



HAL
open science

Les protéines de nucléocapside du VIH-1 : structures, dynamiques, propriétés de fixation et de déstabilisation des acides nucléiques

Anissa Belfetmi

► To cite this version:

Anissa Belfetmi. Les protéines de nucléocapside du VIH-1 : structures, dynamiques, propriétés de fixation et de déstabilisation des acides nucléiques. Biochimie, Biologie Moléculaire. Université Paris Saclay (COmUE), 2017. Français. NNT : 2017SACLN055 . tel-01827242

HAL Id: tel-01827242

<https://theses.hal.science/tel-01827242>

Submitted on 2 Jul 2018

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

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

NNT : 2017SACLN055

**Les protéines de nucléocapside du VIH-1: structures,
dynamiques, propriétés de fixation et de déstabilisation
des acides nucléiques**

Thèse de doctorat de l'Université Paris-Saclay
Préparée à l'Ecole Normale Supérieure de Cachan (ENS Paris-Saclay)

École doctorale n°577

SDSV : Structure et Dynamique des Systèmes Vivants
Spécialité de doctorat: Science de la vie et de la santé

Thèse présentée et soutenue à Cachan, le 18 Décembre 2017, par

Anissa BELFETMI

Composition du Jury :

| | |
|---|--------------------|
| Claudine Mayer Pr, Université Paris Diderot (Institut Pasteur) | Présidente |
| Serena Bernacchi Dr, Université de Strasbourg (IBMC) | Rapporteur |
| Bruno Kieffer Pr, Université de Strasbourg (IGBMC) | Rapporteur |
| Dominique Fourmy Dr, Université Paris Saclay (I2BC) | Examineur |
| Carine Tisné Dr, Université Paris Diderot (IBPC) | Invitée |
| Olivier Mauffret Dr, ENS Paris-Saclay (LBPA) | Directeur de thèse |



Remerciements

Ce travail de thèse a été réalisé au Laboratoire de Biologie et de Pharmacologie Appliquée (LBPA), CNRS UMR 8113, à l'École Normale Supérieure Paris-Saclay (ENS de Cachan).

Je tiens tout d'abord à remercier tous les membres du jury qui m'ont honoré en acceptant d'évaluer mon travail de thèse, en y consacrant leur temps et leur expertise scientifique. Je remercie le Pr Claudine Mayer d'avoir bien voulu présider mon jury de thèse, je suis particulièrement reconnaissante à mes deux rapporteurs, Dr Serena Bernacchi et Pr Bruno Kieffer, pour le temps qu'ils ont consacré à l'évaluation de mon manuscrit et je remercie également les Dr Dominique Fourmy et Dr Carine Tisné qui ont accepté d'examiner mon travail.

Je tiens tout particulièrement à remercier Dr Malcolm Buckle pour m'avoir accueillie au sein de son unité et permis surtout de travailler dans un environnement épanouissant.

Je remercie mon directeur de thèse, Dr Olivier Mauffret, de m'avoir accueillie au sein de son équipe et de son soutien tout au long de ma thèse. Je suis très reconnaissante pour le savoir que vous m'avez transmis et la confiance que vous m'avez accordée. Grâce à votre sens de la pédagogie et à votre gentillesse, j'ai pu m'épanouir et évoluer en tant que jeune chercheuse dans un environnement agréable.

Je remercie chaleureusement le Dr Loussiné Zargarian, qui m'a encadrée tout au long de mon parcours avec beaucoup d'attention et d'humanité. Grâce à son expertise scientifique, elle m'a permis d'avoir confiance en moi. Son sens aigu de la pédagogie m'a permis de corriger mes erreurs et elle a su mettre en valeur mes qualités avec une attention et une gentillesse particulière. Loussiné tu as été mon ange gardien pendant toutes ces années d'encadrement, tant sur le plan professionnel que personnel, et les mots sont trop faibles pour exprimer toute ma gratitude. Encore une fois, *Merci*.

J'ai eu la chance au cours de ma thèse ainsi que mon master de travailler avec le Dr Philippe Fossé que je tiens à remercier pour m'avoir fait partager son savoir-faire, ses connaissances en Biochimie et Biologie Moléculaire dans le partage de sa bonne humeur.

Je tiens également à remercier Dr Brigitte René, Dr Brigitte Hartmann et Mme Françoise Chaminade pour leur gentillesse et leurs encouragements. Un merci au Dr Brahim Heddi, qui, bien qu'il soit arrivé au sein de l'équipe au cours de ma dernière année de thèse, m'a beaucoup encouragée et soutenue.

Au cours de ma thèse, j'ai eu la chance de collaborer avec l'équipe du Dr Carine Tisné, à qui je voudrais exprimer mes sincères remerciements ainsi qu'à tous les

membres de son équipe, j'adresse ma plus profonde gratitude au Dr Carine Tisné pour m'avoir ouvert les portes de son laboratoire ce qui m'a permis de travailler sur les formes immatures de la protéine de nucléocapside. Je remercie sincèrement le Dr Valéry Larue pour tous les échanges que nous avons eus sur l'attribution des protéines, avec toute sa gentillesse. Je suis particulièrement reconnaissante au Dr Pierre Barraud, d'avoir partagé avec moi son savoir-faire en RMN PRE et pour tous les moments qu'il m'a gentilement accordé. Sans oublier que c'est avec plaisir que j'ai travaillé avec la doctorante Assia Mouhand, que je remercie pour les moments que nous avons partagés et pour le travail que nous avons réalisé ensemble.

A tous les étudiants que j'ai vu passer au sein de mon équipe : Je remercie particulièrement le Dr Akli Ben Imeddouren pour tous ses conseils, ses encouragements et le temps qu'il m'a consacré au cours de mes premiers pas en RMN. Je remercie vivement les Dr Ouerdia Maskri, Dr Ahmad Elbahnsi, Cel Areny-Naves et Romain Retureau pour les moments que nous avons partagés au sein de l'équipe. Et aussi, à tous les stagiaires, qui m'ont apporté beaucoup de bonne humeur au cours de ma thèse : Audrey, Florian, Soriya, Rachel, Caroline...

Je remercie également tous les membres du laboratoire pour ces belles années passées en leur compagnie, pour tous les moments scientifiques et aussi de détente (déjeuners, pauses café, pots, barbecue, sport ...) ce qui a permis de créer un climat agréable pendant ma formation au LBPA. Un remerciement à Ilige Youssef que j'ai découvert pendant l'organisation des séminaires internes, à Rahima Chennoufi et à Gladys Mbemba avec qui j'ai partagé l'organisation du repas de Noël et de la journée scientifique Sidaction au laboratoire, mais aussi tous les autres moments toujours plein de bonne humeur. Merci les filles.

Je remercie enfin celles et ceux qui me sont chers et que j'ai quelque peu délaissés ces derniers mois pour achever cette thèse. Leurs attentions et encouragements m'ont accompagnée tout au long de ces années. Je suis redevable à mes parents, pour leur soutien moral et matériel et leur confiance indéfectible dans mes choix. Je remercie également mon frère Naoufel, qui partage mon goût pour la Biologie ainsi que la chasse au trésor que constitue une thèse. Enfin, je remercie tous les membres de ma famille et mes amis qui m'ont soutenue tout au long de mon doctorat.

Abbreviations

| | |
|----------------------------------|--|
| 3'UTR | UnTranslated Region située à l'extrémité 3' de l'ARN génomique |
| 5'UTR | UnTranslated Region située à l'extrémité 5' de l'ARN génomique |
| aa | acides aminés |
| ABCE1 | ATP binding Cassette Subfamily E member 1 |
| ADN | Acide DésoxyriboNucléique |
| ADNss | ADN strong stop |
| AIP | ALG-2 Interacting Protein 1 ou ALIX |
| ALIX | Apoptosis-Linked gene 2-Interacting protein X ou AIP |
| APOBEC | Apolipoprotein B mRNA-editing Enzyme Catalytic Polypeptide |
| APOBEC3G | Apolipoprotein B mRNA-editing Enzyme Catalytic Polypeptide de type G |
| ARL | A-Rich Loop |
| ARN | Acide RiboNucléique |
| ARN pol II | ARN polymérase II |
| ARNg | ARN génomique |
| ARNm | ARN messenger |
| ARN _t ^{lys3} | ARN de transfert de lysine cellulaire |
| ARNv | ARN viral |
| ASBVd | avocado sunblotch viroid |
| ATP | Adénosine-5'-triphosphate |
| AUG | codon start pour la traduction de gag |
| BMH | Branched Multiple Hairpins |
| BST-2 | bone marrow stromal antigen 2 |
| CA | protéine de la CApside |
| CAp24 | Capside mature de poids moléculaire de 24 kDa |
| CCHC | motif Cystéine Cystéine Histidine Cystéine |
| CCR5 | corécepteur des cellules CD4+ |
| CD4 | Cluster de Différenciation de type 4 |
| CMH-1 | Complexe Majeur d'Histocompatibilité de type 1 |
| CPMG | Car Purcell Meiboom Gill |
| cPPT | central Poly Purine Tract |
| CRINEPT | Cross-correlated relaxation-enhanced polarization transfer |
| CRM1 | Chromosomal Maintenance 1 |

| | |
|------------|---|
| CSA | Chemical Shift Anisotropy |
| CSP | Cold Shock Protein |
| CSP | chemical shift perturbations |
| cspA | la Cold Shock Protein majeure chez E. Coli |
| cspE | une forme de Cold Shock Protein chez E. Coli |
| cTAR | complementary TAR (ADN) |
| cTAR0 | séquence cTAR sauvage tronquée |
| cTAR0 loop | séquence cTAR sauvage modifiée au niveau de la boucle interne |
| cTAR0 stem | séquence Ctar sauvage modifiée au niveau de la tige basse |
| cTAR1 | séquence cTAR mutée tronquée |
| cTAR1 loop | séquence cTAR mutante modifiée au niveau de la boucle interne |
| cTAR1 stem | séquence Ctar mutante modifiée au niveau de la tige basse |
| cTAR26 | séquence cTAR 26 nucléotides |
| cTAR33 | séquence cTAR 33 nucléotides |
| CTD | C Terminal Domain |
| CXCR4 | corécepteur des cellules CD4+ |
| CypA | Cyclophiline A |
| db | double brin |
| DD | dipôle-dipôle |
| DDX6 | DEAD-Box helicase 6 |
| DIS | Dimer Initiation Site |
| DLS | Dimer Linkage Structure |
| DLS | Dynamic Light Scattering |
| E.Coli | Escherichia coli |
| Env | gènes des protéines d'enveloppe |
| ESCRT | Endosomal Sorting Complex Required for Transport |
| F | Fluoresceine |
| FRET | Förster/Fluorescence resonance energy transfer |
| G2 | phase du cycle cellulaire G2 |
| Gag | group antigen |
| Gag Pr55 | polyprotéine précurseur Gag au poids moléculaire de 55 kDa |
| Gag-Pol | polyprotéine précurseur |

| | |
|----------|--|
| gp120 | glycoprotéine de surface extramembranaire |
| gp41 | glycoprotéine de surface transmembranaire. |
| HBR | Highly basic region |
| HCV | hepatitis C virus |
| HDV | hepatitis delta virus |
| HMGA1 | High Mobility Group AT-Hook 1 |
| HMQC | Heteronuclear Multiple-Quantum Correlation |
| hnRNP | Heterogeneous nuclear ribonucleoproteins |
| hnRNP A1 | heterogeneous nuclear ribonucleoprotein A1 |
| HOHAHA | HOmonuclear HArtmann HAhn |
| HRV-2 | human rhinovirus-2 |
| HSQC | Heteronuclear Single Quantum Correlation |
| HTLV-1 | Human T-lymphotropic virus type 1 |
| huPrp | human Prion Protein |
| HXB2 | souche du VIH-1 |
| IDD | Intrinsically Disordered Domain |
| IDP | Intrinsically Disordered Protein |
| INEPT | Insensitive Nuclei Enhanced by Polarization Transfer |
| INp32 | intégrase d'un poids moléculaire de 32 kDa |
| ITC | Isothermal Titration Calorimetry |
| kb | kilobases |
| Kd | constante de dissociation |
| LAI | souche du VIH-1 |
| LDI | Long Distance Interaction |
| LTR | Long Terminal Repeat |
| M | Main (sous-groupe du VIH-1) |
| MA | protéine de la MApside |
| MA | protéine de la Matrice |
| MAL | souche du VIH-1 |
| MAp17 | Matrice mature de poids moléculaire de 17 kDa |
| MLV | Murine Leukemia Virus |
| MS2 | phage protéine |

| | |
|-----------|---|
| N | Non-M, non-O (sous-groupe du VIH-1) |
| NC | protéine de nucléocapside |
| NC(11-55) | protéine de nucléocapside tronquée |
| NCp15 | protéine de nucléocapside immature de poids moléculaire de 15 kDa |
| NCp7 | protéine de nucléocapside mature de poids moléculaire de 7 kDa |
| NCp9 | protéine de nucléocapside intermédiaire de poids moléculaire de 9 kDa |
| Nef | negative expression factor |
| NL4.3 | Souche du VIH-1 |
| NOE | Nuclear Overhauser Effect |
| NOESY | Nuclear Overhauser Effect correlation Spectroscopy |
| nonsat | nonsaturé |
| nt | nucléotides |
| NTD | N Terminal Domain |
| O | Outlier (sous-groupe du VIH-1) |
| OB | Oligonucleotide Binding |
| P | (sous-groupe du VIH-1) |
| p2 et p1 | domaine « Spacer Peptide » de la Pr55Gag ou SP1 et SP2 |
| p53 | ou TP53 pour tumor protein 53 |
| p6 | peptide de poids moléculaire de 6 kDa |
| PAS | Primer Activation Signal |
| PBS | Primer Binding Site |
| PDB | Protein Data Bank |
| PEP | Preservation of Equivalent Pathway |
| PI(4,5)P2 | phosphatidylinositol-4,5-bisphosphate (PIP2) |
| PIC | complexe de préintégration |
| PIP2 | phosphatidylinositol-4,5-bisphosphate |
| pol | gène de la polymérase |
| poly(A) | comporte le signal de polyadénylation AAUAAA |
| ppm | partie par million |
| PPT | Poly Purine Tract |
| PRE | Paramagnetic Relaxation Enhancement |
| Proxyl | 3-(2-Iodoacetamido)-2,2,5,5-tetramethyl-1-pyrrolidinyloxy, free radical |

| | |
|------------|---|
| PrP | Prion Protein |
| PRp12 | PRotéase d'un poids moléculaire de 12 kDa |
| PrPc | Cellular Prion Protein |
| Psi Ψ | signal d'encapsidation situé sur l'ARNg |
| PTB | polypyrimidine tract-binding protein |
| R | Repeat, séquence répétée dans les deux extrémités du génome (en 5' et 3') |
| RanGTP | RAs-related Nuclear protein GTP binding |
| Rev | Regulation of expression of virion proteins |
| rf | radiofréquence |
| RMN | Résonance Magnétique Nucléaire |
| RNAse H | RibonucléAse H |
| RNP | complexe ribonucleoprotéique |
| RRE | Rev Responsive Element |
| RT | reverse transcriptase formée d'un hétérodimère p66 et p51 |
| sat | saturé |
| sb | simple brin |
| SdAg | small delta antigen |
| SGSS | syndrome de Gerstmann-Straussler-Scheinker |
| SIDA | Syndrome d'immunodéficience acquise |
| SL1 | Stem loop 1 |
| SL2 | Stem loop 2 |
| SL3 | Stem loop3 |
| SL3 ADN | homologue ADN de la séquence SL3 de 20 nt |
| SL3 ADN20G | SL3 ADN avec une G ajoutée sur sa boucle apicale (21 nt) |
| SL4 | Stem loop 4 |
| SNV | Sin Nombre Virus |
| SP1 et SP2 | domaine « Spacer Peptide » de la Pr55Gag ou p2 et p1 |
| SSHS | motif Serine Serine Histidine Serine |
| SSP | Secondary Structure Propensity |
| stpA | Protéine bacterienne se liant à l'ADN |
| SU | SUrface |
| T CD4+ | cellules immunitaires T exprimant le recepteur CD4 |

| | |
|----------------|--|
| T1 | temps de relaxation longitudinale |
| T1 | Temps relaxation longitudinale |
| T2 | temps de relaxation transversale |
| T2 | relaxation transversale |
| TAR | Transactivation responsive element |
| TAR0 | séquence TAR sauvage tronquée |
| TAR1 | séquence TAR mutée tronquée |
| Tat | TransActivaTor |
| TM | transmembranaire |
| TMR | Tetraméthylrhodamine |
| TOCSY | Total Correlation Spectroscopy |
| TOCSY | Total Correlation Spectroscopy |
| TRIM5 α | Tripartite motif-containing protein 5 |
| TROSY | Transverse relaxation optimized spectroscopy |
| Tsg101 | protéine de l'ESCRT-I |
| T Ψ C | Bras de l'ARNtLys3 |
| U3 | Unique en 3', séquence située en 3' de l'ARN génomique |
| U5 | Unique en 5', séquence située en 5' de l'ARN génomique |
| Unr | upstream of N-ras |
| UTR | untranslated region |
| UV | Ultra Violet |
| Vif | Viral Infectivity Factor |
| VIH | Virus d'Immunodéficience Humaine |
| VIH-1 | Virus d'Immunodéficience Humaine de type 1 |
| VIH-2 | Virus d'Immunodéficience Humaine de type 2 |
| VIS | Virus d'Immunodéficience Simienne |
| VIScpz | Virus d'Immunodéficience Simienne du chimpanzé |
| Vpr | Viral protein R |
| Vpu | Viral protein U |
| Vpx | Viral protein x |
| ZF | Zinc Finger |
| ZF1 | Zinc Finger 1 |

ZF2

Zinc Finger 2

Table des matières

| | | |
|------------|--|-----------|
| I | Introduction | 18 |
| I.1 | Le virus de l'immunodéficience humaine | 19 |
| I.1.1 | Classification et origine du VIH-1..... | 19 |
| I.1.2 | Structure du VIH-1..... | 20 |
| I.1.3 | Génome du VIH-1 | 21 |
| I.1.3.1 | Généralités | 21 |
| I.1.3.2 | Partie non codante | 22 |
| I.1.3.3 | Partie codante | 24 |
| I.1.4 | Le Cycle réplcatif du VIH-1..... | 28 |
| I.1.4.1 | Etapes précoces : | 28 |
| I.1.4.2 | Étapes tardives : | 29 |
| I.2 | La transcription inverse..... | 30 |
| I.2.1 | Le processus de la transcription inverse | 30 |
| I.2.1.1 | Initiation de la RT : | 31 |
| I.2.1.2 | Synthèse de l'ADN strong stop : | 33 |
| I.2.1.3 | Premier transfert de brin : | 33 |
| I.2.1.4 | Synthèse du long brin d'ADN négatif : | 34 |
| I.2.1.5 | Synthèse du long brin d'ADN positif et second transfert de brin : | 35 |
| I.2.1.6 | Synthèse de l'ADN viral double brin possédant deux LTR : | 36 |
| I.2.2 | Mutations et recombinaison génétiques du VIH-1..... | 37 |
| I.2.2.1 | Mutation et fidélité : | 37 |
| I.2.2.2 | Recombinaisons..... | 38 |
| I.2.3 | Modèles de transferts de brins : | 40 |
| I.2.3.1 | Forced copy choice | 41 |
| I.2.3.2 | Copy choice..... | 41 |
| I.2.3.3 | Dynamic copy choice | 41 |
| I.2.4 | Mécanismes du transfert de brin : | 42 |
| I.3 | Le rôle de Gag dans la sélection, l'encapsidation et la migration du génome viral | 44 |
| I.3.1 | Sélection de l'ARN génomique | 45 |
| I.3.2 | Migration et assemblage des particules virales : | 49 |
| I.3.3 | Bourgeoisement et maturation | 51 |
| I.4 | La protéine de nucléocapside du VIH-1..... | 53 |
| I.4.1 | Genèse..... | 53 |
| I.4.2 | Structure..... | 55 |
| | | 13 |

| | | |
|--------------|---|-----------|
| I.4.3 | Propriétés | 56 |
| I.4.3.1 | Reconnaissance et modalités de liaisons aux acides nucléiques | 56 |
| I.4.3.1.1 | Spécificité et affinité de liaison | 56 |
| I.4.3.1.2 | Etudes structurales des complexes NCp7 acides nucléiques | 57 |
| I.4.3.1.3 | Polarité de fixation de la NCp7 aux acides nucléiques..... | 59 |
| I.4.3.2 | La non équivalence des deux doigts de zinc | 60 |
| I.4.3.3 | Activité chaperonne des acides nucléiques..... | 64 |
| I.4.3.3.1 | Déstabilisation des structures secondaires des acides nucléiques..... | 67 |
| I.4.3.3.2 | Hybridation des séquences complémentaires d'acides nucléiques..... | 70 |
| I.4.4 | Comparaison des activités chaperonnes des différentes formes de NC (p15, p9 et p7).72 | |
| I.5 | Les protéines chaperonnes des ARNs | 74 |
| I.5.1 | Le désordre intrinsèque d'une protéine..... | 74 |
| I.5.2 | Les protéines chaperonnes des acides nucléiques | 76 |
| I.5.3 | Le désordre intrinsèque chez les protéines chaperonnes des ARNs | 79 |
| I.5.4 | Mécanisme d'action des protéines chaperonnes des ARNs..... | 79 |
| I.5.4.1 | Protéines bactériennes chaperonnes des ARNs | 81 |
| I.5.4.1.1 | La protéine StpA | 81 |
| I.5.4.1.2 | La protéine CspA | 82 |
| I.5.4.2 | Protéines virales chaperonnes des ARNs..... | 82 |
| I.5.4.3 | Protéines chaperonnes des ARNs du VIH-1 | 83 |
| I.5.4.3.1 | La protéine de nucléocapside | 84 |
| I.5.4.3.2 | La protéine Tat | 84 |
| I.5.4.3.3 | La protéine Vif..... | 86 |
| I.5.4.4 | Protéine de nucléocapside du VIS | 87 |
| I.5.5 | Les prions humains impliqués dans les maladies | 88 |
| II | Objectifs de la thèse | 89 |
| III | Matériels et méthodes | 91 |
| III.1 | Matériels | 92 |
| III.1.1 | Oligonucléotides | 92 |
| III.1.2 | Peptides | 95 |
| III.2 | Méthodes | 96 |
| III.2.1 | Spectroscopie de fluorescence | 96 |
| III.2.1.1 | Principe de la fluorescence..... | 96 |
| III.2.1.2 | Caractéristiques de la fluorescence..... | 97 |

| | | |
|-------------|---|-----|
| III.2.1.3 | Anisotropie de fluorescence | 99 |
| III.2.1.4 | FRET (Förster Resonance Energy Transfer):..... | 101 |
| III.2.2 | Résonance Magnétique Nucléaire (RMN)..... | 105 |
| III.2.2.1 | Méthode d'attribution d'une protéine en RMN | 105 |
| III.2.2.1.1 | Attribution du squelette peptidique | 107 |
| III.2.2.1.2 | Attribution des chaînes latérales | 110 |
| III.2.2.2 | Analyse de la dynamique interne d'une protéine | 110 |
| III.2.2.2.1 | Mouvements moléculaires..... | 111 |
| III.2.2.2.2 | La fonction de densité spectrale | 111 |
| III.2.2.2.3 | Relaxation des spins et mouvements moléculaires | 112 |
| III.2.2.2.4 | L'effet nOe (Nuclear Overhauser Effect)..... | 115 |
| III.2.2.2.5 | La relation entre les paramètres de relaxation et les densités spectrales..... | 116 |
| III.2.2.2.6 | Mouvement lent des protéines et relaxation de dispersion CPMG | 117 |
| III.2.2.3 | Mise en œuvre expérimentale | 120 |
| III.2.2.3.1 | Mesure de R1 et R2..... | 120 |
| III.2.2.3.2 | Mesure de l'effet nOe | 121 |
| III.2.2.3.3 | Mesure de la relaxation CPMG | 122 |
| III.2.2.4 | L'étude des complexes protéines acides nucléiques par RMN..... | 124 |
| III.2.2.4.1 | Les variations de déplacements chimiques :..... | 125 |
| III.2.2.4.2 | La saturation croisée des protons iminos | 125 |
| III.2.2.4.3 | Contraintes de distances intermoléculaires courtes (nOe)..... | 126 |
| III.2.2.4.4 | Contraintes de distances intermoléculaires longues (Paramagnetic Relaxation Enhancement ou PRE)..... | 126 |
| III.2.3 | Techniques de Biochimie et de Biologie moléculaire..... | 128 |
| III.2.3.1 | Purification des oligonucléotides..... | 128 |
| III.2.3.2 | Marquage des oligonucléotides en isotope radioactif..... | 129 |
| III.2.3.3 | Marquage des oligonucléotides par des agents paramagnétiques | 129 |
| III.2.3.4 | Test d'hybridation par gel retard..... | 130 |
| III.2.3.5 | Production et purification de la protéine de nucléocapside du VIH-1 mature NCp7 marqué en ¹⁵ N | 131 |
| III.2.3.5.1 | Surexpression de la protéine NCp7..... | 131 |
| III.2.3.5.2 | Purification de la protéine d'intérêt NCp7..... | 132 |

IV Résultats133

IV.1 Étude du mécanisme d'hybridation des séquences TAR et cTAR impliquées dans le premier transfert de brin du VIH-1 en présence de la NCp7 134

IV.1.1 Contexte..... 134

| | | |
|-------------|---|------------|
| IV.1.2 | Résultats..... | 136 |
| IV.1.2.1 | Analyse de l'hybridation des séquences TAR et cTAR par FRET | 136 |
| IV.1.2.1.1 | Spectre d'émission de l'ADN..... | 137 |
| IV.1.2.1.2 | Cinétique d'hybridation TAR et cTAR sauvages en présence de NCp7 à différentes températures | 138 |
| IV.1.2.1.3 | Cinétique d'hybridation TAR et cTAR sauvages et mutantes en présence de différentes concentrations de NCp7 à 20°C | 139 |
| IV.1.2.1.4 | Cinétique d'hybridation des séquences sauvages "loop0" et "stem0" en présence de NCp7 | 140 |
| IV.1.2.1.5 | Cinétique d'hybridation des séquences mutantes "loop1" et "stem1" en présence de NCp7 | 142 |
| IV.1.2.2 | Analyse de l'hybridation de TAR et cTAR en absence et en présence de NCp7 par gel retard | 144 |
| IV.1.2.2.1 | Hybridation en absence de NCp7 | 145 |
| IV.1.2.2.2 | Hybridation en présence de NCp7 | 145 |
| IV.1.2.2.3 | Analyse de la cinétique d'hybridation..... | 146 |
| IV.1.2.3 | Analyse de la déstabilisation de la structure secondaire de TAR ARN par la NCp7 | 147 |
| IV.1.3 | Discussion..... | 148 |
| IV.2 | Etude des propriétés d'interaction de la NCp7 avec des acides nucléiques | 165 |
| IV.2.1 | Contexte..... | 165 |
| IV.2.2 | Résultats..... | 166 |
| IV.2.2.1 | Etude de l'affinité des complexes par anisotropie de fluorescence | 166 |
| IV.2.2.1.1 | Ètude des complexes NC-acides nucléiques avec des séquences en tiges boucles | 166 |
| IV.2.2.1.2 | Ètude des complexes NC-acides nucléiques avec des séquences ADN courtes et simple brin | 168 |
| IV.2.2.1.3 | Ètude des complexes NC-acides nucléiques avec des séquences ARN courtes et simple brin | 171 |
| IV.2.2.1.4 | Ètude des complexes NC-acides nucléiques à différentes concentrations de sels | 172 |
| IV.2.2.1.5 | Ètude des complexes NC-acides nucléiques avec les deux formes de protéines (complète et tronquée)..... | 173 |
| IV.2.2.2 | Analyse par RMN des complexes..... | 174 |
| IV.2.2.2.1 | Attribution de l'acide nucléique d(TAGAGAT) | 174 |
| IV.2.2.2.2 | Attribution de la protéine NC(11-55) non marquée | 176 |
| IV.2.2.3 | Analyse par RMN PRE | 178 |
| IV.3 | Etude structurale et dynamique des différentes formes de protéines de nucléocapsides du VIH-1 | 182 |

| | | |
|-----------|--|------------|
| V | Conclusions et perspectives | 237 |
| VI | Références bibliographiques..... | 242 |

I Introduction

I.1 Le virus de l'immunodéficience humaine

I.1.1 Classification et origine du VIH-1

Le VIH-1 appartient à la famille des *rétroviridae* et au genre des *lentivirus*. Ce virus se caractérise par une longue période d'incubation, et une évolution lente vers le stade SIDA, pathologie caractérisée par une réplication virale active, une diminution du taux de lymphocytes T CD4+ circulants (<200 cellules/ μ l), ainsi qu'une susceptibilité accrue à des infections opportunistes conduisant à la mort des patients infectés.

C'est en 1983 à l'institut Pasteur de Paris que le virus du VIH-1 a été isolé en tant qu'agent causal du SIDA par Dr Françoise Barré-Sinoussi et Pr Luc Montagnier. Quelques années plus tard, en 1986, un virus similaire au VIH-1 fut cloné et a été nommé VIH-2¹⁻². Les deux virus se distinguent par leur composition génétique et par leur potentiel pathogène, le VIH-2 causant une pathologie moins importante avec une progression extrêmement lente vers le SIDA³⁻⁴.

L'origine du virus remonte à de multiples infections zoonotiques des VIS-primates non humains, permettant ainsi au VIH d'évoluer au sein de la population humaine⁵⁻⁶, souvent pris comme animaux de compagnie ou chassés comme gibier, les contacts par morsures ou par blessures lors du dépeçage des animaux peuvent expliquer la transmission du virus à l'homme^{5,7}.

En se basant sur la classification phylogénétique des lentivirus infectant les primates, il a été montré que les VIH-1 et VIH-2 proviennent de VISs différents, ce qui implique une origine et une évolution différente.

Dans le cas du VIH-1, l'ancêtre le plus proche est le VIScpz, isolé de la sous-espèce du chimpanzé *Pan troglodytes troglodytes*, vivant principalement en Afrique centrale et dans la région équatoriale occidentale^{6, 8-13}. Le VIH-1 provient d'un VIS qui est passé au moins quatre fois à l'homme, qui s'est adapté, qui a donné lieu aux quatre groupes connus M, N, O, P (Main, Non-M non-O, Outlier et P respectivement)¹⁴ et dont l'épidémiologie varie considérablement. Le groupe M est le plus pandémique, avec 99% des infections sur un total de 75 millions d'individus infectés dans le monde. Il est divisé en neuf sous types A (A1, A2), B, C, D, F (F1, F2), G, H, J et K.

Le groupe M du VIH-1 est le plus étudié en laboratoire dont les isolats sont HXB2,

NL4-3, LAI (sous-type B) et MAL (recombinant des sous-types A1, D et K) (<http://www.hiv.lanl.gov/>). Le groupe O est le deuxième groupe le plus pandémique, causant l'infection d'environ 100.000 personnes présentes majoritairement à l'ouest et au centre de l'Afrique, c'est-à-dire, Cameroun, Tchad, Gabon, Niger, Nigeria, Sénégal et Togo ¹⁵⁻¹⁶. De plus, une infection en dehors de l'Afrique a été rapportée chez un navigateur norvégien et sa famille, qui auraient été infectés en 1960 ¹⁷. Il est à noter que les deux groupes M et O possèdent une histoire naturelle similaire ⁵.

En revanche, le VIH-2 descendrait du VISsm que l'on trouve au sein d'une espèce de singes de l'Ouest Africain, le sooty mangabey (*Cercocebus atys*) ¹⁸. Le VIH-2 est composé de huit sous-types : A, B, C, D, E, F, G et H ¹⁹, les groupes A et B étant les plus représentés.

I.1.2 Structure du VIH-1

La particule virale du VIH-1 possède une forme sphérique d'environ 145 nm de diamètre et a été caractérisée par microscopie électronique ²⁰. La particule virale infectieuse est munie d'une enveloppe en bicouche lipidique, issue de la membrane plasmique de la dernière cellule infectée ²¹, dans laquelle sont insérées les glycoprotéines d'enveloppe majeures constituées de deux sous-unités : la gp120 extramembranaire et la gp41 transmembranaire. Ces gp120 et gp41 s'organisent en trimères permettant la fixation aux récepteurs CD4 et corécepteurs CCR5 ou CXCR4 des cellules cibles ²²⁻²³.

Quant à la face interne de l'enveloppe, celle-ci, est tapissée d'une couche de la protéine constituant la matrice (MA), p17 ²⁴⁻²⁷. La particule virale contient également la capsid (CA) de forme conique, constituée par l'assemblage des protéines p24 ²³ avec un agencement hexamérique responsable de la structure conique du core du virus ²⁸⁻²⁹. La capsid renferme l'ARN génomique (ARNg), sous forme de deux copies d'ARN viral simple brin (sb), à polarités positives et étroitement associées aux protéines de la nucléocapsid (NC), p7 ³⁰, aux enzymes virales, la protéase (PR), p12, l'intégrase (IN), p32 et la transcriptase inverse (RT), hétérodimère p66-p51) ³¹.

La particule virale mature est le fruit de la maturation, par la protéase virale, des trois précurseurs polyprotéiques (Env, Gag et Pol) qui dans un premier temps, s'assemblent en formant une particule virale immature (Figure 1A).

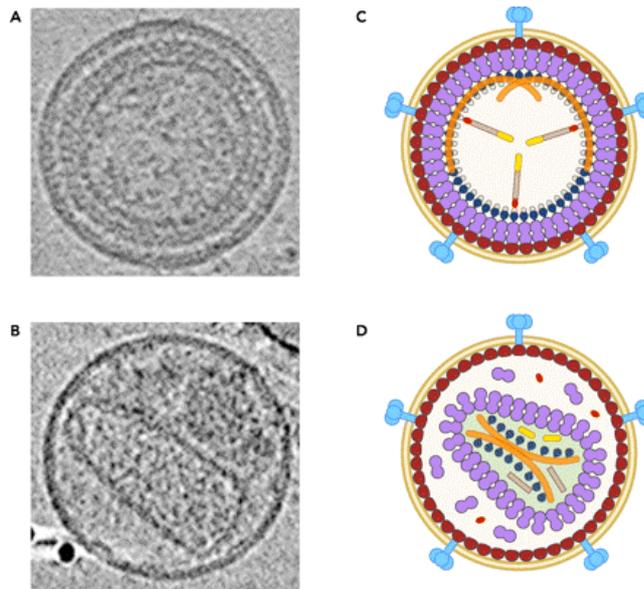


Figure 1A : Structure de la particule virale mature et immature du VIH-1. (A) et (B) Observation des particules virales immature et mature de 130 nm de diamètre par cryotomographie électronique ²⁹. **(C) et (D)** Représentation schématique des particules virales immature et mature. En bleu, gp120 ; en jaune, bicouche lipidique ; en rouge, MA ; en mauve, CA ; en bleu nuit, NC ; filament orange, ARNg ³².

I.1.3 Génome du VIH-1

I.1.3.1 Généralités

Les rétrovirus ont un génome composé de deux brins d'ARN à polarité positive et monocaténaire de 9,2 kb chacun ³³, orientés 5'-3', liés par une liaison non covalente au niveau d'un site majeur d'interaction nommé DLS (Dimer Linkage Structure) de l'extrémité 5' ³⁴⁻³⁸.

Comme ces molécules d'ARN sont issues d'une transcription par l'ARN pol II cellulaire, elles sont donc toutes coiffées en 5' et polyadénylées en 3' (Figure 2A). Le génome est également associé à l'ARNt^{Lys3} cellulaire, aux enzymes virales (RT, IN, PR) ainsi qu'à la protéine de nucléocapside NCp7 qui stabilise sa conformation ³⁹⁻⁴⁰. Au total, le génome du VIH-1 est constitué de 9173 nucléotides ⁴¹, contenant une partie codante bordée par une partie non codante à ses deux extrémités ³¹.

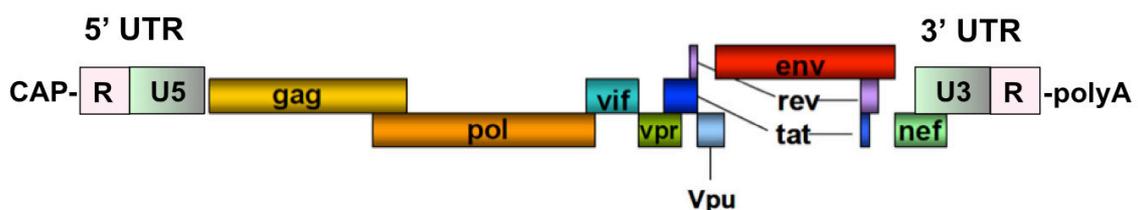


Figure 2A: Structure génomique du VIH-1. Vue d'ensemble de l'organisation du génome du VIH-1 de 9 Kb incluant trois gènes structuraux (*i.e.*, gag, pol et env), 6 gènes accessoires (*i.e.*, vif, vpr, vpu, tat, rev, et nef), les séquences UTR en 5' et 3' coiffée et polyadénylée respectivement (adaptée d'après ³¹).

I.1.3.2 Partie non codante

Les séquences non codantes (5'UTR et 3'UTR) sont des séquences promotrices répétées qui sont caractérisées par des structures secondaires très stables constituant des sites de liaison pour les protéines virales et cellulaires capables de réguler positivement ou négativement la transcription virale (Figure 3A).

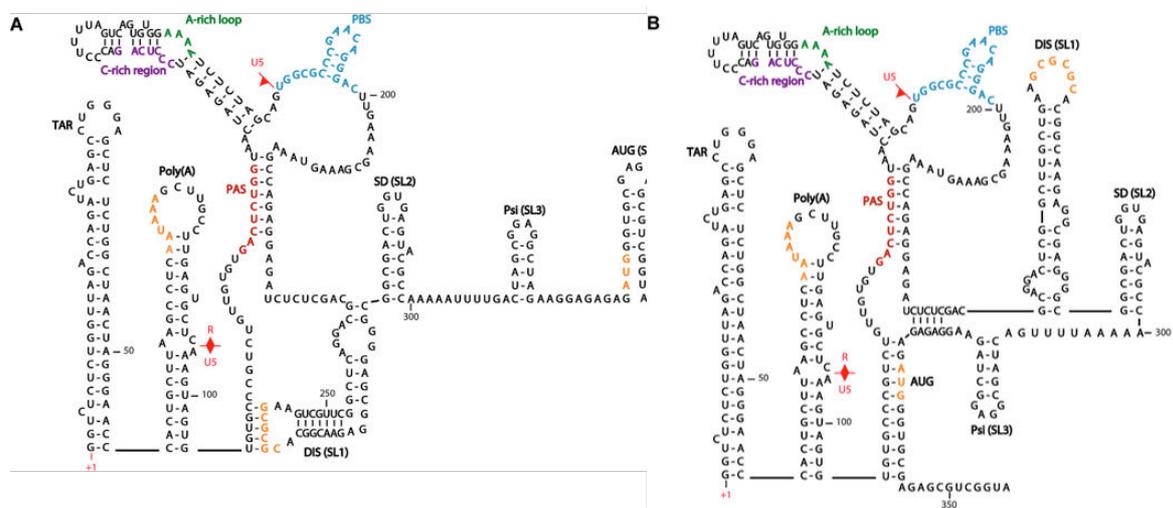


Figure 3A: Structure secondaire de la région 5'-UTR de l'ARNg de la souche NL4.3 du VIH-1. (A) Structure secondaire qui promeut la transcription. (B) Structure secondaire qui promeut la dimérisation de l'ARNg. On y observe les séquences PBS (bleu), PAS (rouge), les régions riches en A (vert) et riches en C (mauve). La région contenant le signal de polyadénylation, les nucléotides de DIS (SL1) impliqués dans la dimérisation de l'ARN viral et le codon start (AUG) de la traduction de gag sont indiquées en orange. ⁴².

La séquence R est répétitive, c'est à dire qu'elle est présente à l'identique aux deux extrémités 5' et 3' de l'ARNg. Cette répétition est absolument nécessaire pour le bon déroulement de la transcription inverse et joue un rôle fondamental dans le premier transfert de brin ⁴³, la séquence R est formée de 97 nt et composée de deux séquences très structurées (Figure 3A):

-La séquence **TAR** « Transactivation responsive element » est composée des 59 premiers nucléotides de l'ARN viral, elle correspond à un site d'interaction avec la protéine Tat ⁴⁴⁻⁴⁶, faisant d'elle un site indispensable pour stimuler la transcription de l'ADN proviral ⁴⁷. TAR intervient aussi lors du premier transfert de brin qui a lieu au

cours de la transcription inverse ⁴⁸. Il semblerait également que TAR participe à l'encapsidation du génome viral lors de l'assemblage des virions ⁴⁹.

