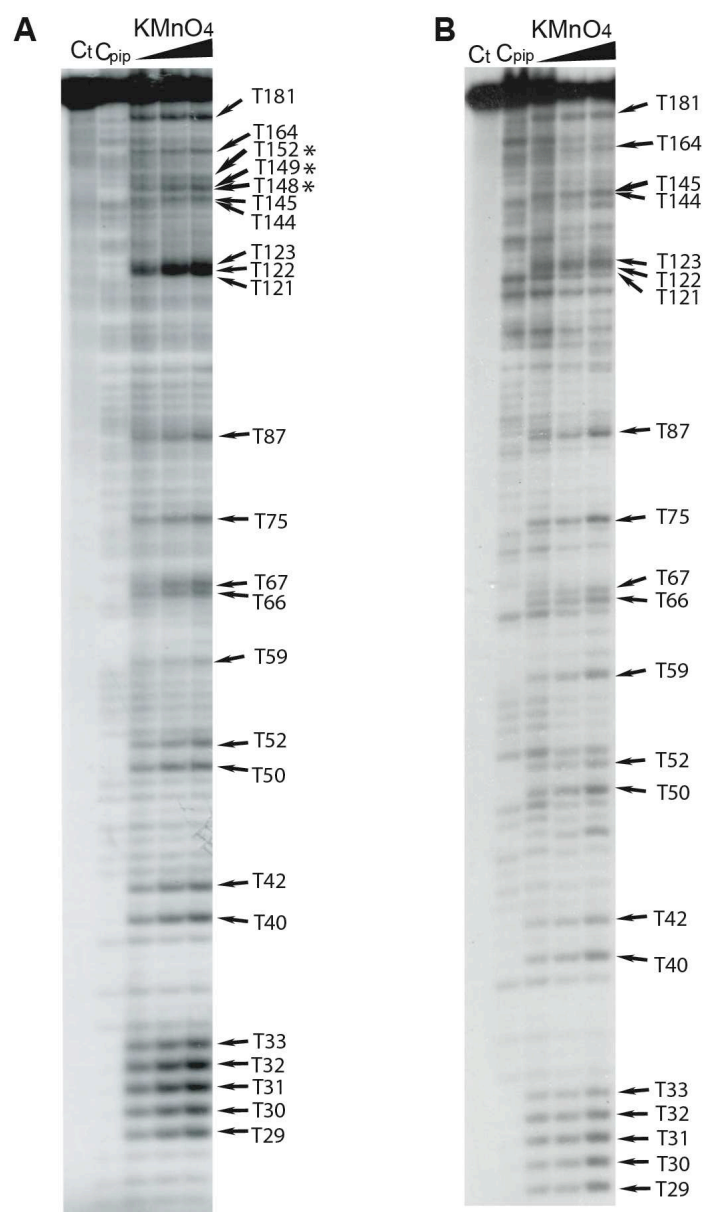
Figure 82. KMnO$_4$ probing of 5' end-labeled ssDNA-L. KMnO$_4$ probing experiments were performed in the presence of 0.2 mM MgCl$_2$ **(A)** or in the presence of 2 mM MgCl$_2$ **(B)** as described in the 'Materials and Methods' section. Lanes C$_t$ are controls without any chemical treatment. Lanes C$_{pip}$ are controls without KMnO$_4$ treatment but with piperidine treatment. The 5' end-labeled ssDNA-L was incubated with KMnO$_4$ (0.5, 1 and 2 mM). Arrows indicate the medium and strong KMnO$_4$ reactive thymine residues. The differences in KMnO$_4$ sensitivity between 0.2 mM and 2 mM MgCl$_2$ are indicated by asterisks.
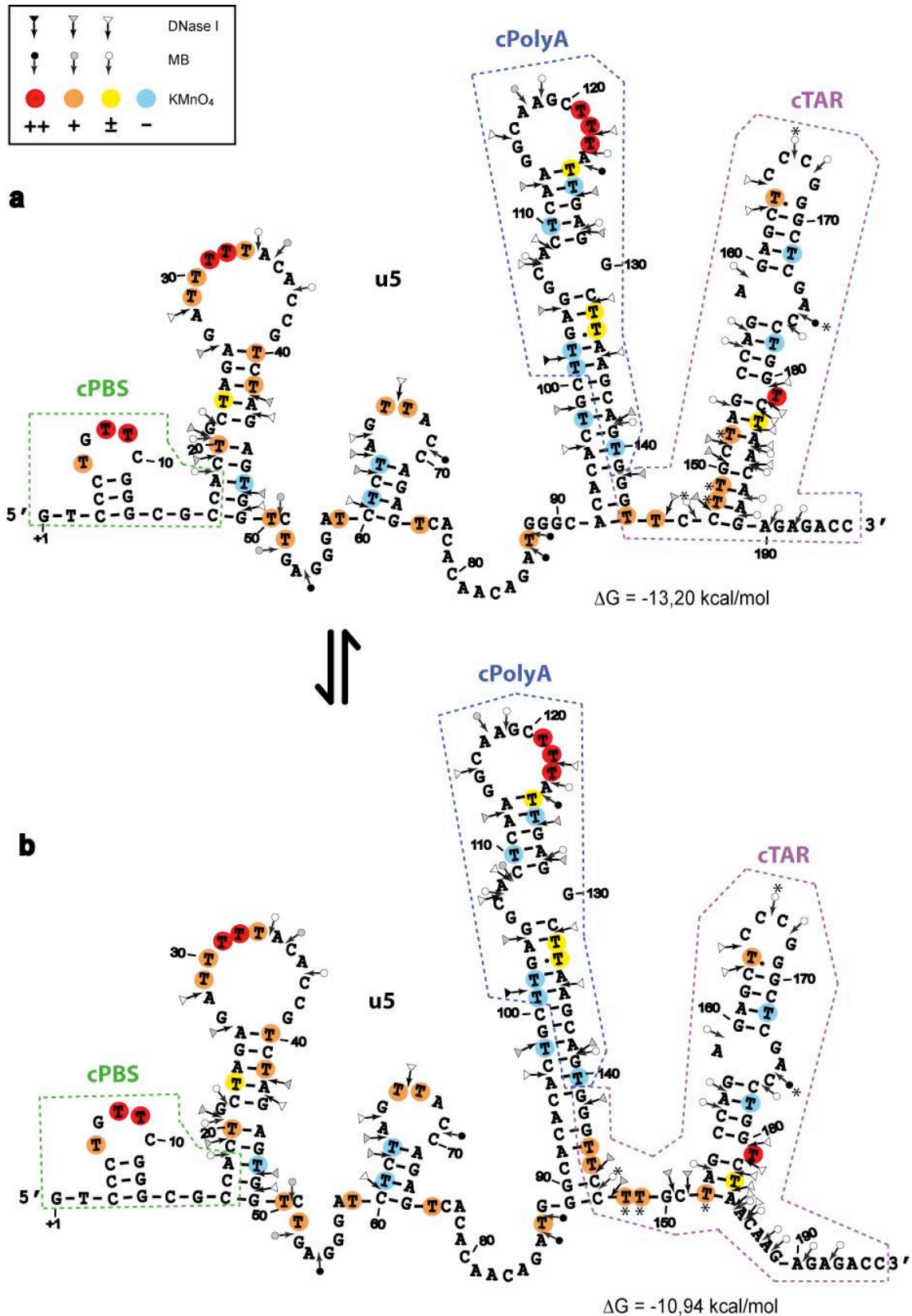
Figure 83. Secondary structure models of ssDNA-L in the presence of 0.2 mM MgCl₂. ssDNA-L may adopt two conformations **(a and b)**. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The differences in KMnO₄ sensitivity and in the cleavage rate between 0.2 mM and 2 mM MgCl₂ are indicated by asterisks.

146

$\Delta G = -17,34$ kcal/mol
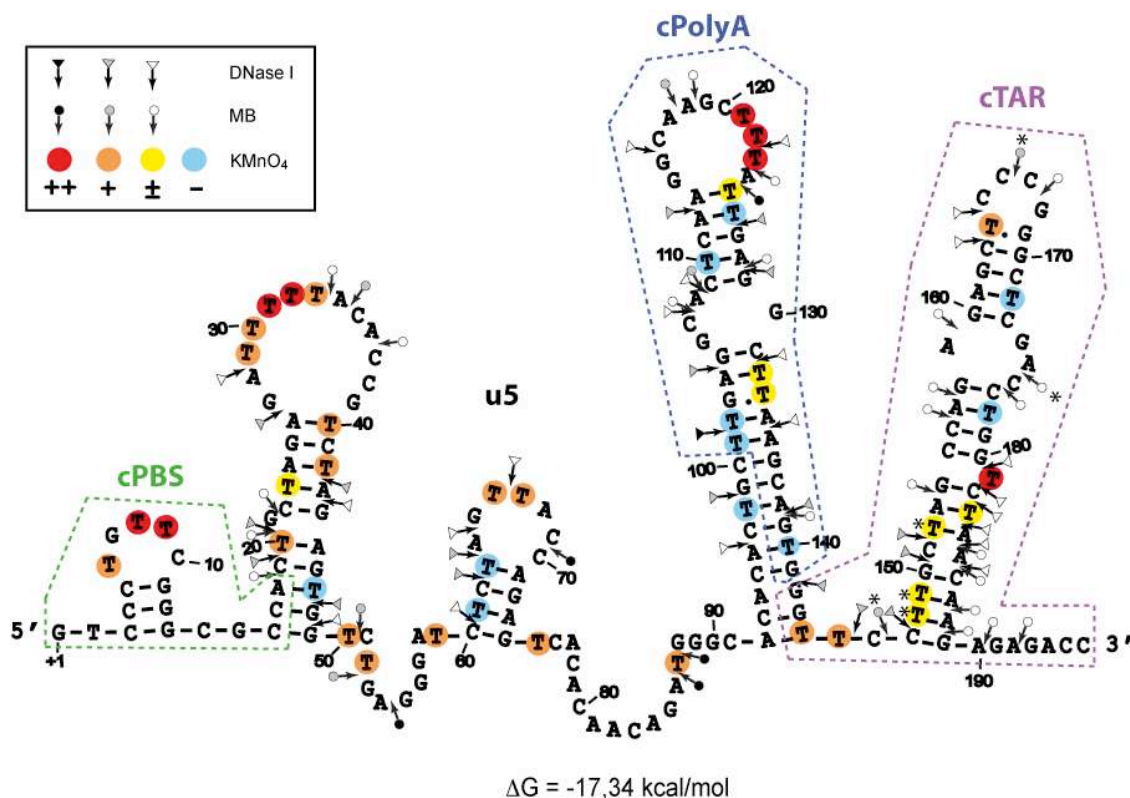
Figure 84. Secondary structure model of ssDNA-L in the presence of 2 mM MgCl$_2$. ssDNA-L forms a stable conformation. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The differences in KMnO$_4$ sensitivity and in the cleavage rate between 0.2 mM and 2 mM MgCl$_2$ are indicated by asterisks.

The KMnO$_4$ probing patterns of ssDNA-L in 0.2 mM and 2 mM MgCl$_2$ exhibited differences for T$_{148}$, T$_{149}$ and T$_{152}$ **(Figures 82-84)**. Interestingly, the differences in the reactivity of T$_{148}$, T$_{149}$ and T$_{152}$ to KMnO$_4$ between 0.2 and 2 mM MgCl$_2$ **(Figure 78)** are consistent with a dynamic structure of the cTAR hairpin in ssDNA-L. Our KMnO$_4$ probing data support the equilibrium between two conformations of the cTAR hairpin in 0.2 mM MgCl$_2$ **(Figure 78)**. Our data also suggest that the cTAR sequence folds into the partially closed conformation in the majority of ssDNA-L molecules that are in the presence of 2 mM MgCl$_2$.

6.1.3 Comparison of secondary structures adopted by the two ssDNAs

Probing data show that there is no difference between ssDNA-S and ssDNA-L for the folding of the r region (nucleotides 101-196) **(Figures 79, 80, 83 and 84)**. In contrast, there are differences between ssDNA-S and ssDNA-L for the folding of the 19-89 sequence (a large part of the u5 region), which is present in both ssDNAs. These differences are due to nucleotides 16-18 in ssDNA-L that allow nucleotides 16-49 to fold to into the stem-loop

containing a single-nucleotide bulge. The 22-49 sequence in ssDNA-S is involved in a stem-loop containing a large internal loop that is less stable (ΔG = -1,09 kcal/mol) than the stem-loop (ΔG = -1,87 kcal/mol) described above.

6.1.4 *Probing of ssDNA in the reverse transcription mixture*
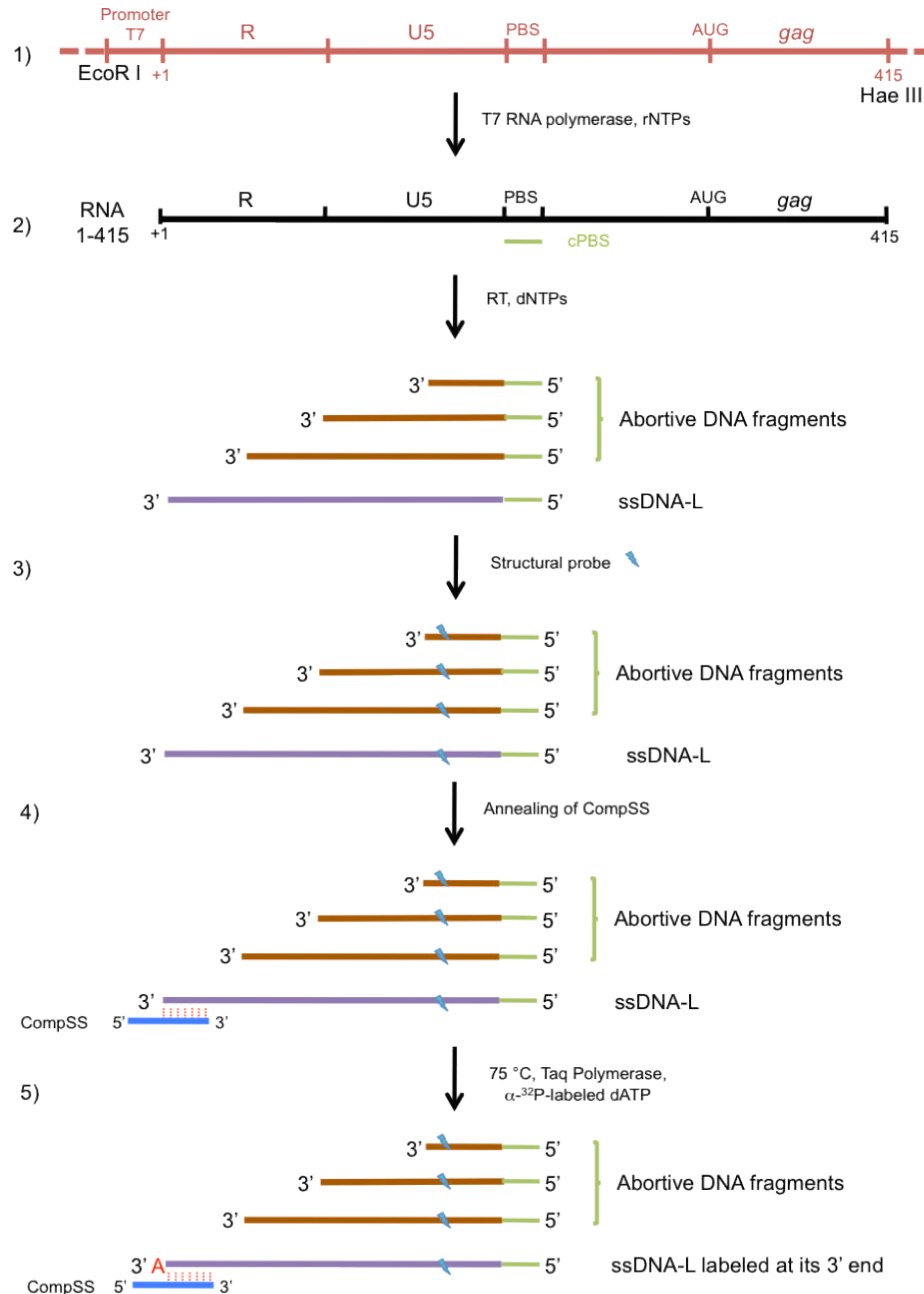
Figure 85. Specific 3' end labeling of ssDNA-L and its cleavage products. 1) *In vitro* transcription of RNA 1-415 via the template generated from plasmid pYC5' digested by HaeIII. 2) *In vitro* reverse transcription of RNA 1-415 from the cPBS primer. 3) Structural probing of the mixture of reverse transcription products. 4) Annealing of oligonucleotide CompSS. Only the probing products containing the complementary sequence of CompSS should be annealed. 5) A labeled adenine is added to the 3' end of ssDNA due to the 5' tail of CompSS. The plasmid is indicated in red, RNA 1-415 in black, abortive DNA fragments in brown and ssDNA-L in purple. The cleavage induced by the probe is indicated by a blue lightning.