-La séquence « **poly(A)** » comporte le signal de polyadénylation AAUAAA ⁵⁰, bien qu'elle soit présente aux deux extrémités du génome, celle-ci n'est fonctionnelle qu'en 3'.

-En 5', la région **U5** « Unique en 5' » (+98 à +182) est la première région à être transcrite, lors de la transcription inverse. Située immédiatement en 3' de cette région U5, on trouve une séquence de 18 nt constituant la tige boucle **PBS** « Primer Binding Site ». Cette dernière est complémentaire aux 18 nt situées à l'extrémité 3' de l'ARNt^{Lys3}, ce dernier ayant été encapsidé dans la particule virale lors d'une infection précédente, cette complémentarité est requise pour l'initiation du processus de transcription inverse ⁵¹⁻⁵³.

En 3' du PBS se trouve la région leader, constituée de trois tiges boucles, celle-ci est également nommée DLS « Dimer Linkage Structure » et comporte l'ensemble des différentes tiges boucles (SL1 à SL4), ce qui en fait une région hautement structurée; ces séquences sont indispensables à la dimérisation de l'ARN viral (ARNv) et contiennent le signal d'encapsidation de l'ARNg ⁵⁴⁻⁵⁵. DLS est située entre la région PBS et le codon d'initiation AUG du gène gag lui-même, situé sur une quatrième tige boucle SL4.

-**SL1**, aussi nommée DIS (Dimer Initiation Site) est une séquence associée à la dimérisation de l'ARNg ³⁶, qui s'est également avérée comme étant le site le plus important pour l'encapsidation de l'ARNg ⁵⁶⁻⁵⁷.

-**SL2**, possède un « site donneur » d'épissage majeur nommé « SD » ⁵⁸ dont l'interaction avec la protéine U1 (qui fait partie des petites ribonucléoprotéines nucléaires (snRNP)) réprimerait le signal de polyadénylation situé en 5' ⁵⁹.

-**SL3**, longtemps considérée comme contenant le signal d'encapsidation majeur à la reconnaissance et l'encapsidation de l'ARNg lors de l'assemblage des nouvelles particules virales ⁶⁰⁻⁶³.

-**SL4**, contient le codon d'initiation AUG, tige boucle assez instable et qui établit des paires de bases alternatives avec la région U5 ⁶⁴⁻⁶⁶.

Par ailleurs, à proximité de l'extrémité 3' du génome, se trouve la séquence **PPT 3'** « Poly Purine Tract », cette région est riche en purines et sert d'amorce à la synthèse du brin d'ADN (+) car au cours de la transcription inverse, le duplexe ADN-ARN qui

se forme dans la région PPT possède des propriétés structurales particulières et n'est pas dégradé par l'activité RNase H de la RT. De plus, une copie de cette séquence appelée cPPT (central Poly Purine Tract) se trouve sur le gène pol et sert également d'amorce pour la synthèse du brin d'ADN (+).

-La séquence **RRE** « Rev Responsive Element »⁶⁷⁻⁶⁸, située sur le gène env est une région structurée en plusieurs tiges boucles, sur laquelle la protéine Rev s'oligomérisse permettant l'export nucléaire de l'ARN non- ou non- épissé après la transcription⁶⁹.

-**U3** « Unique en 3' » (-454 à -1), est une région située à l'extrémité 3', contenant des séquences consensus, nécessaires à l'intégration de l'ADN proviral dans le génome de la cellule hôte⁷⁰ et aident à la régulation de la transcription du provirus. En 3' de U3, se trouve la région R identique à celle en 5'.

Cependant, quand les régions UTRs sont rétrotranscrites, elles deviennent des LTRs (Long Terminal Repeat). La partie du génome ADN codant sera alors bordée par des séquences LTRs identiques aux deux extrémités 5' et 3'. Ces dernières sont cruciales pour l'intégration du génome viral dans celui de la cellule hôte et pour sa transcription⁷¹.

I.1.3.3 Partie codante

La partie codante du génome comprend en tout, neuf gènes codant pour 15 protéines au total^{41, 72}. Les trois gènes *gag*, *pol* et *env*, codant respectivement pour les protéines de structure (matrice, capsid, nucléocapsid), les enzymes impliquées dans la réplication virale (transcriptase inverse, intégrase, protéase) et les deux glycoprotéines d'enveloppe (Tableau 1A).

Tableau 1A. Rôles des protéines essentielles du VIH-1.

| Gène et produit | Nomenclature | Fonctions |
|-----------------|-----------------------------|---|
| Env | Protéines d'enveloppe | |
| gp120 | glycoprotéine de surface SU | Permettent la fixation aux récepteurs CD4 |

| | | |
|-----------------|---|---|
| gp41 | glycoprotéine transmembranaires TM | et co-récepteurs CCR5 ou CXCR4 des cellules cibles ²²⁻²³ . |
| Gag | Protéines de structure | |
| Gag Pr55 | polyprotéine précurseur Gag (Group antigen) | Orchestre l'étape d'encapsidation et de bourgeonnement viral. |
| MA (p17) | matrice | Tapisse la face inférieure de l'enveloppe lipidique du virus ²⁵⁻²⁷ . La MA permet l'interaction avec l'ARNg ²⁴ , la sélection de l'ARNt et la fixation à la membrane plasmique en tant que domaine de Gag ⁷³⁻⁷⁴ . |
| CA(p24) | capside | Responsable de la formation d'une coque au cœur du virus mature ²⁸⁻²⁹ . La CA possède une boucle riche en Pro qui constitue un site de fixation pour la CypA ⁷⁵ et le facteur de restriction TRIM5α ⁷⁶⁻⁷⁷ , qui interfère avec la décapsidation du virus lors de l'infection cellulaire en dirigeant les virus entrant vers le protéasome. |
| NC(p7) | nucléocapside | Reconnaissance et encapsidation de l'ARNg en tant que domaine de Gag. Stabilisation et protection du génome dans le virus mature. Rôle chaperon dans les processus de transfert de brins au cours de la transcription inverse. |
| Pol | Protéines enzymatiques | |
| PR | Protéase | Essentielle pour la maturation de la particule virale et donc pour son infectiosité. Réalise des clivages en des sites spécifiques au niveau des protéines |

| | | |
|-----------|-----------------------|--|
| | | précurseurs Gag et Gag-Pol. |
| RT | Reverse transcriptase | Joue un rôle crucial dans le processus de transcription inverse grâce à ses activités polymérase et RNase H. La première activité permet de synthétiser un brin ADN à partir d'un brin ARN, en utilisant une amorce ARN. La deuxième permet de dégrader l'ARN si et seulement si, il fait parti d'un duplex ARN-ADN. |
| IN | Intégrase | Essentielle pour l'intégration de l'ADN viral dans le génome de la cellule hôte. |

Les gènes codant pour des protéines accessoires et régulatrices sont Tat, Rev, Nef, Vif, Vpr et Vpu. Ces protéines sont importantes pour la régulation de la réplication ainsi que pour la pathogénicité du virus (Tableau 2A).

Tableau 2A. Rôles des protéines régulatrices et accessoires du VIH-1.

| Gène et produit | Nomenclature | Fonctions |
|------------------------|---|--|
| Tat | Transactivating responsive element | Tat se lie au niveau de la région TAR des ARNm naissants ⁷⁸ , activant ainsi l'initiation de la transcription et l'élongation à partir du promoteur des LTRs ⁷⁹⁻⁸⁰ , empêchant le signal de polyadénylation 5'LTR AATAAA de provoquer l'arrêt prématuré de la transcription et de la polyadénylation ⁸¹ . |
| Rev | Regulator of Expression of Viral proteins | Rev régule l'export nucléaire des transcrits viraux non- et mono- épissés en formant un complexe multimérique sur le RRE ⁸² . |

| | | |
|------------|----------------------------|---|
| Vif | Viral infectivity factor | Vif influence indirectement l'infectiosité virale en affectant l'assemblage, la libération et/ou la maturation du virus ⁸³ . De même Vif neutralise des facteurs cellulaires qui inhibent la réplication virale, comme APOBEC3G dont la dégradation et l'ubiquitination sont induites par Vif dans les cellules humaines ⁸⁴⁻⁹¹ . |
| Vpr | Viral protein R | Vpr est importante pour la formation du complexe de préintégration (PIC) et pour la régulation de sa localisation ⁹²⁻⁹⁴ . Elle assure également la transactivation des gènes cellulaires et la différenciation cellulaire ⁹⁵ en arrêtant leur croissance en phase G2 où la réplication virale est très efficace ⁹⁶⁻⁹⁷ . Vpr est aussi indispensable à l'infection de cellules qui ne se divisent pas tels que les macrophages ⁹⁸⁻⁹⁹ . |
| Nef | Negative regulatory Factor | Nef diminue l'expression des récepteurs CD4 et du CMH-1 à la surface des cellules infectées ¹⁰⁰⁻¹⁰² , ce qui favorise la libération des virions ¹⁰³ et l'augmentation de l'infectiosité virale ¹⁰⁴ . |
| Vpu | <i>Viral Protein U</i> | Vpu induit la dégradation des molécules CD4 dans le réticulum endoplasmique ^{90, 105-106} et intervient dans le processus de bourgeonnement ¹⁰⁷⁻¹⁰⁹ en empêchant l'introduction du facteur de séquestration des virions, la tetherin/BST-2 ¹¹⁰⁻¹¹¹ . |

I.1.4 Le Cycle répliatif du VIH-1

Le cycle viral du VIH-1 (Figure 4A) suit un schéma commun à tous les rétrovirus et se divise en deux phases, précoce et tardive.

I.1.4.1 Etapes précoces :

L'infection est initiée par la fixation des glycoprotéines de surface virale gp120 au récepteur membranaire CD4 de la cellule cible ¹¹²⁻¹¹⁵, cette interaction mène à un changement conformationnel des glycoprotéines d'enveloppe, conduisant au démasquage d'une région hydrophobe N-terminal de la gp41 dite "peptide de fusion" ¹¹⁶, qui jusqu'alors était masquée au sein des trimères gp120-gp41. En effet, l'exposition du site de fixation au co-récepteur, appartenant à une famille de récepteurs de chimiokines, CCR5 ou CXCR4 ¹¹⁷, induira une seconde interaction à la surface cellulaire, ce qui est absolument nécessaire pour que la fusion des membranes ait lieu ¹¹⁸. Après fusion de l'enveloppe virale avec la membrane cellulaire et une fois dans le cytoplasme, le génome va alors subir une transcription inverse qui aura lieu majoritairement au niveau des pores nucléaires et durera entre 8 à 12h ¹¹⁹⁻¹²⁰. Au cours de la transcription inverse l'ARN simple brin est converti en ADN double brin susceptible d'être intégré sous forme de provirus au génome cellulaire. Cette réaction est catalysée par la transcriptase inverse (RT), enzyme ARN/ADN dépendante, qui synthétise une copie d'ADN complémentaire (ADNc) à l'ARN viral puis, à l'aide de son activité RNase H, dégrade le brin d'ARN résiduel et termine la synthèse de l'ADN proviral en synthétisant le second brin. La RT utilise au cours de la rétrotranscription les nucléotides cellulaires ⁵². Le mécanisme de la transcription inverse sera détaillé au paragraphe (I.2) ainsi que le rôle de la NCp7 au cours de ce processus. Après la rétrotranscription de l'ARN en ADN double brin, celui-ci doit être intégré dans le génome de la cellule hôte, étape cruciale pour le cycle répliatif du VIH-1, à partir de laquelle le virus peut commencer à proliférer ¹²¹. Cette étape est orchestrée par l'intégrase, au cours de laquelle l'ADN viral est tout d'abord transporté vers le noyau, au sein du complexe de préintégration (PIC), formé de différentes protéines virales et cellulaires dont la transcriptase inverse (RT), la matrice (MA), Vpr et la HMGA1 ¹²²⁻¹²³, qui est par la suite intégré dans le génome de la cellule hôte.

I.1.4.2 Étapes tardives :

Après l'intégration de l'ADN proviral dans le génome de la cellule hôte, celui-ci va être transcrit en ARN par l'ARN polymérase II de la cellule hôte ¹²⁴. L'ARN viral doit être exporté du noyau, pour pouvoir jouer son rôle d'ARNm dans le cytoplasme, permettant la traduction des protéines virales, mais également pour servir d'ARN génomique qui sera encapsidé dans les nouvelles particules virales. Cependant, la machinerie de la cellule hôte empêche l'export de l'ARN viral mono- ou non-épissé, les rétrovirus détournent ce blocage grâce à la protéine Rev, qui se fixe sur le site « RRE » de l'ARN ainsi qu'avec le complexe d'export nucléaire CRM1/RanGTP, permettant une stabilisation de l'export cytoplasmique des ARNs messagers longs, mono- ou non-épissés ^{68, 82, 125-128}. La production des différentes protéines virales est possible, grâce à l'épissage alternatif des transcrits viraux, qui génère plusieurs espèces d'ARNm : les ARNm de pleine longueur (non-épissés) de 9 kb; les ARNm mono-épissés d'environ 4 kb et les ARNm multi-épissés d'environ 2 kb. L'expression de ces messagers est hautement régulée au cours du cycle viral. Dans les premiers stades de la réplication, les ARNm de 2 kb sont les plus abondants et permettent de produire les protéines Tat, Rev et Nef en grande quantité. S'ensuit, une augmentation progressive du nombre d'ARNm non-épissés et mono-épissés, ce qui permet la production des protéines de structure (gag-pol) et des protéines accessoires à la fin du cycle ¹²⁹. Il est important de préciser que dans le cas du VIH-1, celui-ci utilise son ARNm non épissé à la fois pour être traduit et encapsidé contrairement au virus de la leucémie murine (MLV), qui lui sépare son ARNm en deux populations distinctes : l'une associée aux ribosomes pour être traduite et l'autre pour être encapsidé dans les particules virales ¹³⁰⁻¹³². Après l'expression du génome et la synthèse des composants viraux, ceux-ci vont être adressés à la membrane plasmique en vue de leur assemblage. Cette étape est majoritairement orchestrée par la polyprotéine précurseur Gag Pr55 ¹³³, qui va interagir par le biais de son domaine NC avec la séquence contenant le signal d'encapsulation Ψ présente dans la région 5' non codante de l'ARN génomique non épissé, celui-ci va alors être dimérisé et encapsidé ¹³⁴⁻¹³⁷. Le rôle détaillé de Gag au cours cette étape sera exposé au paragraphe (I.3). La genèse de nouvelles particules virales, commence donc, avec l'encapsulation de deux copies d'ARN génomique et certains facteurs viraux et cellulaires. Par la suite, les particules virales néosynthétisées

doivent se détacher et quitter la cellule par un processus de bourgeonnement, à partir de la membrane plasmique de la cellule infectée, laquelle permet au virus d'acquérir son enveloppe qui est constituée d'une bicouche lipidique empruntée à la cellule hôte ²¹. A ce moment, le virus est sous forme immature et sous l'action de la protéase, la particule virale va subir un processus de maturation.

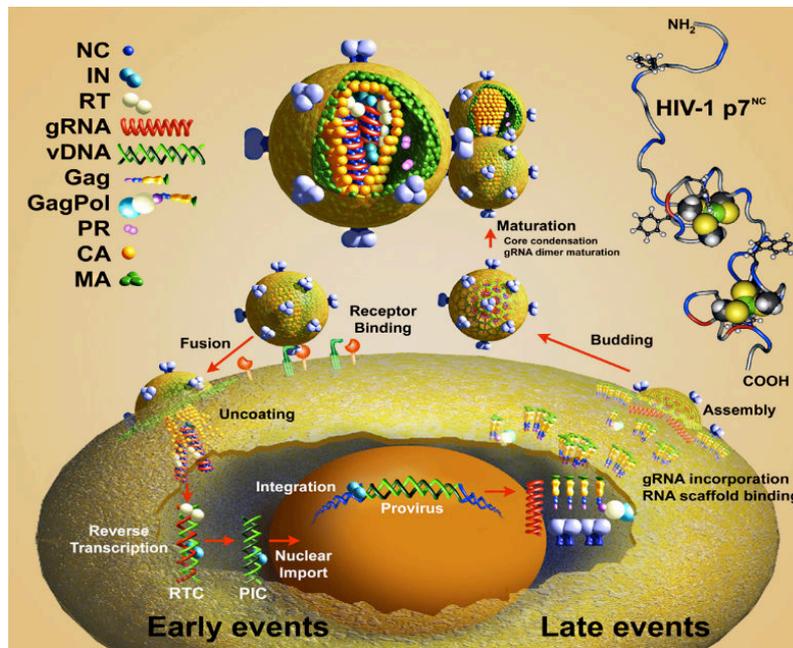


Figure 4A: Schéma représentatif de l'ensemble du cycle réplcatif du VIH-1 où la protéine de nucléocapside joue plusieurs rôles. 1) l'entrée *via* l'interaction de la glycoprotéine gp120 virale avec le CD4 et les co-récepteurs (CCR5 ou CXCR4) à la surface des cellules cibles, 2) la **fusion** *via* la glycoprotéine gp41, 3) la **transcription inverse** de l'ARN en ADN complémentaire, 4) l'**import nucléaire** du complexe de pré-intégration et l'**intégration** de l'ADN proviral, 5) la **transcription** de l'ADN proviral, 6) la **traduction** de l'ARN en protéines, et finalement les étapes de 7) l'**assemblage**, le **bourgeonnement** et la **maturation** des nouvelles particules virales ¹³⁸.

I.2 La transcription inverse

I.2.1 Le processus de la transcription inverse

Le mécanisme de conversion de l'ARNv génomique simple brin en ADN double brin linéaire est un processus complexe commun à tous les rétrovirus, qui comprend huit étapes principales (Figure 8A) ^{33, 139-141}. Plusieurs facteurs viraux et cellulaires participent à cette opération, mais les deux activités essentielles et indispensables à la transcription inverse sont présentes chez la reverse transcriptase rétrovirale (RT) : une activité polymérase et une activité RNase H. *In vitro*, l'activité RNase H est peu

ou pas du tout détectable lorsque la RT synthétise l'ADN ; mais, des clivages apparaissent aux sites où la synthèse d'ADN a réalisé des pauses ¹⁴²⁻¹⁴³.

I.2.1.1 Initiation de la RT :

Comme la majorité des ADN polymérases, la RT a besoin d'une matrice et d'une amorce. La matrice est l'ARN simple brin viral (ARNv) à polarité positive dont la transcription inverse est initiée par l'appariement de l'ARNt^{Lys3} cellulaire ⁵², qui sert donc d'amorce, avec une séquence complémentaire à l'extrémité 5' de l'ARNv nommée PBS ¹⁴⁴⁻¹⁴⁹ (Figure 5A). Le type d'ARN de transfert sélectionné varie selon les différents rétrovirus, pour le VIH-1 et VIH-2, il s'agit de l'ARNt^{Lys3} dont le processus d'empaquetage a été analysé en détail ¹⁵⁰. Des travaux récents, indiquent que la partie matrice (MA) de Gag serait plus particulièrement impliquée, dans le processus de reconnaissance de ces ARNt ¹⁵¹. Un certain nombre d'expériences menées *in vitro*, ont montré que le processus d'initiation de la transcription est particulièrement lent et difficile à démarrer. Le processus de synthèse d'ADN s'accélère considérablement, lorsque cinq ou six résidus ont été ajoutés à l'extrémité 3' du primer ARNt ^{150, 152-153}. Le processus est initié par la formation d'un duplex de 18 paires de bases entre le PBS et l'extrémité 3' de l'ARNt^{Lys3}, néanmoins d'autres séquences sont également essentielles pour une initiation efficace de la transcription inverse ¹⁵⁴. En tout, trois régions sont considérées comme essentielles pour l'interaction ARNt-ARNv: les régions PBS (Primer Binding Site) ; ARL (A-Rich Loop) et PAS (Primer Activation Signal) ¹⁵⁵⁻¹⁶⁴.

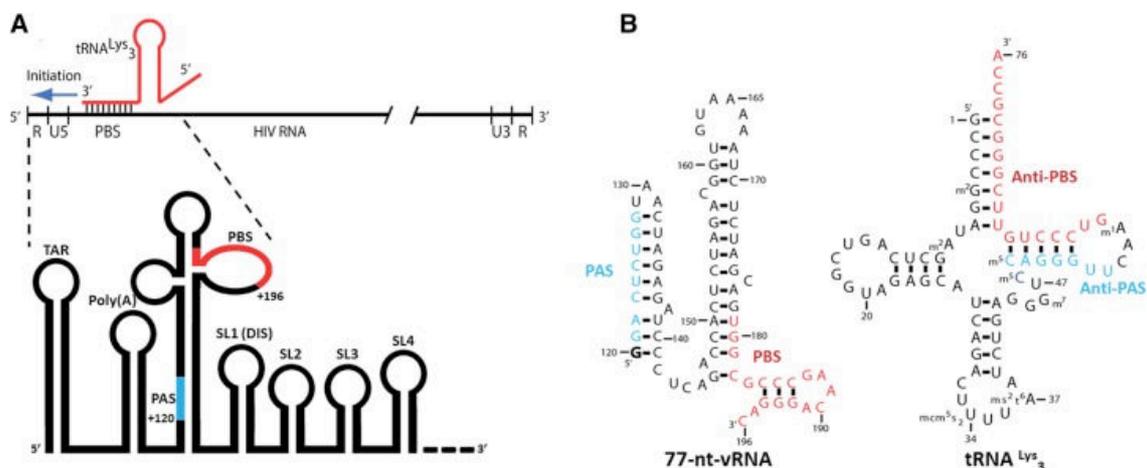


Figure 5A: Complexe d'initiation de la transcription inverse. A) L'amorce ARNt^{Lys3} s'hybride à l'ARNv au niveau du site PBS formant une hélice de 18 nt à partir de laquelle la transcription inverse va s'initier. Les localisations des séquences d'ARNv PBS et PAS sont

indiquées en rouge et bleu respectivement. **B)** Les séquences de l'ARNv et de l'amorce ARNt^{Lys3} contenant les séquences complémentaires PBS/anti-PBS et PAS/anti-PAS sont indiqués en rouge et bleu respectivement ¹⁶⁴.

La formation du complexe d'initiation de la transcription inverse, requiert l'ouverture de la structure tridimensionnelle de l'ARNt^{Lys3} et le réarrangement de l'ARNv, pour former un complexe d'initiation efficace, qui va être reconnu par la RT. Ce processus est chaperonné par la protéine de nucléocapside NCp7, qui joue également un rôle dans l'association (PAS anti-PAS) et (PBS anti-PBS) ^{42, 53, 165}. Dans le détail, la NCp7 déstabilise l'ARNt^{Lys3} au niveau du bras TΨC, ce qui requiert la présence des deux doigts de zinc ⁵³, qui se lient avec le bras TΨC, par les mêmes mécanismes de liaison qu'avec les tiges boucles SL2 et SL3. Au cours de cette interaction, le bras TΨC est déstabilisé par l'intermédiaire de résidus aromatiques tel que le Trp37. D'autre part, si le bras est déjà partiellement ouvert, la NCp7 déplace l'équilibre vers des formes ouvertes du bras ⁴² (Figure 6A).

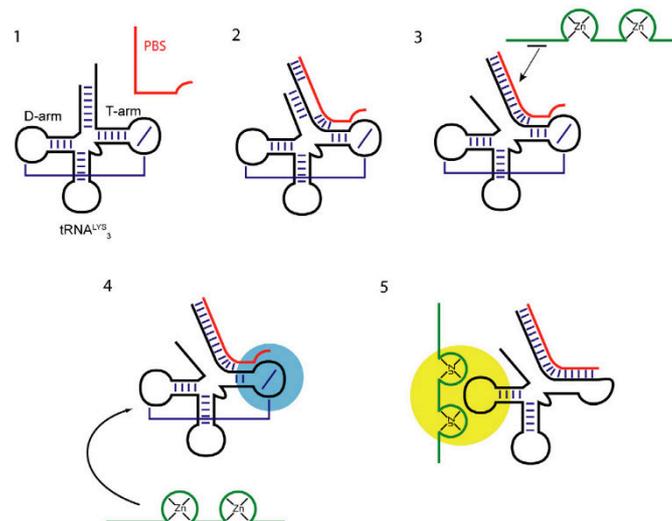


Figure 6A: Mécanisme de l'hybridation de l'ARNt^{Lys3} à la séquence PBS facilitée par la NCp7. 1) L'ARNt^{Lys3} (noir), PBS (rouge) et la NCp7 (vert). 2) L'hybridation s'initie au niveau des faibles paires de bases de l'ARNt^{Lys3} localisées à la jonction de l'accepteur et des tiges TΨC ou sur les bases non appariées à l'extrémité 3' de l'ARNt. 3) L'échange de brin se propage le long de la tige acceptrice, facilité par les extensions basiques de la NCp7. 4) Une fermeture structurale (structural lock) est révélée par la persistance de l'interaction entre les boucles D et T. 5) La NCp7 promeut l'hybridation totale de l'ARNt^{Lys3} à la séquence PBS par une reconnaissance spécifique de la boucle D de l'ARNt par les doigts de zinc ⁴².

Il a également été montré que la RT adoptait plusieurs types d'orientations dans les complexes RT-acides nucléiques. Son activité polymérase (vs. l'activité RNAase H) était reliée à un type bien particulier d'orientation de la RT sur le substrat acide

nucléique, tandis qu'un autre mode de positionnement était associé au mode de coupure des hybrides ADN:ARN ¹⁶⁶. La présence de structures tige-boucle dans l'ARN viral au site de démarrage de la transcription inverse est susceptible de modifier le positionnement de la RT et de la placer dans le mode "RNaseH", ce qui augmente les pauses de la RT dans le processus de copie. La protéine de nucléocapside est importante à ce niveau pour déstabiliser ces structures tiges-boucles et permettre de maintenir le mode "polymérase" de la RT ¹⁵².

I.2.1.2 Synthèse de l'ADN strong stop :

Après la formation du complexe d'initiation, la RT va synthétiser un brin d'ADN dans le sens 5', créant ainsi un duplex ARN-ADN dont la partie ARN sera dégradée par l'activité RNase H de la RT ¹⁵³. Il est à noter que le duplex ARN^{Lys3}-ARNv est résistant à la dégradation par la RNase H, ce qui permettra dans une étape ultérieure de copier le PBS ¹⁶⁷. La RNase H dégrade l'ARN complémentaire de l'ADN néo-synthétisé et élimine la totalité de l'extrémité 5' de l'ARNv (U3, R, U5), cet ADN néo-synthétisé est appelé ADN (-) ou ADN strong stop (ssADN), qui est donc la copie de la région R-U5 de l'ARNg ¹⁶⁸. Les séquences R (répétées aux extrémités 5' et 3' du génome viral) ont une longueur de 97 nt pour la souche « NL4.3 » et 96 nt pour l'isolat « MAL » et l'ensemble du "strong-stop" a une longueur de 178 nt. Les séquences R contiennent, les séquences TAR et polyA, qui sont partiellement auto-complémentaires possédant la capacité de se replier en structure tige boucle. Le clivage par la RNase H se déroule pendant la synthèse de l'ADNss et génère de courts hybrides dont la stabilité thermodynamique est réduite. L'activité déstabilisatrice de NCp7 permet la dissociation de ces courts duplexes et la libération de l'ADNss qui a la possibilité de se replier. De ce fait, la faculté qu'ont les séquences R à se replier semble essentielle dans le processus d'élimination des fragments d'ARN dégradés ¹⁶⁹⁻¹⁷⁰.

I.2.1.3 Premier transfert de brin :

Afin de continuer la synthèse du brin négatif, l'ADNss va migrer de l'extrémité 5' vers l'extrémité 3' du génome ARN de la même molécule d'ARN ou de l'autre molécule présente dans la particule virale, ceci est possible grâce à la présence des

séquences R aux deux extrémités du génome permettant ainsi une complémentarité. Ce transfert est nommé « premier transfert de brin » ou « transfert de brin négatif »^{33, 140, 171-172}. Des études *ex vivo*, ont montré que la synthèse complète de la séquence R était nécessaire pour permettre le transfert de brin¹⁷³ (Figure 7A). Plusieurs études montrent que des mutants de la séquence TAR, altèrent fortement les appariements entre les séquences R complémentaires ainsi que le processus de transfert de brin. Le rôle de la protéine de nucléocapside dans le transfert a bien été montré^{169, 174-175}. Des expériences menées *in vitro* suggèrent que la tige-boucle poly(A) n'est pas critique pour le transfert de brin¹⁷⁵⁻¹⁷⁶.

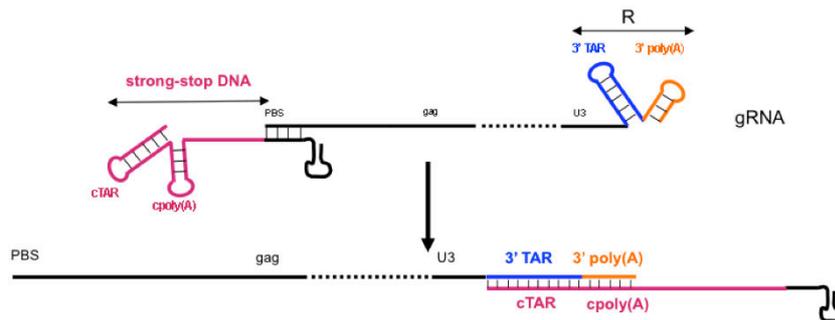


Figure 7A: Processus du premier transfert de brin du VIH-1.

Plus précisément, le premier transfert de brin repose en grande partie sur l'hybridation des tiges boucles TAR ARN et cTAR ADN présentes dans les régions R^{169, 175, 177-178}. Ce processus d'hybridation a fait l'objet de nombreuses études *in vitro*, démontrant qu'il est chaperonné par la protéine de nucléocapside NCp7, mécanisme qui sera détaillé au paragraphe (I.4.3.3).

Du fait, que les rétrovirus possèdent deux copies du génome d'ARNv, le premier transfert de brin peut impliquer les séquences R de deux molécules d'ARN différentes^{139, 179-181}, ce qui est donc susceptible d'induire des recombinaisons génétiques, un processus détaillé au paragraphe (I.2.2).

Dans le contexte du génome viral, plusieurs types de mécanismes notamment une circularisation du génome avec différents types de séquences impliquées, ont été invoqués comme pouvant faciliter le transfert de brin^{160, 182-184}.

I.2.1.4 Synthèse du long brin d'ADN négatif :

Après ce premier transfert de brin, la synthèse du brin d'ADN négatif se poursuit le

long de la chaîne d'ARN et durant la progression de cette synthèse, la dégradation de l'ARN se poursuit également. Néanmoins, une fois arrivée à la séquence ppt (poly purine tract) de l'ARNv, celle-ci résiste au clivage de la RNase H et va servir d'amorce pour l'initiation de la synthèse du brin d'ADN positif ¹⁵³. La synthèse du brin (-) a pu être reconstituée *in vitro* avec des systèmes très simplifiés contenant : l'amorce ; la matrice d'ARN ; la RT et la NC. Dans ces conditions, il a été montré qu'en présence de quantités élevées de NC, les produits de synthèse avec une matrice de 874 nt, sont de pleine longueur, mais présentent toutefois un faible rendement ¹⁸⁵. De même, la synthèse d'un brin de 4 kb peut être synthétisé *in vitro* à partir de préparations semblables mais toujours avec un rendement faible ¹⁸⁶. L'un des points importants durant cette synthèse, est la pause de la RT associée à la rencontre de structures secondaires présentes dans l'ARNv ¹⁸⁷⁻¹⁸⁸, ces événements de pauses sont associés aux événements de transferts de brins qui sont détaillés dans les paragraphes suivants. Comme il a déjà été dit plus haut, la présence de structures secondaires dans l'ARN viral est susceptible de changer le positionnement de la RT sur l'ARN et donc son mode de fonctionnement ¹⁵². Ce type de mécanisme pourrait être à l'origine des transferts de brins. De façon intéressante, il a été montré que la plupart des recombinaisons rétrovirales (plus de 98%), réalisées par le jeu des transferts de brin, se produisent durant la synthèse du brin (-) ¹⁸⁹.

I.2.1.5 Synthèse du long brin d'ADN positif et second transfert de brin :

Quand la RT génère le brin d'ADN (+), bien que celui-ci soit initié à partir des ppt en 3', la RT ne se contente pas de copier seulement le brin d'ADN (-), elle copie aussi les premiers 18 nt de l'amorce ARNt^{Lys3}. Or, du fait que le brin d'ADN (+) possède une copie des 18 nt de l'ARNt^{Lys3}, complémentaires aux 18 nucléotides présents à l'extrémité du brin d'ADN (-), qui ont été copiés à partir de la séquence PBS, les deux séquences complémentaires peuvent s'hybrider, permettant la poursuite de la synthèse totale à la fois des deux brins négatif et positif. Cette hybridation va avoir lieu suite au second transfert de brin ou transfert de brin positif ¹⁶⁷, au cours de ce dernier, la NCp7 joue également un rôle fondamental qui sera détaillé au paragraphe (I.4.3.3.2).

I.2.1.6 Synthèse de l'ADN viral double brin possédant deux LTR :

Le processus de transcription inverse crée un produit ADN plus long que le génome ARN duquel il dérive : les deux extrémités de l'ADN contiennent des séquences de chaque extrémité de l'ARN (U3 de l'extrémité 3' et U5 de l'extrémité 5'). De ce fait, chaque extrémité de l'ADN viral possède la même séquence, U3-R-U5, ce sont ces éléments de séquence qu'on appelle LTR (Long Terminal Repeats) et qui sont essentiels pour l'intégration dans le génome de la cellule hôte.

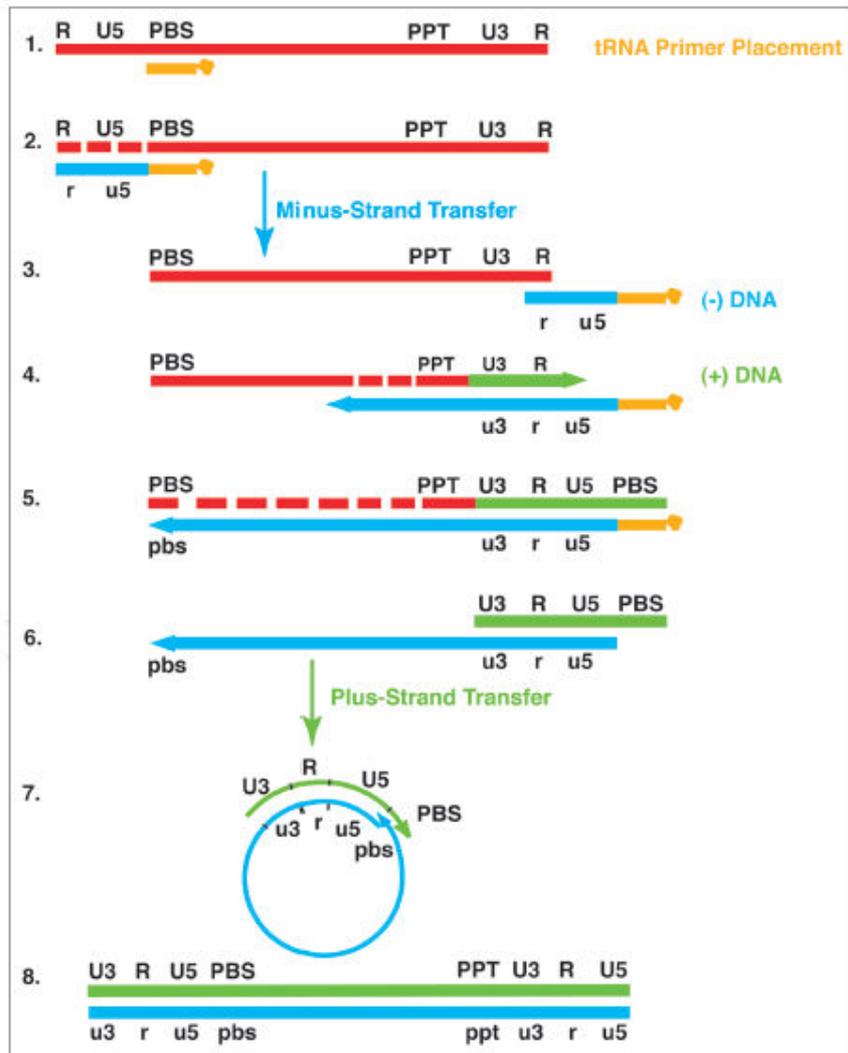


Figure 8A: Schéma des différentes étapes de la transcription inverse du VIH-1. 1) Initiation de la synthèse du brin d'ADN (-) par l'hybridation de l'ARNt^{Lys3} (orange) au PBS ; 2) Synthèse de l'ADN "Strong Stop" (-) (bleu); 3) Saut du brin d'ADN "Strong-Stop" (-); 4) Fin de la synthèse du brin d'ADN (-) et dégradation de l'ARN matrice (rouge) par l'activité RNase H de la RT; 5) Initiation de la synthèse du brin d'ADN (+) (vert) au niveau du PPT3'; 6) Synthèse de l'ADN "Strong-Stop" (+) et initiation de la synthèse du brin d'ADN (+) au niveau du PPTc; 7) Dégradation des amorces ARN et ARNt et saut de brin de l'ADN "Strong-Stop" (+); 8) Terminaison de la synthèse des deux brins d'ADN ¹⁴⁰.

I.2.2 Mutations et recombinaison génétiques du VIH-1

I.2.2.1 Mutation et fidélité :

Le virus du VIH-1 possède la capacité de faire évoluer rapidement son matériel génétique, ce qui lui permet d'échapper au système immunitaire de l'hôte, à la pharmacologie antirétrovirale et de compliquer le développement d'un vaccin efficace. Il en résulte, des souches virales du VIH-1 qui varient considérablement, non seulement entre différents patients mais également au sein d'un seul patient infecté, on peut observer ces variations, juste après l'infection même s'il est infecté par un seul virus ¹¹⁷. La source de cette diversité virale du VIH-1 provient d'une part des mutations génétiques fréquentes qui ont lieu au cours de la réplication virale ¹⁹⁰ et d'autre part de leurs recombinaisons qui sont liées aux mécanismes de transfert de brin ^{171, 191}.

Pendant longtemps, on a majoritairement incriminé la RT comme étant à la source des mutations du VIH-1, notamment parce qu'elle produit des erreurs pendant la conversion du génome et ne possède pas d'activité correctrice (proof reading), c'est-à-dire qu'elle n'est pas capable de corriger les erreurs commises au cours de la rétrotranscription (de l'ordre de 10^{-3} à 10^{-4}) ^{129, 192-194}. Entre autre, dans le cycle viral, une autre enzyme, l'ARN polymérase II de la cellule hôte (RNA pol II), est responsable de la synthèse du génome ARN et ne possède pas d'activité correctrice, ce qui constitue donc un autre candidat à incriminer pour les mutations génétiques du VIH-1. En plus des erreurs réalisées par la RT et l'ARN pol II, la transcription inverse peut être affectée directement ou indirectement par des facteurs cellulaires ¹⁹⁵, tels que les protéines APOBEC (apolipoprotein B mRNA-editing enzyme catalytic polypeptide) qui affectent la fidélité et l'efficacité de la transcription inverse tout en introduisant un grand nombre de mutations dans l'ADN viral (Figure 9A).

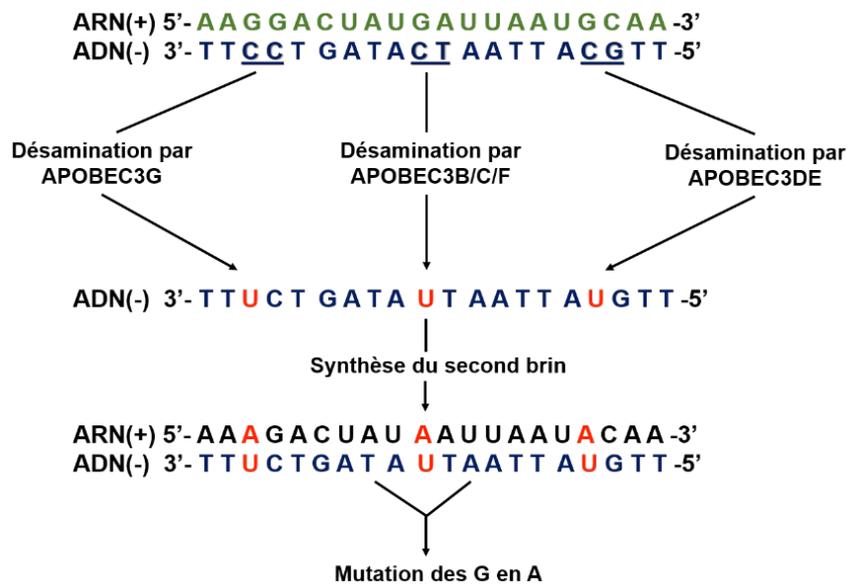


Figure 9A: La mutation des bases G en A médiée par les protéines APOBEC. Les protéines APOBEC sont encapsidées dans les virions VIH-1 et sont associées au complexe de transcription inverse (RTC) au niveau des cellules infectées, où elles y effectuent une désamination des Cytosines pour les convertir en Uraciles sur le brin d'ADN sb (-). La désamination des Cytosines n'est pas aléatoire car les membres de la famille des APOBEC possèdent des préférences à des nucléotides distincts. La désamination des Cytosines résultent en la mutation des G en A sur le brin d'ADN (+) (Adaptée d'après ¹⁹⁵).

I.2.2.2 Recombinaisons

Le taux de recombinaison génétique des rétrovirus est plus élevé que celui des autres virus, idem que pour le VIH-1, son taux de recombinaison est plus élevé que celui d'autres rétrovirus, tels que le virus de la leucémie murine (MLV) et le « sin nombre virus » (SNV) ^{139, 196-202}. La principale raison de ces recombinaisons est : Au cours de la synthèse du brin d'ADN (-), la RT peut changer de matrice entre les deux molécules d'ARN empaquetées dans la particule virale, la RT utilise des portions, des deux matrices ARN, pour former un ADN chimérique contenant des séquences provenant des deux molécules d'ARNg. Ce processus de transfert de brins peut se produire entre deux molécules d'ARN identiques, mais la genèse des virus recombinants, n'a lieu que si les séquences d'ARNg présentes dans la particule virale sont génétiquement différentes ^{139, 171}.

L'exemple suivant (Figure 10A) illustre le processus par lequel une recombinaison génétique entraîne à une résistance aux antirétroviraux. Prenons le cas d'une cellule infectée par deux génomes génétiquement différents : Un génome contenant des mutations dans le gène de la protéase qui permettent à celle-ci de fonctionner,

malgré la présence d'inhibiteurs de la protéase ; un génome possédant des mutations dans le gène de la transcriptase inverse lui permettant d'échapper aux inhibiteurs de la transcriptase inverse. Par l'intermédiaire de transfert de brins lors de la transcription inverse, le brin d'ADN est copié d'une matrice d'ARN vers une autre matrice, les deux génomes se recombinent conduisant à la formation d'un génome unique multirésistant aux agents antirétroviraux ¹⁷¹.

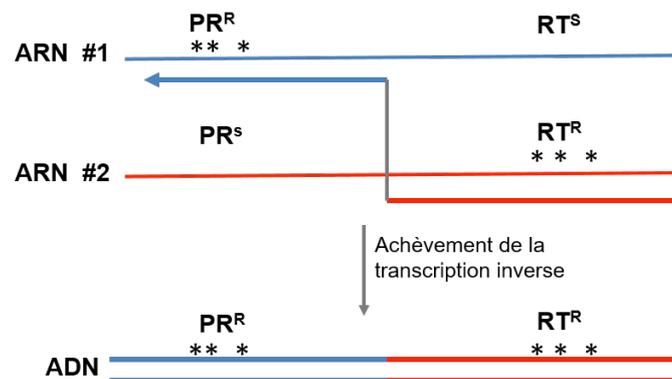


Figure 10A : Schéma illustrant les recombinaisons génétiques dues aux transferts de brins et leurs implications dans la résistance aux antirétroviraux. Les traits fins bleu et rouge représentent deux ARNg distincts mais co-encapsidés dans la même particule virale. Les traits épais représentent l'ADN viral. Les flèches indiquent le sens de la synthèse de l'ADN, et les astérisques indiquent des sites de mutations qui confèrent des résistances aux inhibiteurs de la RT ou de la PR. Dans ce cas, les transferts de brins génèrent un provirus recombinant résistant aux deux types d'inhibiteurs (adaptée d'après ¹⁷¹).

Cependant, plusieurs étapes sont requises pour la production d'un virus recombinant : Premièrement, la cellule hôte doit être infectée par au moins deux particules virales, les ARNs des deux provirus doivent être co-encapsidés dans le même virion ; Deuxièmement, les transferts de brins doivent avoir lieu au cours de la transcription inverse, pour générer une copie d'ADN chimérique, qui va ensuite être intégrée dans le génome de la cellule hôte ; Enfin, ce virus recombinant doit être capable de se répliquer pour que les effets de sa recombinaison puissent avoir lieu ¹⁷¹. La mesure du taux de recombinaison *in vivo* est une tâche difficile, nécessitant des montages expérimentaux complexes, qui doivent tenir compte de la proximité génétique parfois très grandes des génomes impliqués ²⁰³, la mesure de ce taux est néanmoins importante car chez le VIH-1, la recombinaison se produit beaucoup plus fréquemment que la mutation ¹⁹⁸ et constitue la voie principale de l'évolution virale ²⁰⁴. La recombinaison rétrovirale peut être sous-estimée si les possibilités de changement multiples de matrice ne sont pas prises en compte. Des mesures

récentes estiment la vitesse de recombinaison à 12-13 événements par cycle de transcription inverse ce qui est un taux très élevé ^{195, 203}, ces recombinaisons ont lieu tout au long du génome du VIH-1, mais leur fréquence est très largement affectée par la structure de l'ARN dans certaines régions ^{191, 203, 205}; elle est également affectée par l'homologie des séquences d'ARNg, c'est-à-dire, que les recombinaisons sont plus fréquentes quand les deux ARNg empaquetés appartiennent au même sous type ²⁰². Les recombinaisons se produisent beaucoup plus fréquemment au cours de la synthèse du brin d'ADN (-) que dans le second transfert de brin comme il a déjà été dit plus haut ^{181, 197}. La recombinaison génétique du VIH-1 est étroitement associée aux événements de transfert de brins dont nous allons présenter maintenant quelques modèles, qui permettront aussi de comprendre quels en sont leurs principaux déterminants.

I.2.3 Modèles de transferts de brins :

Concernant les transferts de brins, il faut distinguer entre les deux transferts de brin qu'on peut qualifier "d'obligatoires", baptisés premier et second transfert de brin, et les autres transferts de brin, qui se produisent durant la copie du brin (-). Les transferts de brin obligatoires sont conditionnés et déterminés par la présence des séquences R pour le premier et PBS pour le deuxième. Ces deux transferts de brins ne se déroulent pas dans des séquences codantes, ce qui fait qu'ils ne donnent pas lieu à des phénomènes de recombinaison génétique. En contre partie, les autres transferts de brin peuvent se produire dans les phases codantes et donner lieu à des recombinaisons. Comme il a été dit au paragraphe précédent, une douzaine de ce type d'événement peut avoir lieu au cours d'un cycle de transcription inverse. Parmi les causes principales de ces transferts de brins, les coupures fréquemment rencontrées au niveau des molécules d'ARNg, induisant des pauses de la RT, lorsque la RT rencontre ces coupures, elle doit "sauter" sur l'autre molécule d'ARNg pour poursuivre son activité polymérase, ce mécanisme de recombinaison est dénommé « copy choice » ou « forced copy choice » ^{171, 206}. D'autre part, le modèle original a été révisé et renommé en modèle de « dynamic copy choice » ²⁰⁷.

I.2.3.1 Forced copy choice

Dans ce modèle, les coupures dans l'ARNg induisent le transfert de brin pendant la synthèse du brin (-), l'extension d'amorce est initiée sur un ARNg et se poursuit jusqu'à la rencontre d'une coupure au niveau du brin matrice. A ce stade, sur la base de l'homologie de séquences, le brin d'ADN(-) est transféré vers le second ARNg co-encapsidé et la synthèse se poursuit.

I.2.3.2 Copy choice

Ce modèle de transfert de brin, n'est pas associé à la présence de coupure dans l'ARNg mais repose sur la possibilité de changement de matrice pendant la synthèse du brin (-) incluant des régions intactes du génome²⁰⁸. Le changement de matrice se faisant sur la simple base des homologies de séquences, cette dernière condition étant également nécessaire pour le transfert de type « forced copy choice ». Dans le cas de recombinaisons non homologues, le taux de recombinaison moyen est d'environ 0,1 à 1,0 % du taux de la recombinaison homologue¹⁸⁹. *In vivo*, la transcription inverse a lieu sur un complexe ribonucléique, dont une composante majeure est la protéine de nucléocapside NCp7²⁰⁹⁻²¹⁰. Le « copy choice » est fortement augmenté par la présence de NCp7²¹¹ au cours duquel celle-ci interagit spécifiquement avec la RT²¹², ce qui modifierait la capacité de la RT à changer de matrice ARN²¹³⁻²¹⁴.

I.2.3.3 Dynamic copy choice

Ce dernier propose que la balance entre les activités polymérase et RNase H de la RT (détermine la stabilité de l'association entre l'ADN néosynthétisé et la matrice ARN) dont la perturbation affecte la recombinaison (Figure 11A). Il apparaît dans ce modèle, qu'un ralentissement de la vitesse de polymérisation, favorise le clivage de la matrice par la RNase H et donc le transfert²¹⁵. Rappelons aussi, les études structurales effectuées par le groupe de Zhang¹⁶⁶, qui ont permis de mettre en évidence que ces deux modes de fonctionnement de la RT, étaient liés à des orientations de fixations le long des substrats acides nucléiques qui étaient différents. Ces orientations différentes étant dépendantes de la nature du substrat et des éventuelles structures secondaires présentes dans ce substrat¹⁵².

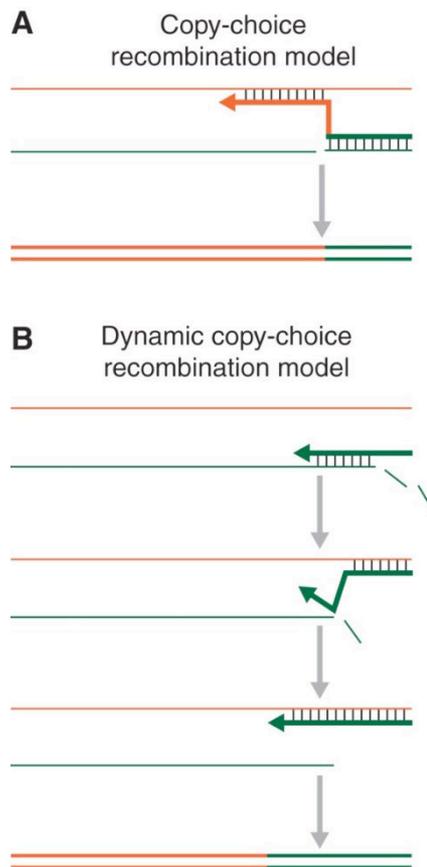


Figure 11A: Modèles de recombinaisons. (A) Modèle copy-choice. Deux brins d'ARN sont représentés en lignes orange et verte. La synthèse du brin d'ADN (-) utilise le brin d'ARN vert comme matrice, cependant ce dernier est coupé, ce qui induit un saut vers le brin orange pour poursuivre la synthèse. Ce saut de brin génère un ADN double brin composé de séquences à partir des deux brins d'ARN (vert et orange). **(B)** Modèle dynamic copy-choice. La synthèse du brin d'ADN (-) utilise également le brin d'ARN vert comme matrice. Quand l'ADNv est synthétisé, la RNase H dégrade le brin d'ARN vert et l'ADN qui a été copié à partir de ce dernier peut alors s'hybrider au brin d'ARN en orange. Ceci facilite le transfert de l'ADN (-) en cours de synthèse de la matrice en vert vers la matrice en orange, ce qui résulte en la synthèse d'un ADN double brin avec des séquences à partir des deux ARNs parentaux

153

I.2.4 Mécanismes du transfert de brin :

Ces mécanismes ont été étudiés principalement *in vitro*, en utilisant des systèmes d'ARN donneur/accepteur, avec comme matrice des séquences variées. Le système comprend en général, outre la reverse transcriptase, la protéine de nucléocapside indispensable au transfert de brin ³³. Dans ces expériences la synthèse est initiée à partir d'un "primer" apparié à l'extrémité 3' de l'ARN "donneur" à partir duquel se fait la synthèse. Lorsque la synthèse d'une région homologue du deuxième ARN appelé ARN accepteur a été effectuée, la synthèse peut basculer du brin donneur vers le brin accepteur et donc se terminer sur le brin accepteur. Ce type

de montage expérimental peut permettre par exemple de décrire le premier transfert de brin dans lequel l'extrémité 5' de l'ARN est le brin donneur et l'extrémité 3' le brin accepteur, la région d'homologie entre les deux brins étant la région R (97nt)²¹⁶, ces expériences *in vitro* sont également utiles pour comprendre de façon plus générale les mécanismes de transfert de brin. L'étude de ces mécanismes *in vitro* a été importante pour mieux comprendre les différents facteurs indispensables à ces transferts de brin en plus du rôle de la structuration des acides nucléiques impliqués dans ces transferts. Il en ressort, trois principaux types de mécanismes : le transfert de brin par invasion ou « docking » ; le transfert pause dépendant ou pause indépendant²¹⁷.

Dans le cas du processus d'invasion de brin (c'est à dire l'invasion du duplexe donneur : ADN par l'accepteur), celui-ci repose sur un mécanisme d'appariement du brin ARN accepteur avec l'ADNc, le brin d'ADN néosynthétisé, réalisant un appariement de l'accepteur avec l'ADNc très en amont de l'extrémité 3' de l'ADNc^{33, 216}. Le modèle est décrit dans la figure 12A, il repose en grande partie sur le fait que l'ARN donneur est clivé par la RT-RNase H et que de ce fait l'appariement de l'ARN accepteur avec l'ADNc devient possible. Les travaux effectués en utilisant des oligonucléotides, possédant les séquences R impliquées dans le premier transfert de brin, montrent que la zone d'appariement comme indiquée sur le schéma se fait en amont de l'extrémité 3' de l'ADNc²¹⁶. La synthèse se poursuit sur le brin accepteur, après qu'un appariement ADNc-ARN accepteur de taille suffisante se soit constitué.

Ces études ont permis de mieux comprendre les mécanismes sous-jacents à ces hybridations, notamment, le site majeur d'invasion de l'ARN accepteur est une région située de 10 à 19 nt de la base de la tige boucle stable de TAR de la région R. Il apparaît donc que c'est la présence de telles structures secondaires qui entraîne des pauses de la transcriptase inverse, de façon similaire à ce qui avait été décrit pour la phase d'initiation de la transcription^{33, 152, 216, 218}. Dans tous ces processus la présence de la protéine de nucléocapside est absolument indispensable, il a été démontré qu'elle favorise largement le mécanisme d'invasion²¹⁶. Ce mécanisme implique en effet des structures secondaires présentes notamment dans le brin ARN accepteur qui doivent être déstabilisées par la protéine pour permettre l'hybridation^{174, 216, 219}. Des études détaillées montrent que la taille de la région R, sa structure et sa séquence sont des éléments déterminants dans la réaction de transfert^{33, 220}. Il a

également été montré que des régions adjacentes à la région R peuvent avoir un rôle important pour favoriser le transfert de brin ²²¹. Ce mécanisme de transfert de brin se déroule en deux étapes dites « dock » et « lock », le schéma d'un tel mécanisme est montré ci-dessous avec l'implication des structures secondaires repliées "hairpins" ²²²⁻²²³.

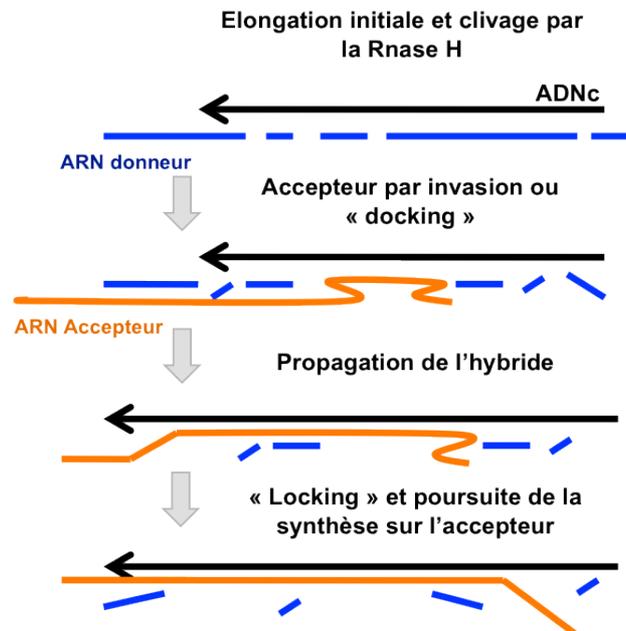


Figure 12A: Mécanisme de transfert de brin par invasion. La RT synthétise l'ADN initiateur (ou primer en noir) et clive périodiquement la matrice ARN dont il a été copié (ARN donneur en bleu). Une série de coupures proches en distances générées par la RNase H crée une région ADNc simple brin et permet à l'accepteur (rouge) de s'hybrider à l'ADNc, cette étape est dite « docking » ou « invasion ». L'ADNc accepteur et hybride se propage par migration de branches, déplaçant les pièces clivées du brin donneur. Éventuellement, le « primer » est libéré du donneur et s'hybride (lock) sur l'accepteur, permettant la synthèse jusqu'à l'extrémité 5' de l'accepteur (adaptée d'après ³³).

1.3 Le rôle de Gag dans la sélection, l'encapsidation et la migration du génome viral

Après l'expression du génome et la synthèse des composants viraux, ceux-ci doivent être adressés à la membrane plasmique en vue de leur assemblage. Cette étape est majoritairement orchestrée par la polyprotéine précurseur Pr55 Gag ^{133, 224-225}, cette polyprotéine est divisée en quatre domaines majeurs (Figure 13A) : en N-terminal la matrice (MA), la capsid (CA), la nucléocapsid (NC) et en C-terminal le domaine p6. Le domaine NC est bordé par deux petits peptides p2 en N-terminal et p1 en C-terminal ou nommés SP1 et SP2 respectivement, chacun de ses domaines

joue un rôle précis dans le processus d'encapsidation du génome et d'assemblage de la particule virale.

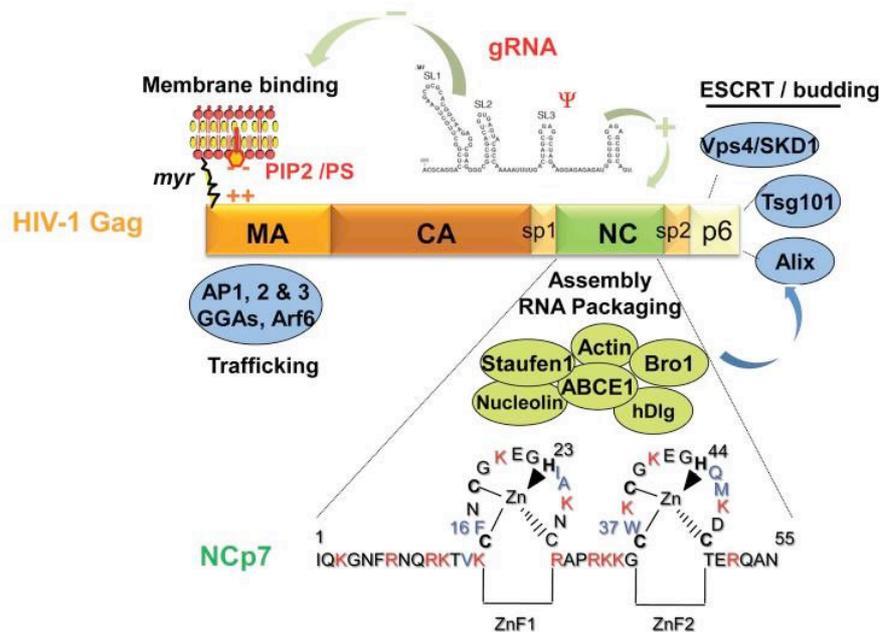


Figure 13A: La polyprotéine précurseur Gag et ses domaines d'interactions avec les protéines cellulaires. Gag est composée de trois domaines majeurs qui sont la MA, la CA et la NC, ainsi qu'un domaine C-terminal nommé peptide p6 et deux peptides espaceurs, sp1 and sp2. Le domaine MA est responsable de la fixation de Gag sur la membrane plasmique via son groupement myristate et sa région basique (HBR) avec les phospholipides PIP2/PS. La MA contient également des motifs qui se lient à différents facteurs cellulaires tels que AP1, AP2, AP3, GGA et Arf6 impliqués dans le trafic de Gag à travers le cytoplasme. Le domaine NC est nécessaire pour le recrutement de l'ARNg via son interaction avec le signal d'encapsidation Psi (ψ) et les résidus basiques de la NC (en rouge) ainsi que son plateau hydrophobe (résidus hydrophobes en bleu) qui impliquent ses deux doigts de zinc hautement conservés. L'ARNg est également apte à se lier au domaine MA par compétition avec PI(4,5)P2. Le domaine NC est un facteur clés pour l'assemblage par la formation d'une multimérisation de Gag sur l'ARNg. La NC et p6 interagissent avec des facteurs cellulaires ; NC avec l'actin, Staufen1, le domaine Bro1 d'ALIX et ABCE1. Le peptide p6 est essentiel pour le bourgeonnement viral vu qu'il recrute les protéines cellulaires de l'ESCRT, tel que tsg101, Alix et Vps4/SDK1. Sp1 et Sp2 sont des régions charnières impliquées dans la morphogénèse du VIH-1 ²²⁴.

I.3.1 Sélection de l'ARN génomique

Gag encapside sélectivement un ARN génomique viral dimérique et non épissé, sélectionné à partir d'un pool d'ARN qui comprend l'ensemble des ARNs dont l'ARN cellulaire et l'ARNm viral épissé ^{66, 226-227}. La sélection de l'ARNg est gouvernée par la région contenant le signal d'encapsidation Psi (Ψ), qui se situe dans la région 5' leader du génome viral (Figure 3A) ^{60, 61, 66, 228-231}, région hautement

structurée^{54-55, 230}. Il s'avère que les séquences situées en avant de cette région Psi (Ψ): TAR, polyA et PBS ont également un impact sur l'encapsidation du génome^{49, 232}.

Le signal d'encapsidation Psi (Ψ) proprement dit est constitué de quatre structures en tiges boucles (SL1 à SL4)^{55, 60-61, 63}. Pendant longtemps, SL3 a été identifiée comme la boucle contenant le signal d'encapsidation le plus important, néanmoins, des expériences où des délétions de SL3 ont été réalisées, tout en préservant le repliement de l'ARNg, montrent que ceci n'affecte que légèrement l'encapsidation du génome viral⁶¹⁻⁶³. Par contre, suite à la délétion de SL1, une plus grande réduction de l'infectiosité virale comparée à SL3 a été observée (200-fold et 4-fold respectivement)⁵⁷. Récemment il a été montré que SL1 possède un site spécifique de liaison à Gag de forte affinité, situé sur la boucle interne de la tige boucle et de motif AGG²³³. Vu que SL1 est un site d'initiation de la dimérisation ceci pourrait également impliquer un rôle de la dimérisation du génome sur l'efficacité de la fixation de Gag à l'ARNg²³³.

Bien qu'il ne soit pas encore tout à fait clair que, la dimérisation du génome viral a lieu avant ou après son encapsidation^{36, 234-236}, la formation d'un dimère stable entre les deux molécules d'ARNg apparaît cruciale pour la maturation du virion²³⁷⁻²³⁹, cependant il est connu que le processus de dimérisation nécessite une séquence palindromique de six nucléotides, nommée DIS « Dimerization initiation site », qui est capable d'établir des interactions boucle-boucle intermoléculaire^{35, 240-246} (Figure 14A). En effet, DIS peut former des « kissing complexes » grâce à l'auto-complémentarité de sa boucle^{35, 247-249} où la protéine de nucléocapside intervient pour permettre de former un duplex étendu (DIS-DIS) dans des conditions physiologiques²⁵⁰⁻²⁵¹ (Figure 14A). La boucle interne au niveau de DIS est le site de liaison de la NCp7²⁵²⁻²⁵³, toute mutation de cette boucle a pour effet de baisser l'affinité de liaison entre SL1 et la NCp7²⁵¹.

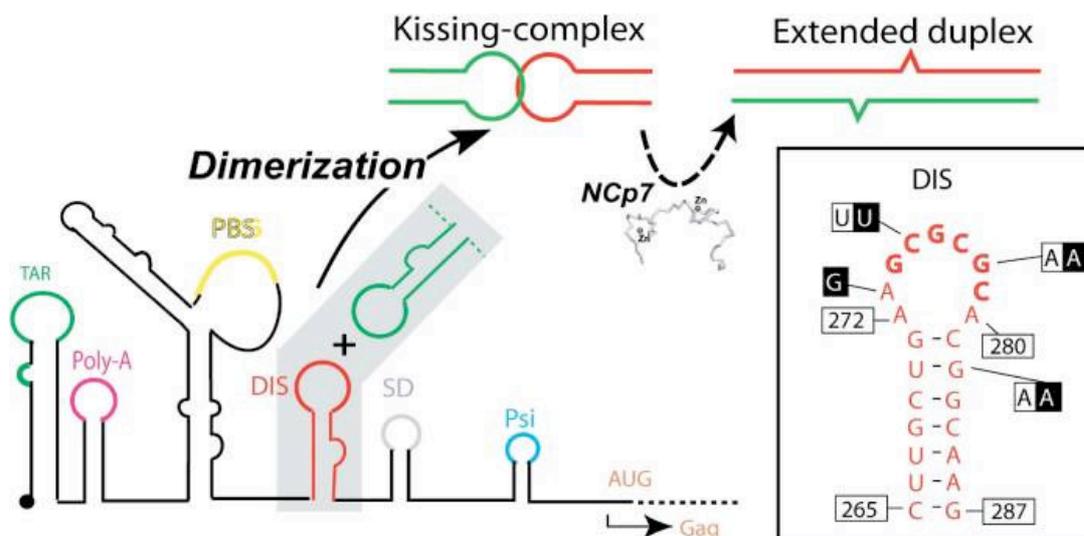


Figure 14A: Mécanisme de la dimérisation de l'ARNg du VIH-1. Schéma illustrant la séquence 5' non codante où se situe le DIS. La dimérisation de l'ARNg du VIH-1 a lieu via un mécanisme de « kissing complex » impliquant les séquences DIS des deux homologues ARN (en rouge et vert), le tout facilité par la NCp7. L'insert montre 23 nt de la séquence DIS du VIH-1 de la souche B, les bases encadrées en noir désignent celles de la souche A et les encadrées en blanc désignent celles de de la souche F ²⁴³.

Comme il a été mentionné plus haut, le site spécifique de reconnaissance de Gag a été identifié au sein de la boucle interne de SL1 ²⁵⁴⁻²⁵⁵, ces résultats étaient inattendus vu que SL1 est situé en 5' du « spliced donor site » qui est présent au niveau de l'ARNg non épissé et des ARNv épissés, suggérant l'existence d'un mécanisme qui préviendrait la fixation de Gag sur le SL1 des ARNv épissés. SL1 promeut donc l'encapsidation exclusive des ARNg non-épissés. Des expériences ont démontré que la délétion de SL1 au niveau des ARNv épissés montrent que ceux-ci sont incorporés avec une efficacité similaire comparée aux ARNv épissés contenant SL1 ⁵⁷. L'image qui en ressort est que la fixation de Gag est contrôlée par un mécanisme de double régulation : La région 5' de SL1, présente chez tous les ARNv, prévient la fixation de Gag à SL1, alors que la région en aval du gène Gag, qui est présente que sur l'ARNg, abolit cet effet négatif ²⁵⁴ (Figure 15A, A). Ce mécanisme nécessite probablement des interactions tertiaires qui coopèrent dans le but de créer un site de haute affinité pour la fixation de Gag, cependant celles-ci restent encore à identifier ²³⁵. Étant donné que le signal de dimérisation se chevauche avec le signal d'encapsidation, il est probable que Gag reconnaisse un dimère d'ARNg plutôt que deux monomères. Cela a été mis en évidence par des expériences où le site d'encapsidation a été dupliqué dans un ARNv non épissé ce qui a conduit à

l'encapsidation d'un monomère avec une interaction intramoléculaire²³⁶. Cependant, les processus de dimérisation et d'encapsidation restent très corrélés^{35, 57, 256-260} ce qui a mené à l'émergence de différents modèles où la conformation de l'ARNg change pour permettre la fixation de Gag (Figure 15A) :

Le premier modèle propose un changement conformationnel entre une forme "LDI" (Long Distance Interaction) qui orienterait l'ARNg non-épissé vers la traduction de ses protéines et "BMH" (Branched Multiple Hairpins) qui permettrait la dimérisation de l'ARNg et son encapsidation^{64, 261} (Figure 15A, B).

Dans la forme LDI, SL1 et poly-A forment des paires de bases entre elles et constituent une longue hélice qui séquestrent la séquence DIS et la rendent de ce fait inaccessible pour la dimérisation.

Par contre, le codon de départ AUG de Gag est situé dans une boucle interne ("bulge") et serait disponible pour la traduction de l'ARN. Dans la forme BMH, SL1, SL2 et SL3 forment des structures en tiges boucles, et la séquence AUG forme des paires de base avec la région (U5), formant l'interaction U5-AUG. Cette conformation expose SL1 pour promouvoir la dimérisation et l'encapsidation plutôt que la traduction²⁶². Cependant, la conformation en LDI n'a pas été détectée dans les cellules²⁶³⁻²⁶⁴ ce qui amène des doutes sur l'existence de cette dernière. Par ailleurs, l'existence de l'interaction U5-AUG est soutenue par des études phylogénétiques et biochimiques^{64, 264-265}. Un deuxième modèle déterminé principalement grâce à des données RMN, stipule que l'interaction U5-AUG existerait, en alternance, avec une interaction entre la région U5 et la boucle riche en GC de SL1⁶⁶ (Figure 15A, C). Dans ce schéma, l'interaction U5-AUG promeut la dimérisation et l'encapsidation de l'ARN suite à la libération du DIS²⁶⁶. Par ailleurs, une étude par RMN de la structure de l'ARNg du VIH-1, de 155 nt où les séquences TAR, Poly-A et PBS ont été délétées, a révélé l'existence d'une jonction à trois voies en tandem⁶⁵ (Figure 15A, C). Dans cette structure l'interaction U5-AUG s'étend pour inclure les résidus du site majeur donneur d'épissage dans SL2, tout en exposant des résidus G non appariés favorisant la fixation de la NCp7. La stabilisation de l'interaction U5-AUG stimule la récupération des dimères d'ARN isolées des cellules²²¹, et la mutation des régions entourant le codon de départ AUG de Gag inhibe l'encapsidation^{65, 267-268}. Par ailleurs, la perturbation de l'interaction U5-AUG fait diminuer le taux de fixation de la NCp7⁶⁵. Cependant, les mutations prédites sur SL4 qui perturbent l'interaction U5-

AUG n'affectent pas toutes la fixation de Gag *in vitro*²⁵⁴⁻²⁵⁵; et les mutations stabilisant SL2 mais perturbant la jonction à trois voies augmentent la fixation de Gag de manière surprenante dans le cadre d'une étude *in vitro*²⁵⁵.

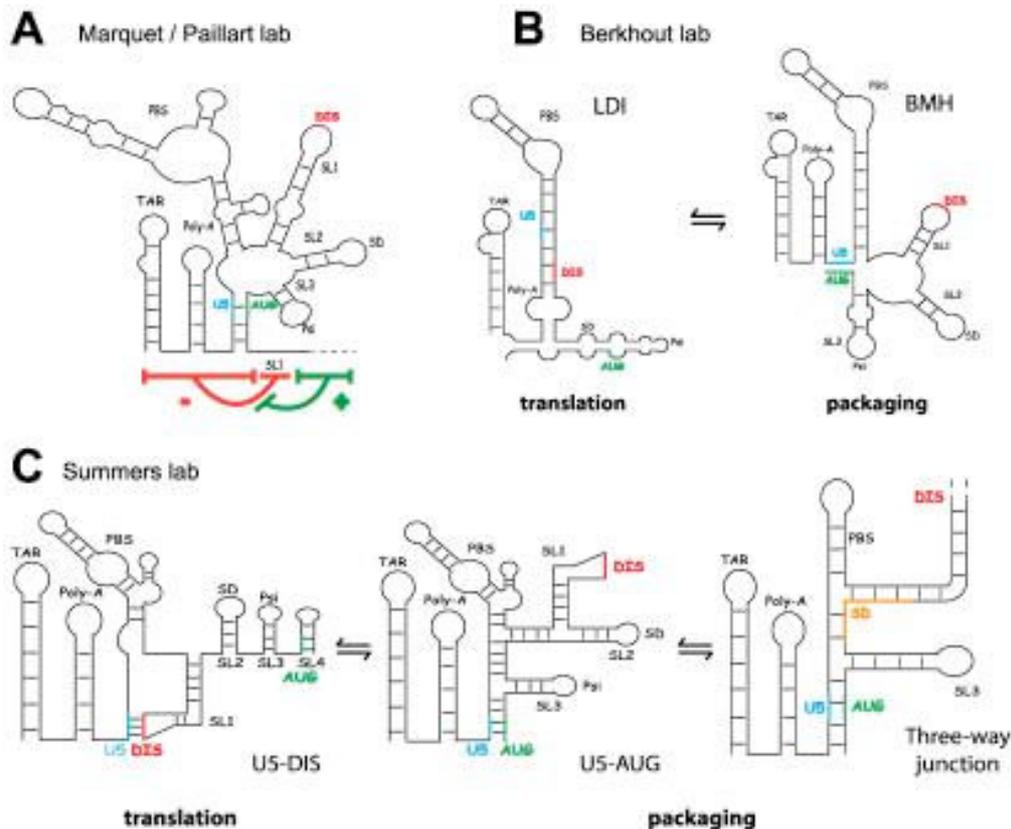


Figure 15A: Modèles proposés pour la structure secondaire de la région 5'UTR régulant la transcription et l'encapsidation de l'ARNg du VIH-1. (A) La fixation de Gag est soumise à un mécanisme de double régulation. La région en amont de SL1 (présente dans tous les ARNv) régule négativement la liaison Gag. Une région en amont de SL1 empêche la liaison de Gag à SL1 et cet effet négatif est contre par un motif en aval de SL4, qui est présent uniquement dans les espèces d'ARNg; (B) La structuration en LDI (The Long Distance Interaction) a été proposée pour promouvoir la traduction, et la structuration en BMH (the Branched Multiple Hairpin) a été proposée pour promouvoir l'encapsidation; (C) La région U5 est impliquée dans des interactions alternatives avec le site d'initiation de la dimérisation (DIS) (qui promeut la traduction) et la région entourant le codon de départ AUG (qui promeut l'encapsidation). Une jonction à trois voies, qui prolonge encore l'interaction U5-AUG et élimine SL2, est proposée pour promouvoir l'encapsidation²³⁵.

I.3.2 Migration et assemblage des particules virales :

Afin de procéder à l'assemblage des particules virales, les complexes Gag-ARNg formés doivent migrer à travers le cytoplasme pour atteindre la membrane plasmique. Plusieurs protéines cellulaires sont impliquées dans le transport des complexes Gag-ARNg, tels que l'ABCE1 (ATP binding cassette subfamily E member

1) ²⁶⁹⁻²⁷⁰, Staufen1 ²⁷¹ et DDX6 (DEAD-Box helicase 6) ²⁷².

Grâce à son domaine MA, Gag va se lier à la membrane plasmique via deux groupements : le groupement myristyl en N-terminal du domaine MA, qui facilite les interactions hydrophobes avec la membrane plasmique et aussi la région hautement basique (HBR) à la surface de la MA, qui médie les interactions électrostatiques avec les lipides cellulaires tel que les phosphatidylinositol-4,5-bisphosphate (PIP2) ^{131, 273}. Dans le cytosol cellulaire, Gag existe sous forme de monomères ou en multimères à faible taux, ce n'est que seulement après fixation à la membrane plasmique que se forment un taux élevé de multimères ²²⁵ créant un point de nucléation pour permettre l'assemblage de milliers de molécules de Gag dans le virion immature ²⁷⁴. Le domaine NC joue un rôle crucial dans la multimérisation de Gag via son association avec l'ARNg ^{224-225, 275-278}, mais également dans le recrutement des facteurs cellulaires pour le déplacement de Gag à travers le cytoplasme vers la membrane ²⁷⁷. Suite à la substitution du domaine NC de Gag par des régions qui peuvent accomplir des interactions protéines-protéines, tel que la MS2 phage protéine ou des motifs "leucines zipper", ont démontré qu'elles imitaient l'efficacité d'assemblage conférée à Gag par la présence du domaine NC ²⁷⁹⁻²⁸⁰. Il a également été démontré que le domaine NC est responsable du regroupement des protéines Gag et de leur accumulation dans les uropodes cellulaires entraînant la formation de synapses virales et la transmission du virus de cellule à cellule ²⁸¹. La mutation de deux arginines essentielles situées à l'extrémité N-terminale de la NC engendre une rupture de la formation des assemblages de Gag au sein de la membrane plasmique en inhibant la production des particules de Gag ²⁸². De même que la délétion du domaine NC de Gag induit une accumulation de la polyprotéine au niveau de la membrane plasmique *in cellulo* et affecte le bourgeonnement viral ²⁸³.

Dans la suite des événements, une fois arrivés à la membrane plasmique, La polyprotéine Gag et l'ARNg s'assemblent en particule virale immature avec le précurseur GagPol. Comme mentionné plus haut, la formation du complexe Gag-ARN crée un point de nucléation pour l'assemblage, cependant le domaine CA constitue le facteur viral le plus déterminant pour la multimérisation de Gag ⁷⁵. Au final, les molécules de Gag seront alignées et encapsidées avec le domaine MA vers la face interne de la membrane alors que le domaine C-terminal de Gag sera orienté vers le centre de la particule. Gag formera un réseau sphérique intercalé avec la

polyprotéine GagPol dans l'enveloppe de la particule virale immature en cours de bourgeonnement ¹³³.

I.3.3 Bourgeonnement et maturation

Lors de la phase de bourgeonnement, Gag interagit avec des facteurs cellulaires et détourne des protéines impliquées dans l'ESCRT (endosomal sorting complex required for transport) pour les convertir en une machinerie permettant le bourgeonnement des virions hors de la cellule infectée ²⁸⁴⁻²⁸⁵.

Le domaine p6 de Gag, inclus dans NCp15, constitue un site majeur d'interaction avec ces protéines cellulaires et virales ²⁸⁶, tels que les protéines Tsg101 (un composant de l'ESCRT-I) ²⁸⁷⁻²⁸⁸, AIP/ALIX (une protéine d'interaction avec ESCRT-I et ESCRT-III) ²⁸⁹⁻²⁹⁰ et la protéine virale accessoire Vpr ²⁹¹⁻²⁹³. Tsg101 et AIP/ALIX jouent des rôles essentiels au cours du bourgeonnement du VIH-1 ²⁸⁷⁻²⁹⁰ alors que Vpr est indispensable pour la réplication du virus dans des cellules qui ne se divisent pas ²⁹¹⁻²⁹³. Il a été démontré que p6 est exclu du cœur de la capsidie après la maturation du virus, alors que le domaine NC et Vpr restent localisés à l'intérieur de ce cœur ²⁹⁴⁻²⁹⁵. Ces résultats soulignent l'importance de p6 pour interagir avec les facteurs cellulaires et viraux lors de la phase de bourgeonnement.

Par ailleurs, il s'est avéré que le motif PTAP au sein de p6 est responsable de son interaction avec Tsg101 ²⁹⁶⁻³⁰⁰. De plus, un second motif terminal de p6, YPLTSL, a été identifié comme étant le domaine de liaison à la protéine AIP-1/ALIX. Cette interaction induit des changements conformationnels qui exposent la charnière centrale de p6 et permettent l'accès à la région riche en proline en C-terminal d'ALIX ²⁹⁹⁻³⁰¹. Il est important de souligner que AIP-1/ALIX, elle-même se lie à Tsg101 et aide au recrutement de Tsg101 sur le domaine p6. Enfin, le motif LRSLF de p6, contenant des leucines, a été identifié par mutagénèse comme étant le site le plus important pour l'interaction entre p6 et Vpr, et c'est probablement grâce à cette interaction que Vpr va être incorporé dans la particule virale ³⁰²⁻³⁰³. Dans le cas du VIH-2, cette région a été identifiée comme site d'encapsidation de la protéine Vpx. Récemment, une seconde séquence, FRFG, en N-terminal de p6 a été suggérée comme étant importante pour l'interaction avec Vpr ³⁰⁴. De plus, ALIX (de même que son domaine Bro1 isolé), se lie à la NC aussi bien qu'au domaine p6. Cette interaction peut être ARN dépendante mais semble être non spécifique ²⁸⁴⁻²⁸⁵. La

mutation du domaine NC inhibe l'interaction avec ALIX, mais contrairement au mutant du motif PTAP situé sur p6, le sauvetage de l'interaction n'a pas été possible en surexprimant ALIX³⁰⁵. Finalement, les particules virales néosynthétisées doivent se détacher et quitter la cellule par un processus de bourgeonnement à partir de la membrane plasmique de la cellule infectée, laquelle permet au virus d'acquérir son enveloppe, constituée d'une bicouche lipidique empruntée à la cellule hôte²¹.