Our 3' end-labeling method was also designed to investigate by chemical and enzymatic probes the secondary structure of ssDNA after reverse transcription, i.e. within a mixture containing abortive DNA fragments and DNA fragments generated by a probe. More precisely, this method should allow to check that ssDNA in the reverse transcription mixture adopts the same folding as the purified ssDNA that has been denatured and renatured. As shown in **Figure 85 (step 3)**, both ssDNA-L and abortive DNA fragments may be sensitive to one structural probe. When DNAs are labeled at their 5' ends, all the probing products can be detected on a denaturing polyacrylamide gel, which complicates the analysis of the secondary structure of ssDNA-L. However, using specific 3' end-labeling **(Figure 85 steps 4 and 5)**, only the probing products derived from ssDNA-L should be detected on the gel.

After the reverse transcription of RNA 1-415 **(Figure 85 step 2)**, we tried to label only the 3' ends of full-length ssDNAs and DNA fragments produced by cleavage of full-length ssDNAs with an enzymatic probe (DNase I or mung bean nuclease). Despite many trials, we did not get reproducible results allowing to determine the secondary structure of ssDNAs. Indeed, the enzymatic probes did not produced dose-dependent cleavages and there were high levels of nonspecific labeling (data not shown). A likely explanation is that DNA and RNA fragments interfered with the enzymatic activity of probes and nonspecific labeling is the result of an increase in the number of DNA molecules exhibiting a 3' end.

6.2   Structural analysis of ssDNA-L in the presence of NC

Both ssDNA-S and ssDNA-L were monomeric after incubation with NC as assessed by native polyacrylamide gel electroporesis in the TBE buffer **(Figure 66)**. Therefore, the structural analysis was not complicated by the presence of homoduplexes. To analyze the NC binding sites, we have taken into account the knowledges that are mentioned here. The binding site size of NC is 5 to 8 nucleotides (239). NC preferentially interacts with single-stranded regions and unpaired guanines. More precisely, NC binds with high affinity the TG, TNG and GNG motifs in DNA (10, 11, 390). Based on NMR structures, our team recently proposed that NC binds to DNAs and RNAs with opposite polarities (21). The binding polarity of NC complexed to DNA would be the following: the C-terminal zinc finger interacts with a guanine residue and the N-terminal zinc finger interacts with a residue (C or T) upstream to the guanine residue. To identify destabilized regions and protections induced by NC in ssDNA, we compared the enzymatic and KMnO$_4$ probing patterns of ssDNA-L in the absence or presence of increasing concentrations of NC. Unreactive thymine residues in stems can become reactive to KMnO$_4$ if these stems are destabilized by NC (see paper in the end of

this part). A close examination of the structures of the NC:DNA complexes (21, 40) shows that the 5, 6 double bond of the nucleobase of the thymine residues is accessible to $KMnO_4$ that is a small probe. Therefore, $KMnO_4$ is expected to interact with the nucleobases of unpaired thymine residues that interact with the N-terminal zinc finger of NC. In other word, $KMnO_4$ allows to identify the destabilized stems but not the thymine residues interacting with NC. After several trials, the $KMnO_4$ assays did not show any significant: 1) NC-mediated destabilization of stems in ssDNA-L; 2) change induced by NC in the ssDNA-L secondary structure.

The rate of DNase I cleavage at the level of T_{98}, A_{138} and G_{141} was not deeply affected by NC at protein to nucleotide molar ratios equal to or lower than 1:12 **(Figure 86)**, showing that the part of the stem containing these nucleotides is not a strong binding site for NC. This is expected since NC preferentially binds the single-stranded regions. Interestingly, the rate of DNase I cleavage was deeply affected at the level of most sites in the presence of increasing concentrations of NC. DNase I cleavage strongly increased at the level of A_{26}, A_{28}, C_{62} and T_{63}, suggesting that the regions containing these nucleotides are more accessible in the presence of NC than in its absence. In other words, the three-dimensional folding of ssDNA-L may be changed by NC. The effect of NC on the three-dimensional folding may depend on the magnesium concentrations. Indeed, in the presence of increasing concentrations of NC, DNase I cleavage at the level of A_{104} increased in 0.2 mM $MgCl_2$, whereas it decreased in 2 mM $MgCl_2$ **(Figure 86)**. In the presence of increasing concentrations of NC, DNase I cleavage at the level of several sites did not decrease at the same rate, i.e. NC induced specific protection against DNase I **(Figure 86)**. In both conditions (0.2 mM and 2 mM $MgCl_2$), there were moderate protections at the level of T_{42}, T_{47}, T_{101}, A_{108} and A_{128} **(Figures 86, 88 and 89)**. In the presence of 2 mM $MgCl_2$, there were also moderate protections at the level of T_{126}, T_{145}, G_{150} and C_{151}. These results are consistent with the notion that NC binds the double-stranded regions of nucleic acids non-specifically through electrostatic interactions of the basic residues with the phosphodiester backbone (229, 391). The stem containing the TG motif with the bulged guanine may be a strong binding site for NC since the protein induced a strong decrease in DNase I sensitivity at the level of C_{19} and G_{20}. This hypothesis is not supported by the MB probing data. Indeed, the sensitivity of A_{18}, T_{20} et G_{21} did not decrease in the presence of NC **(Figures 88 and 89)**. In the presence of 0.2 mM $MgCl_2$, there was a strong protection at the level of T_{126}, suggesting that the stem containing this residue is a strong binding site for NC. Since the binding site size of NC is 5 to 8 nucleotides (239), a likely hypothesis is that the zinc fingers of NC interacted with the unpaired G_{114} and G_{115} of the apical loop, whereas

the basic N-terminal domain of NC interacted with the adjacent stem. This hypothesis is consistent with a moderate protection at the level of T_{126} in the presence of 2 mM $MgCl_2$. Indeed, Mg^{2+} could compete with the basic N-terminal domain of NC for nonspecific binding to the stem. In the presence of 0.2 mM $MgCl_2$, there were strong protections against DNase I within the 145-152 sequence **(Figures 86 and 88)**. These results suggest that NC bound the single-stranded 148-151 sequence and switched the equilibrium toward the open conformation of the cTAR hairpin. The strong protection at the level of C_{147} in 2 mM $MgCl_2$ **(Figure 86)**, suggests that the lower part of the cTAR hairpin is a strong binding site for NC **(Figure 89)**.

The rate of MB cleavage at the level of A_{34} was not deeply affected by NC **(Figure 87)**, showing that the part of the apical loop containing this nucleotide is not a strong binding site for NC. Note that there is no unpaired guanine near A_{34} **(Figures 88 and 89)**. Surprisingly, there was a strong protection at the level of C_{69}, although there is no unpaired guanine near C_{69}. We do not think that the apical loop containing this nucleotide and G_{65} constitute an NC binding site. Indeed, there was no protection against DNase I at the level of T_{63} and A_{64}. Although there is a moderate protection at the level of A_{124}, we think that the surrounding nucleotides do not consitute an NC binding site. Indeed, there is no unpaired guanine near A_{124}. We propose that binding of NC to G_{119} induced structural changes in the apical loop leading to decrease the accessibility of A_{124} to MB. Interestingly, the moderate protections at the level of T_{50}, T_{52}, A_{54}, A_{108} and A_{117} occurred within unpaired regions containing at least a guanine residue **(Figures 87-89)**. Therefore, these regions are probably binding sites for NC. In both conditions (0.2 mM or 2 mM $MgCl_2$), the $_{85}GATGG_{89}$ and $_{146}CCTTG_{150}$ sequences are probably strong binding sites for NC since there were strong protections against MB within these sequences **(Figures 87-89)**. Note that the TG motif is present in both sequences. Interestingly, the MB sensitivity of A_{187} and A_{188} increased in the presence of NC and 2 mM $MgCl_2$, suggesting that the sequence containing these residues became more accessible for MB. In the same conditions, the MB sensitivity of A_{190} and A_{192} decreased, suggesting that the sequence containing these residues may be a NC binding site. This is consistent with the hypothesis that NC bound the 148-151 sequence and slightly destabilized the lower part of the cTAR hairpin. In the presence of 2 mM $MgCl_2$, there was a moderate protection at the level of C_{166}, suggesting that the apical loop containing this residue is a binding site for NC. The C-terminal zinc finger of NC may interact with G_{168} of the apical loop. However, this hypothesis is not compatible with the observation that the sensitivity of C_{167} to MB did not decrease in the presence of NC **(Figure 87B)**. A likely explanation is that NC changed the three-

dimensional folding of ssDNA-L and therefore the accessibility of the cTAR apical loop. In the presence of NC and 2 mM $MgCl_2$, the rate of MB cleavage increased between A_{157}-G_{158}, A_{159}-G_{160}, A_{175}-C_{176} and C_{176}-C_{177}, suggesting that the internal loop became more accessible to MB. Taken together, these data suggest that NC induced changes in the accessibility to the upper part of the cTAR hairpin.
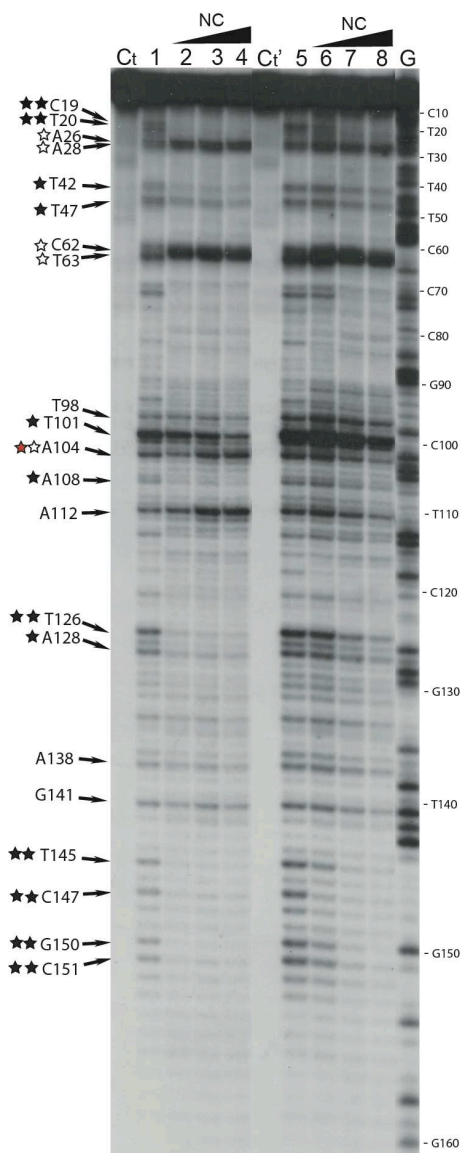
Figure 86. DNase I probing of ssDNA-L in the presence of NC. The NC-mediated probing experiments were performed and analyzed as described in the 'Materials and Methods' section. In the absence (lanes 1 and 5) or presence of NC (lanes 2-4 and 6-8), the 3' end-labeled ssDNA-L was incubated with DNase I (0.15 U for lanes 1-4 and 0.08 U for lanes 5-8). Lanes C_t and C_t' are controls without NC and DNase I. Lanes 1-4, assays performed in the presence of 0.2 mM $MgCl_2$. Lanes 5-8, assays performed in the presence of 2 mM $MgCl_2$. The protein to nucleotide molar ratios were 1:18 (lanes 2 and 6), 1:12 (lanes 3 and 7) and 1:7 (lanes 4 and 8). G refers to Maxam-Gilbert sequence markers. Arrows indicate the medium and strong DNase I cleavage sites. For the assays performed in 0.2 mM $MgCl_2$, moderate and strong protections induced by NC are indicated by one and two black stars, respectively. The white stars indicate the sites where the rate of cleavage increased in the presence of NC. The red star indicates the DNase I site where the cleavage rate in the presence of NC depends on magnesium concentration.