La capacité intrinsèque de Gag à s'assembler en sphères sur les surfaces membranaires contenant les motifs PI(4,5)P2 est suffisante pour initier mais insuffisante pour compléter le processus d'assemblage de la particule virale. C'est pour cette raison que Gag a recours à la machinerie de l'ESCRT dans le but de compléter l'assemblage et de séparer l'enveloppe du virion naissant de la membrane cellulaire²⁹⁹⁻³⁰⁰. A ce moment, le virus est sous forme immature, sous l'action de la protéase (PR) sur Gag, la particule virale va subir un processus de maturation induisant un réarrangement morphologique avec la formation du core CA contenant le dimère d'ARN, dont la stabilité est fortement augmentée grâce à l'activité de protéine chaperon NCp7^{177, 209, 306-307}. Le clivage protéolytique des polyprotéines précurseurs Pr55 Gag et Pr160 Gag-Pol permet la formation des protéines structurales requises pour le réarrangement de l'architecture de la particule virale²³⁸. Ce mécanisme a lieu pendant, ou juste après le bourgeonnement du virion³⁰⁸⁻³⁰⁹, celui-ci est essentiel pour donner naissance à une particule virale mature infectieuse¹³³. Il faut aussi souligner l'importance de la maturation des protéines de nucléocapsides, le passage de la forme immature NCp15 ou NCp9 à la forme mature NCp7 est essentiel pour l'activité de protéine chaperon de la NC dans les étapes précoces du cycle virale. En effet, il a été démontré que le clivage entre le peptide p2 et le domaine NC au sein de Gag est essentiel pour l'initiation de la stabilisation du dimère d'ARNg³¹⁰. Les changements des propriétés de l'activité de protéine chaperon de la NC ont donc un impact direct sur la régulation du processus de transcription inverse et du remodelage de l'ARNg en général^{211, 311-313}.

I.4 La protéine de nucléocapside du VIH-1

La NCp7 est l'une des protéines majeures de l'architecture du cœur du virus. Du fait de son abondance, elle recouvre l'ensemble du génome ARN, permet sa condensation³¹⁴⁻³¹⁶ et le protège contre les attaques des RNases cellulaires³⁹⁻⁴⁰. On compte environ 1500 à 2000 molécules de NC, avec chaque molécules se liant au maximum à 11-12 nucleotides³¹⁷⁻³²¹. La NCp7 joue un rôle crucial, tant dans les étapes précoces que tardives du cycle viral, telles que l'assemblage des nouvelles particules virales^{224, 322-323}, l'encapsidation de l'ARNg⁶³, la dimérisation de l'ARNg³²⁰, la transcription inverse²⁰⁹ et l'intégration³²⁴. La NCp7 est une protéine chaperon des acides nucléiques qui déstabilise les structures secondaires des acides nucléiques pour les restructurer en des conformations thermodynamiquement plus stables^{177, 209-210, 325}. Cette propriété chaperon des acides nucléiques est essentielle au cours de la transcription inverse de l'ARNg⁴¹, par exemple, elle intervient dans le placement de l'ARNt^{Lys3} sur la séquence PBS en tant que domaine de GagPr55³²⁶, elle joue également un rôle dans le premier et le second transfert de brin^{33, 140, 177}. Les mutations dans la NCp7 réduisent considérablement l'infectiosité virale³²⁷⁻³²⁸ et désigne la NCp7 comme une cible thérapeutique potentielle pour le développement d'une nouvelles classe d'antirétroviraux^{48, 140, 329-330}.

I.4.1 Genèse

La protéine de nucléocapside est un produit de clivage de la polyprotéine précurseur Pr55 Gag, qui contient quatre domaines majeurs : la matrice (MA), la capsid (CA), la nucléocapside (NC) et le domaine p6 avec un domaine NC bordé par deux petits peptides p2 en N-terminal et p1 en C-terminal ou nommés SP1 et SP2 respectivement. La protéine Ncp7 est présente, selon, les phases du cycle de réplication, sous forme intégrée à Gag, sous forme immature (NCp15), intermédiaire (NCp9) ou mature (NCp7), désignées par leurs masses moléculaires respectives (Figure 16A).

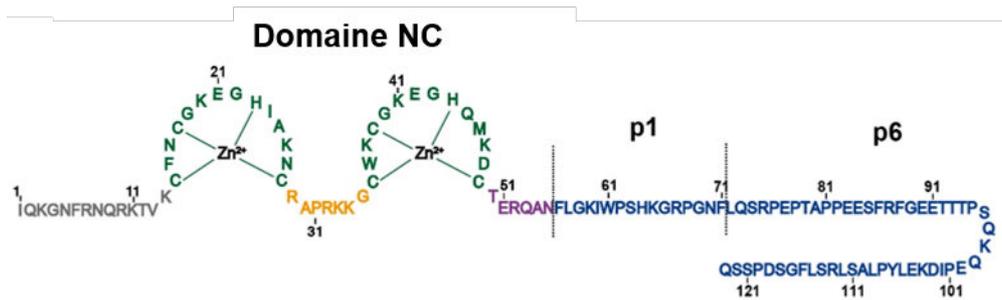


Figure 16A: Séquences des protéines de nucléocapside du VIH-1. Le premier clivage par la protéase libère la forme immature NCp15 (1-123), puis la forme intermédiaire NCp9 (1-71) et finalement la forme mature NCp7 (1-55). Les lignes représentent les deux sites de clivages par la protéase présents dans la NCp15. La partie N-Terminal est en gris, les deux doigts de zinc en vert, le linker en orange, la partie C-terminal en mauve et enfin p1-p6 en bleu.

Au cours de la maturation de la particule virale, la protéine Précurseur Pr55 Gag va subir une série de clivages, réalisée par la protéase (PR) en des sites bien spécifiques³³¹ et qui se déroule dans un ordre séquentiel défini (Figure 17A) : Le clivage initial se produit après le peptide p2, les produits résultant de ce clivage sont le domaine p41 (MA, CA, p2) et le domaine p15 (NC, p1, p6)³³². Le domaine p15 (NC, p1, p6) constitue le domaine NC sous forme immature NCp15 qui doit obligatoirement être lié à l'ARNg viral pour subir un second clivage après le peptide p1³³³⁻³³⁴, libérant le domaine p6 et aboutissant à une forme intermédiaire, la NCp9 (NC, p1). Un dernier clivage après le domaine NC, libère le domaine p1 et aboutit à la forme mature de la NC nommée NCp7. Quant au domaine p41, il va également subir un processus de maturation, la PR va d'abord cliver après la MA pour aboutir à une MA mature (p17) et un domaine p25 (CA, p2), ce dernier subira un clivage après la CA pour la libérer du petit peptide p2 (Figure 17A). Ainsi, au cours de la maturation, la protéine de nucléocapside existe d'abord sous une forme immature ou intermédiaire NCp15 (123 aa) et NCp9 (71 aa) respectivement avant d'adopter sa forme mature NCp7 (55 aa), forme prédominante dans les particules virales matures³³⁵⁻³³⁶. Il faut noter que le clivage correctement fait par la protéase permet la maturation de la NC, qui est essentielle pour l'infectivité du VIH-1³³⁶⁻³³⁸. Effectivement, l'assemblage, le bourgeonnement et le clivage protéolytique de Gag et Gag-pol sont des événements très reliés. Pour ce qui est de la durée des événements, il semblerait que l'activation de la protéase a lieu plus tard, au cours de l'assemblage, à condition que le clivage soit complet au moment du bourgeonnement

ou juste après la libération de la particule virale³³⁹.

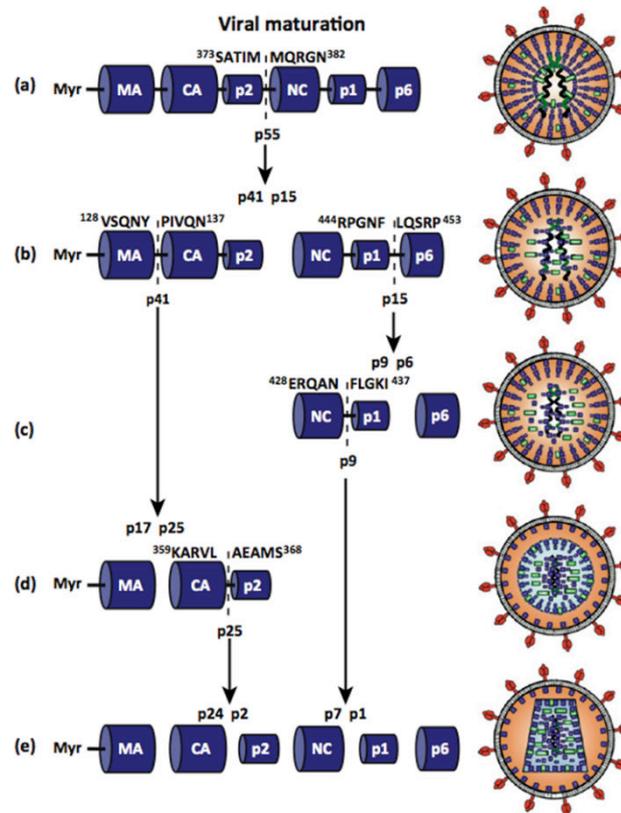


Figure 17A: Illustration du clivage protéolytique de la polyprotéine Gag^{Pr55} par la PR et l'influence de cela sur la maturation de la particule virale²³⁸.

I.4.2 Structure

La protéine de nucléocapside mature NCp7 est constituée de 55 acides aminés et possède une masse moléculaire de 6430 Da. C'est une protéine qui contient une partie N-terminale fortement basique et deux doigts de zinc (ZF) reliés par une région charnière ou "linker" fortement basique (Figure 18A). Ce linker est fortement conservé et il a été montré que le résidu Pro31 joue un rôle critique dans le rapprochement des doigts de zinc³⁴⁰⁻³⁴². Chaque ZF possède un motif conservé CCHC fixant le zinc qui est défini par un arrangement C-X₂-C-X₄-H-X₄-C (C=Cystéine, H=Histidine et X est un acide aminé variable)³⁴³⁻³⁴⁴. Ces motifs CCHC sont capable de chélater le Zn²⁺ de manière stœchiométrique et avec une affinité élevée d'environ 10⁻¹³ M³⁴⁵⁻³⁴⁸. Le cation Zn²⁺ est coordonné de manière tétraédrique par les groupements thiols des cystéines et aminos de l'histidine^{343, 347, 349-350}. La structuration des domaines à doigts de zinc permet la formation de plateaux hydrophobes dans chacun des doigts de zinc : le premier est constitué par

les résidus Val13, Phe16, Ile/Thr24 et Ala25 ; le second est constitué par les résidus Trp37, Gln45 et Met46^{344, 349, 351}. Ce plateau hydrophobe est impliqué de façon essentielle dans la fixation de la NCp7 aux acides nucléiques, via de multiples contacts avec les bases des oligonucléotides, selon des modalités qui seront discutées au chapitre suivant. Il est à noter que la structuration des doigts de zinc est essentielle pour l'activité de la protéine^{344, 349} d'autant plus que la conservation du motif CCHC est très élevée et témoigne donc de son importance biologique. En effet, plusieurs études ont démontré que des mutations au niveau des doigts de zinc avaient des conséquences très négatives sur la réplication virale^{177, 351-352}.

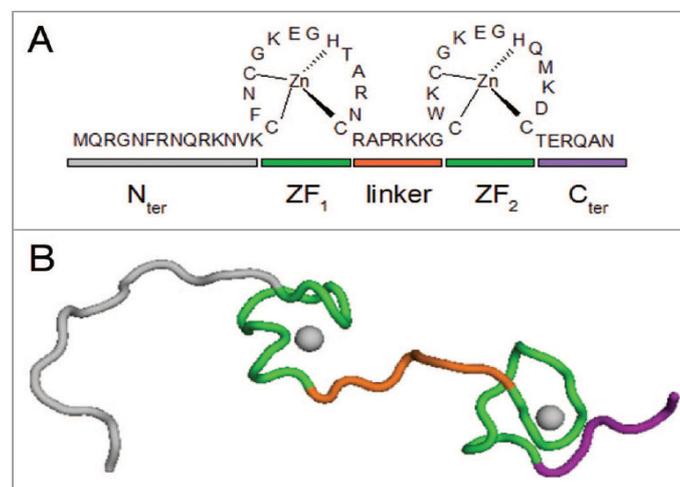


Figure 18A: Structure secondaire (A) et tertiaire (B) de la protéine de nucléocapside NCp7. Le code couleur est le même que pour la figure (16A)¹⁷⁸.

I.4.3 Propriétés

I.4.3.1 Reconnaissance et modalités de liaisons aux acides nucléiques

I.4.3.1.1 Spécificité et affinité de liaison

La protéine de nucléocapside est une protéine qui se lie aux ADN et ARN avec une affinité de l'ordre du millimolaire au nanomolaire. Elle peut se lier de manière spécifique et non spécifique avec un Kd dépendant de la nature, de la séquence et du repliement de l'acide nucléique. Ces variations d'affinités confèrent à la protéine une capacité à exercer différentes fonctions, dépendantes de la nature des séquences d'acides nucléiques impliquées ainsi que des concentrations

respectives de la protéine et de l'acide nucléique ^{151, 353}. Lors du cycle de réplication, les rapports NC:nt (rapport protéine par nucléotide) varient considérablement, la protéine peut se comporter comme une protéine reconnaissant des sites précis ou fixer presque tout l'ARN de manière non spécifique selon les étapes du cycle ³⁵³.

Du fait de son caractère basique, la fixation de la NCp7 aux acides nucléiques est très dépendante de la force ionique ³⁵⁴⁻³⁵⁵. Ainsi, le mode de fixation prédominant est un mode électrostatique non spécifique et de faible affinité. Il est connu que la NCp7 a une préférence de liaison pour les acides nucléiques simple-brin ³⁵⁶. Des études par résonance plasmon de surface (SPR) ou de nano-manipulation de molécule unique ont démontré que la NCp7 se fixe préférentiellement aux ADNs simples brins ³⁵⁷⁻³⁵⁸. La NCp7 se fixe aux acides nucléiques à travers des interactions hydrophobes, électrostatiques et des liaisons hydrogènes ^{357, 359-361}. Parmi les régions simple brin, situées sur les boucles internes ou apicales de structures tige-boucles repliées, la NCp7 a une préférence marquée pour les séquences contenant des guanines et en particulier les séquences TG, TXG, GXG (avec X représentant une base variable) ^{175, 177, 357, 362-364}. On retrouve ces motifs dans les tiges boucles SL1, SL2, SL3, TAR ARN, cTAR ADN ^{175, 363-364} et PBS ADN ³⁶⁵.

Certaines fonctions de la NC impliquent des liaisons spécifiques à certaines séquences d'acides nucléiques, alors que d'autres sont associées à des propriétés de liaisons non spécifiques. La propriété de spécificité de liaison est principalement associée à la reconnaissance sélective du génome, dans ce cas le domaine NC au sein de Gag reconnaît le signal d'encapsidation Psi de l'ARNg et en particulier les boucles SL1, SL2 et SL3, ceci permettant sa reconnaissance spécifique et sa sélection à partir d'un large pool d'ARNs cellulaires ^{63, 278}.

1.4.3.1.2 Etudes structurales des complexes NCp7 acides nucléiques

Les études structurales effectuées par RMN, ont permis de mettre en évidence les modalités principales du mode de fixation de la NCp7 aux acides nucléiques et de décrire les bases structurales de la sélectivité pour les guanines.

A l'heure actuelle, il existe six études de complexes entre la NCp7 (sous formes tronquée ou complète) avec les acides nucléiques, qui ont été déterminés avec les méthodes RMN. Trois de ces complexes comportent des brins ADN, ^{362, 365-366} et trois

autres des brins ARN ^{352, 360, 367}. Seuls cinq structures ont été publiés dans la PDB, les résultats de l'étude de Spriggs et al, 2008 n'ont à ce jour pas été publiés dans la PDB.

D'un point de vue fonctionnel, l'étude de ces complexes a permis d'améliorer la compréhension des mécanismes de reconnaissance spécifique de la NCp7 aux acides nucléiques notamment à travers l'analyse des complexes avec les tiges boucles SL2, SL3 et SL4 contenant le signal d'encapsidation Psi. Une information importante issue de ces études est que la NC se lie aux tiges boucles SL2 et SL3 avec une affinité de l'ordre du nanomolaire ³⁶⁸. Il est important de préciser qu'il n'a pas été observé de forte affinité pour les tiges de ces séquences et encore moins pour la séquence SL4. Cependant, comme décrit au paragraphe (I.3.1) SL1 s'est aussi avéré être un site de liaison de haute affinité de Pr55 Gag avec un motif de reconnaissance AGG ²³³.

D'un point de vue structural, l'analyse des complexes montre que les interactions spécifiques sont principalement dues aux plateaux hydrophobes, permettant des interactions non électrostatiques avec les boucles de séquences GGXG (avec X une base correspondant à A, C ou U) au sein de SL2 ou SL3. L'étude de ces complexes montre aussi que les séquences reconnues sont très courtes : quatre ou cinq résidus sont directement contactés mais seulement deux ou trois résidus sont contactés spécifiquement.

Au niveau moléculaire les interactions entre les acides aminés hydrophobes impliqués dans l'interface (Val13, Phe16, I24, A25, Trp37, Gln45 et Met46) et les sucres riboses et désoxyriboses apparaissent assez différentes ³⁶². Cette dernière observation suggère que la nature des sucres pourrait être importante pour l'orientation de la NCp7 le long des chaînes d'acides nucléiques.

Pour chacun des doigts de zinc, la spécificité de liaison pour les guanines a été bien caractérisée : elle est due au fait que les guanines sont profondément insérées dans une poche hydrophobe pour chaque doigt de zinc, avec en plus un réseau de liaisons hydrogène spécifiques des guanines qui stabilisent leur positionnement. Dans le cas des guanines, quatre liaisons hydrogène spécifiques peuvent s'établir entre la NCp7 et l'acide nucléique, alors qu'avec les autres bases (A, C, T, U) seulement une ou deux liaisons hydrogène peuvent s'établir. Enfin la nucléobase insérée dans chacun des doigts de zinc établit aussi des interactions d'empilement

avec un acide aminé aromatique : Trp37 pour ZF2 et Phe16 pour ZF1 (Figure 19A).

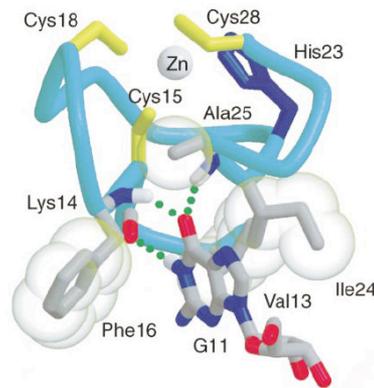


Figure 19A: Structure tridimensionnelle de l'interaction des guanines avec le premier doigt de zinc ³⁵².

Dans le cas des ARNs, une guanine est située dans chacun des deux doigts de zinc (ZF1 et ZF2) et la séquence reconnue est du type G-X-G (où X est un A, C ou U) ³⁶⁵.

Dans le cas des ADNs, une guanine est toujours insérée dans le ZF2, et une insertion seulement partielle de C ou T se produit dans le ZF1.

N'oublions pas de rappeler que, outre ces interactions de nature hydrophobe, de nombreuses autres interactions de type électrostatique, impliquant les acides aminés basiques (se trouvant dans les parties N-terminales, dans les doigts de zinc ou le "linker") sont présentes et jouent un rôle capital dans la stabilisation des complexes.

A côté des études structurales il y a eu des études sur les aspects dynamiques des complexes NC-acides nucléiques ³⁶⁹⁻³⁷¹.

1.4.3.1.3 Polarité de fixation de la NCp7 aux acides nucléiques

Un autre élément intéressant qui émerge de l'analyse des complexes est que la NCp7 paraît être capable de reconnaître la polarité des chaînes d'acides nucléiques simple-brin. Cette polarité est de plus différente pour les brins d'ADN et les brins d'ARN. Le schéma suivant (Figure 20A) explicite cette reconnaissance : (1) dans le schéma du haut (brin ADN) une guanine est fixée par ZF2, comme décrit plus haut, le ZF1 interagit avec un résidu situé en 5' de la guanine interagissant avec le ZF2. (2) dans le schéma du bas (ARN), le ZF1 interagit avec un résidu (qui est une

guanine) situé toujours en 3' du résidu G reconnu par ZF2. Ces cas de figure se retrouvent dans les trois complexes impliquant des ADNs et dans les trois impliquant des ARNs.

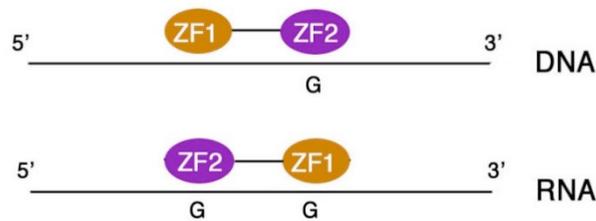


Figure 20A: Schéma descriptif de la polarité de fixation de la NCp7 aux Acides Nucléiques. La NCp7 se fixe dans le sens 5'3' sur l'ADN, c'est à dire de manière parallèle et inversement la fixation est antiparallèle sur l'ARN, avec une préférence des ZFs pour les Guanines.

L'ensemble des nombreuses interactions électrostatiques, de liaisons hydrogènes et d'empilements hydrophobes, qui stabilisent le complexe est conservé quel que soit le type d'orientation de la NCp7 le long de la chaîne d'acide nucléique. Les contacts avec les sucres riboses ou désoxyriboses diffèrent néanmoins significativement entre les deux types de complexes^{353, 362}. Ces résultats, nous suggèrent l'existence d'une relation entre le type d'interaction des résidus protéiques avec les sucres et l'orientation de la NCp7. Notons que ce type de comportement a déjà été observé chez d'autres protéines, tel la transcriptase inverse^{152, 372}. Par ailleurs, le rôle important des interactions avec les sucres riboses/désoxyriboses suggère que ces contacts pourraient intervenir dans la polarité de fixation de la NCp7³⁶² ainsi que dans la discrimination acide nucléique simple-brin/acide nucléique double brin. En effet, les sucres, sont d'une manière générale beaucoup plus accessibles, dans les acides nucléiques simple-brin que dans les double-brin où ils se trouvent enfouis dans le petit sillon (pour les ADNs). Les contacts avec les sucres suggèrent aussi un mode de déstabilisation des portions double-brin adjacentes aux régions fixées par la NCp7. Les aspects dynamiques de la liaison de NCp7 aux acides nucléiques apparaissent critiques et semblent être un déterminant clé de son activité chaperonne.

1.4.3.2 La non équivalence des deux doigts de zinc

Les deux doigts de zinc sont requis pour l'interaction avec les acides nucléiques

et la réplication virale, mais ces derniers semblent jouer des rôles asymétriques^{328, 373}. Les doigts de zinc jouent un rôle majeur dans le processus de déstabilisation des structures secondaires des acides nucléiques et la liaison rapide aux séquences d'acides nucléiques, contrairement à la partie N-Terminale qui est principalement responsable de l'activité agrégative des acides nucléiques^{53, 138, 140, 177-178, 353, 374-376}. Des expériences avec un mutant de NC, où le doigt de zinc en N-Terminal (ZF1) a été dupliqué (mutant ZF1 :ZF1) montrent, que le mutant possède une meilleure efficacité de réplication comparé à un mutant de NC où c'est le doigt de zinc en C-Terminal (ZF2), a été dupliqué (mutant ZF2 :ZF2), où encore par rapport à un mutant où les deux doigts de zinc ont été inversés (mutant ZF2 :ZF1)³⁷³. L'analyse de l'activité de protéine chaperon de ces mêmes mutants, a montré que le premier doigt de zinc (ZF1) était plus important pour cette activité comparé au deuxième doigt de zinc (ZF2)^{358, 377-378}. Dans le détail, la facilitation des transferts de brins par la NC ainsi que l'hybridation des séquences d'acides nucléiques complémentaires n'a été observé que quand le premier doigt de zinc (ZF1) était à sa place native, en d'autres termes les mutants ZF2:ZF1 et ZF2:ZF2 se sont montrés incapables d'assurer ces réactions³⁷⁸⁻³⁷⁹.

Par ailleurs, dans chaque doigt de zinc se trouve un résidu aromatique (Phe16 dans ZF1 et Trp37 dans ZF2), il a été démontré que ces résidus sont impliqués dans la stabilisation des contacts avec les acides nucléiques en établissent des interactions d'empilements (stacking) avec les guanines et dans certains cas avec les thymines^{352, 360, 362, 365-367}. La mutation de ces résidus aromatiques, Phe16 et Trp37, a démontré leur rôle essentiel pour la fixation aux acides nucléiques ainsi que pour l'activité de protéine chaperon^{239, 377, 380-381} avec des conséquences plus dramatiques pour cette dernière quand la mutation touche le premier doigt de zinc^{358, 373, 377-378}.

Comme il l'a déjà été dit plus haut, d'après l'analyse des structures des complexes NC:acides nucléiques déterminés par RMN^{352, 360, 362, 365-367}, on observe qu'à l'échelle atomique, dans tous les complexes une guanine est insérée dans la poche hydrophobe du deuxième doigt de zinc (ZF2), tout en établissant des interactions d'empilements avec le Trp37. Cependant, dans le cas du premier doigt de zinc (ZF1), on observe l'insertion d'une guanine dans sa poche hydrophobe, seulement quand deux guanines sont disponibles dans la séquence d'acide nucléique, sinon le ZF1

peut également établir des interactions d'empilements partielles entre son Phe16 et une thymine. On peut aussi y observer l'interaction avec une cytosine mais celle-ci reste en dehors de la poche hydrophobe du doigt de zinc^{362, 365-366, 380}. Ces données structurales suggèrent que le deuxième doigt de zinc est plus apte à fixer les guanines accessibles que le premier. De ce fait, la délétion du deuxième doigt de zinc (ZF2) semble plus critique pour la fixation du domaine NC au sein de Gag, à ses acides nucléiques cibles, ce qui démontre que le Trp37 est plus important pour la stabilisation des complexes NC:acides nucléiques³⁸². Cependant, le premier doigt de zinc (ZF1) isolé, se lie aux séquences ADN et ARN avec une forte affinité, comparé au deuxième doigt de zinc isolé (ZF2)³⁸³⁻³⁸⁵, alors que les mutants ZF2:ZF1 et ZF1-ZF1 se lient avec une très forte affinité aux acides nucléiques substrats comparé à la protéine sauvage et au mutant ZF2-ZF2³⁸⁶. Ces résultats, démontrent que le premier doigt de zinc (ZF1), sous forme isolée ou en position C-terminal de NCp7, se lie aux acides nucléiques avec une plus forte affinité que le ZF2. Il se trouve que ZF1 apparaît comme étant plus important que le ZF2 pour l'activité de protéine chaperon de la NCp7 et la réplication virale, mais pas pour la fixation aux acides nucléiques, ce qui est assez intrigant. Cette observation nous a amené à reconsidérer la coordination des activités entre les deux doigts de zinc et le rôle de la dynamique interne de la protéine. La dynamique interne des protéines se liant aux acides nucléiques est essentielle pour la reconnaissance de leurs substrats ADN ou ARN³⁸⁷⁻³⁹¹; elle régule aussi probablement la coordination des activités des différentes parties. L'investigation de cette dynamique a fait l'objet de nombreuses études, dont une réalisée au laboratoire, grâce à des expériences RMN mesurant les temps de relaxation ¹⁵N de la NCp7^{370-371, 392-394} à deux champs différents (950 et 500 MHz), ce qui a permis de comprendre l'influence de celle-ci sur la fonction des deux doigts de zinc³⁶⁹. Une étude antérieure par Ramboarina et al., 2002, réalisée avec des expériences de relaxation ¹⁵N par RMN et de spectroscopie de fluorescence, sur la protéine libre (12-53) et liée à un ADN d(ACGCC), a démontré une transition de la dynamique interne et globale de la protéine, celle-ci est dépendante de la température, mène à l'ouverture de la molécule. Cependant une fois le peptide lié, on observe une réduction de la flexibilité conformationnelle du peptide, avec un échange conformationnel qui a été déplacé du ZF2 vers le résidu Ala30 du linker. Cette dynamique est liée au rôle multifonctionnel de la NCp7 dans

le cycle répliatif du VIH-1, qui requiert autant d'interactions spécifiques que non spécifiques avec les acides nucléiques. En effet, ces propriétés dynamiques de la NCp7 confèrent à la protéine une adaptabilité nécessaire pour la reconnaissance de différentes séquences d'ARN.

Lee et al., 1998 ont confirmé la présence de faibles NOE intermoléculaires³⁴⁹ entre les deux doigts de zinc, qui impliquent une proximité entre les chaînes latérales de la Phe16 et du Trp37, ces données ne sont pas en faveur d'une seule conformation de la protéine. Grâce à des expériences de relaxation ¹⁵N, il a été démontré que les deux doigts de zinc possèdent deux différents temps de corrélation rotationnel, ce qui indique que les deux doigts de zinc ne bougent pas tel qu'un domaine globulaire singulier mais que les interactions entre les deux doigts de zinc sont transitoires.

Quant aux résultats obtenus dans notre laboratoire, ils sont en accordances avec les travaux décrits plus haut et suggèrent que les deux doigts de zinc bougent l'un par rapport à l'autre autour d'une charnière moléculaire, localisée au niveau du résidu Gly35 sur le Linker. D'autre part, plusieurs autres résidus au sein du Linker : Arg29, Ala30, Pro31, Arg32 et Lys33, montre des mouvements restreints c'est-à-dire qu'ils s'avèrent être assez rigides et en contact assez étroits (visualisé par le grand nombre d'interactions NOE) avec le ZF1. Ces contacts Linker-ZF1 sont probablement un obstacle pour les interactions entre le domaine ZF1 et les guanines non appariées alors que le ZF2 est plus accessible et susceptible d'interagir avec ces dernières. Rappelons que ZF1 grâce à son large plateau hydrophobe (Val13, Phe16, Thr24 et Ala25) est le plus apte, à déstabiliser les régions doubles brin adjacentes aux guanines insérées au niveau du ZF2. Cela, souligne la non équivalence des deux doigts de zinc et désigne le ZF2 comme celui présentant une activité de fixation "initiale" alors que ZF1 posséderait l'activité déstabilisatrice. Finalement un modèle a été proposé pour illustrer le rôle du "Linker" qui module l'activité de protéine chaperon et la dynamique interne de la protéine (Figure 21A).

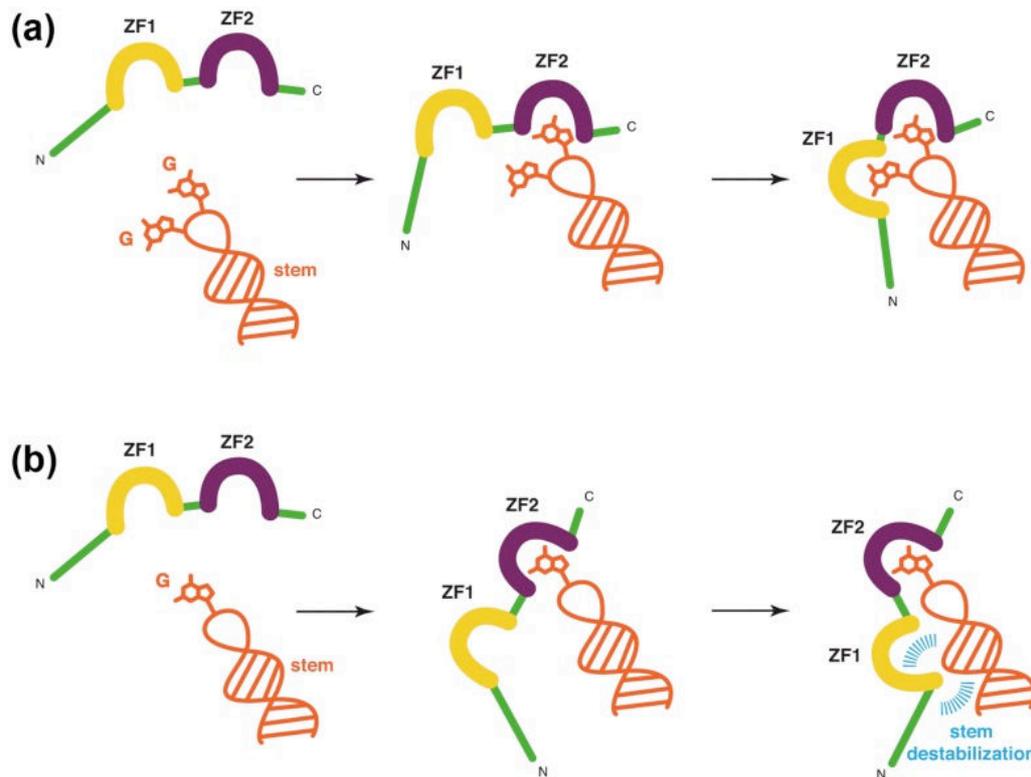


Figure 21A: Représentation schématique de la non équivalence des doigts de zinc de la NCp7 dans son mécanisme de fixation aux acides nucléiques. L'acide nucléique est représenté en orange, Le ZF1 est en jaune, le ZF2 en mauve et le reste de la protéine est en vert. (a) Situation #1 avec deux guanines accessibles. Tout d'abord le ZF2 se lie à une guanine accessible, et le ZF1 se lie à la guanine qui reste. La tige de l'AN n'est pas déstabilisée. (b) Situation #2 avec une seule guanine accessible. Le ZF2 se lie à cette guanine accessible, et le ZF1 reste libre pour contacter la tige double brin via son plateau hydrophobe. La tige est alors déstabilisée (en bleu)³⁶⁹.

I.4.3.3 Activité de protéine chaperon des acides nucléiques

La NCp7 est une protéine chaperon des acides nucléiques, c'est à dire qu'elle oriente la réorganisation de deux molécules d'acides nucléiques (ADN ou ARN) en un complexe de stabilité thermodynamique plus élevée^{177, 395}. La NCp7 accomplit ce processus en réduisant les barrières énergétiques pour dissoudre et reformer des paires de bases. L'interaction entre les molécules d'acides nucléiques et les protéines de nucléocapsides mène à une ouverture transitoire des paires de bases pour leur réassociation en des combinaisons alternatives. De nombreuses études ont été réalisées afin de comprendre les propriétés chaperons de la NCp7 pour ses substrats ADN et ARN³⁹⁶⁻³⁹⁷. Il en ressort que l'activité de protéine chaperon de la NCp7 est essentielle pour les deux transferts de brins obligatoires au processus de transcription inverse ainsi qu'à la plupart des transferts de brin internes^{356, 398}, mais

également pour l'étape d'initiation de la transcription inverse qui consiste en l'hybridation de l'ARNt^{Lys3} cellulaire avec la séquence PBS de l'ARNg⁴².

Néanmoins, la protéine NCp7 et des formes tronquées de la protéine Gag se sont montrées efficaces pour réaliser l'hybridation de la séquence PBS avec l'ARNt^{Lys3}³⁹⁹, ces protéines Gag tronquées se sont même montrées beaucoup plus efficaces dans cette réaction⁴⁰⁰, ce qui pourrait impliquer que l'ARNt^{Lys3} est hybridé au cours du bourgeonnement viral avant le clivage protéolytique de Gag^{Pr55}. Cependant, afin de vérifier cette hypothèse, il est nécessaire de réaliser ces expériences avec la protéine Gag^{Pr55} complète, car de nombreuses études *in vitro* ont été réalisées avec une forme Gag sans le domaine p6. D'autre part, il est clair que seule la forme mature NCp7 peut faciliter le premier transfert de brin⁴⁰¹.

Sur le plan du fonctionnement de l'activité de protéine chaperon de la NCp7, celle-ci est déterminée par deux activités distinctes : la première correspond à une activité déstabilisatrice des structures secondaires des acides nucléiques conféré par les deux doigts de zinc⁴⁰²⁻⁴⁰⁴ ; la deuxième correspond à une activité agrégative ou d'hybridation des séquences complémentaires d'acides nucléiques associée au domaine N-Terminal fortement basique^{177, 395, 398, 405-406} ainsi qu'au domaine linker riche en résidus basiques^{311, 353, 374, 407}. C'est grâce à cette combinaison d'activité déstabilisatrice et agrégative que la NCp7 facilite les processus de transfert de brin au cours de la transcription inverse⁴⁰⁸. La NCp7 peut également déstabiliser des duplexes d'acides nucléiques pour former d'autres duplexes encore plus stables avec un autre brin d'acide nucléiques pour lequel la complémentarité est beaucoup plus étendue⁴⁰⁹, cette dernière opération s'appelle un déplacement de brin. D'autres mécanismes influencent l'activité de protéine chaperon de la NCp7, notamment la cinétique rapide de la liaison de la protéine aux acides nucléiques^{358, 374}.

Il a été montré que pour atteindre une activité de protéine chaperon optimale, le rapport NC:nt doit être compris entre 1:8 à 1:6 nt, ce qui correspond à des proportions similaires à celles existant dans le virion. On peut également dire que ces valeurs mettent la protéine dans des conditions en « mode protéine chaperon » ou « chaperoning mode »³⁹⁵. Cette activité de protéine chaperon peut-être mimée expérimentalement en l'absence de la NCp7, dans une solution fortement saline en chauffant brutalement et en refroidissant par étape⁴¹⁰⁻⁴¹¹. Selon le niveau d'occupation de l'acide nucléique par les molécules de NCp7 et donc, le rapport

NC:nt, trois différents modes d'actions sont proposés ³⁵³ (Figure 22A):

i-Un mode de fixation spécifique: Se produit quand la protéine est en concentration limitante, ce qui induit la formation de simples complexes ribonucleoprotéiques (RNP).

ii-Un mode d'activité de protéine chaperon: Survient quand plusieurs protéines et que toutes les molécules d'acides nucléiques sont disponibles pour interagir, menant à la formation de RNP de masses moléculaires importantes.

iii-Un mode de protection de l'ARN: Qui est observé quand la protéine est à des concentrations de saturation et que toutes les molécules d'ARN se retrouvent recouvertes par la protéine.