152

Figure 87. MB probing of ssDNA-L in the presence of NC. The NC-mediated probing experiments were performed and analyzed by short **(A)** and long **(B)** migrations as described in the 'Materials and Methods' section. **(A)** In the absence (lanes 1 and 5) or presence of NC (lanes 2-4 and 6-8), the 3' end-labeled ssDNA-L was incubated with MB (0.12 U). Lanes C_t and C_t' are controls without NC and MB. Lanes 1-4, assays performed in the presence of 0.2 mM $MgCl_2$. Lanes 5-8, assays performed in the presence of 2 mM $MgCl_2$. The protein to nucleotide molar ratios were 1:18 (lanes 2 and 6), 1:12 (lanes 3 and 7) and 1:7 (lanes 4 and 8). G refers to Maxam-Gilbert sequence markers. **(B)** In the absence (lane 1) or presence of NC (lanes 2-4), the 5' end-labeled ssDNA-L was incubated with MB (0.12 U) in the presence of 2 mM $MgCl_2$. Lane C_t is the control without NC and MB. The protein to nucleotide molar ratios were 1:18 (lane 2), 1:12 (lane 3) and 1:7 (lane 4). Sequence lanes (A, C, G and T) were run in parallel. Arrows indicate the medium and strong DNase I cleavage sites. For the assays performed in 0.2 mM $MgCl_2$, moderate and strong protections induced by NC are indicated by one and two black stars, respectively. The white stars indicate the sites where the rate of cleavage increased in the presence of NC. The red star indicates the DNase I site where the cleavage rate in the presence of NC depends on magnesium concentration.
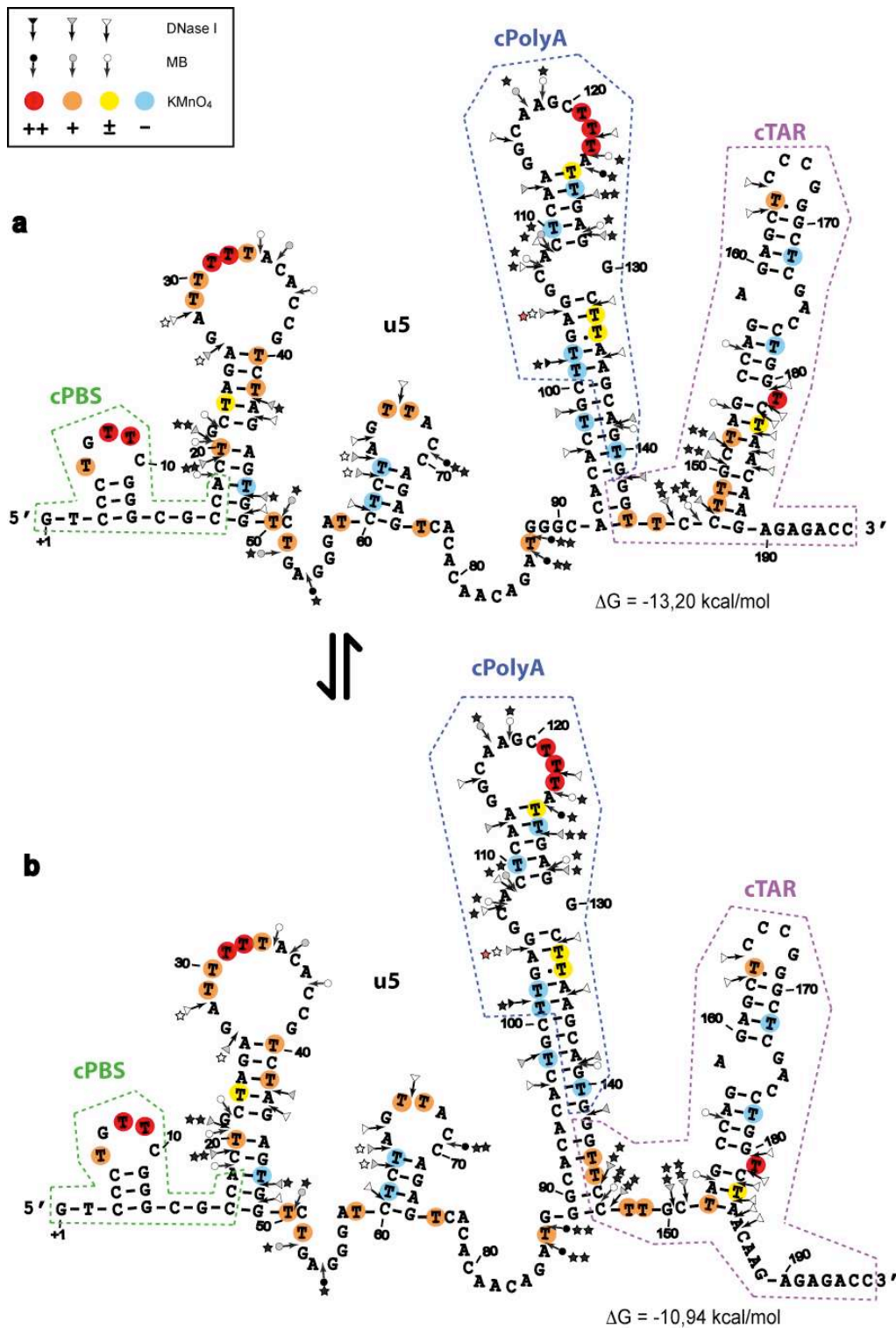
Figure 88. Secondary structure models for ssDNA-L in the presence of NC and in the presence of 0.2 mM MgCl₂. ssDNA-L may adopt two conformations **(a and b)**. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The protection induced by NC at the level of enzymatic cleavage sites are indicated by black stars (the strong protections are indicated by two black stars). The white stars indicate the sites where the rate of cleavage increased in the presence of NC. The red stars indicates the DNase I site where the cleavage rate in the presence of NC depends on magnesium concentrations. The protection or destabilization induced by NC at the level of MB cleavages in the 157-196 sequence has not been determined.
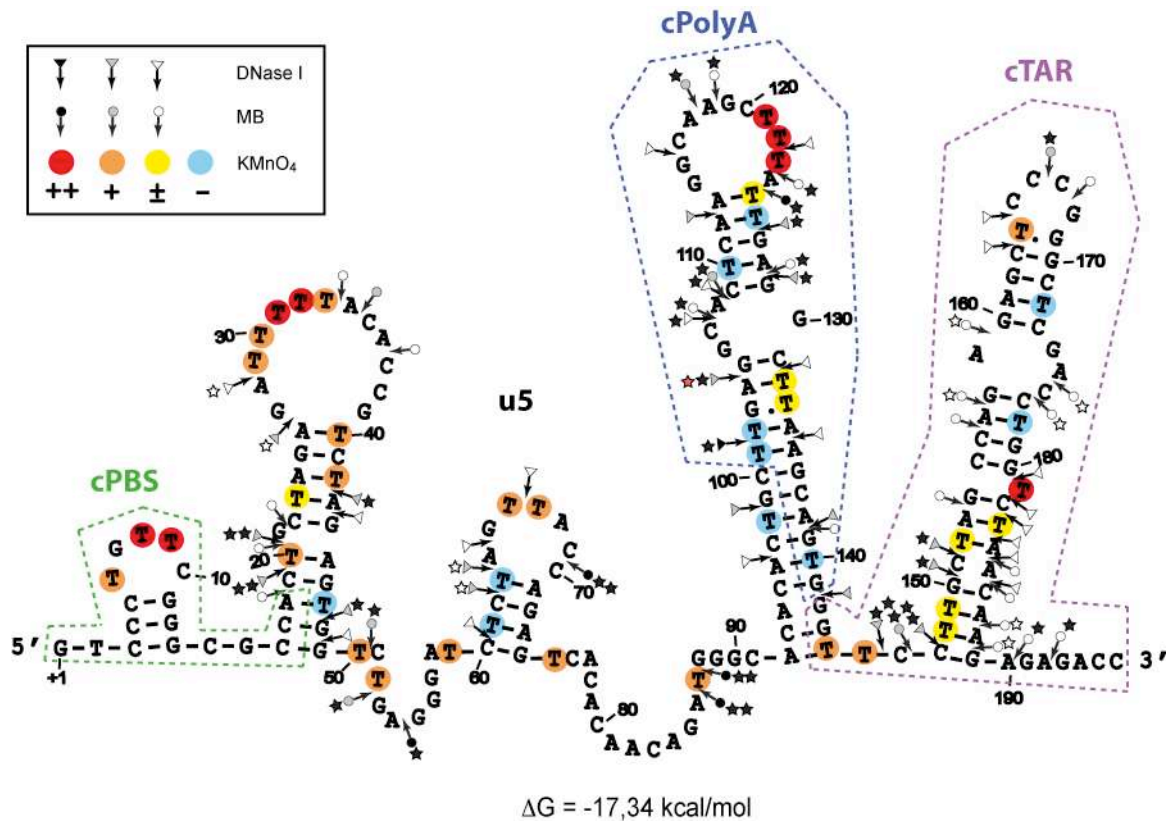
Figure 89. Secondary structure model for ssDNA-L in the presence of NC and in the presence of 2 mM MgCl₂. ssDNA-L forms a stable conformation. Delta G-values were predicted by mfold. Closed, gray and open symbols indicate strong, medium and weak cleavage sites, respectively, for the various enzymes (triangle for DNase I and circle for mung bean nuclease). The color codes used for the reactivity of thymine residues are indicated in the inset. The protections induced by NC at the level of enzymatic probe cleavage sites are indicated by the black stars (the strong protections are indicated by two black stars). The white stars indicate the sites where the rate of cleavage increased in the presence of NC. The red stars indicates the DNase I site where the cleavage rate in the presence of NC depends on magnesium concentrations.
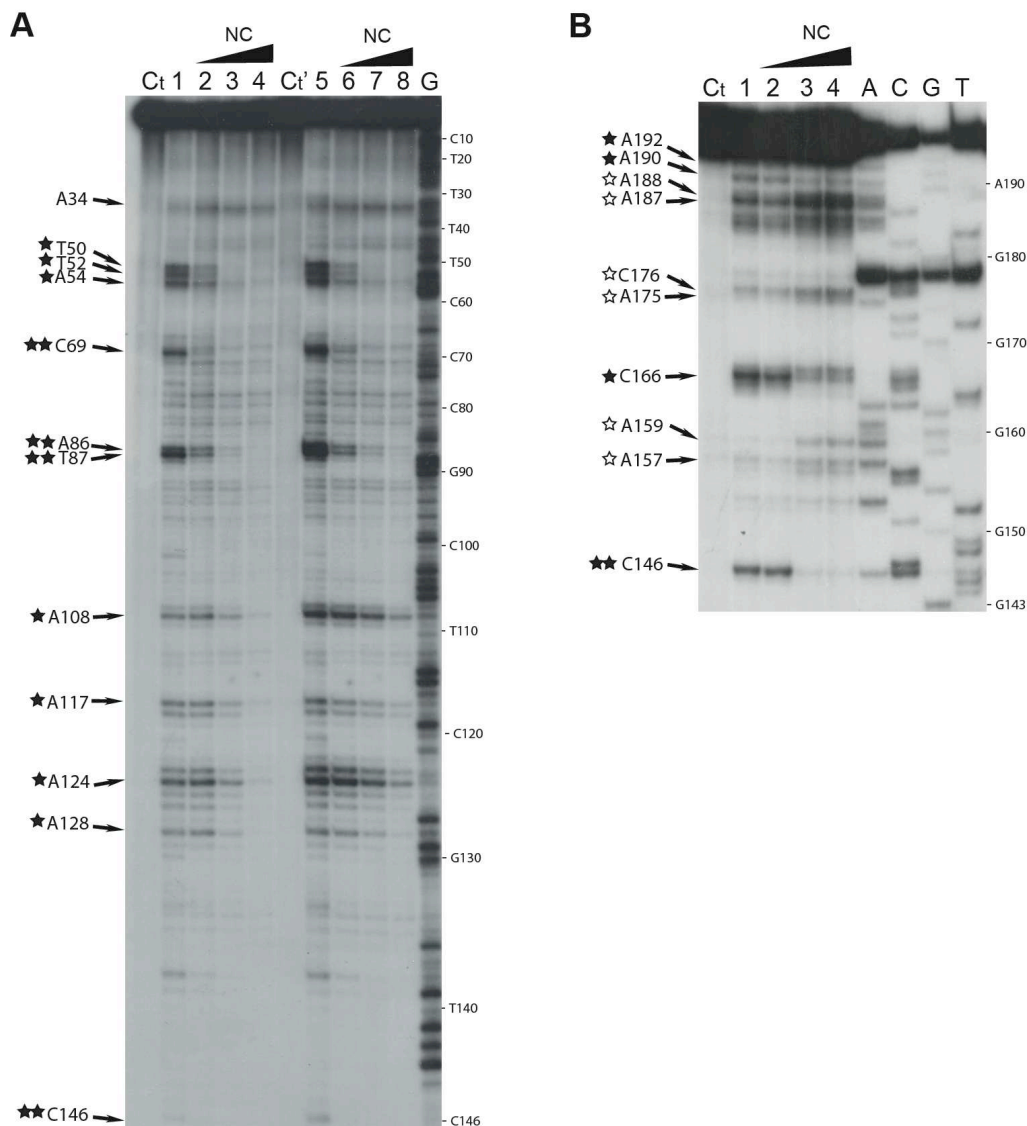
GENERAL CONCLUSIONS

Our study is the first analysis of structure-function relationships in the full-length ssDNA. Until our study, this analysis has not been performed because it requires relatively large amounts of the full-length ssDNA. Therefore, a part of my thesis was devoted to develop a method to prepare sufficient amounts of the pure full-length ssDNA. I prepared and studied ssDNA-S and ssDNA-L that are two full-length ssDNAs **(Figure 56)**. ssDNA-L represents the full-length ssDNA that is not annealed to the PBS region, whereas ssDNA-S represents the full-length ssDNA that is paired with this region. All experiments were performed in 0.2 mM $MgCl_2$ (intracellular concentration) and 2 mM $MgCl_2$ (concentration required for reverse transcription and strand transfer *in vitro*).

1. <u>ssDNA adopts two distinct conformations in 0.2 mM $MgCl_2$</u>

Our results suggest that ssDNA-S or ssDNA-L folds mainly into one conformation in the presence of 2 mM $MgCl_2$. In contrast, analysis of ssDNA-S and ssDNA-L by non-denaturing polyacrylamide gel electrophoresis suggests that ssDNA adopts two distinct conformations in equilibrium in the presence of 0.2 mM $MgCl_2$. Although the cPBS sequence is deleted in ssDNA-S, we observed two conformers for this DNA. Therefore, the cPBS sequence has no role in formation of the two confomers. Our probing data suggest that formation of two conformers is due to the cTAR sequence that can form a long stem-loop or a short stem-loop in ssDNA **(Figure 90B)**. To demonstrate this hypothesis, it will be necessary to design mutations within the cTAR sequence that block one conformation.

2. <u>Folding of the r region</u>

The r region in ssDNA-S and ssDNA-L does not fold as an independent domain. Indeed, 8 base pairs (conformer 1) or 11 base pairs (conformer 2) are formed between the u5 and r sequences **(Figure 90B)**. As predicted by Berkhout *et al.* (34), the r sequence folds into the cpolyA and cTAR hairpins. Our probing data strongly suggest that the secondary structures of the cTAR sequence alone or within ssDNA are not identical **(Figure 90, compare A and B)**. Thus, the closed conformation is not formed in ssDNA and the cTAR hairpin is more open (conformer 2) in ssDNA than the 'Y' conformation of the cTAR sequence alone. In both conformations of ssDNA, the 5' end of the cTAR sequence forms a base-pairing interaction with nucleotides of the u5 region. In contrast, the 3' end of the cTAR sequence is unpaired in both conformations of ssDNA. In the absence of NC and in the presence of 7 mM $MgCl_2$, our data support a dynamic structure of the cTAR hairpin, involving equilibrium between both the closed conformation and the 'Y' conformation. Our data suggest that most of ssDNA

157

molecules fold into conformer 1 in the absence of NC and in the presence of 2 mM MgCl$_2$. It is likely that the magnesium ions stabilize the lower part of the cTAR hairpin in conformer 1. Finally, our results indicate that a DNA oligonucleotide corresponding to the cTAR sequence is not a perfect model for the annealing reaction, since its folding is partially different from the cTAR sequence in ssDNA.
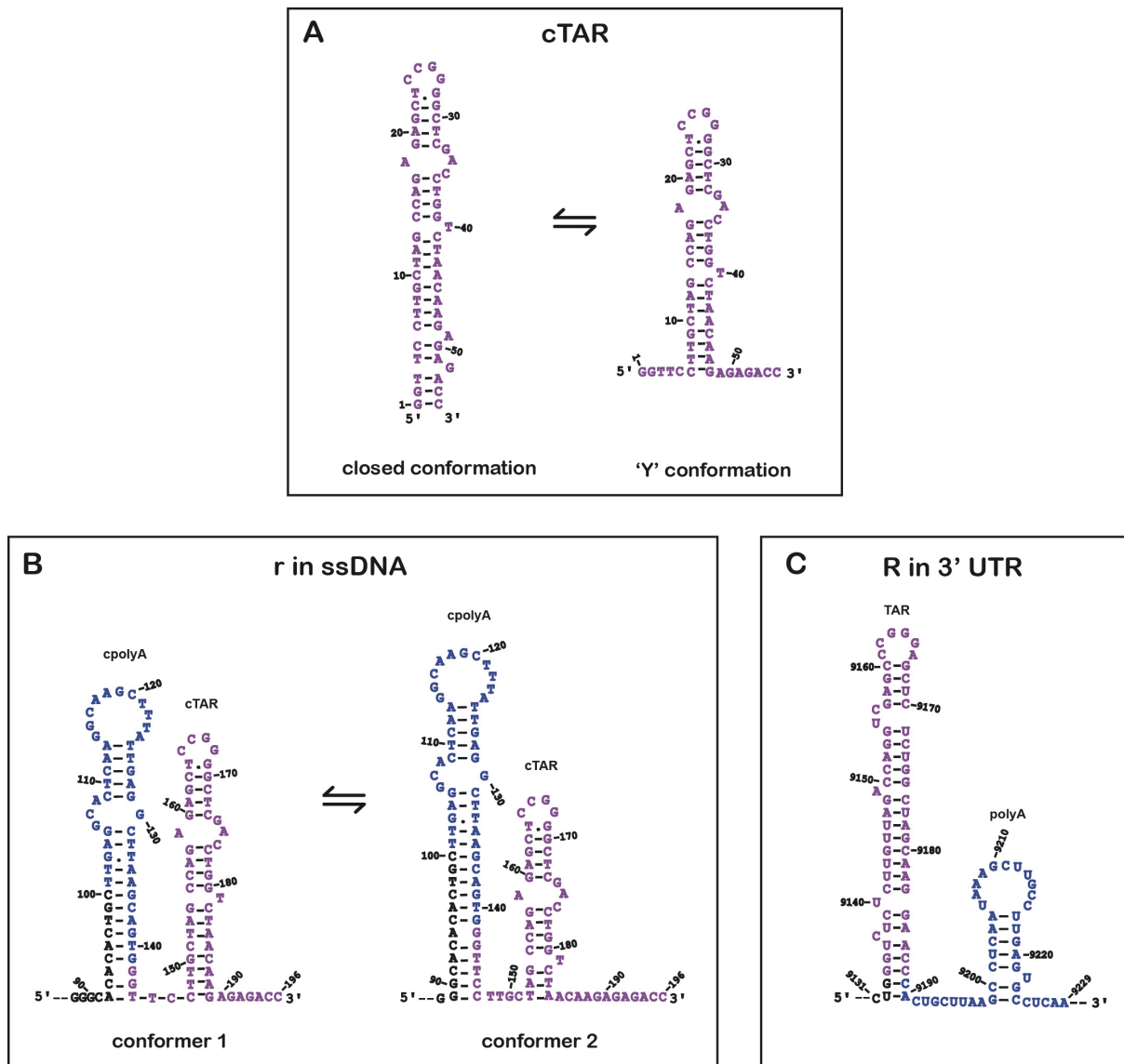
Figure 90. Comparison of secondary structures of cTAR DNA involved in the annealing of ssDNA to 3' UTR. (A) Secondary structures for the closed and 'Y' conformations of cTAR DNA. **(B)** Secondary structure models of the r region in ssDNA. **(C)** Secondary structure of R region in 3' UTR. The cTAR DNA, the cTAR hairpin in ssDNA and the TAR hairpin in 3' UTR are indicated in purple. The cpoly(A) region in ssDNA and the poly(A) region in 3' UTR are indicated in blue. Numbering of ssDNA corresponds to +1 of ssDNA-L and numbering of 3' UTR corresponds to +1 of HIV-1 (MAL isolate).

3. <u>NC binding sites in ssDNA</u>

Our footprinting experiments were performed at NC to nucleotide molar ratio of 1:18, 1:12 and 1:7 that have been found in infectious HIV-1 particles (45, 60). Our probing data suggest that there are at least five preferential binding sites for NC in ssDNA. Taken together, the DNase I and MB probing data suggest that the $_{85}GATGG_{89}$ and $_{146}CCTTG_{150}$ sequences are two strong NC binding sites, whereas the $_{50}TCTGAG_{55}$, $_{106}GCA_{108}$ and $_{117}AAGC_{120}$ sequences are moderate NC binding sites. All NC binding sites contain at least one unpaired guanine. Note that the TG motif is present in both strong NC binding sites. In the u5 region, the two NC binding sites are within single-stranded regions that are located between two stem-loops. In the r region; the three NC binding sites are located within: 1) the internal loop of the cpolyA hairpin; 2) the apical loop of the cpolyA hairpin; 3) the single-stranded region between the cpolyA and cTAR hairpins. Surprisingly, we do not have evidence that NC binds the apical and internal loops of the cTAR hairpin in ssDNA, although we found that NC binds these loops in a DNA oligonucleotide corresponding to the cTAR sequence (**Figure 53** and paper at the end of Part 1 of the 'Results and Discussion' section). A hypothesis is that the unpaired G_{168} and G_{174} are not accessible to NC in structured ssDNA. However, this hypothesis is not supported by the observation that MB cleaves with these loops in structured ssDNA. To date, we cannot give any logical explanation for the discrepancy between the footprinting results obtained with cTAR DNA and ssDNA.