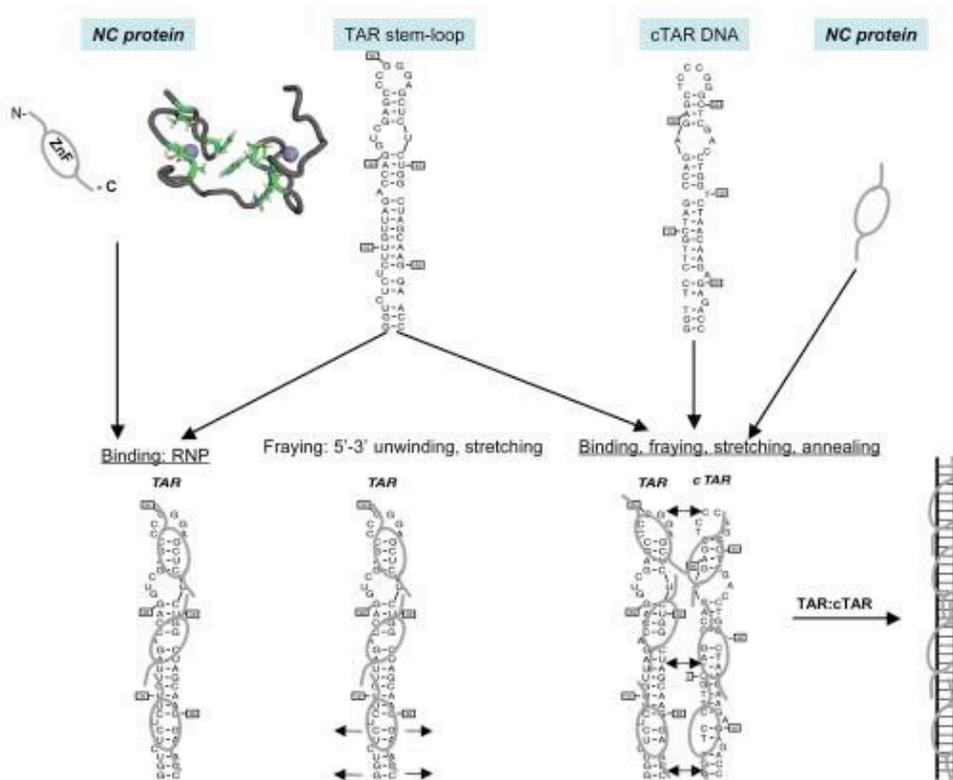


Figure 22A: Illustration schématique des propriétés de fixation, de déstabilisation et d'agrégation des acides nucléiques par la NCp7 du VIH-1. En haut, de gauche vers la droite, la NCp7 est représentée par un cercle ovale correspondant à son plateau hydrophobe (ZF), bordé par les régions désordonnées N et C terminales présentées ici en petits fragments. La structure 3D du plateau hydrophobe à la surface des doigts de zinc de la NCp7 sont représentés en mode "stick" (le mode CPK montre le volume réel relatif des atomes) avec Val13, F16, Ile24, Ala25, Trp37, Gln45 et Met46, et les atomes de zinc apparaissent sous forme de sphères grises. La séquence TAR avec sa boucle et ses bulges sont représentées ici en tige boucle. Son ADN complémentaire cTAR est également représenté ici en tige boucle. En bas, de gauche vers la droite : La fixation de la NCp7 sur TAR résulte en la formation d'un ARN complètement recouvert par les molécules de NCp7,

ce qui cause l'ouverture transitoire des extrémités 5' et 3' (flèches courtes), par un processus de « fraying » et de « strand stretching ». La fixation de la NCp7 à TAR et cTAR mène à la formation de complexes nucléoprotéiques où les ANs sont entièrement recouverts par les molécules de NCp7, cela mène à l'hybridation rapide des séquences complémentaires ce qui aboutit à la formation d'un hybride ADN :ARN²²⁴.

1.4.3.3.1 Déstabilisation des structures secondaires des acides nucléiques

La capacité de déstabilisation des acides nucléiques par la NCp7 est médiée par ces deux doigts de zinc, dépend fortement de la stabilité des acides nucléiques, de la présence de motifs tels que des "bulges", des boucles apicales, des mis-appariements ainsi que des paires de bases terminales^{178, 412}.

L'un des modèles proposés est afin de déstabiliser les structures secondaires des acides nucléiques, la NCp7 progresse le long des brins des acides nucléiques en se fixant sur les régions simples brins, puis en déstabilisant les paires de bases adjacentes. Cela permet aux régions simples brins précédemment "séquestrées" dans les régions doubles brins de devenir accessibles et de s'apparier avec des séquences complémentaires situées sur d'autres molécules (Figure 23A). Ce mode d'action, très schématique et non démontré, permettrait une propagation de la déstabilisation tout en permettant un glissement de la NCp7 ou la fixation d'autres molécules de NCp7¹⁷⁸.

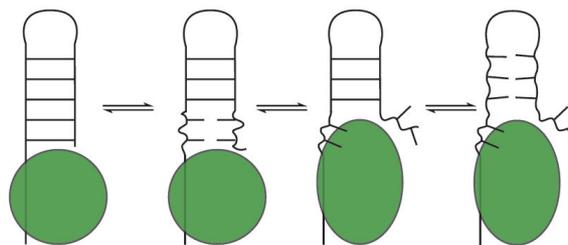


Figure 23A: Mécanisme schématique de la déstabilisation d'une séquence repliée d'acide nucléique par la NCp7. La protéine (vert) se fixe premièrement sur la région simple brin et déstabilise ensuite la région double brin voisine pour pouvoir se fixer aux nucléotides exposés, c'est ainsi qu'elle va permettre l'appariement des séquences complémentaires⁴¹³.

Le mode de fixation "polarisée" de la NCp7 le long des chaînes d'acides nucléiques tel que nous l'avons décrit au paragraphe (1.4.3.1.3), devrait aussi conditionner la localisation des zones double brin qui vont être déstabilisées par la NCp7 et donc « piloter » tout le mécanisme de déstabilisation des structures repliées.

En effet la fixation de la NCp7 au niveau d'une région simple brin va positionner, dans le cas des ARNs, le domaine ZF1 impliqué dans la déstabilisation à proximité des séquences double brin situées en 3' de cette région simple brin, ce seront donc les zones double brin situées en 3' des boucles internes qui seront déstabilisées initialement et non celles situées en 5', la déstabilisation est « polarisée » (Figure 24A).

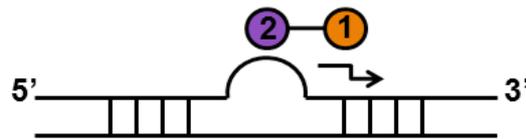


Figure 24A: Schéma descriptif du mode de déstabilisation d'une séquence ARN double brin située à proximité d'une boucle interne. Dans un complexe NCp7:ARN, Le ZF2 est d'avantage impliqué dans la fixation aux régions simple brin et le ZF1 dans la déstabilisation.

De façon frappante, de tels "patterns" de déstabilisation ont été observés dans l'analyse des régions double brin déstabilisées par la NCp7 dans l'ensemble du génome. Dans la région 5' UTR du virus, qui comporte le plus de régions déstabilisées par la NCp7, les séquences déstabilisées sont toujours situées en 3' des boucles internes simple-brin ²⁶⁴. La polarité de fixation de la NCp7 le long des acides nucléiques permet aussi de mieux comprendre le processus de redistribution de la NCp7 des brins ARN aux brins ADN lors de la transcription inverse ⁴¹⁴.

La compréhension du mode d'action de la NCp7 pour procéder à la déstabilisation des structures secondaires est cruciale à l'obtention d'une meilleure compréhension des transferts de brins. Exemple, au cours du premier transfert de brin, l'étape initiale est liée à la déstabilisation des structures secondaires de la séquence TAR et de son ADN complémentaire cTAR suite à leur liaison à la NCp7 ^{209, 355-356}.

Il a été démontré que la NCp7 active l'ouverture transitoire des paires de bases terminales de l'ADN cTAR (Figure 25A) ce qui mène à un "melting" partiel des terminaisons de cTAR jusqu'au bulge T40 ou au mésappariement C.A44 ^{402, 415}, cette ouverture reste imparfaite car elle demande obligatoirement la présence de son complémentaire TAR ARN pour optimiser cette déstabilisation ⁴¹⁶. Ceci indique également que la déstabilisation de cTAR et l'hybridation avec son complémentaire sont étroitement liées.

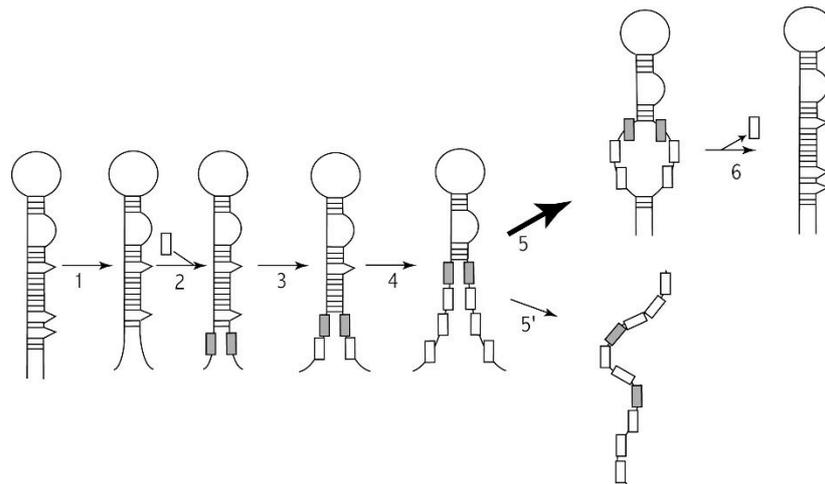


Figure 25A: Modèle illustrant le mécanisme de déstabilisation de cTAR ADN par la NCp7. Les peptides sont représentés par des rectangles. Les rectangles gris correspondent aux premiers peptides qui se lient à cTAR. **1)** En conséquence au « fraying », les extrémités terminales des cTAR sont transitoirement présentes en séquence simple brin. **2)** Vu que la NCp7 a une préférence pour les séquences simple-brin, les extrémités sb de cTAR vont constituer le site initial de fixation de la NCp7. **3)** À partir de là, la NCp7 va déstabiliser les paires de bases adjacentes, permettant le glissement de la protéine et la fixation coopérative de peptides additionnels. **4)** En conséquence à la stabilité limitée de la tige de cTAR, le melting va alors se propager jusqu'au misappariement 10C.A44 ou au bulge T40. À ce moment, il est probable que la stabilité brins double-brin restant peut être assez élevée pour prévenir le « melting » complet de cTAR. En conséquence, il y aurait une compétition entre l'extension du melting (**5'**) et la restructuration des séquences ouvertes (**5 et 6**)⁴⁰².

Il est également important de souligner l'importance des deux doigts de zinc qui confèrent justement à la protéine son activité de déstabilisation³⁵⁸, ceci a été confirmé avec des mutants de NCp7 ne possédant pas de motifs à doigts de zinc (NC(13-61)_{dd}) ou ne présentant pas de structuration ((SSHS)₂NC(1-55) où toutes les Cystéines ont été remplacées par des Sérines pour prévenir la fixation du zinc). En effet, ces derniers ont failli à déstabiliser la structure secondaire de la séquence cTAR, le peptide (SSHS)₂NC (1-55) a été inapte à altérer la coopérativité de la transition helix coil des molécules d'ADN λ ³⁵⁸ et a eu des effets limités sur l'inhibition du « self priming » de cTAR³⁸³. D'autre part, la mutation du Trp37 par une Leu a des conséquences dramatiques sur la fixation et la déstabilisation de cTAR, avec une perte de la fixation de l'ARNt^{Lys3} à la séquence PBS, ce qui bloque l'initiation de la transcription inverse³⁸³. La nécessité de la présence de deux doigts de zinc bien structurés ainsi que du Trp37 au sein du second ZF suggère l'importance du plateau hydrophobe dans le processus de déstabilisation³⁸³.

L'activité déstabilisatrice de la NCp7 apparait inversement proportionnelle à la

stabilité de l'oligonucléotide. Par exemple, la déstabilisation par la NCp7 est plus efficace pour la séquence cTAR ADN que pour son homologue ARN qui est plus stable^{402, 404}. D'ailleurs une stabilisation des parties basses de cTAR a démontré la difficulté de la NCp7 à déstabiliser cTAR ainsi qu'à bloquer le "self priming" de l'ADN strong stop⁴¹⁷. On en conclut, que le "melting" des tiges de cTAR est dû à la présence de bulges terminaux, ce qui aide la NCp7 à déstabiliser cTAR⁴⁰².

Dans le cas du second transfert de brin, la NCp7 promeut l'hybridation entre les séquences complémentaires PBS (+) et PBS (-), ce mécanisme est dépendant de la déstabilisation des tiges de PBS⁴¹⁸. Il a également pu être montré que la fixation de la NCp7 à la séquence 5-CTG-7 de PBS (-), induit une "présentation" des bases présentes dans la boucle apicale ainsi qu'une déstabilisation de la paire C5-G11 adjacente à la boucle³⁶⁵, Cet ensemble d'événements favorise le mécanisme d'hybridation du type "kissing complexe" (interaction boucle-boucle) entre les séquences PBS complémentaires (+) et (-)⁴¹⁹.

1.4.3.3.2 Hybridation des séquences complémentaires d'acides nucléiques

Outre sa propriété déstabilisatrice des structures secondaires, la NCp7 possède également une propriété agrégative des acides nucléiques non spécifique³⁹⁶, conférée par la partie N-terminale^{177, 395, 398, 406, 420} et le domaine linker qui sont fortement basiques^{311, 353, 407}. En effet, cette propriété d'agrégation des acides nucléiques est promue par les résidus basiques de la NCp7 qui vont neutraliser les charges négatives des acides nucléiques, ce qui réduit la répulsion électrostatique entre les molécules d'acides nucléiques. Cependant, bien que l'action principale de la NCp7 reste l'hybridation des séquences complémentaires des acides nucléiques, il faut noter que si le degré d'occupation de l'ARN par les molécules de NCp7 est très élevé, la protéine va alors induire la réaction inverse, c'est-à-dire qu'elle va déstabiliser les duplexes de courtes tailles (<20 nt) ou imparfaits⁴²¹.

Comme décrit précédemment, l'activité de protéine chaperon de la NCp7 est dépendante de la complémentarité entre les séquences mises en jeu. Quand la complémentarité entre les deux séquences à hybrider est faible (inférieure à 13 sur 28 bases), la protéine défavorise l'hybridation en faisant baisser la constante de vitesse d'hybridation ; pour une complémentarité moyenne d'au moins 13 bases

complémentaire sur 28 entre les deux séquences, la protéine n'a aucun effet ; lorsque la complémentarité entre les séquences est importante (supérieure à 13 bases sur 28), la NCp7 favorise l'hybridation en augmentant la constante de vitesse d'hybridation entre les deux séquences ⁴²². Clairement, la NCp7 favorise les structures les plus stables ce qui est parfaitement en adéquation avec son rôle de protéine chaperon. Dans le cas du premier transfert de brin, la NCp7 va hybrider les séquences homologues TAR/cTAR pour permettre la formation d'un hétéro duplex ADN/ARN stable ^{355, 398}. En absence de NCp7, cette réaction d'hybridation est très lente ^{355-356, 406, 423} et en présence de NCp7, elle est 10^5 fois plus rapide, tout en restant dans des conditions physiologiques de sels et de température ⁴²⁴. Ce processus d'hybridation peut se produire à priori selon deux voies principales, l'une impliquant une interaction boucle-boucle par les deux boucles apicales des séquences complémentaires (" kissing complex "), et la seconde impliquant une hybridation par l'extrémité des molécules, ce qui formera des structures de type cruciforme (" zipper complex ") (Figure 26A).

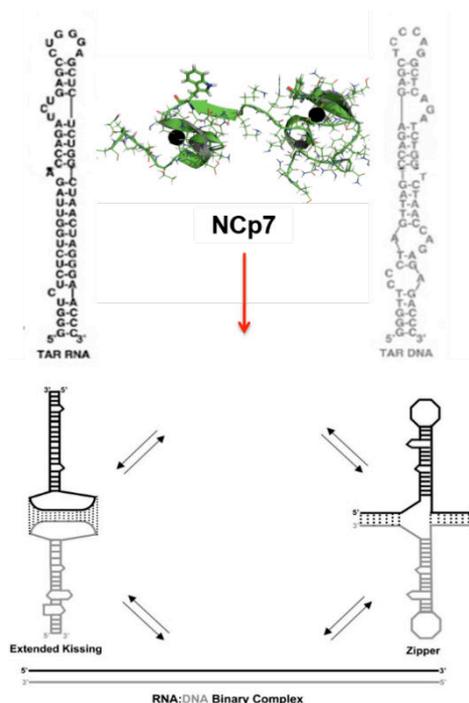


Figure 26A: Schéma représentatif des deux mécanismes d'hybridation des séquences TAR/cTAR facilités par la NCp7. En haut sont représentées les tiges boucles TAR et cTAR libres. À gauche, on observe une voie d'hybridation par les extrémités 3'/5' des tiges, mécanisme « zipper ». À droite, on observe une voie d'hybridation par les boucles, mécanisme « kissing ». Finalement, on obtient un hétéroduplex ADN/ARN (adaptée d'après ⁴²⁴).

En absence de NCp7 le mécanisme majoritaire semble être le complexe boucle-boucle ("kissing complex")⁴²⁴⁻⁴²⁶. D'après les travaux de Vo et al, 2009, effectués avec le protéine NC(1-55), le mécanisme majoritaire évolue lorsque le rapport NC:nt change et plus la concentration de NCp7 est élevée plus le mécanisme de type "zipper" devient majoritaire. Ce basculement de voie indique que le mécanisme d'hybridation est à la fois dépendant de la séquence de l'acide nucléique et de la concentration de la protéine⁴²⁴⁻⁴²⁵. Dans le cas du second transfert de brins, la NCp7 va hybrider les séquences ADN PBS(+)/PBS(-) pour permettre la formation d'un homo-duplex stable ADN/ADN⁴¹⁹. En absence de NCp7, les séquences PBS (+) et (-) s'hybrident par leur tiges basses simples brins nommées « ss overhangs » suivant un mécanisme « zipper complex »³⁵³. En présence de NCp7, la cinétique d'hybridation s'accélère et change de voie vers un mécanisme d'hybridation par les boucles suivant un mécanisme « kissing complex »³⁵³. Par ailleurs, les tiges des séquences PBS sont déstabilisées par la NCp7, confirmant une déstabilisation dépendante de la NCp7 nécessaire pour la réaction d'hybridation⁴¹⁹.

I.4.4 Comparaison des activités de protéines chaperons des différentes formes de NC (p15, p9 et p7)

Il a été démontré que des mutations affectant le clivage protéolytique résultent dans la production de virus contenant la forme NCp15 possédant des déficiences au niveau de l'infectiosité et de la réplication virale³³⁶⁻³³⁸. Plusieurs études ont été réalisées dans le but de comprendre la nécessité du clivage de la NCp15 pour la réplication du VIH-1, en comparant les propriétés de l'activité de protéine chaperon et de l'activité de fixation aux acides nucléiques par les trois formes majeures de la NC : NCp7, NCp9 et NCp15, exhibant des différences importantes, observées *in vitro* concernant leurs activités de protéines chaperons des acides nucléiques³¹¹. La NCp9 a affiché la plus forte activité de protéine chaperon, qu'il s'agisse d'activité de fixation, d'hybridation, d'aggregation et de déstabilisation des séquences d'acides nucléiques³¹³. Cette haute activité est conservée au cours du processus de transcription inverse comparé aux deux autres formes de protéines⁴²⁷, néanmoins, la NCp9 s'est montré plus lente dans la stimulation du processus de dimérisation de l'ARNg immature contrairement à la NCp7 et la NCp15, alors que la NCp9 est tout à

fait apte à faciliter la formation du dimère mature, contrairement à la NCp15 qui ne peut pas du tout réaliser cette opération ²³⁷. Bien que les taux de l'activité de protéine chaperon des trois formes de protéines peuvent être différents selon le type d'expérience utilisée, cette différence observée *in vitro* ne permet pas d'expliquer totalement pourquoi l'infectiosité virale décroît aussi dramatiquement dans les VIH-1 contenant la NCp15. Des tests de sédimentation et des expériences de « Dynamic Light Scattering » ou DLS, ont démontré que la NCp7 et la NCp9, promeuvent toutes les deux la formation de larges agrégats protéines-ARN, alors que les complexes NCp15-ARN n'ont pas pu être décrochés et n'ont pas réussi à disperser (scatter) la lumière par DLS ³¹³. Ceci est en accord avec une étude par microscopie électronique qui a démontré que la NCp15 se lie à l'ADN simple brin tel 'des perles sur une chaîne', contrairement à la NCp7 ou la NCp9, qui forment des agrégats condensés ⁴²⁸⁻⁴²⁹.

Des expériences de RMN en solution suggèrent que le mécanisme d'interaction de la NCp15 aux acides nucléiques est régulé par son domaine p6, comme ce dernier est à caractère acide, il interagirait avec les résidus basiques des doigts de zinc avec une conformation en "fold-back" de la protéine sous forme libre ³¹³, cette interaction serait principalement de nature électrostatique, vu qu'à forte concentration de sels (500 mM NaCl) les déplacements chimiques sont légèrement perturbés. La neutralisation des résidus basiques de p6 améliore l'activité de protéine chaperon de la NCp15, réduit la conformation en "fold-back" et induit des agrégats plus larges suite à la fixation de la protéine aux acides nucléiques ³¹³. Des interactions similaires entre les domaines CTD et NTD ont été proposées pour la protéine de nucléocapside du HTLV-1 ⁴³⁰⁻⁴³², bien que la NC du HTLV-1 est une faible protéine chaperon des acides nucléiques, la délétion de son domaine CTD restaure ses fonctions chaperons ^{430, 433}. De ce fait, que ce soit pour la NC du HTLV-1 ou pour la NCp15, leurs domaines CTD acides régulent négativement leurs activités de protéines chaperons ainsi que leurs cinétiques de fixation aux acides nucléiques ^{313, 430}. Si l'on considère que la conformation de la NCp15 une fois liée aux acides nucléiques est plus ouverte comparée à la protéine libre qui est repliée sur elle-même ³¹³, cela pourrait expliquer la nécessité d'interaction avec l'ARNg pour que la PR puisse réaliser un clivage efficace au niveau du site SP2/p6 ^{333, 428}. D'ailleurs, il a bien été

démontré qu'une variante de la NCp15 incapable de fixer l'ARN ne pouvait être clivée de manière efficace par la PR³³³.

I.5 Les protéines chaperons des ARNs

I.5.1 Le désordre intrinsèque d'une protéine

Pendant longtemps, le paradigme classique de la biologie structurale définissait l'existence d'une relation forte et unique entre la structure et la fonction d'une protéine. Ainsi une séquence d'acides aminés donnée définissait une conformation tridimensionnelle unique, rigide et stable constituant un pré-requis absolu pour la fonction biologique de la protéine. Ce concept a incontestablement permis d'accumuler une quantité importante de données structurales sur les protéines et les enzymes⁴³⁴. Ces données ont été très fructueuses dans la compréhension des fonctions des protéines. Elles ont également profité à la pharmacologie, car en effet, grâce aux études structurales, la synthèse d'agents thérapeutiques ciblant des sites actifs bien définis, inhibant l'activité d'une protéine sont désormais possibles. Cependant, le dogme structural, selon lequel la protéine a besoin d'une structure bien définie pour exercer une fonction, s'est depuis quelques années avéré largement erroné. En effet, il a maintenant été montré que plusieurs protéines ou domaines protéiques ne possèdent pas de structures bien définies et exercent leur fonction dans cet état conformationnel⁴³⁵. Une classification des états conformationnels qui peuvent être adoptés par une protéine a été établie (Figure 27A). Ces états correspondent à : « globule, molten_globule_like, pre_molten_globule_like, et random coil »⁴³⁶⁻⁴³⁸. La fonction d'une protéine peut provenir de n'importe lequel de ces états conformationnels ou alors d'autres formes transitoires entre ces états précédemment décrits⁴³⁹.

| | Paradigme Structure-fonction (Fisher, 1894) | Wright et Dyson, 1999 | Protein Trinity model (Dunker et Obradovic, 2001) | Protein Quartet model (Uversky, 2002) |
|--|---|-----------------------|---|---------------------------------------|
| | Séquence en acides aminés | | | |
| Random coil Absence de structure secondaire ou tertiaire | | | | |
| Pre-molten globule Structure secondaire partiellement native, absence de structure tertiaire | | | | |
| Molten globule Structure secondaire native, absence de structure tertiaire | | | | |
| Structure 3D ordonnée Structure secondaire et tertiaire totalement formée | | | | |
| | Fonction biologique | | | |

Figure 27A: Évolution des modèles illustrant la relation structure-fonction d'une protéine. Le concept de la clé et de la serrure d'Emil Fischer, qui souligne l'importance d'une structure 3D bien définie pour la fonction protéique, a dominé la biologie moléculaire depuis plus d'un siècle. Cependant, la dernière décennie a permis de constater qu'un grand nombre de protéines ne nécessitent pas une structure 3D rigide pour l'acquisition de leur fonction. Celles-ci existent dans un état dit « désordonné », avec des quantités variables de structures secondaires et tertiaires préformées, correspondant à l'un ou l'autre de quatre états thermodynamiques (random coil, pre-molten globule-like, molten globule-like, et ordered) dans des conditions physiologiques. Cette diversité dans les états conformationnels natifs des protéines et les interconversions entre eux peuvent jouer un rôle instrumental dans des processus cellulaires hautement régulés (adaptée d'après ⁴⁴⁰).

Les protéines ou domaines protéiques qui n'adoptent pas une structure tridimensionnelle unique sont désignés comme des protéines ou domaines intrinsèquement désordonnés (IDP ou IDD respectivement) ^{435-437, 441}. Cependant, il est important de souligner que les régions désordonnées ont fréquemment tendance à se structurer suite à leur fixation avec un partenaire physiologique ⁴⁴².

D'autre part, les domaines désordonnés confèrent plusieurs avantages fonctionnels comparés à des structures rigides ou figées ^{435-436, 441}. Tout d'abord les régions flexibles peuvent adopter des conformations différentes pour permettre l'établissement d'interactions stables avec plusieurs substrats hétérogènes d'un point de vue structural ⁴³⁹. Ces régions permettent aussi de réaliser des interactions réversibles avec des substrats de haute spécificité mais de faible affinité. Également, les IDDs fournissent des sites cibles pour des modifications post-transcriptionnelles ⁴⁴³ et des sites spécifiques pour des dégradations par des protéases ⁴⁴⁴. Par ailleurs,

cette flexibilité conformationnelle est requise pour l'assemblage macromoléculaire incluant la formation d'oligomères ⁴⁴². Le désordre natif a également un rôle protecteur sur la fonction de la protéine, en permettant de préserver l'activité de celle-ci sous des conditions extrêmes de pH ou de température, ce qui est le cas des « cold shock proteins » (Csp).

L'analyse de prédictions de structures protéiques a permis de montrer qu'une grande proportion de protéines eucaryotes (30-60%) contiennent de longues régions désordonnées (≥ 40 résidus consécutifs) ⁴⁴⁵⁻⁴⁴⁶. De larges segments non structurés sont fréquents chez les protéines régulatrices tels que celles impliquées dans la signalisation cellulaire et le cancer, les régulateurs de la transcription et les protéines se liant à l'ARN. Cependant, ces segments restent rares chez les protéines avec une fonction catalytique ⁴⁴⁶⁻⁴⁴⁷.

Une meilleure compréhension du rôle du désordre intrinsèque chez les protéines chaperonnes des acides nucléiques permettrait de mieux comprendre les mécanismes d'action de ces protéines et d'arriver à définir leurs fonctions cellulaires précises.

1.5.2 Les protéines chaperons des acides nucléiques

Afin de pouvoir assurer leur rôle biologique, les acides nucléiques ont besoin d'adopter une structure secondaire ou tertiaire définie. Ce sont des protéines chaperons qui les aident à atteindre un repliement moléculaire correct, ce qui leur permet d'exercer leur fonction. En effet, le repliement des molécules d'acides nucléiques peut parfois se révéler être difficile à réaliser, car ces molécules peuvent adopter des conformations tridimensionnelles parfois très stables mais non fonctionnelles. Ces états sont nommés des pièges cinétiques ou « kinetic traps » (Figure 28A).

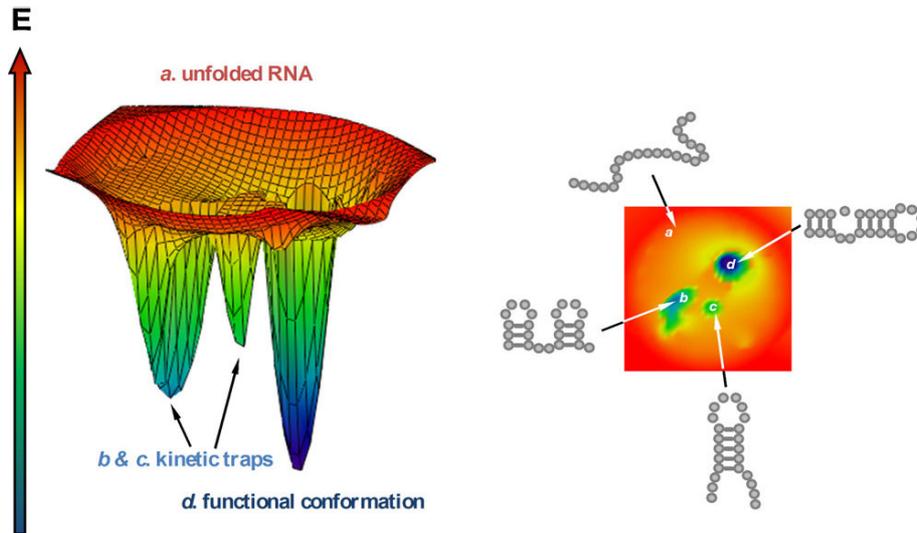


Figure 28A: Schéma illustrant le paysage énergétique d'un ARN. Le paysage énergétique de l'ARN est accidenté par plusieurs minima locaux qui correspondent à des repliements défavorables de l'ARN « suboptimal kinetic trap foldings » (**b et c**). Les protéines chaperons des ARNs lissent ce paysage en déstabilisant les conformations métastables et en réduisant les barrières énergétiques entre les différentes conformations, en dirigeant le repliement de l'ARN vers sa conformation native ⁴⁰⁷.

Cependant, le cas des ARNs est encore plus particulier, vu qu'*in vitro*, ces derniers perdent leurs structures facilement ⁴⁴⁸⁻⁴⁴⁹, alors qu'*in cellulo*, leur repliement se fait de manière efficace, ceci souligne d'avantage le rôle des interactions protéines-ARNs dans ce processus de repliement ^{410, 450-451}.

Les protéines chaperons des ARNs sont ubiquitaires dans tous les organismes vivants et les virus. Elles guident les molécules d'ARN dans chaque étape du métabolisme cellulaire, jouant un rôle pivot dans la transcription, la maturation de l'ARN, son transport et sa transcription. Ces facteurs protéiques sont également abondants et contiennent de larges domaines flexibles ^{409, 452}. Ces protéines chaperons se lient aux ARNs avec une spécificité de séquence assez vaste. Elles aident les molécules d'ARN à atteindre leurs conformations natives, tout en évitant les pièges cinétiques, ou en tout cas en les corrigeant. En d'autres termes, les protéines chaperons aident les ARNs à atteindre leur conformation thermodynamiquement la plus stable en réduisant les barrières énergétiques entre les différents états conformationnelles (Figure 28A) ^{409-410, 450}. Plusieurs exemples de protéines chaperons ont été reportés : les protéines ribosomales, les hnRNP (heterogeneous nuclear ribonuclear proteins), les protéines à « cold shock » domaine, les « histone-like » protéines bactériennes, les « SM-like proteins », la

FMRP, p53, PrPc, ainsi que les protéines de nucléocapside virales. Ces protéines sont très hétérogènes et leur nombre s'accroît constamment du fait du développement de nouvelles techniques d'investigations ^{409-410, 450, 453-454}. Beaucoup de ces protéines ont été caractérisées pour leurs propriétés de fixation aux ARNs et pour leur activité de protéine chaperon des ARNs. Cependant, leur mode d'action reste encore peu clair. De même, il est important de souligner le manque de similarité entre les différentes protéines chaperons des ARNs, ce qui implique des mécanismes d'actions divers. Cela dit, des principes généraux peuvent être dégagés à partir de leurs fonctions communes. D'ailleurs, une classification a été établie pour définir si une protéine possède toutes les caractéristiques pour être classée comme protéine chaperon (Tableau 3A). En effet, Rajkowitsch et al, 2007 ainsi que Doetsch et al, 2011 définissent une chaperon à partir du principe que celle-ci possède une activité de déstabilisation des structures secondaires des ARNs, une activité d'hybridation des séquences complémentaires d'ARN et une activité de déplacement de brins. Le tout sans avoir recours à une source d'énergie externe.

Tableau 3A. Classification des protéines assistant les ARNs dans leur structuration (adapté d'après ³²⁵).

| |
|---|
| RNA chaperone (Protéine avec une activité chaperon) |
| Une protéine qui se fixe de manière transitoire et non spécifique aux ARNs pour résoudre les pièges cinétiques et les conformères non structurés. L'activité de protéine chaperon englobe la déstabilisation des interactions ARN-ARN et le relâchement des structures des ARNs. Cependant, l'interaction avec la protéine est nécessaire pour la déstabilisation des ARNs et non pour maintenir leurs structures secondaires. La protéine n'a pas besoin d'une source d'énergie pour son activité (fixation d'ATP ou hydrolyse). |
| RNA annealer (Protéine avec une activité d'hybridation) |
| Une protéine qui accélère l'hybridation des séquences complémentaires d'ARN. Elle peut agir en se liant à l'une ou aux deux molécules d'ARN impliquées augmentant ainsi la concentration locale et la probabilité des interactions ARN-ARN (molecular crowding). |
| RNA helicases (Protéine avec une activité hélicase) |
| Un membre de la famille des « DExD/H-box protein » qui utilise l'énergie dérivée de l'hydrolyse de l'ATP pour résoudre les structures des ARNs et déplacer les ARNs liés aux protéines. Seulement quelques unes sont processives, alors que la plupart des hélicases qui sont bien caractérisées sont dites non processives ou "unwindases". |
| Specific RNA binding proteins (Protéine avec une spécificité de liaison) |
| Une protéine ligand (ou cofacteur protéique) qui se lie de manière spécifique à l'ARN et contribue à sa restructuration native ainsi qu'à sa stabilité thermodynamique. Une interaction continue avec la protéine permet de maintenir la structure fonctionnelle de |

I.5.3 Le désordre intrinsèque chez les protéines chaperons des ARNs

Une des limites à la classification des protéines chaperons des ARNs, est que leur activité de protéine chaperon ne peut être prédite du fait de la présence de domaines ou motifs structuraux spécifiques. Cependant, elles présentent presque toutes une fréquence plus élevée de régions désordonnées, avec plus de la moitié des acides aminés situés dans des segments non structurés⁴⁵². Basé sur cette observation, un modèle a été proposé pour la fonction des chaperons des ARNs, nommé « the entropy exchange model ». Suivant ce dernier, l'interaction d'une chaperon avec un ARN non replié est accompagnée par un repliement local de la protéine, avec un « melting » concomitant (unfolding) d'un segment d'ARN. La molécule d'ARN est alors libre pour explorer son espace conformationnel tout en libérant la protéine chaperonne lors des dernières étapes (Figure 29A). Des cycles de transitions successives désordre-à-ordre et ordre-à-désordre dirigé par un transfert entropique réciproque entre une molécule d'ARN et une molécule chaperon se produisent. Ceci va mener à la formation d'une conformation d'ARN la plus stable et requise pour sa fonction^{435, 440, 452}.

I.5.4 Mécanisme d'action des protéines chaperons des ARNs

Pour ce qui est de leurs activités précises, les protéines chaperons des ARNs catalysent des réactions relativement diverses. Telle est l'hybridation de deux structures d'acides nucléiques complémentaires pour aboutir à un homo- ou hétéro-duplex ou bien la dissociation d'un duplex ou son sauvetage lorsqu'il est mal replié^{409, 455}. Une des caractéristiques les plus importantes de ces protéines, est qu'elles n'ont guère besoin d'adénosine triphosphate (ATP) ou d'une autre source d'énergie externe pour exercer leur activité. Ces protéines interagissent de manière transitoire et non spécifique avec leurs substrats ARNs. Par ailleurs, une fois que l'ARN a atteint sa conformation thermodynamiquement la plus stable, les protéines peuvent se dissocier sans que la structure de l'ARN en question en soit affectée^{325, 409}. Il a été démontré que les protéines possédant une activité chaperon se lient aux ARNs avec une faible spécificité. Ceci suggère que l'interaction de ces protéines

avec les ARNs se fait de manière transitoire et principalement de nature électrostatique ⁴⁵⁶. Cela a été confirmé avec un mutant de la protéine StpA d'*E. coli*, pour lequel la capacité de liaison à l'ARN est réduite, mais qui affiche néanmoins une activité de protéine chaperon plus élevée, comparée à la protéine sauvage. On peut supposer qu'une forte liaison mène à une stabilisation de l'ARN et que celle-ci est désavantageuse pour l'activité de protéine chaperon. Il a également été démontré pour les protéines StpA ⁴⁵⁷ et S12 ⁴⁵⁸, avec des expériences *in vitro*, que le mode d'action de telles protéines est transitoire. En effet, une fois que l'ARN a adopté son repliement fonctionnel, les protéines chaperons ne sont plus requises pour maintenir la structure native de l'ARN (Figure 28A). Plusieurs chaperons des ARNs ont montré une aptitude à l'oligomérisation, comme dans le cas des « histone-like protein H-NS » et de « StpA » ^{335, 459}. En conclusion, un repliement efficace des molécules d'ARN au sein de la cellule est d'une importance majeure pour une large variété de fonctions biologiques. Les protéines chaperons qui assistent les ARNs dans l'acquisition de leurs structures fonctionnelles actives, sont abondantes chez tous les organismes vivants.

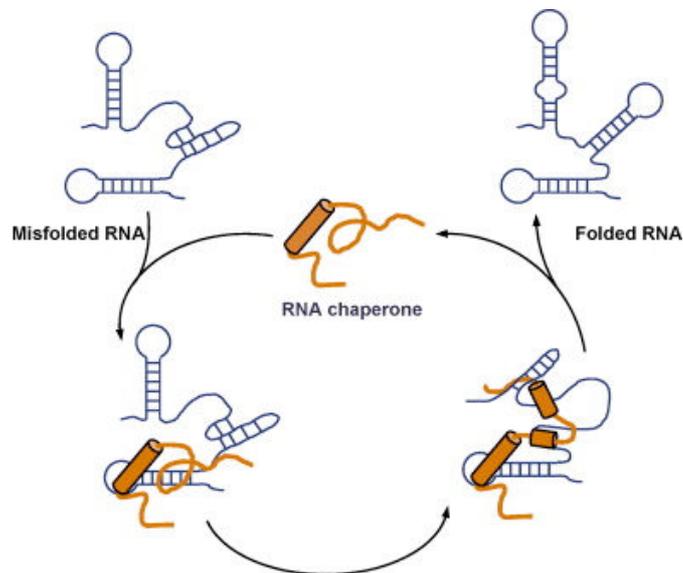


Figure 29A: Modèle illustrant le mécanisme d'action d'une protéine chaperon des ARNs selon « The entropy transfer model » ⁴⁵². Les protéines chaperons des ARNs (orange) possèdent de longues régions intrinsèquement désordonnées. **A)** La première étape consiste en la fixation de ce type de protéine aux ARNs non repliés généralement de manière non spécifique. **B)** Les régions désordonnées de la protéine ont tendance à se structurer quand les domaines du substrat sont non repliés (transfert d'entropie de la protéine au substrat). **C)** L'ARN est correctement replié quand la protéine est également désordonnée (transfert d'entropie du substrat ARN à la protéine). **D)** Finalement, après plusieurs cycles d'ordre-désordre, le substrat ARN, correctement replié, est libéré et la protéine est également libre pour interagir avec d'autres substrats.

Quelques exemples de protéines chaperons des ARNs de différents organismes sont donnés ci-dessous :

I.5.4.1 Protéines bactériennes chaperons des ARNs

I.5.4.1.1 La protéine StpA

La protéine StpA d'*E. coli* est connue comme étant une protéine chaperon des ARNs avec une activité d'hybridation de l'ARN et de déplacement de brins ⁴⁵⁵⁻⁴⁵⁷.

Récemment, il a été démontré que StpA se fixe de manière transitoire à l'ARN et préférentiellement à des régions non structurées. Cette capacité de fixation diminue en présence de fortes forces ioniques, ce qui démontre que ces interactions sont de nature électrostatique ⁴⁵⁶. Par ailleurs, des analyses de prédictions de structures de StpA montrent que cette protéine contient plus de 70% de désordre. Ses régions non structurées joueraient un rôle important dans son activité de protéine chaperon, ce qui lui confère justement une flexibilité de liaison non spécifique ⁴⁵². D'autre part, une autre caractéristique de StpA, est que pour accélérer l'hybridation de deux courts ARNs, celle-ci se lie simultanément aux deux molécules, ce qui démontre que l'hybridation est promue par un encombrement « crowding » des molécules.

D'autre part, une étude réalisée avec une protéine StpA mutante, où une glycine située dans le domaine de liaison aux acides nucléiques a été changée par une valine, a démontré que d'une part la protéine affichait une capacité de fixation plus faible aux ARNs mais que d'autre part son activité de protéine chaperon était plus élevée. En conclusion, l'activité de protéine chaperon de StpA provient d'interactions faibles et transitoires au lieu d'une fixation spécifique et de haute affinité. Cela démontre une relation inverse entre l'efficacité de fixation aux ARNs et l'activité de protéine chaperon. En suggestion, l'activité de protéine chaperon des ARNs provient d'interactions faibles et transitoires plutôt que d'une forte fixation aux ARNs.

1.5.4.1.2 La protéine CspA

Les Csp (Cold-shock proteins) sont de petites protéines de 70 acides aminés, qui sont ubiquitaires au sein des bactéries et des Archées. La bactérie *E. coli* exprime neuf membres de la famille des Csp, dont CspA et CspE représentent ses Csp majeures. Ces deux dernières se lient de manière non spécifique à l'ARN et possèdent une activité de déstabilisation des acides nucléiques⁴⁶⁰⁻⁴⁶³. Ces protéines sont transitoirement exprimées dans des conditions de stress, liées à de faibles températures, ceci afin d'assurer différentes fonctions cellulaires. CspA est une protéine chaperon des ARNs qui se lie aux acides nucléiques simples brins. Cette protéine déstabilise les structures en tiges boucles des ARNm qui se constituent à basse température, permettant ainsi une transcription des gènes même dans des conditions stressantes liées à une chute de température⁴⁶³.

Le site de fixation comprend six résidus aromatiques qui sont accompagnés de plusieurs lysines chargées positivement. D'après les structures des complexes Csp-acides nucléiques⁴⁶⁴⁻⁴⁶⁶, on observe que cinq nucléotides peuvent être insérés dans le site de fixation et que les trois résidus aromatiques F18, F31 et H33 de Csp1 sont essentiels ; la mutation de l'un d'entre eux entraînant une perte totale de l'activité de protéine chaperon alors même que la fixation à l'ARN ne semble pas être affectée par ces mutations⁴⁶⁷⁻⁴⁶⁹. Des travaux récents de RMN sur ce système ont permis d'obtenir de nouvelles informations, les expériences ont permis de suivre l'association en duplexe de deux structures ARN tiges boucles repliées et complémentaires, tout en mesurant les vitesses de dépliement des structures initiales repliées et la formation du duplex au niveau de chaque paire de bases, suggérant que la protéine exerce son activité en déstabilisant les paires proches du complexe boucle-boucle puis en progressant de proche en proche vers les extrémités. Les interactions entre les résidus aromatiques de CspA et les bases acides nucléiques ont été bien mises en évidence et l'effet de ces interactions dans le processus de déstabilisation des paires de bases est également souligné.

1.5.4.2 Protéines virales chaperons des ARNs

La première des protéines virales chaperons des ARNs à avoir été reportée est la protéine de nucléocapside des rétrovirus aviaires et murins⁴⁷⁰. D'autres

protéines virales chaperons ont été reportées par la suite et sont listées au tableau ci-dessous :

Tableau 4. Protéines chaperons impliquées dans la réplication virale ⁴⁷¹.

| Protéines ^a | Virus ^b | Fonctions | Références |
|-----------------------------|---|---|------------|
| Protéines virales | | | |
| NC | Rétrovirus | Transcription inverse, transfert de brin | 477 |
| SdAg | HDV | Réplication | 472 |
| Core | Flaviridae | Réplication | 440 |
| N | Coronavirus | Synthèse d'ARN | 473 |
| N | Hantavirus | Réplication | 474 |
| 3AB | Poliovirus | Réplication, recombinaisons | 475 |
| Vif | VIH-1 | Régulation temporaire de la transcription inverse | 476 |
| Tat | VIH-1 | Transcription de l'ADN viral | 477 |
| Facteurs cellulaires | | | |
| PTB | Picornavirus, HCV, coronavirus, calicivirus | Traduction, synthèse d'ARN | 478-481 |
| hnRNP A1 | Coronavirus | Synthèse d'ARN | 479 |
| La | HCV, poliovirus | Traduction, réplication | 482 |
| Unr | HRV-2 | Traduction | 483 |
| PARBP33 | ASBVd | Réplication | 484 |

a) NC, nucleocapsid protein; SdAg, small delta antigen; N, nucleocapsid protein; Vif, viral infectivity factor; Tat, transcription activation factor; PTB, polypyrimidine tract-binding protein; hnRNP A1, heterogeneous nuclear ribonucleoprotein A1; Unr, upstream of N-ras.
b) HDV, hepatitis delta virus; VIH-1, Virus de l'immunodéficience humaine-1; HCV, hepatitis C virus; HRV-2, human rhinovirus-2; ASBVd, avocado sunblotch viroid.

Néanmoins, le rôle précis de ces chaperons des ARNs dans le cycle viral reste peu clair du fait de la difficulté d'analyse de ces activités de protéines chaperons dans un contexte *in vivo*. De même que pour les protéines chaperons des ARNs en général, celles des virus ne présentent pas d'homologies de motifs structuraux. Néanmoins, elles suivent la règle de la présence de longues régions désordonnées. Ce constat fait donc du degré de désordre d'une protéine (déterminable par des programmes de prédiction) un critère de sélection pour rechercher des activités de protéines chaperons chez plusieurs candidats viraux.

I.5.4.3 Protéines chaperons des ARNs du VIH-1

Le virus du VIH-1 peut être un modèle de choix dans l'étude des protéines

chaperons des ARNs. En effet, le VIH-1 possède plusieurs protéines chaperons des ARNs sans compter les chaperons cellulaires qui sont impliquées dans son processus de réplication. La protéine de nucléocapside du VIH-1 en tant que chaperon a été décrite pour la première fois en 1994^{396, 485} et reste extensivement étudiée vu que celle-ci joue un rôle crucial dans les processus de transfert de brins¹⁷⁷. Récemment, il a été démontré que les protéines virales Tat et Vif possédaient également une activité de protéine chaperon des acides nucléiques⁴⁷⁶⁻⁴⁷⁷.

1.5.4.3.1 La protéine de nucléocapside

L'activité de protéine chaperon de la protéine de nucléocapside du VIH-1 est décrite au paragraphe (IV).

1.5.4.3.2 La protéine Tat

Tat appartient à la famille des IDPs, elle possède une organisation hautement déstructurée (pas d'hélice α ni de feuillet β)⁴⁸⁶, c'est justement ce désordre intrinsèque et sa flexibilité conformationnelle qui confèrent à Tat l'aptitude à interagir avec différents facteurs viraux et cellulaires^{436, 487}. Le fait que Tat soit capable de se lier aux acides nucléiques et qu'elle possède une nature désordonnée, suggère fortement que Tat est également une protéine chaperon des acides nucléiques⁴⁷⁷. Son rôle chaperon des ARNs a été récemment démontré^{436, 477, 487} avec des propriétés semblables à celles de la NCp7. La région 44-61 serait le support principal de cette activité de protéine chaperon et Tat semble se structurer lorsqu'elle interagit avec ses partenaires⁴⁸⁸. Il a été démontré que Tat active efficacement l'hybridation de l'ADN, le clivage de l'ARN médié par le ribozyme et le trans-épissage de l'ARN *in vitro*⁴⁷⁷. D'autre part, Tat induit le remplacement d'un brin ADN imparfait par un autre brin parfaitement complémentaire⁴⁷⁷. Cependant, cette activité n'a pas été observée avec des séquences d'ARN complémentaires⁴⁸⁹. D'après cette dernière étude, Tat ne remplit donc pas tous les critères pour pouvoir être désignée comme une chaperon des ARNs, puisqu'elle ne possède pas d'activité de remplacement de brin, on parle donc plutôt de « RNA annealer »⁴⁸⁹. Cette incapacité à effectuer le remplacement de brin est due au fait que Tat est incapable de déstabiliser les

structures secondaires des ARNs. Néanmoins Tat promeut correctement les réactions d'hybridation non spécifiques des acides nucléiques, quand la déstabilisation des structures n'est que peu ou pas requise. Les acides aminés 44-61 sont les principaux responsables de l'activité « annealer » de Tat et constitue la séquence la plus courte connue possédant cette activité.

Les propriétés d'hybridation de Tat font que cette protéine peut promouvoir l'hybridation de l'ARNt^{Lys3} à l'ARNv et intervenir dans les processus de transferts de brin⁴⁹⁰. Tout cela fait de Tat un facteur stimulant la transcription inverse comme la NCp7⁴⁹¹⁻⁴⁹³. En effet, la comparaison des activités de ces deux protéines a démontré une grande similarité avec toutefois quelques différences. Dans le cas du premier transfert de brin, l'efficacité d'hybridation des séquences TAR/cTAR est comparable entre la NCp7 et Tat. Cela dit, la réaction suit un mécanisme « zipper » en présence de Tat⁴⁹⁰ vu que cette dernière ne peut pas réarranger les structures secondaires de TAR ou cTAR. Il en ressort que, Tat et NCp7 activent de manière coopérative l'hybridation entre les séquences TAR et cTAR, soulignant un rôle possible et accessoire de Tat dans la stimulation du processus de transcription inverse⁴⁰⁷.

Dans le cas du second transfert de brin, *in vitro* les séquences PBS (+) et (-) peuvent s'hybrider spontanément via un mécanisme « zipper », néanmoins en présence de NCp7, le mécanisme d'hybridation devient majoritairement un mécanisme d'hybridation boucle-boucle (Figure 30A) et l'efficacité de la réaction est augmentée de 60 fois⁴¹⁹. Ce basculement de voies d'hybridation est lié avec l'aptitude de la NCp7 à réarranger les boucles de PBS (+) et (-) tout en induisant des restrictions sur la dynamique des bases situées au niveau des boucles^{365, 407}. En comparaison, Tat accélère l'hybridation de PBS (+) et (-) mais ne parvient pas à changer les voies d'hybridation vers un mécanisme boucle-boucle. De ce fait, Tat stimule le mécanisme d'hybridation des séquences PBS déjà présent c'est-à-dire le « zipper »⁴⁰⁷.

En conclusion, les deux protéines montrent des capacités proches d'hybridation et d'agrégation des acides nucléiques. La différence la plus importante étant l'inaptitude de Tat, contrairement à la NCp7, à être capable de déstabiliser transitoirement les structures secondaires des acides nucléiques et à modifier la dynamique locale des nucléobases.

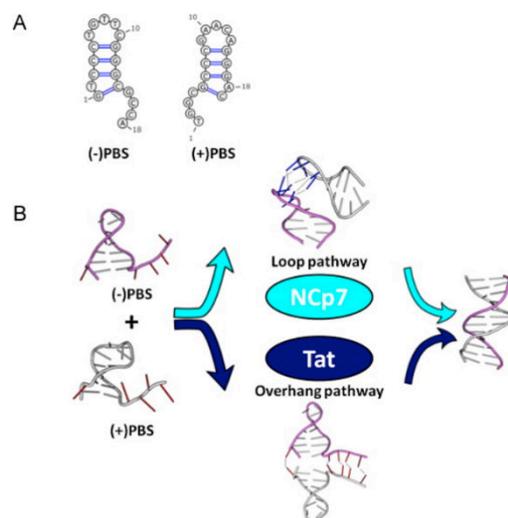


Figure 30A: Mécanismes d'hybridation des séquences PBS(+)/PBS(-). **A)** Séquences PBS(+)/PBS(-). **B)** En présence de NCp7, la réaction d'hybridation des PBS(+)/(-) a lieu selon un mode d'hybridation boucle boucle (cyan). En absence de NCp7, cette voie d'hybridation est négligeable et les séquences PBS(+)/(-) s'hybrident spontanément par leurs extrémités simples brins (bleu). Ce dernier mécanisme est également dominant en présence de la protéine Tat⁴⁰⁷.

1.5.4.3.3 La protéine Vif

Vif est une protéine basique de 192 acides aminés (riche en aa basiques 'Arg et Lys' et hydrophobes 'Trp'), d'une masse moléculaire de 23 kDa. Son activité de protéine chaperon est impliquée dans les étapes précoces de la transcription inverse : Vif augmente l'efficacité de l'hybridation de l'ARN^{t^{Lys}3} à la séquence PBS⁴⁹⁴, réduit les pauses de la RT et augmente la synthèse de l'ADN strong stop. Rappelons que les pauses de la RT se produisent du fait des structures secondaires stables présentes dans la matrice ARN⁴⁹⁵⁻⁴⁹⁶. Vif se lie aux structures secondaires dans la région R et U5' de l'ARNg⁴⁹⁷ avec une forte affinité de liaison⁴⁹⁸. Cependant Vif a un effet assez modeste sur les transferts de brins, qui sont médiés habituellement par la NCp7. Récemment, il a été proposé que Vif joue le rôle d'un régulateur négatif de l'initiation de la transcription inverse de manière temporaire, ceci afin d'éviter que le processus ait lieu de façon immature⁴⁷⁶.

Il en ressort que Vif possède des différences et des similitudes avec la protéine de nucléocapside NCp7. Alors que l'affinité de liaison de la NCp7 pour l'ARN diminue fortement avec des concentrations croissantes de sels^{177, 375}, il apparaît, qu'en revanche, le complexe Vif:ARN résiste à de fortes concentrations en

sels ⁴⁹⁷⁻⁴⁹⁹. De même que pour la NCp7, Vif possède une forte affinité pour les régions simples brins riches en guanines et cytosines ⁴⁹⁸⁻⁴⁹⁹.

I.5.4.4 Protéine de nucléocapside du VIS

L'activité de protéine chaperon de la nucléocapside du VIH-1 a été bien caractérisé *in vitro*, cependant, il est intéressant de comparer les différents mécanismes d'action entre la NCp7 du VIH-1 et celle du VIS (Figure 31A). Récemment, l'activité de protéine chaperon de la NCp7 du VIS a été caractérisée ⁵⁰⁰, celle-ci n'a montré que de légères différences, comparée à la NCp7 du VIH-1, dans sa capacité à faciliter le premier ou les autres transferts de brins qui ont lieu au cours de la transcription inverse. Cependant la NCp7 du VIH-1 s'est montrée beaucoup plus efficace en tant que protéine chaperon avec un taux plus élevé de transferts de brins et d'efficacité d'hybridation, cette différence est due à une énergie de compaction de l'ADN plus élevée et une activité agrégative des acides nucléiques plus importante de la NCp7 du VIH-1 comparée à celle du VIS ⁵⁰⁰. En effet, il s'est avéré que la NCp7 du VIH-1 possède un caractère plus basique que la NC du VIS, par exemple, à pH= 7, la charge estimée pour la NCp7 du VIH-1 est de 11.2 alors que celle du VIS est de 10.2.

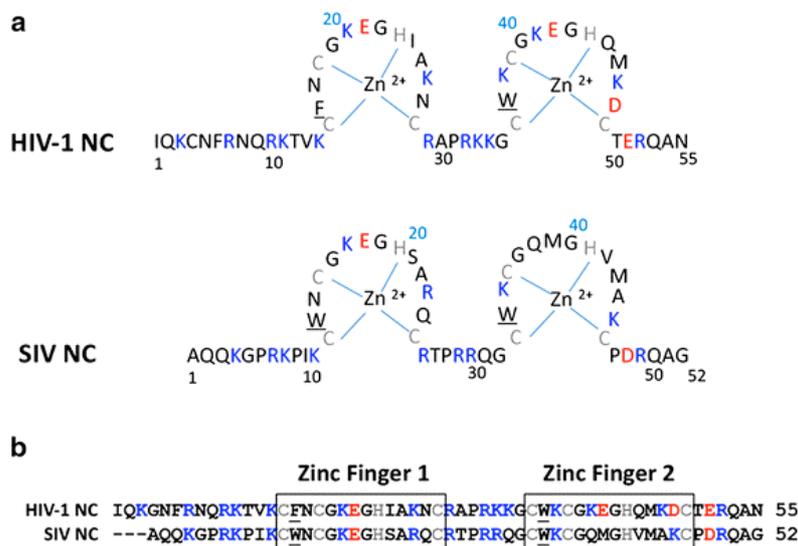


Figure 31A: Séquences et caractéristiques structurales des protéines de nucléocapside du VIH-1 et du VIS. A) Diagramme schématique des protéines de nucléocapsides de la souche NL4-3 du VIH-1 et du VISmne. Résidus basiques (bleu), résidues acides (rouge), les résidus CCHC qui coordonnent les ions Zn^{2+} des deux ZFs (gris), les résidus aromatiques de chaque doigt de zinc sont soulignés. **B)** Alignement de séquences des protéines de NC (VIH-1 et VIS). Les couleurs et les soulignements sont les mêmes qu'en (a). Les cadres indiquent les séquences comprenant chaque doigt de zinc ⁵⁰⁰.

I.5.5 Les prions humains impliqués dans les maladies

Un prion ou particule protéique infectieuse (PROteinaceous INfectious particle) est un agent pathogène constitué d'une protéine dont la conformation ou le repliement est anormal. Les prions sont responsables de maladies neurodégénératives fatales telles que la maladie de Creutzfeldt-Jakob, kuru, l'insomnie fatale familiale (IFF) et le syndrome de Gerstmann-Sträussler-Scheinker (SGSS). Les prions possèdent un domaine N-terminal intrinsèquement désordonné, caractéristique des protéines chaperons des ARNs ⁵⁰¹. Bien que les PrPs soient conservées chez les vertébrés, leurs fonctions restent peu claires.

Une étude avec une huPrP (human PrP), composée de 245 acides aminés, localisée au niveau du cytoplasme et impliquée dans une variante de SGSS ⁵⁰², a démontré que les PrPs possédaient une activité d'hybridation des acides nucléiques (ARN et ADN) similaire à celle de la protéine de nucléocapside du VIH-1 ⁵⁰³. En effet, Les PrPs imitent largement le rôle de la NCp7 dans le cycle de réplication, tant au cours de la dimérisation de l'ARNg, sa stabilisation, l'initiation de la transcription inverse par l'hybridation de l'ARNt^{Lys} à la séquence de l'ARNg en 5' et l'hybridation des séquences complémentaires des acides nucléiques au cours de la transcription inverse. Cette analyse, souligne le fait que les PrPs seraient impliquée dans plusieurs fonctions cellulaires via leurs activités de protéines chaperons des acides nucléiques. Les fonctions connues jusqu'ici englobe leur implication dans l'activité d'une superoxyde dismutase, dans le métabolisme de Copper ⁵⁰⁴, ainsi qu'une participation dans le signal de transduction au cours de la différenciation neuronale ⁵⁰⁵. Egalement, les PrPs sont impliquées dans l'établissement d'interactions avec des glycanes sulfatés ⁵⁰⁶, des aptamères ARN ⁵⁰⁷, de larges acides nucléiques ⁵⁰⁸⁻⁵⁰⁹ menant à la formation de complexes nucléoprotéiques similaires aux complexes NCp7-ARN formés *in vitro* ⁵¹⁰. Par ailleurs, une étude récente a rapporté que le virus du MLV accélère le processus infectieux de la tremblante ⁵¹¹, ce qui suggère une interaction entre les rétrovirus et les prions *in vitro*, qui pourrait expliquer les résultats précédents sur le huPrP et le VIH-1.

II Objectifs de la thèse

Etant donné le rôle crucial que joue la protéine de nucléocapside du VIH-1 dans le cycle viral, l'investigation de ses propriétés structurales et fonctionnelles est d'une importance majeure d'un point de vue fondamental et appliqué. En effet, bien que plusieurs agents pharmacologiques anti-NC aient été développés, des tests cliniques n'ont pas pu être effectués du fait d'un manque de spécificité, ce qui engendre une toxicité importante, notamment pour la classe des éjecteurs de zinc. De ce fait, nous pensons qu'il est nécessaire de mieux comprendre les mécanismes de fixation de la NC aux acides nucléiques afin de pouvoir élaborer des inhibiteurs dotés d'une haute spécificité pour la protéine. La NC constitue une cible thérapeutique attrayante vu que celle-ci facilite les mécanismes de transferts de brin et de ce fait les recombinaisons génétiques du VIH-1. Également, celle-ci joue un rôle déterminant dans l'assemblage et le bourgeonnement des particules virales, ce qui implique que son inhibition bloquerait le processus de production virale à plusieurs étapes du cycle de réplication. Sa conservation très élevée entre les différentes souches est un autre facteur justifiant le ciblage de la NC.

L'objectif principal de ma thèse est de caractériser les mécanismes structuraux et moléculaires qui régissent le mode d'action de la protéine de nucléocapside. En d'autres termes, caractériser par des méthodes de biophysiques (Anisotropie de fluorescence, FRET, RMN) et de biologie moléculaire :

- a) L'affinité de liaison de la NCp7 à différentes séquences d'acides nucléiques et de différentes tailles dans le but de déceler les motifs préférentiels de la protéine.
- b) Les modalités de déstabilisation des séquences TAR/cTAR qui sont impliquées dans le premier transfert de brin au cours de la transcription inverse et ce dans le but de saisir les voies d'hybridation empruntées en fonction du rapport protéine:nucléotide, de la température, du pH ainsi que des structures secondaires des séquences.
- c) La polarité de fixation de la NCp7 aux acides nucléiques en fonction de leurs natures (ADN ou ARN), s'il en existe une.
- d) La structure et la dynamique interne comparée des trois formes majeures de NC: Immature NCp15, intermédiaire NCp9 et mature NCp7.

III Matériels et méthodes

Dans cette partie sont décrits les matériels et méthodes qui ne sont pas mentionnés dans les articles publiés.

III.1 Matériels

III.1.1 Oligonucléotides

Tous les oligonucléotides ont été fournis par Eurogentec (Belgique). Pour les expériences d'anisotropie de fluorescence, la fluorescéine (un marqueur fluorescent) a été greffée à l'extrémité 5' de ces acides nucléiques (Tableau 1B).

Tableau 1B. Oligonucléotides utilisés pour l'étude par anisotropie de fluorescence.

| Oligonucléotide | Séquence |
|-----------------|---|
| SL3 ADN | 5' F*GGACTAGCGGAGGCTAGTCC 3' |
| SL3 ARN | 5' F*GGACUAGCGGAGGCUAGUCC 3' |
| SL3 ADN20G | 5' F*GGACTAGCGGGAGGCTAGTCC 3' |
| cTAR26 | 5' F*CCAGAGAGCTCCCGGGCTCGACCTGG 3' |
| cTAR33 | 5' F*TAGCCAGAGAGCTCCCGGGCTCGACCTGGTCTA 3' |
| GCG | 5' F*TAGCGAT 3' |
| GCG11 | 5' F*TCTAGCGATCT 3' |
| GCG13 | 5' F*TACTAGCGATCAT 3' |
| GGC | 5' F*TAGGCAT 3' |
| TCG | 5' F*TATCGAT 3' |
| AGC | 5' F*TAAGCAT 3' |
| TGC | 5' F*TATGCAT 3' |
| TAC | 5' F*TATACAT 3' |
| ACG | 5' F*TAACGAT 3' |
| CCG | 5' F*TACCGAT 3' |
| GAG | 5' F*TAGAGAT 3' |
| caGAG | 5' F*CAGAGAC 3' |
| caGCT | 5' F*CAGCTAC 3' |
| caGGC | 5' F*CAGGCAC 3' |
| caUCG | 5' F*CAUCGAC 3' |
| caTCG | 5' F*CATCGAC 3' |
| caGCU | 5' F*CAGCUAC 3' |
| caGUC | 5' F*CAGUCAC 3' |
| caGTC | 5' F*CAGTCAC 3' |
| caTGC | 5' F*CATGCAC 3' |
| caUGC | 5' F*CAUGCAC 3' |
| atGAG | 5' F*ATGAGAT 3' |

Nous avons choisi d'étudier des structures en tige boucle (cTAR26, cTAR33, SL3 ADN20, SL3 ARN20, SL3 ADN20G) (Figure 1B) ainsi que de courtes séquences linéaires incapables de se replier et contenant différentes séquences avec des sites de fixation préférentiels ou non de la NCp7. Le choix de la taille et de la séquence a été réalisé en tenant compte des propriétés connues de reconnaissance des séquences d'acides nucléiques par la NCp7.

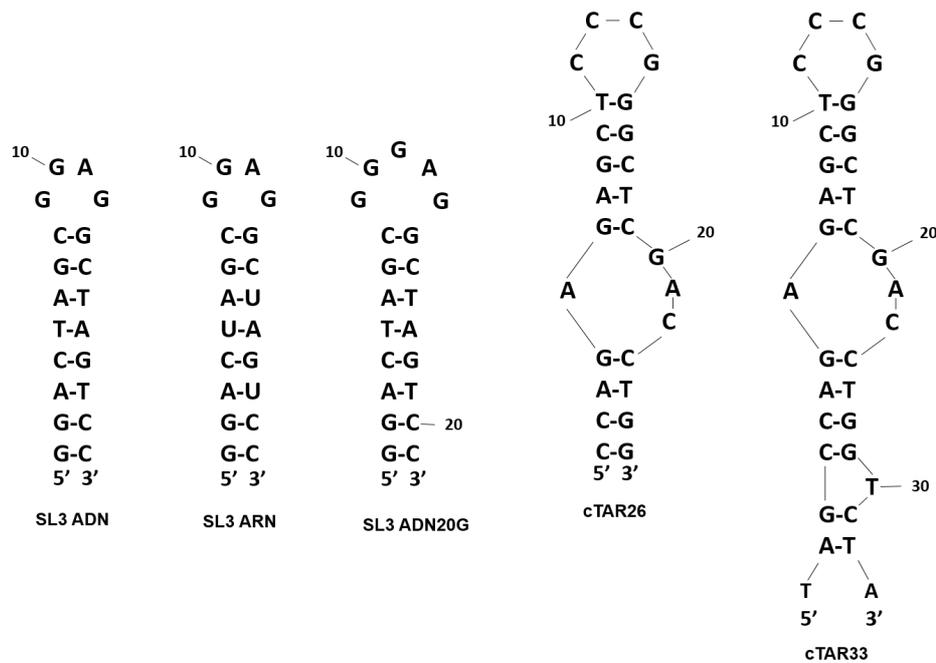


Figure 1B: Séquences d'oligonucléotides en tige boucle utilisées pour l'étude de l'affinité de liaison de la NCp7 aux acides nucléiques. Les structures secondaires ont été prédites par mfold. De gauche à droite, la tige boucle SL3 ADN (20 nt) et son homologue ARN (SL3 ARN), SL3 ADN avec une guanine rajoutée sur sa boucle apicale en position 11 (SL3 ADN20G), les séquences tronquées cTAR ADN (26 nt) et (33 nt).

Pour les expériences de FRET, nous avons étudié l'hybridation entre les séquences TAR et cTAR tronquées (29 nt) afin de simplifier le modèle expérimental et d'étudier correctement les effets de l'inversion des boucles internes dans les molécules. Une tetramethylrhodamine (TMR) a été greffée en 5' et une fluorescéine a été greffée en 3' sur les séquences ADN (Tableau 2B). Nous avons tout d'abord réalisé un couple des séquences sauvages nommées TAR0/cTAR0. Un deuxième couple mutant a été réalisé en inversant les positions des boucles internes simple brin (Figure 2B), car ces dernières constituent des sites où la protéine NCp7 se fixe préférentiellement, puis déstabilise les régions adjacentes ⁴¹². Si la NCp7 se fixe puis déstabilise les séquences double brins avec une polarité définie comme nous l'avons exposée dans l'introduction (Figure 19A), la modification de la position des boucles internes devrait influencer de façon importante sur le positionnement de la protéine et donc sur la déstabilisation associée. Pour chacune des deux familles de molécules (cTAR0/TAR0) et (cTAR1/TAR1), d'autres types de mutants ont été réalisés, le but des expériences avec ces mutants est de caractériser pour chacune des familles le chemin d'hybridation préféré : "kissing complex" ou "zipper mechanism". Les mutants au niveau des boucles apicales permettent de tester si le chemin "kissing complex" est affecté tandis que les mutants au niveau des extrémités des tiges rendent plus difficile le passage par le mécanisme "zipper". La comparaison des expériences effectuées, à l'intérieur de chaque famille, avec les molécules sauvages, les

mutants "loop" et les mutants "stem" devraient permettre de déduire dans chacun des cas quel est le mécanisme d'hybridation majoritaire.

Les mutants destinés à empêcher l'hybridation via un mécanisme "kissing complex" possèdent deux mutations au niveau des boucles apicales des ADNs ce qui affecte la complémentarité de ces dernières avec leur homologue ARN (cTAR0_29nt_loop et cTAR1_29nt_loop) (Figure 2B).

Pour les mutants "stem" destinés à affecter le chemin "zipper", les mutations (au nombre de deux) se trouvent au niveau des extrémités des ADNs ce qui affecte également la complémentarité de ces dernières avec leur homologue ARN (cTAR0_29nt_stem et cTAR1_29nt_stem) (Figure 2B).

Tableau 2B. Oligonucléotides utilisés pour l'étude par FRET.

| Oligonucléotide | Séquence |
|-----------------|--|
| TAR0_29nt | 5' ACA GAG UCU GAG CCU GGG AGC UCC UCU GU 3' |
| cTAR0_29nt | 5' TMR *ACA GAG GAG CTC CCA GGC TCA GAC TCT GT* F 3' |
| TAR1_29nt | 5' ACA GAG GAG CCU GGG AGC UCU CUC UCU GU 3' |
| cTAR1_29nt | 5' TMR *ACA GAG AGA GAG CTC CCA GGC TCC TCT GT* F 3' |
| cTAR0_29nt_stem | 5' TMR *CAA GAG GAG CTC CCA GGC TCA GAC TCT TG* F 3' |
| cTAR0_29nt_loop | 5' TMR *ACA GAG GAG CTC AAA GGC TCA GAC TCT GT* F 3' |
| cTAR1_29nt_stem | 5' TMR *CAA GAG AGA GAG CTC CCA GGC TCC TCT TG* F 3' |
| cTAR1_29nt_loop | 5' TMR *ACA GAG AGA GAG CTC AAA GGC TCC TCT GT* F 3' |

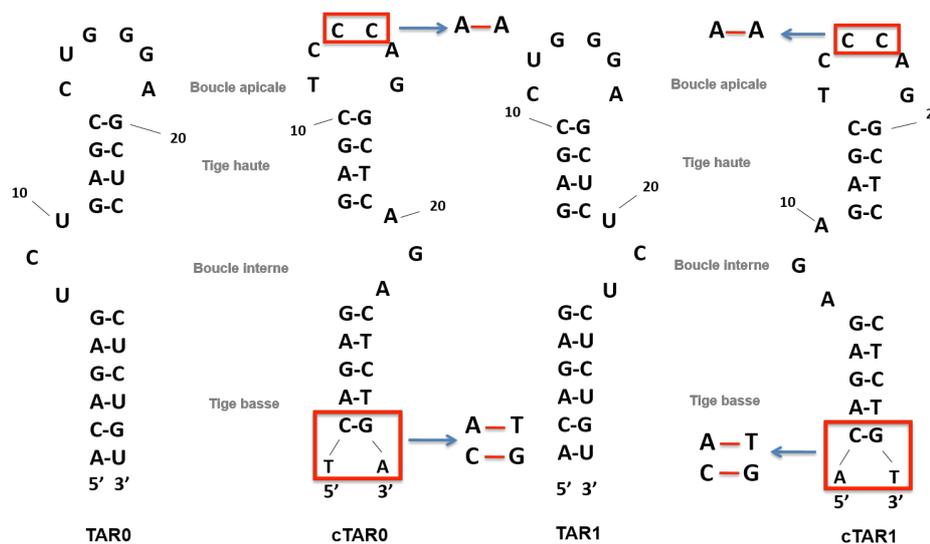


Figure 2B: Séquences d'oligonucléotides TAR/cTAR utilisées pour l'étude par FRET. À gauche est représenté le couple sauvage (TAR0 et cTAR0), à droite est représenté le couple mutant (TAR1 et cTAR1) dont la position des boucles internes a été inversé. Les bases encadrées représentent les modifications effectuées sur les ADNs, sur la boucle apicale pour empêcher l'hybridation par les boucles (loop0 et loop1) ou par les extrémités 5'/3' (stem0 et stem1).

Pour les expériences de RMN PRE, les heptaoligonucléotides (Tableau 3B) sont modifiés soit en 5' ou 3' par une base 4-thio-dT (S4dT) ou 4-thio-U (S4U) pour les ADNs et les ARNs respectivement (Figure 3B). Par la suite, les oligonucléotides seront marqués par un agent paramagnétique, le proxyl. Le protocole de marquage est détaillé au paragraphe chapitre méthodes (III.2.3.3).

Tableau 3B. Oligonucléotides utilisés pour l'étude par RMN PRE.

| Oligonucléotide modifié | Noms | | |
|----------------------------------|-------------|----------------------|---------------------|
| 5' S4-TAGAGAT 3' | 5'S4T | | |
| 5' TAGAGAT-S4 3' | 3'S4T | | |
| 5' S4-UAGAGAT 3' | 5'S4U | | |
| 5' TAGAGAU-S4 3' | 3'S4U | | |
| Oligonucléotide marqué au Proxyl | | Forme paramagnétique | Forme diamagnétique |
| 5' Proxyl-S4-TAGAGAT 3' | 5'Proxyl-D | 5'Ox-D | 5'Red-D |
| 5' TAGAGAT-S4 -Proxyl 3' | 3' Proxyl-D | 5'Ox-D | 3'Red-D |
| 5' Proxyl-S4-UAGAGAU 3' | 5'Proxyl-D | 5'Ox-R | 5'Red-R |
| 5' UAGAGAU-S4 -Proxyl 3' | 3' Proxyl-D | 5'Ox-R | 3'Red-R |

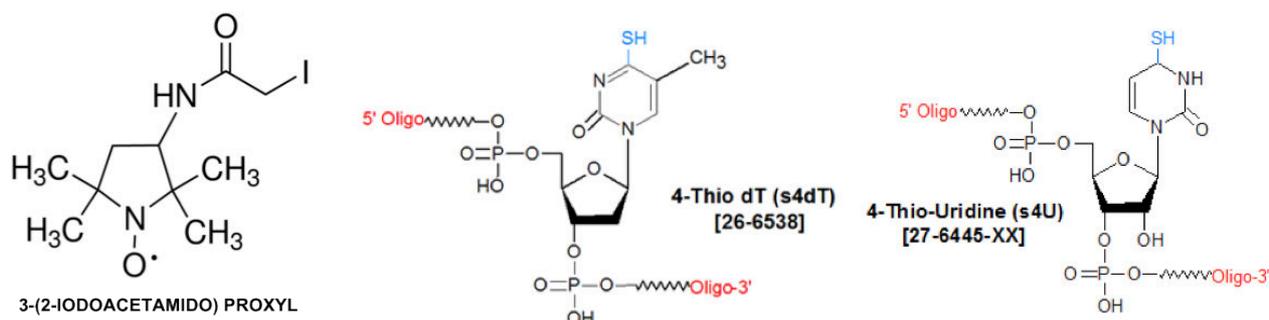


Figure 3B: Sonde paramagnétique et bases modifiées utilisées pour le marquage des oligonucléotides pour l'étude par RMN PRE.

III.1.2 Peptides

Dans un premier temps, les expériences d'anisotropie de fluorescence ont été réalisées avec la protéine tronquée NC(11-55) afin d'éviter les problèmes d'agrégations liés à la partie N-terminale fortement basique. La NC(11-55) a été synthétisée chimiquement (méthode Fmoc/opfp et purifiée par HPLC) par l'équipe du Pr. Yves MELY

de l'Université de Strasbourg à un degré de pureté de >98%. Par la suite, nous avons également utilisé la protéine de forme complète NC(1-55) (Figure 4B). Pour les études de RMN, nous avons procédé à la surproduction de la protéine NC(1-55) enrichie en ^{15}N dans le laboratoire du Dr Carine Tisé (IBPC) et dont le protocole est détaillé au paragraphe méthodes (III.2.3.5).

La concentration de la protéine est déterminée par mesure de l'intensité du signal UV en utilisant un coefficient d'extinction molaire de 6410 M $^{-1}$.cm $^{-1}$. Il est important de préciser que le peptide, dans le cas des synthèses chimiques uniquement, a été préparé avec 3 équivalent de ZnCl_2 afin d'assurer la saturation des motifs en doigts de zinc.

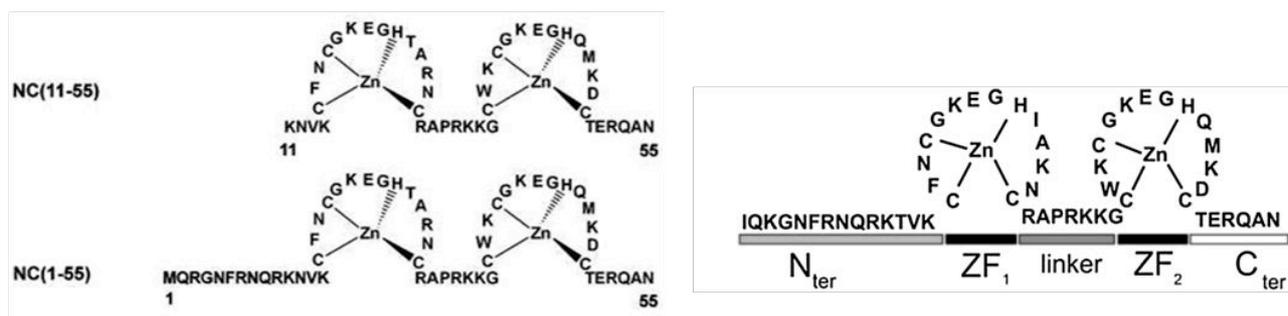


Figure 4B: Structure des peptides utilisés dans nos travaux. Structure de la forme tronquée (11-55) et complète (1-55) de la protéine, souche (MAL) à gauche, utilisées pour les études d'anisotropie de fluorescence. Séquence de la NC(1-55) souche (NL4.3) à droite, utilisée pour les études de FRET et de RMN, où les différents domaines de la NC sont indiqués.

III.2 Méthodes

III.2.1 Spectroscopie de fluorescence

La spectroscopie de fluorescence est une technique versatile qui permet d'analyser, à l'équilibre ou en temps réel, les variations de l'environnement de sondes fluorescentes intrinsèques (notamment le tryptophane pour les protéines) ou extrinsèques (fluorophores synthétiques). Dans mon travail, j'ai utilisé la fluorescence pour déterminer les affinités des interactions des différents partenaires du système, à savoir la NCp7 avec différentes séquences d'acides nucléiques, et également pour explorer les mécanismes d'hybridation entre TAR et cTAR.

III.2.1.1 Principe de la fluorescence

En absorbant un photon, un chromophore passe de son état électronique fondamental (S_0) à un état excité (S_n). Après s'être partiellement désexcité jusqu'au niveau S_1 , il peut retourner à l'état fondamental en émettant un photon : c'est le phénomène de *luminescence*. Ce phénomène englobe deux processus : la fluorescence et la

phosphorescence, qui dépendent de la nature des états fondamentaux et excités de la molécule considérée. La désactivation des molécules excitées s'effectue suivant trois grands processus (Jablonski, 1935):

- a) par désactivation non radiative, $S_n \rightarrow S_0$. L'énergie est alors dissipée par les chocs moléculaires ;
- b) par désactivation radiative. Après désexcitation $S_n \rightarrow S_1$, la molécule, en passant du premier niveau excité S_1 au niveau S_0 , émet un photon de fluorescence : $S_1 \rightarrow S_0 + h\nu_F$;
- c) par conversion inter-systèmes : passage d'un état singulet S_j à un état triplet T_k . Le retour de T_1 à l'état fondamental peut se faire avec émission d'un photon de phosphorescence (Figure 5B).

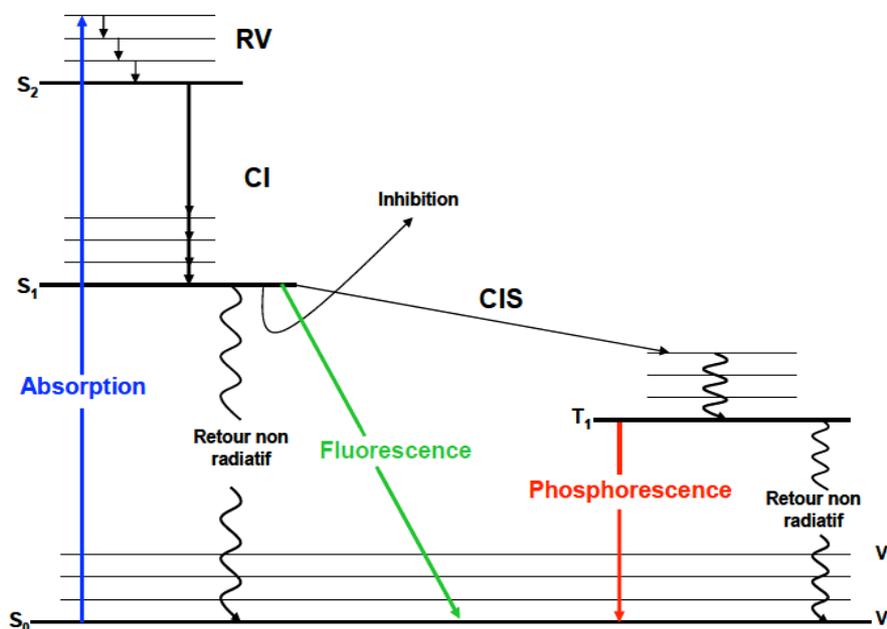


Figure 5B: Diagramme de conversion d'énergie selon Jablonski (Jablonski, 1935) indiquant les chemins permettant la production et la désexcitation d'un état excité (S_0 = état Fondamental ; S_1, S_2 = état Singulet ; T_1 = état Triplet ; RV = rotation vibrationnelle ; CI = Conversion Interne ; CIS = conversion inter-système).