4. <u>Relationships between the ssDNA structure and the annealing reaction</u>

We showed that mutations targeting the apical loop and the 3' end of the cTAR hairpin did not prevent the annealing of ssDNA to 3' UTR in the presence of 0.2 mM $MgCl_2$. These results support the notion that NC-mediated annealing of ssDNA to 3' UTR can be initiated through different regions of the R sequence in 0.2 mM $MgCl_2$ (intracellular concentration). Our results suggest that the r region folds into two conformations in equilibrium in the presence of 0.2 mM $MgCl_2$ **(Figure 90B)**. Under these conditions, the initiation sites for the annealing reaction are probably unpaired regions, i.e. the apical loops, the 3' end of the cTAR sequence and the single-stranded regions between the hairpins **(Figure 90 B and C)**. The complementary apical loops are present in both partners, whereas the complementary sequences of single-stranded regions are in stems (nucleotides 132-139 in ssDNA and 9134-9145 in 3' UTR). Opening of these stems is required for the annealing reaction and is probably more efficient in 0.2 mM $MgCl_2$ than 2 mM $MgCl_2$. The NC binding site at the junction of the cpolyA and cTAR hairpins suggests that NC facilitates the opening

of stems. Our group did not find a NC binding site between the TAR and polyA hairpins; although this single-stranded region contains the UG motif (205). Since this study was performed in the presence of 7 mM MgCl$_2$, we cannot exclude the possibility that NC binds the junction of the TAR and polyA hairpins at low magnesium concentrations. Interestingly, we showed that approximately half of ssDNA molecules annealed through the 'kissing' pathway in 2 mM MgCl$_2$ (concentration required for reverse transcription and strand transfer *in vitro*). In contrast, our results do not support the 'zipper' pathway in 2 mM MgCl$_2$. The lower stems of the TAR and cpolyA elements are probably too stable in 2 mM MgCl$_2$ to be used as initiation sites for the annealing reaction. Therefore, we propose that the apical loops of the polyA and cpolyA hairpins are the second initiation sites for the annealing reaction in 2 mM MgCl$_2$. Note that these apical loops contain 11 unpaired nucleotides **(Figure 90 B and C)**. Finally, our results showed that NC plays an essential role in annealing of ssDNA to 3' UTR, since no full-length heteroduplexes were formed in the absence of NC.

REFERENCES

1. **Abbink, T. E., M. Ooms, P. C. Haasnoot, and B. Berkhout.** 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. Biochemistry **44:**9058-66.
2. **Aldovini, A., and R. A. Young.** 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. J Virol **64:**1920-6.
3. **Amarasinghe, G. K., R. N. De Guzman, R. B. Turner, and M. F. Summers.** 2000. NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. J Mol Biol **299:**145-56.
4. **An, P., and C. A. Winkler.** 2010. Host genes associated with HIV/AIDS: advances in gene discovery. Trends Genet **26:**119-31.
5. **Andersen, E. S., S. A. Contera, B. Knudsen, C. K. Damgaard, F. Besenbacher, and J. Kjems.** 2004. Role of the trans-activation response element in dimerization of HIV-1 RNA. J Biol Chem **279:**22243-9.
6. **Andersen, E. S., R. E. Jeeninga, C. K. Damgaard, B. Berkhout, and J. Kjems.** 2003. Dimerization and template switching in the 5' untranslated region between various subtypes of human immunodeficiency virus type 1. J Virol **77:**3020-30.
7. **Anthony, R. M., and J. J. Destefano.** 2007. In vitro synthesis of long DNA products in reactions with HIV-RT and nucleocapsid protein. J Mol Biol **365:**310-24.
8. **Ashe, M. P., A. Furger, and N. J. Proudfoot.** 2000. Stem-loop 1 of the U1 snRNP plays a critical role in the suppression of HIV-1 polyadenylation. RNA **6:**170-7.
9. **Ashe, M. P., P. Griffin, W. James, and N. J. Proudfoot.** 1995. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. Genes Dev **9:**3008-25.
10. **Avilov, S. V., J. Godet, E. Piemont, and Y. Mely.** 2009. Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. Biochemistry **48:**2422-30.
11. **Avilov, S. V., E. Piemont, V. Shvadchak, H. de Rocquigny, and Y. Mely.** 2008. Probing dynamics of HIV-1 nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. Nucleic Acids Res **36:**885-96.
12. **Azoulay, J., J. P. Clamme, J. L. Darlix, B. P. Roques, and Y. Mely.** 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.
13. **Bachelin, M., G. Hessler, G. Kurz, J. G. Hacia, P. B. Dervan, and H. Kessler.** 1998. Structure of a stereoregular phosphorothioate DNA/RNA duplex. Nat Struct Biol **5:**271-6.
14. **Balakrishnan, M., P. J. Fay, and R. A. Bambara.** 2001. The kissing hairpin sequence promotes recombination within the HIV-I 5' leader region. J Biol Chem **276:**36482-92.
15. **Balakrishnan, M., B. P. Roques, P. J. Fay, and R. A. Bambara.** 2003. Template dimerization promotes an acceptor invasion-induced transfer mechanism during human immunodeficiency virus type 1 minus-strand synthesis. J Virol **77:**4710-21.
16. **Bampi, C., S. Jacquenet, D. Lener, D. Decimo, and J. L. Darlix.** 2004. The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance. Curr HIV Res **2:**79-92.
17. **Bannwarth, S., and A. Gatignol.** 2005. HIV-1 TAR RNA: the target of molecular interactions between the virus and its host. Curr HIV Res **3:**61-71.

18. **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220:**868-71.

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

20. **Baudin, F., R. Marquet, C. Isel, J. L. Darlix, B. Ehresmann, and C. Ehresmann.** 1993. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. J Mol Biol **229:**382-97.

21. **Bazzi, A., L. Zargarian, F. Chaminade, C. Boudier, H. De Rocquigny, B. Rene, Y. Mely, P. Fosse, and O. Mauffret.** 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:**3903-16.

22. **Beerens, N., F. Groot, and B. Berkhout.** 2001. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. J Biol Chem **276:**31247-56.

23. **Beerens, N., and J. Kjems.** 2010. Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. RNA **16:**1226-35.

24. **Beilhartz, G. L., and M. Gotte.** 2010. HIV-1 Ribonuclease H: Structure, Catalytic Mechanism and Inhibitors. Viruses **2:**900-26.

25. **Beilhartz, G. L., M. Wendeler, N. Baichoo, J. Rausch, S. Le Grice, and M. Gotte.** 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:**462-74.

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

27. **Beltz, H., C. Clauss, E. Piemont, D. Ficheux, R. J. Gorelick, B. Roques, C. Gabus, J. L. Darlix, H. de Rocquigny, and Y. Mely.** 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:**1113-26.

28. **Ben Ali, M., F. Chaminade, I. Kanevsky, E. Ennifar, L. Josset, D. Ficheux, J. L. Darlix, and P. Fosse.** 2007. Structural requirements for nucleocapsid protein-mediated dimerization of avian leukosis virus RNA. J Mol Biol **372:**1082-96.

29. **Berglund, J. A., B. Charpentier, and M. Rosbash.** 1997. A high affinity binding site for the HIV-1 nucleocapsid protein. Nucleic Acids Res **25:**1042-9.

30. **Berkhout, B., A. T. Das, and J. L. van Wamel.** 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:**211-8.

31. **Berkhout, B., B. Klaver, and A. T. Das.** 1995. A conserved hairpin structure predicted for the poly(A) signal of human and simian immunodeficiency viruses. Virology **207:**276-81.

32. **Berkhout, B., J. van Wamel, and B. Klaver.** 1995. Requirements for DNA strand transfer during reverse transcription in mutant HIV-1 virions. J Mol Biol **252:**59-69.

33. **Berkhout, B., and J. L. van Wamel.** 2000. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. RNA **6:**282-95.

34. **Berkhout, B., N. L. Vastenhouw, B. I. Klasens, and H. Huthoff.** 2001. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. RNA **7:**1097-114.

35. **Berkowitz, R., J. Fisher, and S. P. Goff.** 1996. RNA packaging. Curr Top Microbiol Immunol **214:**177-218.

36. **Bernacchi, S., S. Stoylov, E. Piemont, D. Ficheux, B. P. Roques, J. L. Darlix, and Y. Mely.** 2002. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. J Mol Biol **317:**385-99.

37. **Biasin, M., L. Piacentini, S. Lo Caputo, Y. Kanari, G. Magri, D. Trabattoni, V. Naddeo, L. Lopalco, A. Clivio, E. Cesana, F. Fasano, C. Bergamaschi, F. Mazzotta, M. Miyazawa, and M. Clerici.** 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:**960-4.

38. **Bishop, K. N., R. K. Holmes, and M. H. Malim.** 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. J Virol **80:**8450-8.

39. **Biswas, N., T. Wang, M. Ding, A. Tumne, Y. Chen, Q. Wang, and P. Gupta.** 2012. ADAR1 is a novel multi targeted anti-HIV-1 cellular protein. Virology **422:**265-77.

40. **Bourbigot, S., N. Ramalanjaona, C. Boudier, G. F. Salgado, B. P. Roques, Y. Mely, S. Bouaziz, and N. Morellet.** 2008. How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. J Mol Biol **383:**1112-28.

41. **Boyer, P. L., S. G. Sarafianos, E. Arnold, and S. H. Hughes.** 2000. Analysis of mutations at positions 115 and 116 in the dNTP binding site of HIV-1 reverse transcriptase. Proc Natl Acad Sci U S A **97:**3056-61.

42. **Brady, J., and F. Kashanchi.** 2005. Tat gets the "green" light on transcription initiation. Retrovirology **2:**69.

43. **Briggs, J. A., K. Grunewald, B. Glass, F. Forster, H. G. Krausslich, and S. D. Fuller.** 2006. The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. Structure **14:**15-20.

44. **Briggs, J. A., and H. G. Krausslich.** 2011. The molecular architecture of HIV. J Mol Biol **410:**491-500.

45. **Briggs, J. A., M. N. Simon, I. Gross, H. G. Krausslich, S. D. Fuller, V. M. Vogt, and M. C. Johnson.** 2004. The stoichiometry of Gag protein in HIV-1. Nat Struct Mol Biol **11:**672-5.

46. **Brule, F., G. Bec, G. Keith, S. F. Le Grice, B. P. Roques, B. Ehresmann, C. Ehresmann, and R. Marquet.** 2000. In vitro evidence for the interaction of tRNA(3)(Lys) with U3 during the first strand transfer of HIV-1 reverse transcription. Nucleic Acids Res **28:**634-40.

47. **Brunel, C., and P. Romby.** 2000. Probing RNA structure and RNA-ligand complexes with chemical probes. Methods Enzymol **318:**3-21.

48. **Buchschacher, G. L., Jr., and A. T. Panganiban.** 1992. Human immunodeficiency virus vectors for inducible expression of foreign genes. J Virol **66:**2731-9.

49. **Bukrinskaya, A. G.** 2004. HIV-1 assembly and maturation. Arch Virol **149:**1067-82.

50. **Cameron, C. E., M. Ghosh, S. F. Le Grice, and S. J. Benkovic.** 1997. Mutations in HIV reverse transcriptase which alter RNase H activity and decrease strand

transfer efficiency are suppressed by HIV nucleocapsid protein. Proc Natl Acad Sci U S A **94:**6700-5.

51. **Carlson, J. M., Z. L. Brumme, C. M. Rousseau, C. J. Brumme, P. Matthews, C. Kadie, J. I. Mullins, B. D. Walker, P. R. Harrigan, P. J. Goulder, and D. Heckerman.** 2008. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol **4:**e1000225.

52. **Carteau, S., R. J. Gorelick, and F. D. Bushman.** 1999. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. J Virol **73:**6670-9.

53. **Cen, S., A. Khorchid, H. Javanbakht, J. Gabor, T. Stello, K. Shiba, K. Musier-Forsyth, and L. Kleiman.** 2001. Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. J Virol **75:**5043-8.

54. **Chan, D. C., and P. S. Kim.** 1998. HIV entry and its inhibition. Cell **93:**681-4.

55. **Charneau, P., G. Mirambeau, P. Roux, S. Paulous, H. Buc, and F. Clavel.** 1994. HIV-1 reverse transcription. A termination step at the center of the genome. J Mol Biol **241:**651-62.

56. **Chen, J., O. Nikolaitchik, J. Singh, A. Wright, C. E. Bencsics, J. M. Coffin, N. Ni, S. Lockett, V. K. Pathak, and W. S. Hu.** 2009. High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. Proc Natl Acad Sci U S A **106:**13535-40.

57. **Chen, L. F., J. Hoy, and S. R. Lewin.** 2007. Ten years of highly active antiretroviral therapy for HIV infection. Med J Aust **186:**146-51.

58. **Chen, Y., M. Balakrishnan, B. P. Roques, and R. A. Bambara.** 2003. Steps of the acceptor invasion mechanism for HIV-1 minus strand strong stop transfer. J Biol Chem **278:**38368-75.

59. **Chen, Y., M. Balakrishnan, B. P. Roques, P. J. Fay, and R. A. Bambara.** 2003. Mechanism of minus strand strong stop transfer in HIV-1 reverse transcription. J Biol Chem **278:**8006-17.

60. **Chertova, E., O. Chertov, L. V. Coren, J. D. Roser, C. M. Trubey, J. W. Bess, Jr., R. C. Sowder, 2nd, E. Barsov, B. L. Hood, R. J. Fisher, K. Nagashima, T. P. Conrads, T. D. Veenstra, J. D. Lifson, and D. E. Ott.** 2006. Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. J Virol **80:**9039-52.

61. **Chiu, Y. L., and W. C. Greene.** 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol **26:**317-53.

62. **Cimarelli, A., and J. L. Darlix.** 2002. Assembling the human immunodeficiency virus type 1. Cell Mol Life Sci **59:**1166-84.

63. **Cimarelli, A., S. Sandin, S. Hoglund, and J. Luban.** 2000. Basic residues in human immunodeficiency virus type 1 nucleocapsid promote virion assembly via interaction with RNA. J Virol **74:**3046-57.

64. **Clavel, F., F. Brun-Vezinet, D. Guetard, S. Chamaret, A. Laurent, C. Rouzioux, M. Rey, C. Katlama, F. Rey, J. L. Champelinaud, and et al.** 1986. [LAV type II: a second retrovirus associated with AIDS in West Africa]. C R Acad Sci III **302:**485-8.

65. **Clavel, F., D. Guetard, F. Brun-Vezinet, S. Chamaret, M. A. Rey, M. O. Santos-Ferreira, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, and et al.** 1986. Isolation of a new human retrovirus from West African patients with AIDS. Science **233:**343-6.

66. **Coffin, J. M.** 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. J Gen Virol **42:**1-26.

67. **Coffin, J. M., and W. A. Haseltine.** 1977. Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. Proc Natl Acad Sci U S A **74:**1908-12.

68. **Coiras, M., M. R. Lopez-Huertas, M. Sanchez del Cojo, E. Mateos, and J. Alcami.** 2010. Dual role of host cell factors in HIV-1 replication: restriction and enhancement of the viral cycle. AIDS Rev **12:**103-12.

69. **Conte, M. R., and S. Matthews.** 1998. Retroviral matrix proteins: a structural perspective. Virology **246:**191-8.

70. **Cosa, G., E. J. Harbron, Y. Zeng, H. W. Liu, D. B. O'Connor, C. Eta-Hosokawa, K. Musier-Forsyth, and P. F. Barbara.** 2004. Secondary structure and secondary structure dynamics of DNA hairpins complexed with HIV-1 NC protein. Biophys J **87:**2759-67.

71. **Cowan, J. A., T. Ohyama, K. Howard, J. W. Rausch, S. M. Cowan, and S. F. Le Grice.** 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. J Biol Inorg Chem **5:**67-74.

72. **Craig, M. L., O. V. Tsodikov, K. L. McQuade, P. E. Schlax, Jr., M. W. Capp, R. M. Saecker, and M. T. Record, 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:**741-56.

73. **Cristofari, G., and J. L. Darlix.** 2002. The ubiquitous nature of RNA chaperone proteins. Prog Nucleic Acid Res Mol Biol **72:**223-68.