III.2.1.2 Caractéristiques de la fluorescence

Le coefficient d'extinction molaire ϵ ($\text{mol.L}^{-1}.\text{cm}^{-1}$) ou absorbance spécifique reflète la probabilité d'absorption d'un composé à une longueur d'onde donnée. Il obéit à la loi de Beer-Lambert :

$$A = \epsilon . l . c \quad (1)$$

Où A est l'absorbance, c est la concentration de l'échantillon et l la longueur du trajet optique.

L'intensité de fluorescence ou « brillance » d'une sonde fluorescente est déterminée par le produit du coefficient d'extinction molaire et le rendement quantique.

Le temps de déclin ou de durée de vie de fluorescence τ (s) correspond à la durée de vie moyenne de l'état excité avant le retour à l'état fondamental. Il peut être mesuré par l'analyse du déclin de fluorescence, suivant la loi exponentielle :

$$I_{(t)} = I_0 e^{-\frac{t}{\tau}} \quad (2)$$

Où I_0 est l'intensité de fluorescence initiale.

τ est caractérisé par la dépopulation du niveau excité par les processus radiatifs et non radiatifs et est relié aux constantes cinétiques de désexcitation correspondantes, par la relation:

$$\tau = \frac{1}{k_r + k_{nr}} \quad (3)$$

Où k_r est la constante cinétique de désexcitation radiative et k_{nr} est la constante cinétique de désexcitation non radiative.

Le rendement quantique et les propriétés spectrales de fluorescence d'une sonde sont sensibles à leur environnement. Une des conséquences d'une modification des propriétés du milieu peut être une modification du rendement quantique de la sonde qui se traduit par une atténuation de la fluorescence. Cette atténuation peut être causée par différents mécanismes comme (i) un transfert d'énergie de la sonde vers un composé environnant, (ii) une influence d'un facteur exogène (pH, température...) ou (iii) une interaction avec d'autres molécules de la même espèce si la densité du fluorophore est trop importante. L'atténuation de fluorescence peut aussi refléter l'agrégation du peptide : les résidus fluorescents sont enfouis dans les oligomères et deviennent moins accessibles à la lumière excitante. On peut définir un rendement quantique photochimique (ϕ_f) par le bilan entre le nombre de photons émis par le nombre de photons absorbés. Il représente donc la fraction des phénomènes radiatifs dans la désexcitation de la molécule :

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

Lorsque la désexcitation non radiative est négligeable, le ϕ_f est proche de 1.

Le déplacement de Stokes représente la différence entre les maxima des longueurs d'onde d'absorption et d'émission.

III.2.1.3 Anisotropie de fluorescence

L'anisotropie de fluorescence est une méthode utilisée pour étudier les interactions des macromolécules en permettant de déterminer les paramètres de liaison impliqués : constante de dissociation, coopérativité et stœchiométrie de liaison⁵¹²⁻⁵¹³. Plus précisément, le phénomène d'anisotropie repose sur le maintien d'une polarisation de la lumière absorbée par l'échantillon (Figure 6B). L'anisotropie de fluorescence est une technique qui enregistre la dépolarisation partielle ou totale de la fluorescence d'un fluorophore localisé dans une macromolécule (protéine ou acide nucléique). La dépolarisation de fluorescence peut être causée par plusieurs phénomènes selon l'échantillon, mais la principale cause est la rotation par diffusion rotationnelle des fluorophores. Les molécules qui ont la bonne orientation par rapport à la lumière excitatrice polarisée linéairement sont donc excitées préférentiellement : c'est le principe de la photosélection (Figure 5B)⁵¹⁴.

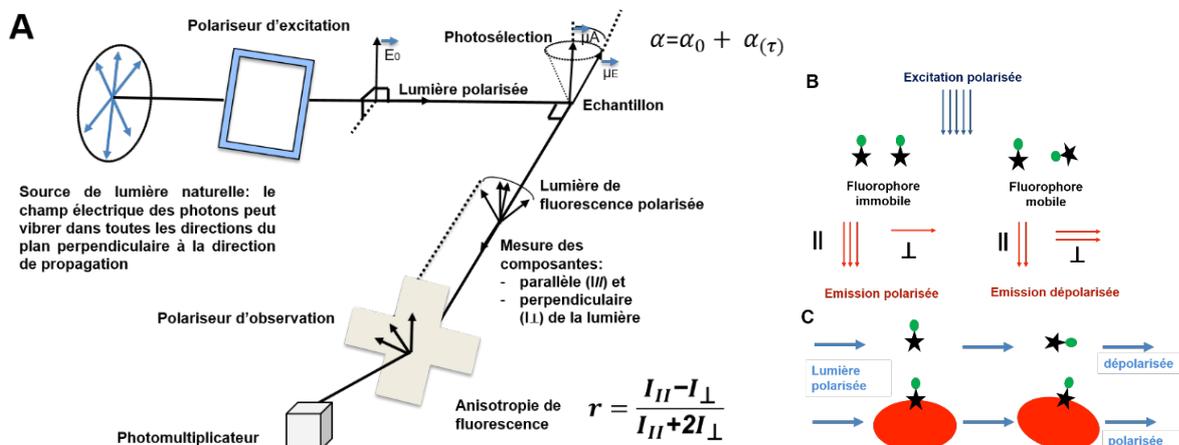


Figure 6B: Représentation schématique de l'analyse de la liaison d'un ligand par anisotropie de fluorescence. (A) Principe de mesure par anisotropie de fluorescence (B) Un fluorophore est excité par une lumière polarisée. Si pendant la durée de vie de l'état excité, le fluorophore est immobile, la lumière émise sera majoritairement orientée parallèlement au rayon incident. Par contre, si l'orientation du dipôle des molécules fluorescentes change (mouvement du fluorophore), la polarisation P mesurée diminue. (C) Lorsqu'un ligand fluorescent (D*) libre en solution est excité par une lumière polarisée, sa rotation rapide (en raison de son faible poids moléculaire) provoque l'émission d'une lumière dépolarisée et une anisotropie faible est mesurée. Quand D* se lie à un récepteur R, son volume effectif augmente et sa rotation est ralentie. La lumière émise est dans ce cas-là polarisée dans le même plan que la lumière d'excitation et A augmente (adaptée d'après⁵¹⁵).

La direction de l'émission de fluorescence dépend de la liberté de rotation du fluorophore dans le milieu. Le phénomène se décompose ainsi : l'onde incidente excitatrice est polarisée linéairement, puis elle va interagir avec l'échantillon. L'interaction

entre le champ électrique de l'onde incidente et le dipôle d'absorption d'un fluorophore est maximum quand ils sont parallèles. Les molécules qui ont la bonne orientation sont donc excitées préférentiellement : c'est le principe de la photosélection. Le champ électrique de l'onde réémise est orienté parallèlement au dipôle d'émission, pour les mêmes raisons qu'à l'excitation. Le degré de polarisation de fluorescence est défini par l'anisotropie (A ou le taux de polarisation (P). On peut alors définir ces deux termes par les relations suivantes :

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \quad (5)$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad (6)$$

P et r peuvent s'interchanger en utilisant les équations suivantes :

$$A = \frac{2P}{3-P} \quad (7)$$

$$P = \frac{3A}{2+A} \quad (8)$$

La mesure de l'anisotropie de fluorescence a été utilisée pour quantifier un grand nombre de réactions d'association entre molécules biologiques. Celles-ci incluent la formation de complexes de type antigène-anticorps⁵¹⁶, protéine-protéine⁵¹⁷⁻⁵¹⁸, ligands-protéines ou encore protéine-acide nucléique^{512, 519}.

Soit un ligand fluorescent, qui peut exister sous forme libre (l) ou associé (a) à une macromolécule (m), l'équilibre $l+m \rightleftharpoons a$ produit une anisotropie (A) de fluorescence observée en présence des deux formes :

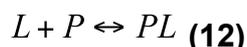
$$A = f_l \times A_l + f_a \times A_a \quad (9)$$

Où A_l et A_a sont respectivement les anisotropies de la forme libre et de la forme associée, et f_l et f_a sont les fractions de l'intensité totale de fluorescence correspondantes. Si le rendement quantique de fluorescence est le même pour le fluorophore libre et le fluorophore lié, alors f_l et f_a représentent aussi les fractions des formes libres et liées.

Il est ainsi possible de déterminer la valeur de la constante de dissociation de la réaction étudiée. Comme $f_a + f_l = 1$ (10)

$$\text{On a : } f_a = \frac{A - A_l}{A_a - A_l} \quad (11)$$

Considérons l'équilibre de fixation d'un ligand (L) fluorescent sur une protéine (P) :



La constante de dissociation de cette réaction est donnée par :

$$K_D = \frac{[P] \times [L]}{[PL]} \quad (13)$$

La fraction liée du fluorophore est donnée par :

$$f_a = \frac{[PL]}{[L]_t} = \frac{[P]}{[P] + K_D} = \frac{A - A_l}{A_a - A_l} \quad (14)$$

Où $[L]_t = [L] + [PL]$ (15) est la concentration totale de ligand fluorescent.

Généralement, la valeur de A_l est obtenue en mesurant l'anisotropie du fluorophore en l'absence de protéine et celle de A_a en mesurant l'anisotropie du fluorophore à une concentration de protéine suffisamment grande pour que la totalité du fluorophore soit complexé à la protéine. Dans ces conditions, l'anisotropie de fluorescence peut être utilisée pour déterminer la constante de dissociation du complexe à partir de la valeur calculée f_a .

Nos études d'anisotropie de fluorescence ont été effectuées sur un fluorimètre FluoroMax Jobin Yvon doté d'un système amovible permettant des mesures d'anisotropie (équipé d'un polariseur au niveau de la lumière d'excitation et d'un autre polariseur au niveau de la lumière d'émission). Les titrages ont été réalisés à 5 °C dans une cuve en quartz thermostatée 1 cm x 0.5 cm. Les oligonucléotides marqués ont été dilués à la concentration désirée (20 nM) dans 800 µl de tampon. Les échantillons d'oligonucléotides ont été titrés par l'addition de concentrations croissantes de NC(11-55) ou NC(1-55). Pour chaque point de titrage, aux moins 8 valeurs d'anisotropie ont été enregistrées avec un temps d'intégration de 1 s. Le traitement des courbes de titrage par GraphPad Prism selon la méthode « Non linear regression (curve fit) Least squares (ordinary fit) » a permis de déterminer les K_d des interactions.

III.2.1.4 FRET (Förster Resonance Energy Transfer):

Un fluorophore à l'état excité peut transmettre son énergie à un autre par transfert non radiatif, sans qu'il soit en contact avec lui. Au cours du transfert, le fluorophore D donneur (dans notre cas la fluorescéine, F) d'énergie revient à l'état fondamental, tandis que l'accepteur d'énergie A (tétraméthylrhodamine, TMR) passe à l'état excité. La fluorescence de D est partiellement éteinte, cependant qu'apparaît la fluorescence spécifique de A (Figure 7B). Le transfert d'énergie résulte d'une interaction entre les dipôles électriques portés par le donneur et l'accepteur. Cette interaction dépend de la distance et de l'orientation de ces dipôles. Il faut de plus pour qu'elle ait lieu, que les deux chromophores aient des niveaux d'énergie appropriés : le spectre d'émission du donneur doit recouvrir le spectre d'absorption de l'accepteur. Quand on observe l'émission de

fluorescence à une longueur d'onde λ_{em} caractéristique de l'accepteur, le transfert d'énergie fait apparaître dans le spectre une bande qui correspond à l'absorption par le donneur, à des longueurs d'onde plus courtes que celles auxquelles l'accepteur absorbe. La bande d'émission du donneur est décrite par l'intensité de fluorescence $F_D(\lambda)$ et la bande d'absorption de l'accepteur par le coefficient molaire d'extinction $\varepsilon_A(\lambda)$. On définit l'intégrale de recouvrement :

$$J = \int_0^{\infty} F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad (15)$$

Le transfert a lieu à une vitesse k_t . On montre que cette vitesse varie comme $1/r^6$ avec la distance r du donneur à l'accepteur. Soit :

$$k_t = \frac{1}{\tau_D} \left(\frac{r_0}{r} \right)^6 \quad (16)$$

Où T_D est le temps de déclin de fluorescence du donneur en l'absence de l'accepteur. Le paramètre r_0 s'appelle la distance Förster. Elle dépend de l'intégrale de recouvrement J et de l'orientation des dipôles de fluorophores. Quand le recouvrement des bandes d'absorption de D et de A est important, la distance de Förster peut atteindre 50 Å. Elle est donc de l'ordre de grandeur des dimensions des macromolécules biologiques.

L'efficacité du transfert d'énergie E est la fraction du nombre de photons absorbés par D et qui sont transférés à A. Elle est égale au rapport de la constante de vitesse k_t du processus de transfert, à celle du processus de désexcitation. En présence d'accepteur, la vitesse de désexcitation est, puisque D peut se désexciter soit par lui-même à la vitesse, soit par transfert à la vitesse k_t :

$$E = \frac{k_t}{k_t + \tau_D^{-1}} = \frac{r_0^6}{r_0^6 + r^6} \quad (17)$$

L'intensité de fluorescence émise par A quand D est excité est proportionnelle à l'efficacité de transfert E . Elle varie donc selon la loi décrite précédemment, ou la loi de Förster, lorsque la distance r entre D et A varie.

En résumé l'efficacité de transfert est proportionnelle à la distance entre le chromophore accepteur et donneur, rendant la technique du FRET très sensible aux changements de distance entre les deux molécules. La distance à partir de laquelle les intensités de fluorescence varient significativement est néanmoins fortement dépendante du couple de fluorophore qui est utilisée. Dans le cas du couple Fluorescéine/TMR (carboxytetramethylrhodamine) que nous avons utilisé la valeur calculée R_0 est de 60 Angstroms, ce qui signifie qu'un accroissement de l'intensité de fluorescence de la fluorescéine (en cas d'excitation à 480 nm) est détectée lorsque la distance interchromophore excède 30 Angstroms⁵²⁰.

Pour le couple de fluorescence qui nous intéresse, il a également été noté que quand les

deux molécules sont très proches, il se produit un couplage excitonique (type de couplage fort) qui se traduit par l'existence d'une très faible intensité de fluorescence pour les deux chromophores. Ce couplage fort n'est pas un couplage de type FRET et n'est visible que quand les deux chromophores sont très proches dans l'espace ^{404, 520}.

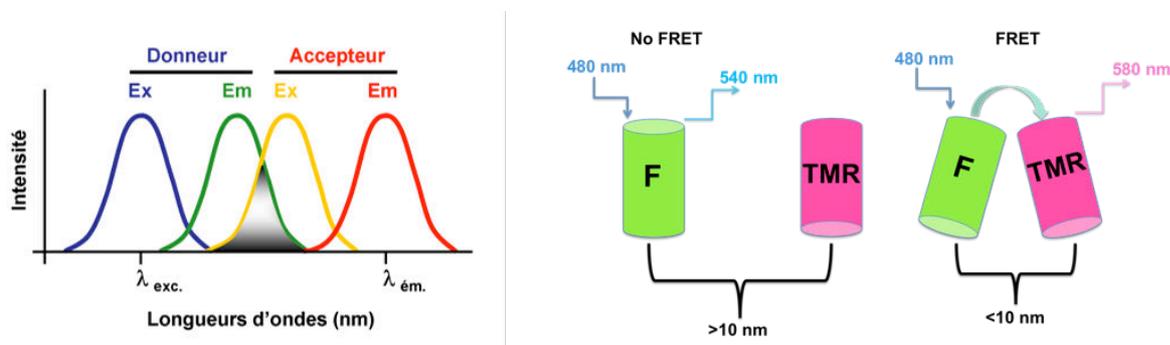
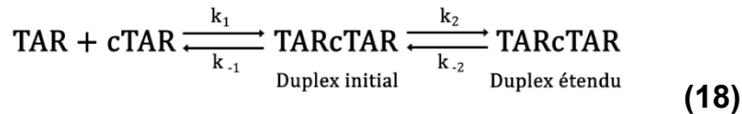


Figure 7B: Schémas illustrant le principe du FRET. A) Les spectres d'émission et d'excitation des deux chromophores avec le spectre d'émission du donneur (F en vert) qui recouvre partiellement le spectre d'excitation de l'accepteur (TMR en rose). **B)** Influence de la distance sur l'efficacité du FRET : Lorsque les deux chromophores sont trop éloignés, il ne peut pas y avoir de transfert d'énergie du donneur à l'accepteur alors que lorsqu'ils sont rapprochés, le transfert est rendu possible.

Dans notre système, lorsque la molécule d'ADN marquée est repliée (en forme tige boucle), les chromophores situés aux extrémités 3' et 5' sont proches. Néanmoins dans nos conditions, le couplage excitonique n'est pas visible probablement du fait d'une ouverture des bases terminales (paire AT peu stable). De ce fait le transfert de type FRET (qui est un cas de couplage faible contrairement au couplage excitonique qui est un cas de couplage fort) est rendu possible et la fluorescence émise par F est donc très faible car l'énergie est transférée à la TMR par mécanisme FRET. Dans le cas d'un appariement total entre la molécule d'ADN et d'ARN en hétéroduplex étendu par la NCp7, les deux marqueurs fluorescents vont se retrouver éloignés, le transfert par FRET ne sera plus possible et donc l'intensité de fluorescence de la fluorescéine augmentera. C'est cette vitesse d'augmentation de l'intensité de fluorescence de la fluorescéine qui nous permettra de calculer la vitesse d'hybridation des molécules par la NCp7.

Les expériences de FRET ont été réalisées avec un fluorimètre FluoroMax-2 Jobin Yvon et les données obtenues ont ensuite été traitées, à l'aide du programme GraphPad Prism, afin d'extraire les paramètres cinétiques de la réaction. L'utilisation de quantité d'ARN en large excès nous permet de nous situer dans des conditions de pseudo-premier ordre, ce qui permet de simplifier les équations cinétiques et l'extraction des vitesses d'hybridation des séquences. Le logiciel « Graph Pad » nous permet d'extraire les trois valeurs expérimentales k_s , k_f et f . Avec ces valeurs, en utilisant le formalisme développé

dans Vo et al, 2006 et 2009, on peut calculer les constantes de vitesses microscopiques dans le schéma réactionnel suivant. Notons également que sur la base de l'existence de deux phases cinétiques, d'autres modèles réactionnels sont possibles, voir par exemple Godet et al, 2006 :



Les constante k_1 et k_{-1} sont liés à la formation d'un complexe initial dans lequel les deux molécules TAR et cTAR sont associées mais ne sont pas complètement appariées. La constante k_2 est une constante de vitesse associée à l'apparition du complexe étendu. L'utilisation de deux exponentielles pour réaliser l'ajustement ou fit (Equation 19) des données est significativement meilleure (sur un plan statistique) qu'avec une seule exponentielle, ce qui signifie que deux phases cinétiques, une rapide et une lente, sont impliquées ^{178, 424-426}. A partir des « fits » des courbes on peut distinguer deux phases reliées par les équations présentées dans l'article Vo et al, 2006, nous permettant de déterminer les constantes de vitesse. La quantité de complexe apparaissant au cours du temps ($Pr(t)$) peut être écrite comme la somme d'un processus lent (avec une constante de vitesse k_s , s pour slow) et d'un processus rapide (avec une constante de vitesse k_f , f pour fast). De plus, il n'y a que ces deux phases donc elles ont les coefficients f et $1-f$, f étant la fraction du composant rapide :

$$Pr(t) = Pr_{\infty} (1 - f e^{-k_f t} - (1 - f) e^{-k_s t}) \quad (19)$$

$$k_f = k_1 \cdot R + k_{-1} \quad (20)$$

$$k_s = f \cdot k_2 \quad (21)$$

Où $k_1 = k_f \left(\frac{f}{R} \right)$ (22) et $k_{-1} = k_f (1 - f)$ (23) sont liées à la formation d'un complexe initial, la constante $k_2 = \frac{k_s}{f}$ (24) est associée à l'apparition du complexe étendu avec R qui correspond à la concentration en ARN. Le K_d représente la constante de dissociation du complexe initial :

$$k_d = \frac{k_{-1}}{k_1} = \left(\frac{1-f}{f} \right) \times R \quad (25)$$

III.2.2 Résonance Magnétique Nucléaire (RMN)

La résonance magnétique nucléaire (RMN) est une technique utilisée pour l'analyse des structures et de la dynamique de nombreuses molécules chimiques et biologiques. En biologie structurale, elle permet la détermination de la structure tridimensionnelle des macromolécules biologiques (protéines, acides nucléiques, sucres, complexes protéine/protéine ou protéine/acide nucléique) et constitue également une méthode puissante pour l'étude des interactions des macromolécules entre elles et avec leurs ligands. La RMN permet non seulement la résolution de structure tridimensionnelle des molécules, mais donne aussi des informations sur la dynamique interne des molécules sur une très large échelle de temps (de la picoseconde à la seconde) (Figure 12B).

En effet, grâce aux progrès technologiques et méthodologiques : développement de spectromètres à très haut champ (1200 MHz), spectromètre à double champ ⁵²¹, marquage isotopique ²H, marquage ¹⁵N, ¹³C sélectif ⁵²² développement de séquences de pulse RMN 3D et 4D ⁵²³⁻⁵²⁴, la mise au point de la séquence TROSY et CRINEPT (Transverse Relaxation Optimized Spectroscopy et Cross-correlated Relaxation-Enhanced Polarization Transfer respectivement) ⁵²⁵⁻⁵²⁷ et l'utilisation des constantes de couplages dipolaires ⁵²⁸⁻⁵²⁹. Ces progrès ont permis d'augmenter la résolution et la sensibilité de la technique. La spectroscopie RMN est de plus en plus performante dans la détermination de structures tridimensionnelles et dans l'étude de la dynamique de macromolécules de grande taille (>100 KDa) ^{392, 530}.

Dans la suite de cette partie, je vais présenter tout d'abord les méthodes d'attribution RMN des protéines puis les méthodes d'étude de la dynamique interne des protéines que j'ai utilisées dans ce travail pour étudier les protéines de nucléocapside du VIH-1. Avant de mener l'étude par RMN d'une protéine, il convient d'exprimer et de purifier la dite protéine à haute concentration (0.1-2 mM) et stable pour plusieurs jours. Le protocole de purification de la NCp7 est détaillé dans la partie Méthodes (III.2.3.5).

III.2.2.1 Méthode d'attribution d'une protéine en RMN

La première étape de l'étude d'une protéine par RMN consiste en l'attribution des fréquences de résonance (déplacements chimiques) des noyaux observables (¹H, ¹³C, ¹⁵N). Dans le cas des protéines non-marquées, l'attribution des résonances des noyaux d'hydrogène se fait principalement en 2 étapes : 1) l'attribution des signaux provenant de chaque type d'acide aminé avec des expériences de type COSY (COrrélation SpectroscopY), TOCSY (TOtal Correlation SpectroscopY ou encore appelé

HOHAHA (HOmonuclear HArtman HAhn spectroscopy)); et 2) l'attribution séquentielle qui permet de déterminer l'enchaînement des acides aminés avec des expériences de type NOESY⁵³¹. Cette approche n'est applicable que sur des petites protéines, en général de taille inférieure à une centaine d'acides aminés et dont les résonances sont peu ou pas superposées. Au-delà, la superposition des signaux dans les spectres augmente l'ambiguïté de l'attribution voir rend impossible l'attribution. De plus, cette méthode repose sur des corrélations séquentielles spatiales (NOEs), qui dépendent de la structure et peuvent être limitées, voir pire mener à des erreurs d'attributions.

Dans le cas d'une protéine doublement marquée (¹⁵N, ¹³C), l'attribution des signaux ¹H, ¹⁵N, ¹³C est grandement facilitée et permet une attribution non-ambiguë, grâce notamment aux expériences multidimensionnelles de type 3D ou 4D. Ces méthodes reposent principalement sur l'ajout d'une troisième voir quatrième dimension et sur les transferts de cohérence par les couplages scalaires utilisant les noyaux ¹⁵N, et/ou ¹³C (à une ou plusieurs liaisons J_{H-N} et J_{H-C} J_{C-N} et J_{C-C}). La méthodologie d'étude d'une protéine que j'ai suivie au cours de mon travail est schématisée sur la figure ci-dessous (8B).

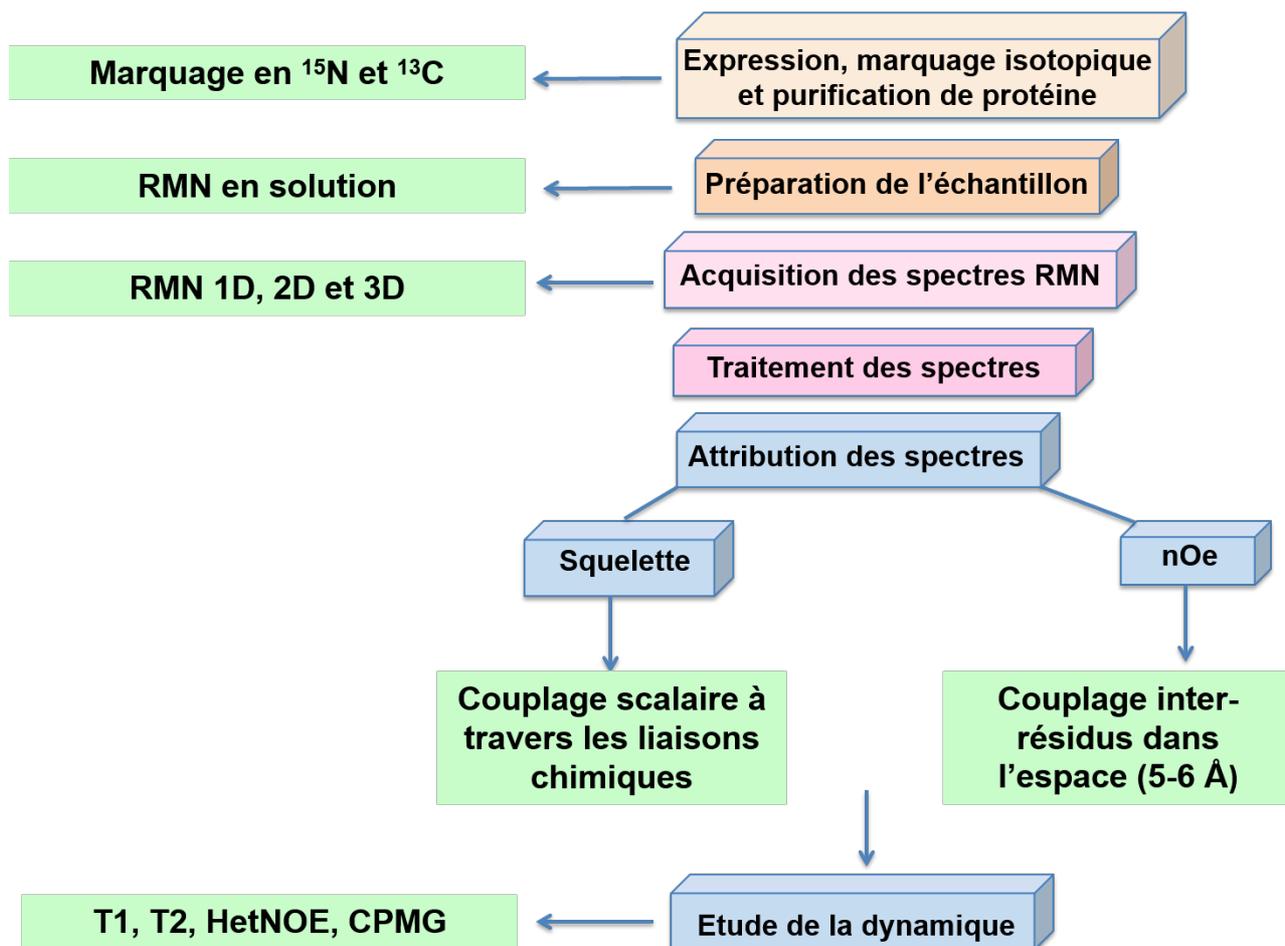


Figure 8B: Schéma illustrant la stratégie d'étude structurale et dynamique d'une protéine par RMN.

L'attribution des pics de resonances se fait classiquement avec une série d'expériences 3D en triple résonance (^1H , ^{13}C et ^{15}N) nécessitant une protéine doublement marquée (^{15}N et ^{13}C). L'attribution se fait généralement en deux étapes i) l'attribution des resonances du squelette peptidique et ii) celles des chaînes latérales.

III.2.2.1.1 Attribution du squelette peptidique

La première étape pour l'étude d'une protéine est d'enregistrer un spectre 2D de type ^{15}N - ^1H HSQC (Heteronuclear Single Quantum Coherence). Chaque pic sur ce spectre connecte le déplacement chimique de l'azote (^{15}N) et du proton (^1H) de tous (sauf des prolines) les groupements amides (HN) présents dans la protéine. Cela signifie que chaque pic sur le spectre représente un acide aminé de la protéine, et le spectre HSQC l'empreinte de la protéine. Il faut noter que les groupes NH/NH_2 des chaînes latérales Asn, Gln, Trp, Lys et Arg peuvent aussi donnée lieu à un pic.

Cependant pour attribuer chaque pic sur la 2D HSQC, il nous faut enregistrer une série d'expériences 3D, qui sont interprétées par paires : HNCA et HN(CO)CA, HNCO et HN(CA)CO, HNCACB et HN(CO)CACB ⁵²⁴. Le nom des expériences provient du type de noyau édité, correspondant donc à une des trois dimensions et des chemins de cohérences suivis. Par exemple, la HNCO corrèle les noyaux HN du résidu i , N du résidu i et CO du résidu $i-1$. Les parenthèses autour d'un noyau signifient que le chemin de cohérence passe par ce noyau mais qu'il n'est pas édité (c'est à dire que son déplacement chimique n'est pas enregistré durant l'expérience): l'expérience HN(CA)CO corrèle les noyaux HN du résidu i , N du résidu i et CO du résidu $i-1$ ainsi que HN du résidu i , N du résidu i et CO du résidu i (Figure 9B).

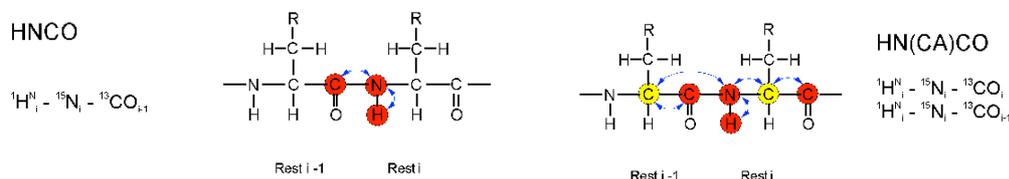


Figure 9B: Stratégie d'attribution d'une protéine marquée en ^{15}N et ^{13}C basée sur des expériences 3D HNCO et HN(CA)CO. Dans une HNCO, l'aimantation est transférée (flèches bleues) du proton amide $\text{H}^{\text{N}}(i)$ via l'atome $\text{N}(i)$ à l'atome de carbone qui est directement attaché, le CO ($i-1$) et retourne de la même manière au noyau $\text{H}^{\text{N}}(i)$ qui est directement détecté. Les fréquences des trois noyaux en rouge sont détectées. Dans une HN(CA)CO, l'aimantation est transférée (flèches bleues) du proton amide $\text{H}^{\text{N}}(i)$ via l'atome $\text{N}(i)$ et du noyau CA ($\text{C}^{\alpha}(i)$) à l'atome de carbone, le CO (i) et retourne de la même manière. Le C^{α} en jaune joue le rôle d'un noyau de relais et sa fréquence n'est pas détectée. On détecte seulement les fréquences des H^{N} , N et CO (rouge).

Un autre exemple est l'expérience HNCACB qui permet d'obtenir l'information sur le déplacement chimique de chaque amide sur l'HSQC du résidu i et les informations sur le déplacement chimique des carbones CA et CB du résidu i et du résidu $i-1$.

Le second spectre HN(CO)CACB permet également d'obtenir l'information sur le déplacement chimique de chaque amide sur le spectre HSQC du résidu i mais uniquement des informations sur le déplacement chimique des carbones CA et CB du résidu $i-1$. De ce fait, on utilise ce spectre, pour définir quel CA et CB corresponde au résidu i et $i-1$ sur le spectre HNCACB (Figure 10B).

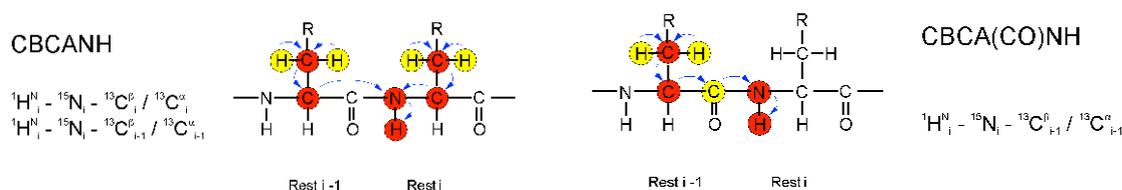
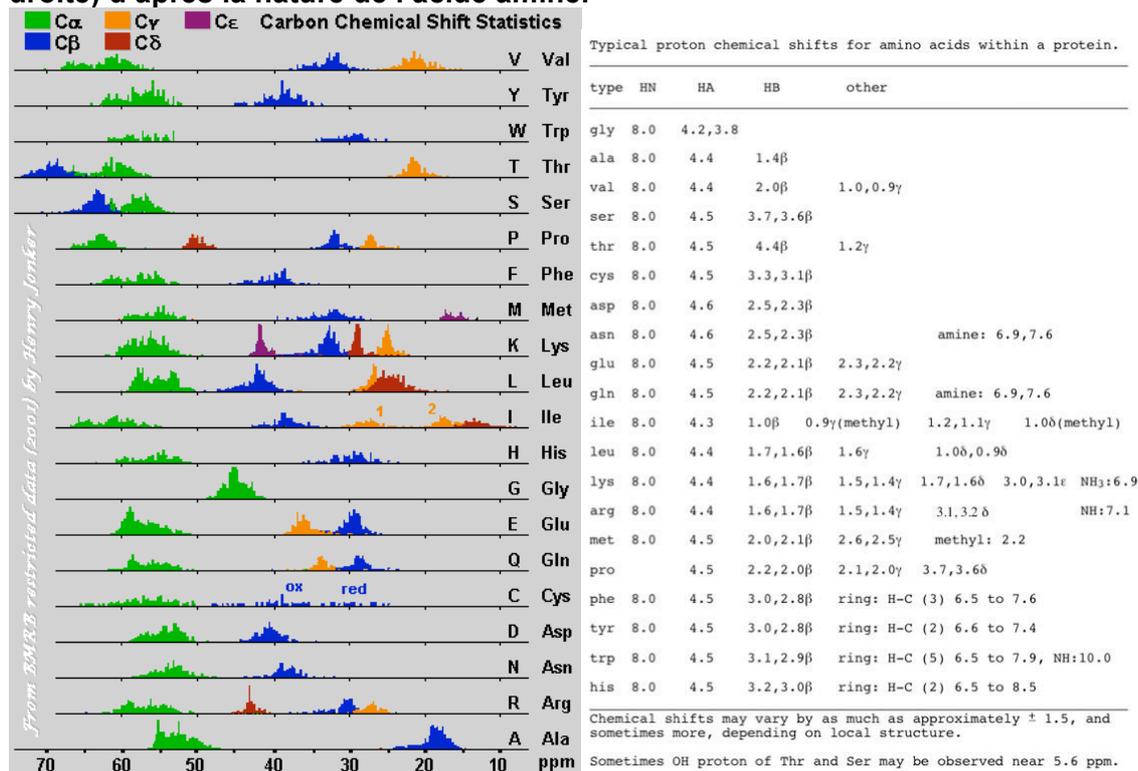


Figure 10B: Stratégie d'attribution d'une protéine marquée en ${}^{15}N$ et ${}^{13}C$ basée sur des expériences 3D HNCACB et HN(CO)CACB. Dans une HNCACB, l'aimantation est transférée (flèches bleues) du proton amide $H^N(i)$ via l'atome $N(i)$ à l'atome de carbone qui est directement attaché, le $CA(i)$ vers le $CB(i)$. L'aimantation retourne de la même manière au noyau $CA(i)$ vers $N(i)$ à $H^N(i)$ qui est directement détecté. De même, l'aimantation est transférée du proton amide $H^N(i)$ via l'atome $N(i)$ à l'atome de carbone $CA(i-1)$ vers le $CB(i-1)$ et l'aimantation retourne ensuite. Dans une HN(CO)CACB, l'aimantation est transférée (flèches bleues) du proton amide $H^N(i)$ via l'atome $N(i)$ et du noyau $CO(i-1)$ à l'atome de carbone $CA(i-1)$ vers le $CB(i-1)$ et retourne de la même manière. Les noyaux en jaune jouent le rôle de noyaux de relais et leur fréquence n'est pas détectée. On détecte seulement les fréquences des atomes en rouge.

De ce fait, l'utilisation de ces expériences rend l'attribution du squelette beaucoup plus facile pour deux raisons : la première est que la dimension ${}^{13}C$ réduit le repliement des pics qui peuvent se produire sur les spectres 2D; d'autres parts, vu que les déplacements chimiques des carbones α et β sont spécifiques à chaque acide aminé (Tableau 4B), il est possible de les utiliser pour identifier la nature du résidu et ainsi procéder à l'attribution du squelette peptidique. Il faut cependant prendre des précautions car la structure secondaire peut influencer considérablement et faire sortir un carbone de la zone décrite dans le tableau ci-dessous.

Tableau 4B. Liste des déplacements chimiques des carbones (à gauche) et des protons (à droite) d'après la nature de l'acide aminé.



Un autre type d'expérience 3D est une ^1H - ^{15}N NOESY-HSQC qui combine deux types d'expériences 2D. Le schéma de base de la séquence d'impulsions de cette expérience est illustré dans la figure ci dessous. Schématiquement l'aimantation originelle du spin K_i (^1H) est transférée sur le spin I_j (HN) pendant le temps de mélange τ_m de la séquence NOESY. L'expérience HSQC transfère ensuite l'aimantation du spin I_j vers le spin S_j (^{15}N) via le couplage scalaire $^1J_{IS}$. Pendant le temps t_1 l'aimantation évolue sous l'effet du seul déplacement chimique de S ; enfin on observe le retour de l'aimantation sur le spin I_j via le couplage scalaire $^1J_{IS}$ pour la détection (Figure 11B).

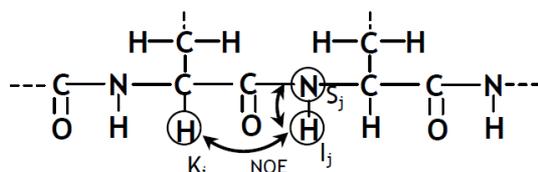
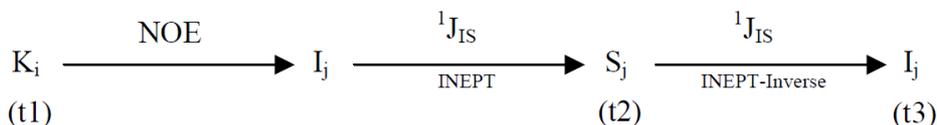


Figure 11B: Transfert d'aimantation dans une expérience 3D ^1H - ^{15}N NOESY-HSQC.

III.2.2.1.2 Attribution des chaînes latérales

L'attribution des chaînes aliphatiques sont accessibles par deux expériences tridimensionnelles : la ^{15}N -TOCSY-HSQC et la ^{13}C -HCCH-TOCSY-HSQC.