74. **Cruceanu, M., M. A. Urbaneja, C. V. Hixson, D. G. Johnson, S. A. Datta, M. J. Fivash, A. G. Stephen, R. J. Fisher, R. J. Gorelick, J. R. Casas-Finet, A. Rein, I. Rouzina, and M. C. Williams.** 2006. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. Nucleic Acids Res **34:**593-605.

75. **D'Souza, V., and M. F. Summers.** 2005. How retroviruses select their genomes. Nat Rev Microbiol **3:**643-55.

76. **da Silva, Z. J., I. Oliveira, A. Andersen, F. Dias, A. Rodrigues, B. Holmgren, S. Andersson, and P. Aaby.** 2008. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? AIDS **22:**1195-202.

77. **Daecke, J., O. T. Fackler, M. T. Dittmar, and H. G. Krausslich.** 2005. Involvement of clathrin-mediated endocytosis in human immunodeficiency virus type 1 entry. J Virol **79:**1581-94.

78. **Damond, F., M. Worobey, P. Campa, I. Farfara, G. Colin, S. Matheron, F. Brun-Vezinet, D. L. Robertson, and F. Simon.** 2004. Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. AIDS Res Hum Retroviruses **20:**666-72.

79. **Dang, Q., and W. S. Hu.** 2001. Effects of homology length in the repeat region on minus-strand DNA transfer and retroviral replication. J Virol **75:**809-20.

80. **Darlix, J. L., C. Gabus, M. T. Nugeyre, F. Clavel, and F. Barre-Sinoussi.** 1990. Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. J Mol Biol **216:**689-99.

81. **Darlix, J. L., J. L. Garrido, N. Morellet, Y. Mely, and H. de Rocquigny.** 2007. Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. Adv Pharmacol **55:**299-346.

82. **Darlix, J. L., J. Godet, R. Ivanyi-Nagy, P. Fosse, O. Mauffret, and Y. Mely.** 2011. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. J Mol Biol **410:**565-81.

83. **Darlix, J. L., A. Vincent, C. Gabus, H. de Rocquigny, and B. Roques.** 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:**763-71.

84. **Das, A. T., B. Klaver, and B. Berkhout.** 1999. A hairpin structure in the R region of the human immunodeficiency virus type 1 RNA genome is instrumental in polyadenylation site selection. J Virol **73:**81-91.

85. **de Bethune, M. P.** 2010. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: a review of the last 20 years (1989-2009). Antiviral Res **85:**75-90.

86. **De Guzman, R. N., Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, and M. F. Summers.** 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science **279:**384-8.

87. **de Rocquigny, H., D. Ficheux, C. Gabus, M. C. Fournie-Zaluski, J. L. Darlix, and B. P. Roques.** 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:**1010-8.

88. **de Silva, T. I., M. Cotten, and S. L. Rowland-Jones.** 2008. HIV-2: the forgotten AIDS virus. Trends Microbiol **16:**588-95.

89. **Deforges, J., N. Chamond, and S. Bruno.** 2012. Structural investigation of HIV-1 genomic RNA dimerization process reveals a role for the Major Splice-site Donor stem loop. Biochimie.

90. **Demirov, D. G., and E. O. Freed.** 2004. Retrovirus budding. Virus Res **106:**87-102.

91. **Deng, H., R. Liu, W. Ellmeier, S. Choe, D. Unutmaz, M. Burkhart, P. Di Marzio, S. Marmon, R. E. Sutton, C. M. Hill, C. B. Davis, S. C. Peiper, T. J. Schall, D. R. Littman, and N. R. Landau.** 1996. Identification of a major co-receptor for primary isolates of HIV-1. Nature **381:**661-6.

92. **Derebail, S. S., and J. J. DeStefano.** 2004. Mechanistic analysis of pause site-dependent and -independent recombinogenic strand transfer from structurally diverse regions of the HIV genome. J Biol Chem **279:**47446-54.

93. **Derebail, S. S., M. J. Heath, and J. J. DeStefano.** 2003. Evidence for the differential effects of nucleocapsid protein on strand transfer in various regions of the HIV genome. J Biol Chem **278:**15702-12.

94. **Desai, N. A., and V. Shankar.** 2003. Single-strand-specific nucleases. FEMS Microbiol Rev **26:**457-91.

95. **DeStefano, J. J., R. A. Bambara, and P. J. Fay.** 1994. The mechanism of human immunodeficiency virus reverse transcriptase-catalyzed strand transfer from internal regions of heteropolymeric RNA templates. J Biol Chem **269:**161-8.

96. **DeStefano, J. J., R. G. Buiser, L. M. Mallaber, T. W. Myers, R. A. Bambara, and P. J. Fay.** 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:**7423-31.

97. **DeStefano, J. J., L. M. Mallaber, L. Rodriguez-Rodriguez, P. J. Fay, and R. A. Bambara.** 1992. Requirements for strand transfer between internal regions of

heteropolymer templates by human immunodeficiency virus reverse transcriptase. J Virol **66:**6370-8.

98. **Diaz-Griffero, F., X. R. Qin, F. Hayashi, T. Kigawa, A. Finzi, Z. Sarnak, M. Lienlaf, S. Yokoyama, and J. Sodroski.** 2009. A B-box 2 surface patch important for TRIM5alpha self-association, capsid binding avidity, and retrovirus restriction. J Virol **83:**10737-51.

99. **Didierlaurent, L., L. Houzet, Z. Morichaud, J. L. Darlix, and M. Mougel.** 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:**4745-53.

100. **Didierlaurent, L., P. J. Racine, L. Houzet, C. Chamontin, B. Berkhout, and M. Mougel.** 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:**8915-27.

101. **Do, H., A. Vasilescu, G. Diop, T. Hirtzig, S. C. Heath, C. Coulonges, J. Rappaport, A. Therwath, M. Lathrop, F. Matsuda, and J. F. Zagury.** 2005. Exhaustive genotyping of the CEM15 (APOBEC3G) gene and absence of association with AIDS progression in a French cohort. J Infect Dis **191:**159-63.

102. **Driscoll, M. D., M. P. Golinelli, and S. H. Hughes.** 2001. In vitro analysis of human immunodeficiency virus type 1 minus-strand strong-stop DNA synthesis and genomic RNA processing. J Virol **75:**672-86.

103. **Driscoll, M. D., and S. H. Hughes.** 2000. Human immunodeficiency virus type 1 nucleocapsid protein can prevent self-priming of minus-strand strong stop DNA by promoting the annealing of short oligonucleotides to hairpin sequences. J Virol **74:**8785-92.

104. **Druillennec, S., A. Caneparo, H. de Rocquigny, and B. P. Roques.** 1999. Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1. J Biol Chem **274:**11283-8.

105. **Drummond, J. E., P. Mounts, R. J. Gorelick, J. R. Casas-Finet, W. J. Bosche, L. E. Henderson, D. J. Waters, and L. O. Arthur.** 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:**533-43.

106. **Dube, M., M. G. Bego, C. Paquay, and E. A. Cohen.** 2010. Modulation of HIV-1-host interaction: role of the Vpu accessory protein. Retrovirology **7:**114.

107. **Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davies.** 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science **266:**1981-6.

108. **Egele, C., E. Schaub, N. Ramalanjaona, E. Piemont, D. Ficheux, B. Roques, J. L. Darlix, and Y. Mely.** 2004. HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. J Mol Biol **342:**453-66.

109. **Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J. P. Ebel, and B. Ehresmann.** 1987. Probing the structure of RNAs in solution. Nucleic Acids Res **15:**9109-28.

110. **Engelman, A., and P. Cherepanov.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol **10:**279-90.

111. **Fabris, D., J. P. Marino, and S. F. Le Grice.** 2009. Revisiting plus-strand DNA synthesis in retroviruses and long terminal repeat retrotransposons: dynamics of enzyme: substrate interactions. Viruses **1:**657-77.

112. **Faure, A., C. Calmels, C. Desjobert, M. Castroviejo, A. Caumont-Sarcos, L. Tarrago-Litvak, S. Litvak, and V. Parissi.** 2005. HIV-1 integrase crosslinked oligomers are active in vitro. Nucleic Acids Res **33:**977-86.

113. **Fedoroff, O., M. Salazar, and B. R. Reid.** 1993. Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. J Mol Biol **233:**509-23.

114. **Feng, S., and E. C. Holland.** 1988. HIV-1 tat trans-activation requires the loop sequence within tar. Nature **334:**165-7.

115. **Fernandes, J., B. Jayaraman, and A. Frankel.** 2012. The HIV-1 rev response element: An RNA scaffold that directs the cooperative assembly of a homo-oligomeric ribonucleoprotein complex. RNA Biol **9**.

116. **Fisher, R. J., M. J. Fivash, A. G. Stephen, N. A. Hagan, S. R. Shenoy, M. V. Medaglia, L. R. Smith, K. M. Worthy, J. T. Simpson, R. Shoemaker, K. L. McNitt, D. G. Johnson, C. V. Hixson, R. J. Gorelick, D. Fabris, L. E. Henderson, and A. Rein.** 2006. Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. Nucleic Acids Res **34:**472-84.

117. **Fisher, R. J., A. Rein, M. Fivash, M. A. Urbaneja, J. R. Casas-Finet, M. Medaglia, and L. E. Henderson.** 1998. Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. J Virol **72:**1902-9.

118. **Fisher, T. S., T. Darden, and V. R. Prasad.** 2003. Substitutions at Phe61 in the beta3-beta4 hairpin of HIV-1 reverse transcriptase reveal a role for the Fingers subdomain in strand displacement DNA synthesis. J Mol Biol **325:**443-59.

119. **Fosse, P., N. Motte, A. Roumier, C. Gabus, D. Muriaux, J. L. Darlix, and J. Paoletti.** 1996. A short autocomplementary sequence plays an essential role in avian sarcoma-leukosis virus RNA dimerization. Biochemistry **35:**16601-9.

120. **Foster, J. L., S. J. Denial, B. R. Temple, and J. V. Garcia.** 2011. Mechanisms of HIV-1 Nef function and intracellular signaling. J Neuroimmune Pharmacol **6:**230-46.

121. **Fox, K. R., and M. J. Waring.** 2001. High-resolution footprinting studies of drug-DNA complexes using chemical and enzymatic probes. Methods Enzymol **340:**412-30.

122. **Fu, T. B., and J. Taylor.** 1992. When retroviral reverse transcriptases reach the end of their RNA templates. J Virol **66:**4271-8.

123. **Fu, W., Q. Dang, K. Nagashima, E. O. Freed, V. K. Pathak, and W. S. Hu.** 2006. Effects of Gag mutation and processing on retroviral dimeric RNA maturation. J Virol **80:**1242-9.

124. **Fu, W., R. J. Gorelick, and A. Rein.** 1994. Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. J Virol **68:**5013-8.

125. **Fu, W., and A. Rein.** 1993. Maturation of dimeric viral RNA of Moloney murine leukemia virus. J Virol **67:**5443-9.

126. **Fuller, S. D., T. Wilk, B. E. Gowen, H. G. Krausslich, and V. M. Vogt.** 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. Curr Biol **7:**729-38.

127. **Furfine, E. S., and J. E. Reardon.** 1991. Reverse transcriptase.RNase H from the human immunodeficiency virus. Relationship of the DNA polymerase and RNA hydrolysis activities. J Biol Chem **266:**406-12.

128. **Galetto, R., and M. Negroni.** 2005. Mechanistic features of recombination in HIV. AIDS Rev **7:**92-102.

129. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, and et al.** 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science **224:**500-3.

130. **Gamble, T. R., F. F. Vajdos, S. Yoo, D. K. Worthylake, M. Houseweart, W. I. Sundquist, and C. P. Hill.** 1996. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell **87:**1285-94.

131. **Ganser-Pornillos, B. K., M. Yeager, and O. Pornillos.** 2012. Assembly and architecture of HIV. Adv Exp Med Biol **726:**441-65.

132. **Ganser-Pornillos, B. K., M. Yeager, and W. I. Sundquist.** 2008. The structural biology of HIV assembly. Curr Opin Struct Biol **18:**203-17.

133. **Gao, F., E. Bailes, D. L. Robertson, Y. Chen, C. M. Rodenburg, S. F. Michael, L. B. Cummins, L. O. Arthur, M. Peeters, G. M. Shaw, P. M. Sharp, and B. H. Hahn.** 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature **397:**436-41.

134. **Gao, F., L. Yue, A. T. White, P. G. Pappas, J. Barchue, A. P. Hanson, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 1992. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature **358:**495-9.

135. **Gao, K., R. J. Gorelick, D. G. Johnson, and F. Bushman.** 2003. Cofactors for human immunodeficiency virus type 1 cDNA integration in vitro. J Virol **77:**1598-603.

136. **Gao, L., M. Balakrishnan, B. P. Roques, and R. A. Bambara.** 2007. Insights into the multiple roles of pausing in HIV-1 reverse transcriptase-promoted strand transfers. J Biol Chem **282:**6222-31.

137. **Garrett, E. D., L. S. Tiley, and B. R. Cullen.** 1991. Rev activates expression of the human immunodeficiency virus type 1 vif and vpr gene products. J Virol **65:**1653-7.

138. **Gatignol, A.** 2007. Transcription of HIV: Tat and cellular chromatin. Adv Pharmacol **55:**137-59.

139. **Gatignol, A., and K. T. Jeang.** 2000. Tat as a transcriptional activator and a potential therapeutic target for HIV-1. Adv Pharmacol **48:**209-27.

140. **Gherghe, C. M., Z. Shajani, K. A. Wilkinson, G. Varani, and K. M. Weeks.** 2008. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. J Am Chem Soc **130:**12244-5.

141. **Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell **18:**93-100.

142. **Gilmartin, G. M., E. S. Fleming, J. Oetjen, and B. R. Graveley.** 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.

143. **Godet, J., H. de Rocquigny, C. Raja, N. Glasser, D. Ficheux, J. L. Darlix, and Y. Mely.** 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:**1180-92.

144. **Godet, J., and Y. Mely.** 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. RNA Biol **7:**687-99.

145. **Goldschmidt, V., J. Didierjean, B. Ehresmann, C. Ehresmann, C. Isel, and R. Marquet.** 2006. Mg2+ dependency of HIV-1 reverse transcription, inhibition by nucleoside analogues and resistance. Nucleic Acids Res **34:**42-52.

146. **Goldschmidt, V., C. Ehresmann, B. Ehresmann, and R. Marquet.** 2003. Does the HIV-1 primer activation signal interact with tRNA3(Lys) during the initiation of reverse transcription? Nucleic Acids Res **31:**850-9.

147. **Goldschmidt, V., J. C. Paillart, M. Rigourd, B. Ehresmann, A. M. Aubertin, C. Ehresmann, and R. Marquet.** 2004. Structural variability of the initiation complex of HIV-1 reverse transcription. J Biol Chem **279:**35923-31.

148. **Goldschmidt, V., M. Rigourd, C. Ehresmann, S. F. Le Grice, B. Ehresmann, and R. Marquet.** 2002. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. J Biol Chem **277:**43233-42.

149. **Golinelli, M. P., and S. H. Hughes.** 2001. Self-priming of retroviral minus-strand strong-stop DNAs. Virology **285:**278-90.

150. **Gopalakrishnan, V., J. A. Peliska, and S. J. Benkovic.** 1992. Human immunodeficiency virus type 1 reverse transcriptase: spatial and temporal relationship between the polymerase and RNase H activities. Proc Natl Acad Sci U S A **89:**10763-7.

151. **Gorelick, R. J., T. D. Gagliardi, W. J. Bosche, T. A. Wiltrout, L. V. Coren, D. J. Chabot, J. D. Lifson, L. E. Henderson, and L. O. Arthur.** 1999. Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. Virology **256:**92-104.

152. **Greene, W. C., and B. M. Peterlin.** 2002. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. Nat Med **8:**673-80.

153. **Griffiths, D. J.** 2001. Endogenous retroviruses in the human genome sequence. Genome Biol **2:**REVIEWS1017.

154. **Grigorov, B., D. Decimo, F. Smagulova, C. Pechoux, M. Mougel, D. Muriaux, and J. L. Darlix.** 2007. Intracellular HIV-1 Gag localization is impaired by mutations in the nucleocapsid zinc fingers. Retrovirology **4:**54.

155. **Guiot, E., K. Carayon, O. Delelis, F. Simon, P. Tauc, E. Zubin, M. Gottikh, J. F. Mouscadet, J. C. Brochon, and E. Deprez.** 2006. Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. J Biol Chem **281:**22707-19.

156. **Guo, F., J. Saadatmand, M. Niu, and L. Kleiman.** 2009. Roles of Gag and NCp7 in facilitating tRNA(Lys)(3) Annealing to viral RNA in human immunodeficiency virus type 1. J Virol **83:**8099-107.

157. **Guo, J., L. E. Henderson, J. Bess, B. Kane, and J. G. Levin.** 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:**5178-88.

158. **Guo, J., T. Wu, J. Anderson, B. F. Kane, D. G. Johnson, R. J. Gorelick, L. E. Henderson, and J. G. Levin.** 2000. Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. J Virol **74:**8980-8.

159. **Guo, J., T. Wu, B. F. Kane, D. G. Johnson, L. E. Henderson, R. J. Gorelick, and J. G. Levin.** 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:**4370-8.

160. **Hahn, B. H., G. M. Shaw, K. M. De Cock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. Science **287:**607-14.

161. **Hameau, L., J. Jeusset, S. Lafosse, D. Coulaud, E. Delain, T. Unge, T. Restle, E. Le Cam, and G. Mirambeau.** 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:**3301-13.

162. **Hampshire, A. J., D. A. Rusling, V. J. Broughton-Head, and K. R. Fox.** 2007. Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. Methods **42:**128-40.

163. **Hanson, M. N., M. Balakrishnan, B. P. Roques, and R. A. Bambara.** 2005. Effects of donor and acceptor RNA structures on the mechanism of strand transfer by HIV-1 reverse transcriptase. J Mol Biol **353:**772-87.

164. **Harrich, D., C. W. Hooker, and E. Parry.** 2000. The human immunodeficiency virus type 1 TAR RNA upper stem-loop plays distinct roles in reverse transcription and RNA packaging. J Virol **74:**5639-46.

165. **Harrich, D., C. Ulich, and R. B. Gaynor.** 1996. A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. J Virol **70:**4017-27.

166. **Heilman-Miller, S. L., J. Pan, D. Thirumalai, and S. A. Woodson.** 2001. Role of counterion condensation in folding of the Tetrahymena ribozyme. II. Counterion-dependence of folding kinetics. J Mol Biol **309:**57-68.

167. **Heilman-Miller, S. L., D. Thirumalai, and S. A. Woodson.** 2001. Role of counterion condensation in folding of the Tetrahymena ribozyme. I. Equilibrium stabilization by cations. J Mol Biol **306:**1157-66.

168. **Heilman-Miller, S. L., T. Wu, and J. G. Levin.** 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:**44154-65.

169. **Heinzinger, N. K., M. I. Bukinsky, S. A. Haggerty, A. M. Ragland, V. Kewalramani, M. A. Lee, H. E. Gendelman, L. Ratner, M. Stevenson, and M. Emerman.** 1994. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proc Natl Acad Sci U S A **91:**7311-5.

170. **Herschhorn, A., and A. Hizi.** 2010. Retroviral reverse transcriptases. Cell Mol Life Sci **67:**2717-47.

171. **Hill, C. P., D. Worthylake, D. P. Bancroft, A. M. Christensen, and W. I. Sundquist.** 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. Proc Natl Acad Sci U S A **93:**3099-104.

172. **Hiscott, J., H. Kwon, and P. Genin.** 2001. Hostile takeovers: viral appropriation of the NF-kappaB pathway. J Clin Invest **107:**143-51.

173. **Holmes, R. K., F. A. Koning, K. N. Bishop, and M. H. Malim.** 2007. APOBEC3F can inhibit the accumulation of HIV-1 reverse transcription products in the absence of hypermutation. Comparisons with APOBEC3G. J Biol Chem **282:**2587-95.

174. **Holstege, F. C., and H. T. Timmers.** 1997. Analysis of open complex formation during RNA polymerase II transcription initiation using heteroduplex templates and potassium permanganate probing. Methods **12:**203-11.

175. **Hong, M. K., E. J. Harbron, D. B. O'Connor, J. Guo, P. F. Barbara, J. G. Levin, and K. Musier-Forsyth.** 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.

176.   **Hooker, C. W., and D. Harrich.** 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:**229-38.

177.   **Houzet, L., Z. Morichaud, L. Didierlaurent, D. Muriaux, J. L. Darlix, and M. Mougel.** 2008. Nucleocapsid mutations turn HIV-1 into a DNA-containing virus. Nucleic Acids Res **36:**2311-9.

178.   **Houzet, L., J. C. Paillart, F. Smagulova, S. Maurel, Z. Morichaud, R. Marquet, and M. Mougel.** 2007. HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. Nucleic Acids Res **35:**2695-704.

179.   **Hu, C., D. T. Saenz, H. J. Fadel, W. Walker, M. Peretz, and E. M. Poeschla.** 2010. The HIV-1 central polypurine tract functions as a second line of defense against APOBEC3G/F. J Virol **84:**11981-93.

180.   **Hu, W. S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci U S A **87:**1556-60.

181.   **Hu, W. S., and H. M. Temin.** 1990. Retroviral recombination and reverse transcription. Science **250:**1227-33.

182.   **Huang, H., R. Chopra, G. L. Verdine, and S. C. Harrison.** 1998. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science **282:**1669-75.

183.   **Huang, Y., J. Wang, A. Shalom, Z. Li, A. Khorchid, M. A. Wainberg, and L. Kleiman.** 1997. Primer tRNA3Lys on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. J Virol **71:**726-8.

184.   **Huang, Z., and J. W. Szostak.** 1996. A simple method for 3'-labeling of RNA. Nucleic Acids Res **24:**4360-1.

185.   **Huber, H. E., and C. C. Richardson.** 1990. Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus 1 reverse transcriptase. J Biol Chem **265:**10565-73.

186.   **Hwang, C. K., E. S. Svarovskaia, and V. K. Pathak.** 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 U S A **98:**12209-14.

187.   **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253:**71-4.

188.   **Isel, C., C. Ehresmann, G. Keith, B. Ehresmann, and R. Marquet.** 1995. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). J Mol Biol **247:**236-50.

189.   **Isel, C., J. M. Lanchy, S. F. Le Grice, C. Ehresmann, B. Ehresmann, and R. Marquet.** 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:**917-24.

190.   **Isel, C., R. Marquet, G. Keith, C. Ehresmann, and B. Ehresmann.** 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:**25269-72.

191. **Isel, C., E. Westhof, C. Massire, S. F. Le Grice, B. Ehresmann, C. Ehresmann, and R. Marquet.** 1999. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. EMBO J **18:**1038-48.

192. **Iwatani, Y., A. E. Rosen, J. Guo, K. Musier-Forsyth, and J. G. Levin.** 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:**14185-95.

193. **Jackson, P. K., A. G. Eldridge, E. Freed, L. Furstenthal, J. Y. Hsu, B. K. Kaiser, and J. D. Reimann.** 2000. The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. Trends Cell Biol **10:**429-39.

194. **Jacob, D. T., and J. J. DeStefano.** 2008. A new role for HIV nucleocapsid protein in modulating the specificity of plus strand priming. Virology **378:**385-96.

195. **Jacobo-Molina, A., J. Ding, R. G. Nanni, A. D. Clark, Jr., X. Lu, C. Tantillo, R. L. Williams, G. Kamer, A. L. Ferris, P. Clark, and et al.** 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc Natl Acad Sci U S A **90:**6320-4.

196. **Javanbakht, H., W. Yuan, D. F. Yeung, B. Song, F. Diaz-Griffero, Y. Li, X. Li, M. Stremlau, and J. Sodroski.** 2006. Characterization of TRIM5alpha trimerization and its contribution to human immunodeficiency virus capsid binding. Virology **353:**234-46.

197. **Jern, P., G. O. Sperber, and J. Blomberg.** 2005. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology **2:**50.

198. **Jiang, M., J. Mak, A. Ladha, E. Cohen, M. Klein, B. Rovinski, and L. Kleiman.** 1993. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. J Virol **67:**3246-53.

199. **Jin, X., A. Brooks, H. Chen, R. Bennett, R. Reichman, and H. Smith.** 2005. APOBEC3G/CEM15 (hA3G) mRNA levels associate inversely with human immunodeficiency virus viremia. J Virol **79:**11513-6.

200. **Jin, X., H. Wu, and H. Smith.** 2007. APOBEC3G levels predict rates of progression to AIDS. Retrovirology **4:**20.

201. **Johnson, P. E., R. B. Turner, Z. R. Wu, L. Hairston, J. Guo, J. G. Levin, and M. F. Summers.** 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. Biochemistry **39:**9084-91.

202. **Jones, K. A., and B. M. Peterlin.** 1994. Control of RNA initiation and elongation at the HIV-1 promoter. Annu Rev Biochem **63:**717-43.

203. **Joyce, C. M., and S. J. Benkovic.** 2004. DNA polymerase fidelity: kinetics, structure, and checkpoints. Biochemistry **43:**14317-24.

204. **Julias, J. G., M. J. McWilliams, S. G. Sarafianos, W. G. Alvord, E. Arnold, and S. H. Hughes.** 2004. Effects of mutations in the G tract of the human immunodeficiency virus type 1 polypurine tract on virus replication and RNase H cleavage. J Virol **78:**13315-24.

205. **Kanevsky, I., F. Chaminade, D. Ficheux, A. Moumen, R. Gorelick, M. Negroni, J. L. Darlix, and P. Fosse.** 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. J Mol Biol **348:**1059-77.

206. **Kao, S. Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin.** 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature **330:**489-93.

207. **Kaplan, A. H., M. Manchester, and R. Swanstrom.** 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:**6782-6.

208. **Kar, A. K., F. Diaz-Griffero, Y. Li, X. Li, and J. Sodroski.** 2008. Biochemical and biophysical characterization of a chimeric TRIM21-TRIM5alpha protein. J Virol **82:**11669-81.

209. **Karkkainen, S., M. Hiipakka, J. H. Wang, I. Kleino, M. Vaha-Jaakkola, G. H. Renkema, M. Liss, R. Wagner, and K. Saksela.** 2006. Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. EMBO Rep **7:**186-91.

210. **Karn, J., and C. M. Stoltzfus.** 2012. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. Cold Spring Harb Perspect Med **2:**a006916.

211. **Kati, W. M., K. A. Johnson, L. F. Jerva, and K. S. Anderson.** 1992. Mechanism and fidelity of HIV reverse transcriptase. J Biol Chem **267:**25988-97.

212. **Kaumanns, P., I. Hagmann, and M. T. Dittmar.** 2006. Human TRIM5alpha mediated restriction of different HIV-1 subtypes and Lv2 sensitive and insensitive HIV-2 variants. Retrovirology **3:**79.

213. **Kelleher, C. D., and J. J. Champoux.** 1998. Characterization of RNA strand displacement synthesis by Moloney murine leukemia virus reverse transcriptase. J Biol Chem **273:**9976-86.

214. **Kenyon, J. C., S. J. Tanner, M. Legiewicz, P. S. Phillip, T. A. Rizvi, S. F. Le Grice, and A. M. Lever.** 2011. SHAPE analysis of the FIV Leader RNA reveals a structural switch potentially controlling viral packaging and genome dimerization. Nucleic Acids Res **39:**6692-704.

215. **Kjems, J., J. Egebjerg, and J. Christiansen.** 1998. Preparation and analysis of RNA-protein complexes *in vitro*., p. 83-179. *In* P. C. Van der Vliet (ed.), *Analysis of RNA-Protein Complexes In Vitro*. Elsevier, Utrecht, The Netherlands.

216. **Klumpp, K., J. Q. Hang, S. Rajendran, Y. Yang, A. Derosier, P. Wong Kai In, H. Overton, K. E. Parkes, N. Cammack, and J. A. Martin.** 2003. Two-metal ion mechanism of RNA cleavage by HIV RNase H and mechanism-based design of selective HIV RNase H inhibitors. Nucleic Acids Res **31:**6852-9.

217. **Koculi, E., N. K. Lee, D. Thirumalai, and S. A. Woodson.** 2004. Folding of the Tetrahymena ribozyme by polyamines: importance of counterion valence and size. J Mol Biol **341:**27-36.

218. **Kogan, M., and J. Rappaport.** 2011. HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. Retrovirology **8:**25.

219. **Kohlstaedt, L. A., J. Wang, J. M. Friedman, P. A. Rice, and T. A. Steitz.** 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science **256:**1783-90.

220. **Kuiken, C., B. Foley, E. Freed, B. Hahn, P. Marx, F. McCutchan, J. W. Mellors, S. Wolinsky, and K. B.** 2002. In *HIV sequence Compendium 2002* (Theoretical Biology and Biophysics Group, ed.).

221. **Kuiken, C., T. Leitner, B. Foley, B. Hahn, P. Marx, F. McCutchan, S. Wolinsky, and B. Korber (ed.).** 2008. HIV Sequence Compendium 2008. Los Alamos National Laboratory, Theoretical Biology and Biohysics, Los Alamos, New Mexico.

222. **Kulpa, D., R. Topping, and A. Telesnitsky.** 1997. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription

and identification of strand transfer-associated reverse transcriptase errors. EMBO J **16:**856-65.

223. **Kvaratskhelia, M., S. R. Budihas, and S. F. Le Grice.** 2002. Pre-existing distortions in nucleic acid structure aid polypurine tract selection by HIV-1 reverse transcriptase. J Biol Chem **277:**16689-96.

224. **Lahouassa, H., W. Daddacha, H. Hofmann, D. Ayinde, E. C. Logue, L. Dragin, N. Bloch, C. Maudet, M. Bertrand, T. Gramberg, G. Pancino, S. Priet, B. Canard, N. Laguette, M. Benkirane, C. Transy, N. R. Landau, B. Kim, and F. Margottin-Goguet.** 2012. SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. Nat Immunol **13:**223-8.

225. **Lanchy, J. M., C. Isel, C. Ehresmann, R. Marquet, and B. Ehresmann.** 1996. Structural and functional evidence that initiation and elongation of HIV-1 reverse transcription are distinct processes. Biochimie **78:**1087-96.

226. **Lanchy, J. M., C. Isel, G. Keith, S. F. Le Grice, C. Ehresmann, B. Ehresmann, and R. Marquet.** 2000. Dynamics of the HIV-1 reverse transcription complex during initiation of DNA synthesis. J Biol Chem **275:**12306-12.

227. **Lanciault, C., and J. J. Champoux.** 2005. Effects of unpaired nucleotides within HIV-1 genomic secondary structures on pausing and strand transfer. J Biol Chem **280:**2413-23.

228. **Langelier, C. R., V. Sandrin, D. M. Eckert, D. E. Christensen, V. Chandrasekaran, S. L. Alam, C. Aiken, J. C. Olsen, A. K. Kar, J. G. Sodroski, and W. I. Sundquist.** 2008. Biochemical characterization of a recombinant TRIM5alpha protein that restricts human immunodeficiency virus type 1 replication. J Virol **82:**11682-94.

229. **Lapadat-Tapolsky, M., H. De Rocquigny, D. Van Gent, B. Roques, R. Plasterk, and J. L. Darlix.** 1993. Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. Nucleic Acids Res **21:**831-9.

230. **Lapadat-Tapolsky, M., C. Gabus, M. Rau, and J. L. Darlix.** 1997. Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability. J Mol Biol **268:**250-60.

231. **Lapadat-Tapolsky, M., C. Pernelle, C. Borie, and J. L. Darlix.** 1995. Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. Nucleic Acids Res **23:**2434-41.

232. **Laughrea, M., and L. Jette.** 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:**1589-98.

233. **Laughrea, M., N. Shen, L. Jette, J. L. Darlix, L. Kleiman, and M. A. Wainberg.** 2001. Role of distal zinc finger of nucleocapsid protein in genomic RNA dimerization of human immunodeficiency virus type 1; no role for the palindrome crowning the R-U5 hairpin. Virology **281:**109-16.

234. **Le Cam, E., D. Coulaud, E. Delain, P. Petitjean, B. P. Roques, D. Gerard, E. Stoylova, C. Vuilleumier, S. P. Stoylov, and Y. Mely.** 1998. Properties and growth mechanism of the ordered aggregation of a model RNA by the HIV-1 nucleocapsid protein: an electron microscopy investigation. Biopolymers **45:**217-29.

235. **Lecossier, D., F. Bouchonnet, F. Clavel, and A. J. Hance.** 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science **300:**1112.

236. **Lee, C. H., K. Saksela, U. A. Mirza, B. T. Chait, and J. Kuriyan.** 1996. Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell **85:**931-42.

237. **Lener, D., V. Tanchou, B. P. Roques, S. F. Le Grice, and J. L. Darlix.** 1998. Involvement of HIV-I nucleocapsid protein in the recruitment of reverse transcriptase into nucleoprotein complexes formed in vitro. J Biol Chem **273:**33781-6.

238. **Lever, A., H. Gottlinger, W. Haseltine, and J. Sodroski.** 1989. Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. J Virol **63:**4085-7.

239. **Levin, J. G., J. Guo, I. Rouzina, and K. Musier-Forsyth.** 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:**217-86.

240. **Levin, J. G., M. Mitra, A. Mascarenhas, and K. Musier-Forsyth.** 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. RNA Biol **7:**754-74.

241. **Li, X., and J. Sodroski.** 2008. The TRIM5alpha B-box 2 domain promotes cooperative binding to the retroviral capsid by mediating higher-order self-association. J Virol **82:**11495-502.

242. **Liu, H. W., G. Cosa, C. F. Landes, Y. Zeng, B. J. Kovaleski, D. G. Mullen, G. Barany, K. Musier-Forsyth, and P. F. Barbara.** 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:**3470-9.

243. **Liu, H. W., Y. Zeng, C. F. Landes, Y. J. Kim, Y. Zhu, X. Ma, M. N. Vo, K. Musier-Forsyth, and P. F. Barbara.** 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:**5261-7.

244. **Lobritz, M. A., A. N. Ratcliff, and E. J. Arts.** 2010. HIV-1 Entry, Inhibitors, and Resistance. Viruses **2:**1069-105.

245. **Lori, F., F. di Marzo Veronese, A. L. de Vico, P. Lusso, M. S. Reitz, Jr., and R. C. Gallo.** 1992. Viral DNA carried by human immunodeficiency virus type 1 virions. J Virol **66:**5067-74.

246. **Lu, K., X. Heng, L. Garyu, S. Monti, E. L. Garcia, S. Kharytonchyk, B. Dorjsuren, G. Kulandaivel, S. Jones, A. Hiremath, S. S. Divakaruni, C. LaCotti, S. Barton, D. Tummillo, A. Hosic, K. Edme, S. Albrecht, A. Telesnitsky, and M. F. Summers.** 2011. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. Science **334:**242-5.

247. **Lu, K., X. Heng, and M. F. Summers.** 2011. Structural determinants and mechanism of HIV-1 genome packaging. J Mol Biol **410:**609-33.

248. **Luban, J., and S. P. Goff.** 1994. Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. J Virol **68:**3784-93.

249. **Luo, G. X., and J. Taylor.** 1990. Template switching by reverse transcriptase during DNA synthesis. J Virol **64:**4321-8.

250. **Mak, J., and L. Kleiman.** 1997. Primer tRNAs for reverse transcription. J Virol **71:**8087-95.

251. **Mansky, L. M., and H. M. Temin.** 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:**5087-94.

252. **Marchand, B., and M. Gotte.** 2003. Site-specific footprinting reveals differences in the translocation status of HIV-1 reverse transcriptase. Implications for polymerase translocation and drug resistance. J Biol Chem **278:**35362-72.

253. **Marchand, B., E. P. Tchesnokov, and M. Gotte.** 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:**3337-46.

254. **Marquet, R., C. Isel, C. Ehresmann, and B. Ehresmann.** 1995. tRNAs as primer of reverse transcriptases. Biochimie **77:**113-24.

255. **Martin-Serrano, J., and S. J. Neil.** 2011. Host factors involved in retroviral budding and release. Nat Rev Microbiol **9:**519-31.

256. **Massiah, M. A., M. R. Starich, C. Paschall, M. F. Summers, A. M. Christensen, and W. I. Sundquist.** 1994. Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. J Mol Biol **244:**198-223.

257. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol **65:**499-560.

258. **McColl, D. J., and X. Chen.** 2010. Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy. Antiviral Res **85:**101-18.

259. **McLaren, M., K. Marsh, and A. Cochrane.** 2008. Modulating HIV-1 RNA processing and utilization. Front Biosci **13:**5693-707.

260. **Meiering, C. D., and M. L. Linial.** 2001. Historical perspective of foamy virus epidemiology and infection. Clin Microbiol Rev **14:**165-76.

261. **Merino, E. J., K. A. Wilkinson, J. L. Coughlan, and K. M. Weeks.** 2005. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc **127:**4223-31.

262. **Miles, L. R., B. E. Agresta, M. B. Khan, S. Tang, J. G. Levin, and M. D. Powell.** 2005. Effect of polypurine tract (PPT) mutations on human immunodeficiency virus type 1 replication: a virus with a completely randomized PPT retains low infectivity. J Virol **79:**6859-67.

263. **Miller, M. D., C. M. Farnet, and F. D. Bushman.** 1997. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol **71:**5382-90.

264. **Miller, W. A., and K. A. White.** 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. Annu Rev Phytopathol **44:**447-67.

265. **Mirambeau, G., S. Lyonnais, D. Coulaud, L. Hameau, S. Lafosse, J. Jeusset, I. Borde, M. Reboud-Ravaux, T. Restle, R. J. Gorelick, and E. Le Cam.** 2007. HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. PLoS One **2:**e669.

266. **Mirambeau, G., S. Lyonnais, D. Coulaud, L. Hameau, S. Lafosse, J. Jeusset, A. Justome, E. Delain, R. J. Gorelick, and E. Le Cam.** 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.

267. **Mirambeau, G., S. Lyonnais, and R. J. Gorelick.** 2010. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. RNA Biol **7:**724-34.

268. **Mische, C. C., H. Javanbakht, B. Song, F. Diaz-Griffero, M. Stremlau, B. Strack, Z. Si, and J. Sodroski.** 2005. Retroviral restriction factor TRIM5alpha is a trimer. J Virol **79:**14446-50.

269. **Miyauchi, K., Y. Kim, O. Latinovic, V. Morozov, and G. B. Melikyan.** 2009. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. Cell **137:**433-44.

270. **Mortimer, S. A., and K. M. Weeks.** 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc **129:**4144-5.

271. **Moumen, A., L. Polomack, B. Roques, H. Buc, and M. Negroni.** 2001. The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. Nucleic Acids Res **29:**3814-21.

272. **Munk, C., S. M. Brandt, G. Lucero, and N. R. Landau.** 2002. A dominant block to HIV-1 replication at reverse transcription in simian cells. Proc Natl Acad Sci U S A **99:**13843-8.

273. **Muriaux, D., and J. L. Darlix.** 2010. Properties and functions of the nucleocapsid protein in virus assembly. RNA Biol **7:**744-53.

274. **Muriaux, D., J. L. Darlix, and A. Cimarelli.** 2004. Targeting the assembly of the human immunodeficiency virus type I. Curr Pharm Des **10:**3725-39.

275. **Muriaux, D., H. De Rocquigny, B. P. Roques, and J. Paoletti.** 1996. NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. J Biol Chem **271:**33686-92.

276. **Muriaux, D., P. Fosse, and J. Paoletti.** 1996. A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. Biochemistry **35:**5075-82.

277. **Muriaux, D., J. Mirro, D. Harvin, and A. Rein.** 2001. RNA is a structural element in retrovirus particles. Proc Natl Acad Sci U S A **98:**5246-51.

278. **Muriaux, D., J. Mirro, K. Nagashima, D. Harvin, and A. Rein.** 2002. Murine leukemia virus nucleocapsid mutant particles lacking viral RNA encapsidate ribosomes. J Virol **76:**11405-13.

279. **Murphy F.A., Fauquet C.M., Biship D.H.L, Ghabrial S.A., Javis A.W., Martelli G.P., Mayo M.A., and S. M.D.** 1994. Virus taxonomy: The classification and nomenclature of viruses, Retroviridae, Springer-Verlag, Vienna.

280. **Nakayama, E. E., H. Maegawa, and T. Shioda.** 2006. A dominant-negative effect of cynomolgus monkey tripartite motif protein TRIM5alpha on anti-simian immunodeficiency virus SIVmac activity of an African green monkey orthologue. Virology **350:**158-63.

281. **Nakayama, E. E., and T. Shioda.** 2012. TRIM5alpha and Species Tropism of HIV/SIV. Front Microbiol **3:**13.

282. **Navia, M. A., P. M. Fitzgerald, B. M. McKeever, C. T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke, and J. P. Springer.** 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature **337:**615-20.

283. **Negroni, M., and H. Buc.** 2000. Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. Proc Natl Acad Sci U S A **97:**6385-90.

284. **Negroni, M., and H. Buc.** 2001. Mechanisms of retroviral recombination. Annu Rev Genet **35:**275-302.

285. **Negroni, M., and H. Buc.** 1999. Recombination during reverse transcription: an evaluation of the role of the nucleocapsid protein. J Mol Biol **286:**15-31.

286. **Negroni, M., M. Ricchetti, P. Nouvel, and H. Buc.** 1995. Homologous recombination promoted by reverse transcriptase during copying of two distinct RNA templates. Proc Natl Acad Sci U S A **92:**6971-5.

287. **Nermut, M. V., and A. Fassati.** 2003. Structural analyses of purified human immunodeficiency virus type 1 intracellular reverse transcription complexes. J Virol **77:**8196-206.

288. **Newman, E. N., R. K. Holmes, H. M. Craig, K. C. Klein, J. R. Lingappa, M. H. Malim, and A. M. Sheehy.** 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol **15:**166-70.

289. **Nikolenko, G. N., E. S. Svarovskaia, K. A. Delviks, and V. K. Pathak.** 2004. Antiretroviral drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase increase template-switching frequency. J Virol **78:**8761-70.

290. **Nisole, S., C. Lynch, J. P. Stoye, and M. W. Yap.** 2004. A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. Proc Natl Acad Sci U S A **101:**13324-8.

291. **Nora, T., C. Charpentier, O. Tenaillon, C. Hoede, F. Clavel, and A. J. Hance.** 2007. Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol **81:**7620-8.

292. **Ohi, Y., and J. L. Clever.** 2000. Sequences in the 5' and 3' R elements of human immunodeficiency virus type 1 critical for efficient reverse transcription. J Virol **74:**8324-34.

293. **Onafuwa-Nuga, A., and A. Telesnitsky.** 2009. The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. Microbiol Mol Biol Rev **73:**451-80, Table of Contents.

294. **Ooms, M., T. E. Abbink, C. Pham, and B. Berkhout.** 2007. Circularization of the HIV-1 RNA genome. Nucleic Acids Res **35:**5253-61.

295. **Ooms, M., D. Cupac, T. E. Abbink, H. Huthoff, and B. Berkhout.** 2007. The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. Nucleic Acids Res **35:**1649-59.

296. **Pace, C., J. Keller, D. Nolan, I. James, S. Gaudieri, C. Moore, and S. Mallal.** 2006. Population level analysis of human immunodeficiency virus type 1 hypermutation and its relationship with APOBEC3G and vif genetic variation. J Virol **80:**9259-69.

297. **Paillart, J. C., M. Dettenhofer, X. F. Yu, C. Ehresmann, B. Ehresmann, and R. Marquet.** 2004. First snapshots of the HIV-1 RNA structure in infected cells and in virions. J Biol Chem **279:**48397-403.

298. **Paillart, J. C., M. Shehu-Xhilaga, R. Marquet, and J. Mak.** 2004. Dimerization of retroviral RNA genomes: an inseparable pair. Nat Rev Microbiol **2:**461-72.

299. **Panganiban, A. T., and D. Fiore.** 1988. Ordered interstrand and intrastrand DNA transfer during reverse transcription. Science **241:**1064-9.

300. **Pari, K., G. A. Mueller, E. F. DeRose, T. W. Kirby, and R. E. London.** 2003. Solution structure of the RNase H domain of the HIV-1 reverse transcriptase in the presence of magnesium. Biochemistry **42:**639-50.

301. **Parolin, C., T. Dorfman, G. Palu, H. Gottlinger, and J. Sodroski.** 1994. Analysis in human immunodeficiency virus type 1 vectors of cis-acting sequences that affect gene transfer into human lymphocytes. J Virol **68:**3888-95.

302. **Paulson, B. A., M. Zhang, S. J. Schultz, and J. J. Champoux.** 2007. Substitution of alanine for tyrosine-64 in the fingers subdomain of M-MuLV reverse transcriptase impairs strand displacement synthesis and blocks viral replication in vivo. Virology **366:**361-76.

303. **Pavon-Eternod, M., M. Wei, T. Pan, and L. Kleiman.** 2010. Profiling non-lysyl tRNAs in HIV-1. RNA **16:**267-73.

304. **Peliska, J. A., S. Balasubramanian, D. P. Giedroc, and S. J. Benkovic.** 1994. Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. Biochemistry **33:**13817-23.

305. **Peliska, J. A., and S. J. Benkovic.** 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science **258:**1112-8.

306. **Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science **271:**1582-6.

307. **Pettit, S. C., M. D. Moody, R. S. Wehbie, A. H. Kaplan, P. V. Nantermet, C. A. Klein, and R. Swanstrom.** 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:**8017-27.

308. **Pfeiffer, J. K., and A. Telesnitsky.** 2001. Effects of limiting homology at the site of intermolecular recombinogenic template switching during Moloney murine leukemia virus replication. J Virol **75:**11263-74.

309. **Piekna-Przybylska, D., and R. A. Bambara.** 2011. Requirements for efficient minus strand strong-stop DNA transfer in human immunodeficiency virus 1. RNA Biol **8:**230-6.

310. **Piekna-Przybylska, D., L. DiChiacchio, D. H. Mathews, and R. A. Bambara.** 2010. A sequence similar to tRNA 3 Lys gene is embedded in HIV-1 U3-R and promotes minus-strand transfer. Nat Struct Mol Biol **17:**83-9.

311. **Piekna-Przybylska, D., C. Dykes, L. M. Demeter, and R. A. Bambara.** 2011. Sequences in the U3 region of human immunodeficiency virus 1 improve efficiency of minus strand transfer in infected cells. Virology **410:**368-74.

312. **Polge, E., J. L. Darlix, J. Paoletti, and P. Fosse.** 2000. Characterization of loose and tight dimer forms of avian leukosis virus RNA. J Mol Biol **300:**41-56.

313. **Poljak, L., S. M. Batson, D. Ficheux, B. P. Roques, J. L. Darlix, and E. Kas.** 2003. Analysis of NCp7-dependent activation of HIV-1 cDNA integration and its conservation among retroviral nucleocapsid proteins. J Mol Biol **329:**411-21.

314. **Pommier, Y., A. A. Johnson, and C. Marchand.** 2005. Integrase inhibitors to treat HIV/AIDS. Nat Rev Drug Discov **4:**236-48.

315. **Pontius, B. W., and P. Berg.** 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 U S A **88:**8237-41.

316. **Pontius, B. W., and P. Berg.** 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 U S A **87:**8403-7.

317. **Poon, D. T., J. Wu, and A. Aldovini.** 1996. Charged amino acid residues of human immunodeficiency virus type 1 nucleocapsid p7 protein involved in RNA packaging and infectivity. J Virol **70:**6607-16.

318. **Pope, M., and A. T. Haase.** 2003. Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. Nat Med **9:**847-52.

319. **Pornillos, O., B. K. Ganser-Pornillos, and M. Yeager.** 2011. Atomic-level modelling of the HIV capsid. Nature **469:**424-7.

320. **Post, K., B. Kankia, S. Gopalakrishnan, V. Yang, E. Cramer, P. Saladores, R. J. Gorelick, J. Guo, K. Musier-Forsyth, and J. G. Levin.** 2009. Fidelity of plus-

strand priming requires the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. Nucleic Acids Res **37:**1755-66.

321. **Poulsen, A. G., P. Aaby, A. Gottschau, B. B. Kvinesdal, F. Dias, K. Molbak, and E. Lauritzen.** 1993. HIV-2 infection in Bissau, West Africa, 1987-1989: incidence, prevalences, and routes of transmission. J Acquir Immune Defic Syndr **6:**941-8.

322. **Powell, M. D., and J. G. Levin.** 1996. Sequence and structural determinants required for priming of plus-strand DNA synthesis by the human immunodeficiency virus type 1 polypurine tract. J Virol **70:**5288-96.

323. **Pullen, K. A., L. K. Ishimoto, and J. J. Champoux.** 1992. Incomplete removal of the RNA primer for minus-strand DNA synthesis by human immunodeficiency virus type 1 reverse transcriptase. J Virol **66:**367-73.

324. **Purcell, D. F., and M. A. Martin.** 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol **67:**6365-78.

325. **Rambaut, A., D. Posada, K. A. Crandall, and E. C. Holmes.** 2004. The causes and consequences of HIV evolution. Nat Rev Genet **5:**52-61.

326. **Rausch, J. W., and S. F. Le Grice.** 2004. 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. Int J Biochem Cell Biol **36:**1752-66.

327. **Reardon, J. E.** 1992. Human immunodeficiency virus reverse transcriptase: steady-state and pre-steady-state kinetics of nucleotide incorporation. Biochemistry **31:**4473-9.

328. **Rein, A.** 1994. Retroviral RNA packaging: a review. Arch Virol Suppl **9:**513-22.

329. **Roda, R. H., M. Balakrishnan, J. K. Kim, B. P. Roques, P. J. Fay, and R. A. Bambara.** 2002. Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. J Biol Chem **277:**46900-11.

330. **Roldan, A., O. U. Warren, R. S. Russell, C. Liang, and M. A. Wainberg.** 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:**17488-96.

331. **Saadatmand, J., M. Niu, L. Kleiman, and F. Guo.** 2009. The contribution of the primer activation signal to differences between Gag- and NCp7-facilitated tRNA(Lys3) annealing in HIV-1. Virology **391:**334-41.

332. **Sarafianos, S. G., K. Das, C. Tantillo, A. D. Clark, Jr., J. Ding, J. M. Whitcomb, P. L. Boyer, S. H. Hughes, and E. Arnold.** 2001. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J **20:**1449-61.

333. **Sarafianos, S. G., B. Marchand, K. Das, D. M. Himmel, M. A. Parniak, S. H. Hughes, and E. Arnold.** 2009. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol **385:**693-713.

334. **Sayah, D. M., E. Sokolskaja, L. Berthoux, and J. Luban.** 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature **430:**569-73.

335. **Schatz, O., F. V. Cromme, F. Gruninger-Leitch, and S. F. Le Grice.** 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:**311-4.

336. **Schroeder, R., A. Barta, and K. Semrad.** 2004. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol **5:**908-19.

337. **Schultz, S. J., and J. J. Champoux.** 2008. RNase H activity: structure, specificity, and function in reverse transcription. Virus Res **134:**86-103.

338. **Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyo, and G. N. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol **64:**2519-29.

339. **Sebastian, S., and J. Luban.** 2005. TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. Retrovirology **2:**40.

340. **Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418:**646-50.

341. **Shen, W., L. Gao, M. Balakrishnan, and R. A. Bambara.** 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:**33883-93.

342. **Shvadchak, V. V., A. S. Klymchenko, H. de Rocquigny, and Y. Mely.** 2009. Sensing peptide-oligonucleotide interactions by a two-color fluorescence label: application to the HIV-1 nucleocapsid protein. Nucleic Acids Res **37:**e25.

343. **Simsek, S., C. Ozcetin, and I. Balkaya.** 2011. Detection of polymorphism in AgB1 gene from sheep, cattle and human isolates of echinococcus granulosus by SSCP. Vet Parasitol **184:**352-5.

344. **Skripkin, E., J. C. Paillart, R. Marquet, B. Ehresmann, and C. Ehresmann.** 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:**4945-9.

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

346. **Smith, C. M., J. S. Smith, and M. J. Roth.** 1999. RNase H requirements for the second strand transfer reaction of human immunodeficiency virus type 1 reverse transcription. J Virol **73:**6573-81.

347. **Sodroski, J., M. Trus, D. Perkins, R. Patarca, F. Wong-Staal, E. Gelmann, R. Gallo, and W. A. Haseltine.** 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:**4617-21.

348. **Song, M., M. Balakrishnan, Y. Chen, B. P. Roques, and R. A. Bambara.** 2006. Stimulation of HIV-1 minus strand strong stop DNA transfer by genomic sequences 3' of the primer binding site. J Biol Chem **281:**24227-35.

349. **Song, M., V. P. Basu, M. N. Hanson, B. P. Roques, and R. A. Bambara.** 2008. Proximity and branch migration mechanisms in HIV-1 minus strand strong stop DNA transfer. J Biol Chem **283:**3141-50.

350. **Song, R., J. Kafaie, and M. Laughrea.** 2008. Role of the 5' TAR stem--loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. Biochemistry **47:**3283-93.

351. **Song, R., J. Kafaie, L. Yang, and M. Laughrea.** 2007. HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process; the DIS of stem-loop 1 initiates viral RNA dimerization. J Mol Biol **371:**1084-98.

352. **Spinardi, L., R. Mazars, and C. Theillet.** 1991. Protocols for an improved detection of point mutations by SSCP. Nucleic Acids Res **19:**4009.

353. **Steitz, T. A.** 1999. DNA polymerases: structural diversity and common mechanisms. J Biol Chem **274:**17395-8.

354. **Steitz, T. A.** 1998. A mechanism for all polymerases. Nature **391:**231-2.

355. **Stewart-Maynard, K. M., M. Cruceanu, F. Wang, M. N. Vo, R. J. Gorelick, M. C. Williams, I. Rouzina, and K. Musier-Forsyth.** 2008. Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. J Virol **82:**10129-42.

356. **Stoylov, S. P., C. Vuilleumier, E. Stoylova, H. De Rocquigny, B. P. Roques, D. Gerard, and Y. Mely.** 1997. Ordered aggregation of ribonucleic acids by the human immunodeficiency virus type 1 nucleocapsid protein. Biopolymers **41:**301-12.

357. **Strack, B., A. Calistri, S. Craig, E. Popova, and H. G. Gottlinger.** 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell **114:**689-99.

358. **Strebel, K., D. Daugherty, K. Clouse, D. Cohen, T. Folks, and M. A. Martin.** 1987. The HIV 'A' (sor) gene product is essential for virus infectivity. Nature **328:**728-30.

359. **Stremlau, M., C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, and J. Sodroski.** 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature **427:**848-53.

360. **Stremlau, M., M. Perron, M. Lee, Y. Li, B. Song, H. Javanbakht, F. Diaz-Griffero, D. J. Anderson, W. I. Sundquist, and J. Sodroski.** 2006. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proc Natl Acad Sci U S A **103:**5514-9.

361. **Subbramanian, R. A., and E. A. Cohen.** 1994. Molecular biology of the human immunodeficiency virus accessory proteins. J Virol **68:**6831-5.

362. **Svarovskaia, E. S., K. A. Delviks, C. K. Hwang, and V. K. Pathak.** 2000. Structural determinants of murine leukemia virus reverse transcriptase that affect the frequency of template switching. J Virol **74:**7171-8.

363. **Swanstrom, R., and J. W. Wills.** 1997. Synthesis, Assembly, and Processing of Viral Proteins.

364. **Szyperski, T., M. Gotte, M. Billeter, E. Perola, L. Cellai, H. Heumann, and K. Wuthrich.** 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:**343-55.

365. **Tanchou, V., T. Delaunay, M. Bodeus, B. Roques, J. L. Darlix, and R. Benarous.** 1995. Conformational changes between human immunodeficiency virus type 1 nucleocapsid protein NCp7 and its precursor NCp15 as detected by anti-NCp7 monoclonal antibodies. J Gen Virol **76 (Pt 10):**2457-66.

366. **Tanese, N., and S. P. Goff.** 1988. Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. Proc Natl Acad Sci U S A **85:**1777-81.

367. **Taylor, B. S., and S. M. Hammer.** 2008. The challenge of HIV-1 subtype diversity. N Engl J Med **359:**1965-6.

368. **Thomas, J. A., and R. J. Gorelick.** 2008. Nucleocapsid protein function in early infection processes. Virus Res **134:**39-63.

369. **Till, B. J., C. Burtner, L. Comai, and S. Henikoff.** 2004. Mismatch cleavage by single-strand specific nucleases. Nucleic Acids Res **32:**2632-41.

370. **Tintori, C., N. Veljkovic, V. Veljkovic, and M. Botta.** 2010. Computational studies of the interaction between the HIV-1 integrase tetramer and the cofactor LEDGF/p75: insights from molecular dynamics simulations and the informational spectrum method. Proteins **78:**3396-408.

371. **Tisne, C., B. P. Roques, and F. Dardel.** 2004. The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. J Biol Chem **279:**3588-95.

372. **Tisne, C., B. P. Roques, and F. Dardel.** 2001. Heteronuclear NMR studies of the interaction of tRNA(Lys)3 with HIV-1 nucleocapsid protein. J Mol Biol **306:**443-54.

373. **Tisne, C., B. P. Roques, and F. Dardel.** 2003. Specific recognition of primer tRNA Lys 3 by HIV-1 nucleocapsid protein: involvement of the zinc fingers and the N-terminal basic extension. Biochimie **85:**557-61.

374. **Towers, G. J., T. Hatziioannou, S. Cowan, S. P. Goff, J. Luban, and P. D. Bieniasz.** 2003. Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat Med **9:**1138-43.

375. **Trono, D.** 1995. HIV accessory proteins: leading roles for the supporting cast. Cell **82:**189-92.

376. **Trono, D.** 1992. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. J Virol **66:**4893-900.

377. **Tsuchihashi, Z., and P. O. Brown.** 1994. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. J Virol **68:**5863-70.

378. **Urbaneja, A., R. Hinarejos, E. Llacer, A. Garrido, and J. A. Jacas.** 2002. Effect of temperature on life history of Cirrospilus vittatus (Hymenoptera: Eulophidae), an ectoparasitoid of Phyllocnistis citrella (Lepidoptera: Gracillariidae). J Econ Entomol **95:**250-5.

379. **Urbaneja, M. A., M. Wu, J. R. Casas-Finet, and R. L. Karpel.** 2002. HIV-1 nucleocapsid protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. J Mol Biol **318:**749-64.

380. **Valente, S. T., G. M. Gilmartin, K. Venkatarama, G. Arriagada, and S. P. Goff.** 2009. HIV-1 mRNA 3' end processing is distinctively regulated by eIF3f, CDK11, and splice factor 9G8. Mol Cell **36:**279-89.

381. **Van Maele, B., K. Busschots, L. Vandekerckhove, F. Christ, and Z. Debyser.** 2006. Cellular co-factors of HIV-1 integration. Trends Biochem Sci **31:**98-105.

382. **van Wamel, J. L., and B. Berkhout.** 1998. The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. Virology **244:**245-51.

383. **Varmus, H. E., S. Heasley, H. J. Kung, H. Oppermann, V. C. Smith, J. M. Bishop, and P. R. Shank.** 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.

384. **Vasa, S. M., N. Guex, K. A. Wilkinson, K. M. Weeks, and M. C. Giddings.** 2008. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA **14:**1979-90.

385. **Vo, M. N., G. Barany, I. Rouzina, and K. Musier-Forsyth.** 2009. Effect of Mg(2+) and Na(+) on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. J Mol Biol **386:**773-88.

386. **Vo, M. N., G. Barany, I. Rouzina, and K. Musier-Forsyth.** 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.

387. **Vo, M. N., G. Barany, I. Rouzina, and K. Musier-Forsyth.** 2006. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. J Mol Biol **363:**244-61.

388. **Vogt, V. M.** 1997. Retroviral Virions and Genomes.

389. **von Schwedler, U. K., M. Stuchell, B. Muller, D. M. Ward, H. Y. Chung, E. Morita, H. E. Wang, T. Davis, G. P. He, D. M. Cimbora, A. Scott, H. G. Krausslich, J. Kaplan, S. G. Morham, and W. I. Sundquist.** 2003. The protein network of HIV budding. Cell **114:**701-13.

390. **Vuilleumier, C., E. Bombarda, N. Morellet, D. Gerard, B. P. Roques, and Y. Mely.** 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. Biochemistry **38:**16816-25.

391. **Wang, H., Y. S. Yeh, and P. F. Barbara.** 2009. HIV-1 nucleocapsid protein bends double-stranded nucleic acids. J Am Chem Soc **131:**15534-43.

392. **Wang, S. W., and A. Aldovini.** 2002. RNA incorporation is critical for retroviral particle integrity after cell membrane assembly of Gag complexes. J Virol **76:**11853-65.

393. **Warrilow, D., and D. Harrich.** 2007. HIV-1 replication from after cell entry to the nuclear periphery. Curr HIV Res **5:**293-9.

394. **Warrilow, D., L. Meredith, A. Davis, C. Burrell, P. Li, and D. Harrich.** 2008. Cell factors stimulate human immunodeficiency virus type 1 reverse transcription in vitro. J Virol **82:**1425-37.

395. **Watts, J. M., K. K. Dang, R. J. Gorelick, C. W. Leonard, J. W. Bess, Jr., R. Swanstrom, C. L. Burch, and K. M. Weeks.** 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature **460:**711-6.

396. **Welker, R., H. Hohenberg, U. Tessmer, C. Huckhagel, and H. G. Krausslich.** 2000. Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. J Virol **74:**1168-77.

397. **Whitcomb, J. M., R. Kumar, and S. H. Hughes.** 1990. Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. J Virol **64:**4903-6.

398. **Wilk, T., I. Gross, B. E. Gowen, T. Rutten, F. de Haas, R. Welker, H. G. Krausslich, P. Boulanger, and S. D. Fuller.** 2001. Organization of immature human immunodeficiency virus type 1. J Virol **75:**759-71.

399. **Wilkinson, K. A., R. J. Gorelick, S. M. Vasa, N. Guex, A. Rein, D. H. Mathews, M. C. Giddings, and K. M. Weeks.** 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biol **6:**e96.

400. **Willey, R. L., F. Maldarelli, M. A. Martin, and K. Strebel.** 1992. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. J Virol **66:**7193-200.

401. **Williams, M. C., I. Rouzina, J. R. Wenner, R. J. Gorelick, K. Musier-Forsyth, and V. A. Bloomfield.** 2001. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. Proc Natl Acad Sci U S A **98:**6121-6.

402. **Wissing, S., N. L. Galloway, and W. C. Greene.** 2010. HIV-1 Vif versus the APOBEC3 cytidine deaminases: an intracellular duel between pathogen and host restriction factors. Mol Aspects Med **31:**383-97.

403. **Wlodawer, A., M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. Kent.** 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science **245:**616-21.

404. **Wu, T., J. Guo, J. Bess, L. E. Henderson, and J. G. Levin.** 1999. Molecular requirements for human immunodeficiency virus type 1 plus-strand transfer: analysis in reconstituted and endogenous reverse transcription systems. J Virol **73:**4794-805.

405. **Wu, T., S. L. Heilman-Miller, and J. G. Levin.** 2007. Effects of nucleic acid local structure and magnesium ions on minus-strand transfer mediated by the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. Nucleic Acids Res **35:**3974-87.

406. **Wu, W., L. E. Henderson, T. D. Copeland, R. J. Gorelick, W. J. Bosche, A. Rein, and J. G. Levin.** 1996. Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. J Virol **70:**7132-42.

407. **Wurtzer, S., A. Goubard, F. Mammano, S. Saragosti, D. Lecossier, A. J. Hance, and F. Clavel.** 2006. Functional central polypurine tract provides downstream protection of the human immunodeficiency virus type 1 genome from editing by APOBEC3G and APOBEC3B. J Virol **80:**3679-83.

408. **Wyatt, R., and J. Sodroski.** 1998. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science **280:**1884-8.

409. **Yi-Brunozzi, H. Y., R. G. Brinson, D. M. Brabazon, D. Lener, S. F. Le Grice, and J. P. Marino.** 2008. High-resolution NMR analysis of the conformations of native and base analog substituted retroviral and LTR-retrotransposon PPT primers. Chem Biol **15:**254-62.

410. **You, J. C., and C. S. McHenry.** 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. J Biol Chem **269:**31491-5.

411. **Yu, H., A. E. Jetzt, Y. Ron, B. D. Preston, and J. P. Dougherty.** 1998. The nature of human immunodeficiency virus type 1 strand transfers. J Biol Chem **273:**28384-91.

412. **Yu, X., Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao, and X. F. Yu.** 2003. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science **302:**1056-60.

413. **Zargarian, L., I. Kanevsky, A. Bazzi, J. Boynard, F. Chaminade, P. Fosse, and O. Mauffret.** 2009. Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. Nucleic Acids Res **37:**4043-54.

414. **Zhang, H., O. Bagasra, M. Niikura, B. J. Poiesz, and R. J. Pomerantz.** 1994. Intravirion reverse transcripts in the peripheral blood plasma on human immunodeficiency virus type 1-infected individuals. J Virol **68:**7591-7.

415. **Zhang, H., G. Dornadula, P. Alur, M. A. Laughlin, and R. J. Pomerantz.** 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 U S A **93:**12519-24.

416. **Zhang, H., G. Dornadula, and R. J. Pomerantz.** 1996. Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological

microenviroments: an important stage for viral infection of nondividing cells. J Virol **70:**2809-24.

417. **Zhang, H., Y. Zhang, T. P. Spicer, L. Z. Abbott, M. Abbott, and B. J. Poiesz.** 1993. Reverse transcription takes place within extracellular HIV-1 virions: potential biological significance. AIDS Res Hum Retroviruses **9:**1287-96.

418. **Zhang, J., and C. S. Crumpacker.** 2002. Human immunodeficiency virus type 1 nucleocapsid protein nuclear localization mediates early viral mRNA expression. J Virol **76:**10444-54.

419. **Zhang, J., and H. M. Temin.** 1993. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science **259:**234-8.

420. **Zheng, Y. H., N. Lovsin, and B. M. Peterlin.** 2005. Newly identified host factors modulate HIV replication. Immunol Lett **97:**225-34.

421. **Zhu, P., E. Chertova, J. Bess, Jr., J. D. Lifson, L. O. Arthur, J. Liu, K. A. Taylor, and K. H. Roux.** 2003. Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. Proc Natl Acad Sci U S A **100:**15812-7.

422. **Zhu, T., B. T. Korber, A. J. Nahmias, E. Hooper, P. M. Sharp, and D. D. Ho.** 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature **391:**594-7.

423. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res **31:**3406-15.