La première corrèle par couplage scalaire isotrope dans les conditions de Hartmann-Hahn⁵³²⁻⁵³³ chaque proton amide avec les hydrogènes de la chaîne latérale de chaque résidu. Le spin-lock est centré sur les carbones aliphatiques, et en suivant les conditions de Hartmann-Hahn, l'aimantation portée par un carbone va se transmettre sur toute la chaîne latérale. La première étape consiste à éditer les hydrogènes de la chaîne aliphatique, transférer l'aimantation par INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) vers les carbones aliphatiques, puis de répartir cette aimantation sur toute la chaîne latérale. Une INEPT retour fait revenir l'aimantation sur les hydrogènes et l'acquisition est effectuée. Le résultat est un spectre 3D contenant la même information qu'une 2D- ^1H - ^1H TOCSY éclatée suivant le déplacement chimique des carbones aliphatiques. Cette méthode présente l'avantage que le transfert HCCH-TOCSY est bien plus efficace qu'un transfert par les hydrogènes.

III.2.2.2 Analyse de la dynamique interne d'une protéine

Outre les mouvements de diffusion rotationnelle et translationnelle, une protéine en solution possède des degrés de liberté internes. On distingue donc les mouvements globaux de la protéine (diffusion d'un corps rigide) et les mouvements des vecteurs internes (liaisons atomiques). Etant donné que chaque spin nucléaire agit comme une sonde extrêmement sensible aux modifications temporelles de son environnement, la RMN est donc un outil puissant et très fin pour analyser l'ensemble de ces mouvements sur plusieurs ordres de grandeur et à l'échelle atomique⁵³⁴⁻⁵³⁶.

La dynamique interne des protéines est étudiée essentiellement à partir de la relaxation des spins nucléaires ^{15}N ou ^{13}C qui sont liés à un seul proton. L'analyse des mouvements internes des protéines de nucléocapside du VIH-1 a été centrée sur l'étude de la relaxation du noyau ^{15}N sur un échantillon uniformément marqué en ^{15}N . Ce sont les mouvements du vecteur ^{15}N -H et par conséquent ceux du squelette peptidique qui ont été explorés. Dans la suite de ce chapitre, je décrirai brièvement les principes théoriques qui sont sous-jacents à l'étude de la dynamique interne d'une protéine par relaxation nucléaire.

III.2.2.2.1 Mouvements moléculaires

Une macromolécule en solution présente divers degrés de liberté de mouvements internes qui sont liés à la forme de l'énergie potentielle du système. Il peut exister plusieurs conformations de basse énergie avec des barrières d'énergie suffisamment faibles pour qu'une molécule passe d'une forme à l'autre facilement (par exemple, la rotation d'un cycle aromatique), certains noyaux peuvent alors passer d'un site à un autre et d'un déplacement chimique à un autre avec une vitesse donnée : c'est l'échange chimique. L'échange chimique met en jeu en général des mouvements de l'ordre de μs – ms . Sur une échelle de temps plus petite (ps – ns), les liaisons atomiques disposent d'une liberté de mouvement autour de leurs orientations moyennes. Les motifs structuraux tels que les boucles ont souvent une dynamique plus importante que les structures en hélice α et en feuillet β . La figure (12B) rassemble les différents types de mouvements et leur échelle de temps observable par RMN. En effet, La cinétique de ces mouvements moléculaires s'étale sur une large gamme de temps allant de la femtoseconde à la minute. Les mouvements de faible amplitude nécessitant peu d'énergie sont extrêmement rapides alors qu'inversement, ceux de grandes amplitudes sont plus lents et plus coûteux en énergie. Ce sont les mouvements rapides (de la pico à la nanoseconde) qui peuvent être sondé par l'étude de la relaxation nucléaire. De nombreuses études portent sur l'analyse des corrélations entre la dynamique de la protéine et sa fonction.

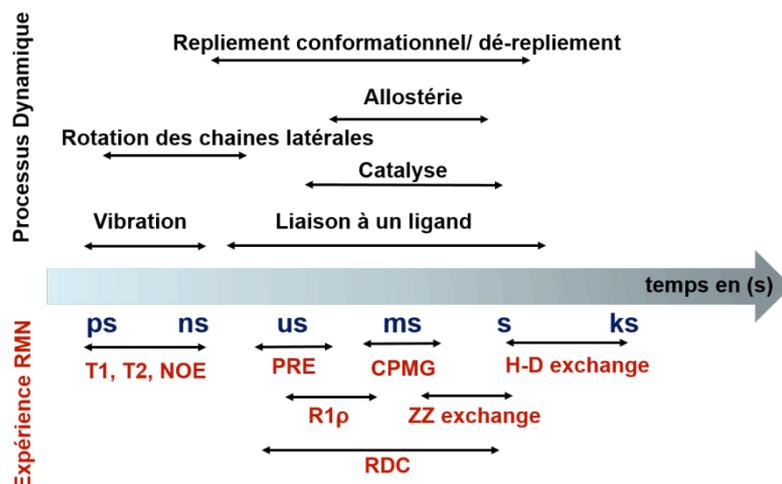


Figure 12B: Schéma représentatif de l'échelle des temps des mouvements accessibles par RMN.

III.2.2.2.2 La fonction de densité spectrale

Les mouvements atomiques induisent des fluctuations locales du champ magnétique. La distribution des mouvements d'une protéine se réorientant en solution se

traduit ainsi par une distribution du champ magnétique fluctuant à différentes fréquences. Cette répartition est exprimée en terme de densité spectrale $J(\omega)$ disponible à chaque fréquence (Figure 13B). La fonction de densité spectrale peut s'exprimer pour une molécule sphérique rigide en rotation globale isotrope et aléatoire (mouvement brownien) :

$$J(\omega) = \frac{\tau_c}{1 + \omega^2 \tau_c^2} \quad (26)$$

Le temps de corrélation τ_c est défini comme la durée que met un vecteur fixe X-H (X= ^{15}N ou ^{13}C) par rapport à la sphère pour balayer un angle de 1 radian.

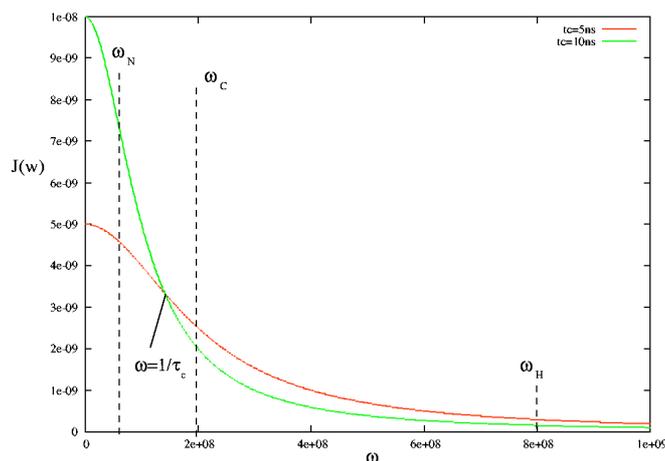


Figure 13B: Fonction de densité spectrale d'une molécule en diffusion rotationnelle isotrope sans mouvement interne. A un champ magnétique donné, plus la protéine étudiée a une masse moléculaire faible, plus le temps de corrélation associé est court et plus la courbe de densité spectrale est étalée. Les deux courbes rouge et verte représentent les temps de corrélations à 5 ns et à 10 ns respectivement.

III.2.2.2.3 Relaxation des spins et mouvements moléculaires

La RMN est un phénomène qui se produit quand les noyaux de certains atomes sont immergés dans un champ magnétique statique (B_0), ils se répartissent sur différents niveaux d'énergie, le système est alors à l'équilibre thermodynamique (Figure 14B). Après une excitation provoquée par un second champ magnétique oscillant (B_1), le système est déplacé hors de sa position d'équilibre. C'est le retour à cette position qui illustre le phénomène de relaxation et dépend de la fréquence à laquelle les champs magnétiques localement ressentis par les spins fluctuent en magnitude et en direction.

De nombreux mécanismes sont à l'origine de la relaxation nucléaire, les deux plus importants étant d'une part l'interaction dipôle-dipôle (DD) qui sont significatives pour des spins proches dans l'espace et d'autre part l'anisotropie de déplacement chimique (CSA pour Chemical Shift Anisotropy) associée aux modulations de l'orientation des noyaux par rapport au champ magnétique statique (^{15}N et ^{13}C possèdent des anisotropies de valeur importante). Les deux phénomènes se traduisent par l'existence de champs magnétiques

locaux autour des noyaux de la protéine. La fluctuation des champs magnétiques localement induits dépend du mouvement global de la molécule ainsi que des mouvements, plus rapides, associés à sa dynamique interne. Les mesures des paramètres de relaxation constituent donc des données très puissantes pour sonder la dynamique interne des macromolécules, protéines et acides nucléiques.

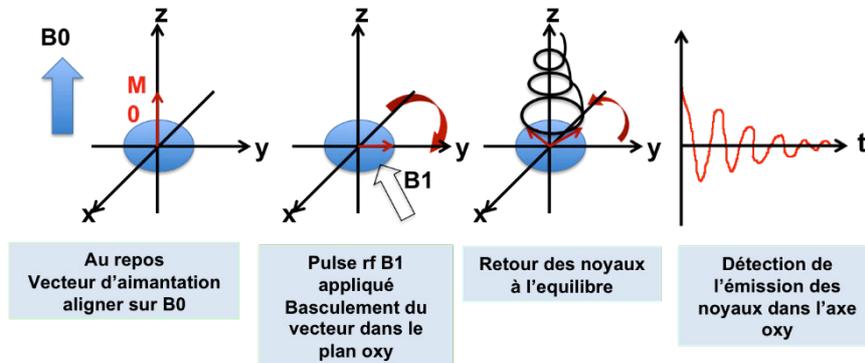


Figure 14B: Représentation schématique du principe d'une expérience RMN.

Sur le plan phénoménologique, le retour à l'équilibre peut être décrit avec une loi mono-exponentielle avec des temps caractéristiques T_1 et T_2 ⁵³⁷⁻⁵³⁸.

$$M_z(t) = M_{z0} \cdot e^{-t/T_1} \quad (27)$$

$$M_{xy}(t) = M_{xy0} \cdot e^{-t/T_2} \quad (28)$$

La magnétisation M est décomposée en deux vecteurs $M_z(t)$ et $M_{xy}(t)$ qui sont les composantes longitudinale et transversale de l'aimantation au temps t , M_{z0} et M_{xy0} sont ces composantes au temps $t=0$ (Figure 15B).

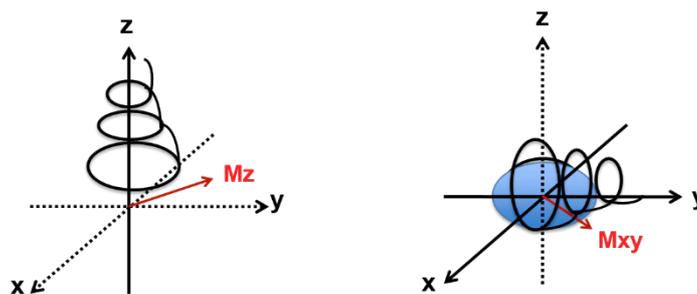


Figure 15B: Représentation schématique de la relaxation longitudinale (à gauche) et transversale (à droite).

La relaxation longitudinale T_1 , liée aux échanges d'énergie entre les spins et leur environnement (appelé réseau : on parle parfois de relaxation spin/réseau), elle décrit le retour à l'équilibre de l'aimantation M_z le long de l'axe z , associé au passage des spins de l'état β à l'état α (et réciproquement). La vitesse de relaxation sera notée $R_X(X_z)$ où X_z désigne l'aimantation longitudinale du noyau X .

Cette relaxation est favorisée par des fluctuations de champ magnétique à haute fréquence ($\omega \approx \omega_0$). Seules celles qui fluctuent à des fréquences de l'ordre de ω_0 sont notables dans le repère tournant. Le temps de relaxation longitudinal T_1 est donc minimum pour $\tau c = 1/\omega_0$ (Figure 16B).

On définit aussi la vitesse de la relaxation longitudinale exprimée en s^{-1} :

$$R_1 = \frac{1}{T_1} \quad (29)$$

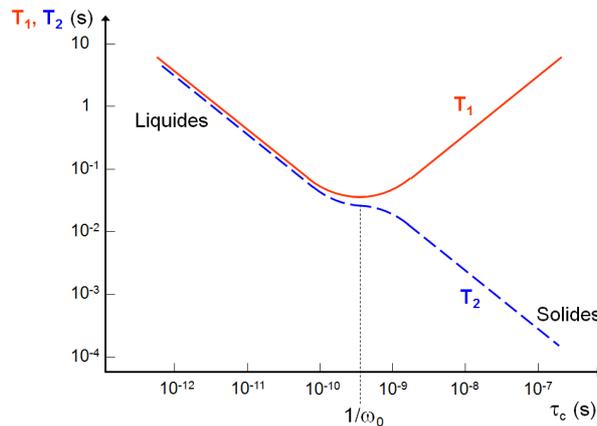


Figure 16B: Relation entre les temps de relaxations T_1 , T_2 et le temps de corrélation τc . Pour des valeurs élevées du temps de corrélation τc (faibles mobilité), la relaxation transversale est très efficace, alors que la relaxation longitudinale est peu efficace. Cette situation est caractéristique de l'état solide. Pour des valeurs faibles de τc (forte mobilité moléculaire), les deux composantes de la relaxation sont peu efficaces et identiques. Cette situation est caractéristique de l'état liquide. Lorsque $\tau c \omega_0$ est proche de 1, on observe un minimum pour T_1 . La valeur de τc correspondant à ce minimum dépend donc de B_0 .

La relaxation transversale T_2 , provient d'un échange d'énergie des contacts spin-spin. Elle décrit le retour à zéro de l'aimantation transversale $M_{xy}(t)$ située dans le plan xy perpendiculairement au champ magnétique B_0 . Le T_2 correspond au déphasage des aimantations nucléaires sous l'effet des fluctuations du champ local à toutes les fréquences : les mouvements « lents » (devant les fréquences de transition RMN) y contribuent cette fois aussi bien que les mouvements haute fréquence. T_2 est inférieur à T_1 puisqu'un plus grand nombre de phénomènes contribuent au T_2 . La courbe décrivant les variations de T_2 en fonction de τc est décroissante, indiquant que la relaxation est d'autant plus efficace que les mouvements sont lents (figure 16B).

On définit aussi la vitesse de la relaxation transversale par R_2 et exprimée en s^{-1} :

$$R_2 = \frac{1}{T_2} \quad (30)$$

III.2.2.2.4 L'effet nOe (Nuclear Overhauser Effect)

Initialement prévu par Overhauser en 1953, l'effet nOe (ou NOE) a été mis en évidence expérimentalement par Solomon en 1955, qui l'a aussi décrit de manière théorique et en a posé l'équation fondamentale. De manière générale, l'interaction impliquée est un couplage dipôle-dipôle entre deux noyaux suffisamment proches dans l'espace (5 à 6 Å) et dont la perturbation de l'un influence sur l'évolution de l'autre (effet nOe). Le nOe implique donc une distance intermoléculaire dans l'espace ainsi que des mouvements moléculaires rapides. L'intensité du nOe est proportionnelle à r^{-6} où r représente la distance entre les deux spins (Figure 17B).

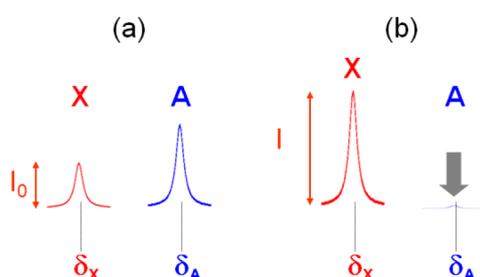


Figure 17B: Effet nOe. (a) spectre d'un système de deux spins $\frac{1}{2}$ non équivalents, non couplés mais suffisamment proches pour être en interaction dipolaire. (b) Spectre du même système après une irradiation continue et longue du spin A ; le noyau A est saturé et son signal disparaît, l'intensité du signal du noyau X est modifiée.

Le processus physique du nOe implique une relaxation nucléaire via une saturation de certains signaux (élimination de la différence de populations, à travers certaines transitions en les irradiant avec un pulse rf faible) tout en observant les signaux de l'autre. Le nOe est donc une tentative du système pour rester à l'équilibre.

Dans le cas des protéines, la mesure de l'effet nOe consiste à comparer l'intensité du signal issu du spin S en présence et en absence d'irradiation du spin I. L'irradiation a pour effet d'équilibrer les populations du spin. La mesure du paramètre nOe $^1\text{H}-^{15}\text{N}$ correspondant à la liaison amide. L'analyse de l'effet nOe le long de la chaîne peptidique donne alors une estimation qualitative de la dynamique interne de la protéine. Plus les vecteurs $^1\text{H}-^{15}\text{N}$ sont rigides (ne montrent pas de dynamique interne), plus le paramètre nOe sera proche de 1 et inversement une valeur de l'effet nOe plus faible indique que ce vecteur est plus mobile.

III.2.2.2.5 La relation entre les paramètres de relaxation et les densités spectrales

Dans notre étude, nous avons mesuré pour le spin nucléaire ^{15}N les paramètres de relaxation suivants : la relaxation longitudinale (R_1), la relaxation transversale (R_2) et le NOE hétéronucléaire (HetNOE) à l'état stationnaire. Ces trois paramètres dépendent des fonctions de densité spectrale $J(\omega)$ de la façon suivante ⁵³⁹ :

$$R_1 = \frac{1}{4} d^2 \{3J(\omega_N) + 7J(\omega_H)\} + c^2 J(\omega_N) \quad (31)$$

$$R_2 = \frac{1}{8} d^2 \{4J(0) + 3J(\omega_N) + 13J(\omega_H)\} + \frac{1}{6} c^2 \{4J(0) + 3J(\omega_N)\} \quad (32)$$

$$NOE = 1 + \frac{1}{4} T_1 d^2 \left(\frac{\gamma_H}{\gamma_N} \right) \times \{5J(\omega_H)\} \quad (33)$$

$$J(0) = \frac{1}{3d^2 + 4c^2} (6R_2 - R_1 (3 + \frac{18}{5} \left(\frac{\gamma_N}{\gamma_H} \right) (NOE - 1))) \quad (34)$$

Les constantes d^2 et c^2 sont définies de la façon suivante :

$$d^2 = \left(\frac{1}{10} \right) \left(\frac{h\gamma_H\gamma_N}{2\pi r_{NH}^3} \right)^2 \quad (35)$$

$$c^2 = \frac{2}{15} \omega_N^2 (\sigma_{\parallel} - \sigma_{\perp})^2 \quad (36)$$

Le terme d^2 représente la constante d'interaction dipolaire et c^2 la constante d'anisotropie de déplacement chimique.

Les fréquences ω_H , ω_N représentent respectivement les fréquences de résonances des noyaux ^1H et ^{15}N , γ_H et γ_N sont leurs rapports gyromagnétiques, h est la constante de Planck dont la valeur est 6.626×10^{-34} J.s, r_{NH} est la distance interatomique de 1,02 Å entre ^1H et ^{15}N , enfin σ_{\parallel} et σ_{\perp} sont les composantes parallèles et perpendiculaires du tenseur de déplacement chimique de ^{15}N . La quantité $\sigma_{\parallel} - \sigma_{\perp}$ qui constitue l'anisotropie de déplacement chimique de ^{15}N est proche de -170 ppm.

Il faut noter que les paramètres de relaxations (R_1 , R_2 , NOE) sont divisé en composantes DD et CSA mais que chacune dépend d'une combinaison linéaire de la fonction de densité spectrale évaluée à cinq fréquences $J(0)$, $J(\omega_N)$, $J(\omega_H)$, $J(\omega_H + \omega_N)$ et $J(\omega_H - \omega_N)$. Le terme $J(0)$, représente le déphasage de la magnétisation transversale par un champ magnétique local non homogène. Les termes $J(\omega_N)$ représente la relaxation vers l'état fondamental qui détruit également la magnétisation transversale. La présence du

terme $J(\omega_H)$ indique que la transition du proton couplé affecte la cohérence de la magnétisation transversale de l'azote.

III.2.2.2.6 Mouvement lent des protéines et relaxation de dispersion CPMG

Les protéines peuvent adopter une grande multiplicité de conformations tridimensionnelles. Cependant, les techniques structurales permettent de détecter majoritairement l'état le plus peuplé, néanmoins il est important de souligner que d'autres éléments peuplant le paysage énergétique (même moins peuplés) peuvent aussi avoir une importance biologique (Figure 18B). La présence de tels conformères peut être identifiée s'il y a un échange conformationnel avec la conformation principale dans certains régimes d'échange. L'échange conformationnel entre deux ou plusieurs environnements entraîne des effets importants sur les vitesses de relaxation et module la forme et la position des résonances des noyaux en échange. Il est possible d'obtenir par la mesure de ces effets, des informations sur les paramètres thermodynamiques, cinétiques et structuraux associées aux états en échange conformationnel.

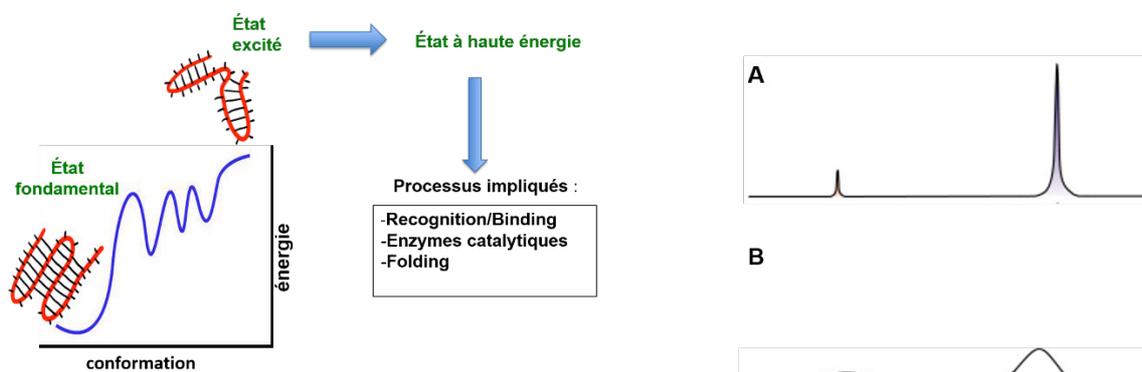
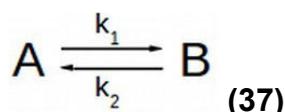


Figure 18B: Illustration de l'échange conformationnel d'une protéine en deux sites. A gauche : L'état de haute énergie (excité) est plus rare que celui de basse énergie mais peut présenter des fonctions biologiques importantes. À droite : L'état de faible population n'est pas directement observable en RMN ; **(A)** En absence d'échange les pics sont non élargis ce qui signifie que la magnétisation est en phase, **(B)** en présence d'échange de l'ordre de la milliseconde, mène au déphasage du signal ce qui fait que les pics sont élargis.

Lorsqu'un spin s'échange entre deux environnements, son déplacement chimique et sa vitesse de relaxation peuvent différer dans chaque environnement.



La cinétique de l'échange est définie par les constantes de vitesse d'échange, qui sont les constantes de vitesse d'échange de l'état A vers l'état B (k_1) et la réaction inverse (k_2).

Les fractions du système dans chaque conformation sont définies par p_A et p_B

respectivement :

$$pA = \frac{k_2}{k_1 + k_2} \quad (38)$$

$$pB = \frac{k_1}{k_1 + k_2} \quad (39)$$

Dans les études RMN des échanges conformationnels, on considère la vitesse d'échange ; $k_{ex} = k_1 + k_2$ (40) et la différence des fréquences de précession des deux états en échange ω_A et ω_B ; $\Delta\omega = \omega_A - \omega_B$ (41).

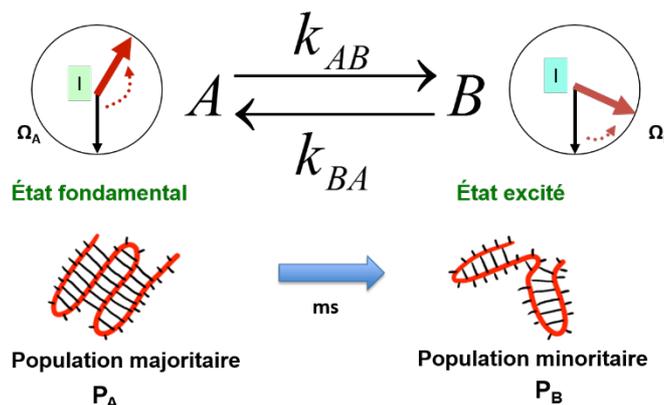


Figure 19B: Schématisation de l'échange conformationnel, entre deux conformations A et B, dont les populations sont p_A et p_B , les fréquences de précession du spin I sont Ω_A et Ω_B .

La RMN s'avère être la technique la plus adaptée aux études dynamiques car les fréquences de résonance des spins sont extrêmement sensibles à l'environnement électronique (Figure 19B). Une modification de structure peut donc provoquer un changement des spectres. En effet, la dispersion de relaxation CPMG est une technique qui permet la caractérisation des mouvements moléculaires de l'ordre de la milliseconde. Ainsi, elle permet d'extraire les paramètres physico-chimiques et structuraux associés aux processus d'échange conformationnel, notamment dans le cas d'un échange entre un état majoritairement peuplé et ceux faiblement peuplés (population supérieure à 0,5%)⁵⁴⁰.

La séquence impulsionnelle des expériences de dispersion de relaxation est une HSQC ou TROSY à laquelle est incorporée une période de relaxation Carr-Purcell-Meiboom-Gill (CPMG)⁵⁴¹⁻⁵⁴². Elle peut se décomposer en plusieurs étapes : excitation du proton, transfert de la polarisation sur l'azote, relaxation en présence de séquence CPMG, marquage des déplacements chimiques de l'azote, transfert de la polarisation sur le proton et acquisition. Pendant la relaxation CPMG, l'échange provoque une non-refocalisation de la polarisation semblable à une perte de d'aimantation transversale (Figure 20B).

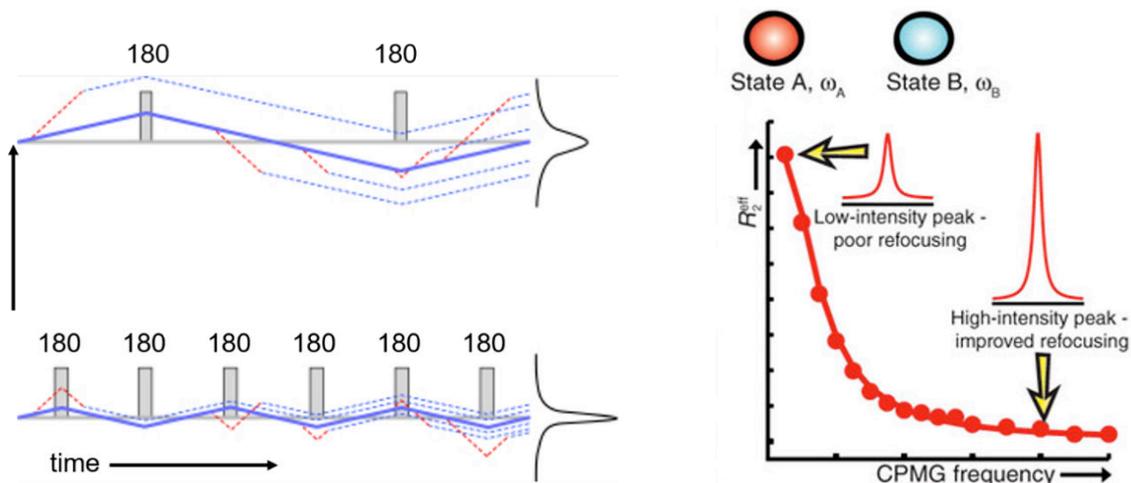


Figure 20B: Schéma simplifié d'une expérience CPMG. À gauche est représenté un train d'impulsions CPMG, les pulses rf à 180° inverse la direction effective de la précession et augmentant le taux de répétitions des pulses (V_{CPMG}); on diminue le déphasage du signal et la perte de celui-ci. A droite, est représenté le profil de dispersion de relaxation à partir duquel les paramètres d'échange peuvent être extraits⁵⁴³.

La vitesse de relaxation transversale observée R_{2eff} est composée de la perte d'aimantation intrinsèque à la protéine R_2 ainsi que d'une composante due à l'échange R_{ex} . Cette dernière dépend de la différence de fréquence de précession des spins dans les deux conformations $\Delta\omega_{AB}$, de la vitesse d'échange entre les conformations k_{ex} (fréquence des changements de vitesse de précession), des populations relatives des deux conformations p_A et p_B (probabilité d'évoluer sous chaque déplacement chimique) ainsi que du délai τ des échos de spin du CPMG :

$$R_{2eff} = R_2 + R_{ex}(\tau, k_{ex}, p_B, \Delta\omega_{AB}) \quad (42)$$

Lors d'une expérience de dispersion de relaxation, plusieurs spectres sont enregistrés en faisant varier le nombre d'échos de spin pendant un délai de relaxation T_{relax} fixe (40 ms par exemple). Plus la fréquence des échos de spin est grande, plus la probabilité d'avoir un échange pendant l'écho diminue et plus l'aimantation est refocalisée.

Les facteurs R_{2eff} sont obtenus pour chaque pic de l'HSQC, en comparant pour chaque délai τ l'intensité du pic de l'expérience avec celle d'une mesure de référence (sans CPMG). Les courbes de dispersion de relaxations sont obtenues en reportant pour chaque pic la valeur du R_{2eff} en fonction de la fréquence des échos de spin $\nu_{CPMG} = \frac{1}{\tau}$ (43).

L'ajustement des courbes obtenues pour chaque pic est réalisé avec les modèles théoriques qui permettent d'obtenir les paramètres physico-chimiques de l'échange :

- i) La fonction de Meiboom⁵⁴⁴ pour un échange rapide ($k_{ex} > \Delta\omega_{AB}$),

On obtient ainsi pour le modèle de Meiboom les paramètres ($p_A p_B \Delta\omega^2$, τ_{ex} , R_2) pour chaque résidu :

$$R_2 = R_2^0 + \frac{p_A p_B \Delta\omega^2}{k_{ex}} \left[1 - \frac{4v_{CPMG}}{k_{ex}} \tanh\left(\frac{k_{ex}}{4v_{CPMG}}\right) \right] \quad (44)$$

- ii) La fonction de Carver⁵⁴⁵ pour un échange intermédiaire ($k_{ex} = \Delta\omega_{AB}$) et (p_B , T_{ex} , $|\Delta\delta|$, R_2) où $\Delta\delta$ est la différence de déplacement chimique du noyau considéré dans les deux conformations :

$$R_2\left(\frac{1}{\tau_{cp}}\right) = \frac{1}{2} (R_{2A}^0 + R_{2B}^0 + k_{ex} - \frac{1}{\tau_{cp}} \cosh^{-1}[D_+ \cosh(\eta_+) - D_- \cos(\eta_-)]) \quad (45)$$

$$D_{\pm} = \frac{1}{2} \left[\pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \zeta^2)} \right] \quad (46)$$

$$\eta_{\pm} = \frac{\tau_{cp}}{\sqrt{2}} \left[\pm \psi + (\psi^2 + \zeta^2)^{1/2} \right]^{1/2} \quad (47)$$

$$\psi = (R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex})^2 - \Delta\omega^2 + 4p_A p_B k_{ex}^2 \quad (48)$$

$$\zeta = 2\Delta\omega (R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex}) \quad (49)$$

- iii) La fonction de Tolliger⁵⁴⁶ pour un échange lent ($k_{ex} < \Delta\omega_{AB}$):

$$R_2 = R_2^0 + k_{AB} \left(1 - \frac{\sin(\Delta\omega T_{CP})}{\Delta\omega T_{CP}} \right) \quad (50)$$

Le calcul des R_{2eff} correspondant à chaque pic et à chaque délai à partir des intensités grâce à la relation :

$$R_{2eff} = \frac{1}{T_{relax}} \ln\left(\frac{I}{I_0}\right) \quad (51)$$

Dont le calcul d'erreur se fait suivant la relation :

$$\Delta_{R2} = \frac{\delta_e}{I_{CP} \times T_{CP}} \quad (52)$$

Où δ_e représente la valeur du bruit rms mesuré sur une boîte vide du spectre de référence, I_{CP} l'intensité du pic mesuré, T_{CP} le temps de relaxation mesuré.

III.2.2.3 Mise en œuvre expérimentale

III.2.2.3.1 Mesure de R1 et R2

Les séquences d'impulsions utilisées pour mesurer les vitesses de relaxation longitudinale $R1$ et transversale $R2$ sont basées sur la séquence d'impulsions HSQC. Celle-ci commence par une séquence INEPT pour transférer l'aimantation du proton vers le noyau ^{15}N . Ensuite une série d'impulsions amène le système hors de son état d'équilibre, dans l'un des états dont on veut mesurer la relaxation. Suit un délai tr de relaxation, avec une série d'impulsions qui permettent de conserver le simple processus de relaxation en supprimant les effets de relaxation croisée dipôle-dipôle et d'anisotropie

de déplacement chimique. Enfin l'aimantation du noyau ^{15}N est traitée afin de transférer sa polarisation vers le proton à travers une séquence rétro-INEPT. Une série d'expériences est enregistrée en faisant varier le délai tr afin d'échantillonner au mieux la courbe de relaxation. On obtient donc une série de spectres de type ^1H - ^{15}N HSQC pour les différentes valeurs de tr . L'intensité des pics croisés décroît de manière mono-exponentielle et tend vers zéro quand les temps de relaxation tendent vers l'infini. Les valeurs de $R1$ et $R2$ sont déterminées en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction monoexponentielle décroissante :

$$I(tr) = i_0 \cdot e^{-\left(\frac{tr}{T}\right)} \quad (53)$$

Où $I(tr)$ est l'intensité du pic croisé au délai de relaxation tr , i_0 l'intensité au temps $tr=0$ et R correspond à la constante de vitesse de relaxation $R1$ ou $R2$.

L'optimisation des valeurs de i_0 et R est obtenue en minimisant la fonction cible χ^2 dont l'expression est la suivante :

$$\chi^2 = \sum_{i=1}^N \frac{(I(tr)_{calc} - I(tr)_{exp})^2}{\sigma_I^2} \quad (54)$$

Où $I(tr)_{calc}$ et $I(tr)_{exp}$ sont respectivement les intensités des pics calculées et mesurées expérimentalement et σ_I correspond à l'incertitude de mesure sur l'intensité du pic.

Plus la valeur du χ^2 est faible plus la correspondance entre le modèle et les valeurs expérimentales est correcte.

Les incertitudes sur les valeurs des paramètres de relaxation $R1$ et $R2$ sont évaluées par la méthode de Monte Carlo⁵⁴⁷ en utilisant curvfit (Palmer). Les intensités calculées grâce à l'équation (52) sont perturbées aléatoirement selon une distribution normale avec un écart type correspondant à celui de l'erreur sur la mesure. Cette dernière a été évaluée en dupliquant plusieurs points de la courbe de relaxation.

III.2.2.3.2 Mesure de l'effet nOe

Le paragraphe (III.2.2.2.4) introduisant l'effet nOe, laisse clairement entendre que cette mesure impose deux expériences : une expérience de référence et une expérience pendant laquelle les protons amides sont préalablement saturés (irradiation des ^1H). Cette irradiation est assurée par un train d'impulsions qui permet de saturer les protons. Lorsque les protons sont saturés, le système se trouve à l'état stationnaire. Il relaxe alors vers l'équilibre thermique aux transitions zéro quantum et double quanta avec un transfert de polarisation du proton vers le noyau ^{15}N . On obtient deux spectres de type ^1H - ^{15}N HSQC correspondant aux deux expériences avec et sans irradiation. Il faut noter que la

sensibilité de mesure de l'effet NOE est moins importante que celle des paramètres de relaxation R1 et R2 car on ne peut bénéficier du premier transfert de polarisation du ^1H vers ^{15}N réalisé par la première série INEPT. Les valeurs du NOE hétéronucléaire $^1\text{H} - ^{15}\text{N}$ sont déterminées par le rapport suivant :

$$\eta_i(s) = \frac{I}{I_0} \quad (55)$$

Où I et I_0 représentent l'intensité des pics de corrélation avec et sans saturation respectivement. $\eta_i(s)$ indique que c'est le nOe du noyau i quand le noyau s est saturé.

Une fois qu'on dispose des expériences saturée et non saturée, on extrait les intensités de chaque pic et on applique le rapport (55).

L'incertitude sur l'effet nOe est calculée sur la base du bruit déterminé dans une région vide du spectre (commande rms) et en appliquant l'équation suivante :

$$\sigma_{NOE} = \left(\frac{I_{sat}}{I_{nonsat}} \right) \times \sqrt{\left(\left(\frac{\sigma_{sat}}{I_{sat}} \right)^2 + \left(\frac{\sigma_{nonsat}}{I_{nonsat}} \right)^2 \right)} \quad (56)$$

En combinant les valeurs de R1, R2 et NOE, on procède au calcul de la valeur de $J(0)$ en appliquant l'équation (34)⁵⁴⁸. En utilisant les déviations standards (SD) obtenus sur les trois paramètres, 500 simulations de Monte Carlo ont été effectuées pour évaluer l'incertitude sur les $J(0)$ ⁵⁴⁷ en utilisant des scripts R et python. Les valeurs de $J(0)$ obtenus nous fournissent une quantification plus précise des mouvements moléculaires de la protéine, à la fréquence $J(0)$.

III.2.2.3.3 Mesure de la relaxation CPMG

La séquence impulsionnelle des expériences de dispersion de relaxation est une HSQC à laquelle est incorporée une période de relaxation Carr-Purcell-Meiboom-Gill (CPMG)⁵⁴¹⁻⁵⁴² (Figure 21B le CPMG est encadré en rouge). Elle peut se décomposer en plusieurs étapes: excitation du proton, transfert de la polarisation sur l'azote, relaxation CPMG, marquage des déplacements chimiques de l'azote, transfert de la polarisation sur le proton et acquisition.

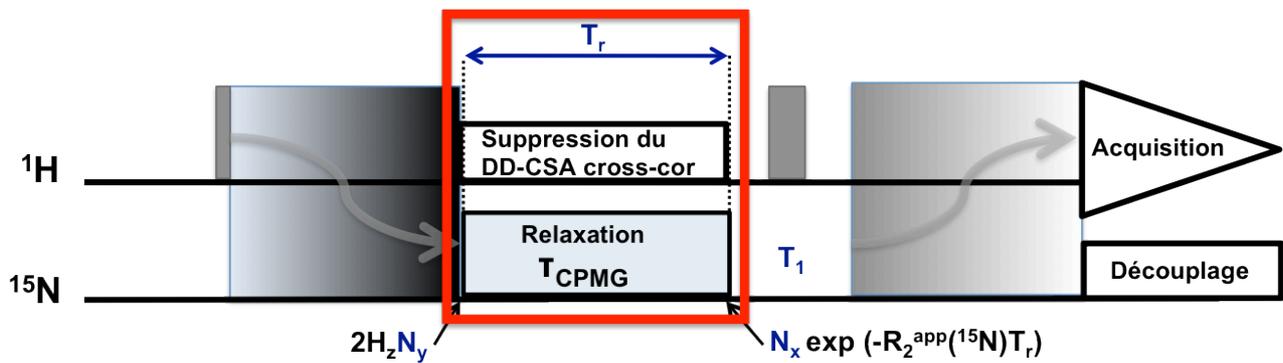


Figure 21B: Représentation schématique de la séquence de dispersion de relaxation.

Pendant la relaxation, l'échange provoque une non-refocalisation de la polarisation semblable à une perte de l'aimantation transversale. La vitesse de relaxation transversale observée R_{2eff} est composée de la perte d'aimantation intrinsèque à la protéine, ainsi que d'une composante due à l'échange R_{ex} . Cette dernière dépend de la différence de fréquence de précession des spins dans les deux conformations $\Delta\omega_{AB}$, de la vitesse d'échange entre les conformations k_{ex} (fréquence des changements de vitesse de précession), des populations relatives des deux conformations p_A et p_B (probabilité d'évoluer sous chaque déplacement chimique) ainsi que du délai τ des échos de spin du CPMG : $R_{2eff} = R_2 + R_{ex}(\tau, k_{ex}, p_B, \Delta\omega_{AB})$.

La figure 22B présente un extrait de la séquence complète, correspondant à la relaxation dans la séquence compensée. Il est précédé par la polarisation du proton et un transfert INEPT vers l'azote. Il présente la relocalisation des cohérences (jusqu'à (b)), la relaxation entre (c)-(d). Pendant les deux CPMG, le découplage proton empêche un couplage scalaire proton-azote et permet l'observation d'une relaxation de la cohérence N_X seule. Il est suivi par le marquage des déplacements chimiques, le retour INEPT et l'acquisition. Les séquences (b)-(c) et (d)-(e) sont utilisées pour préserver l'aimantation de l'eau des effets de l'offset lors du découplage en l'alignant sur son champ magnétique effectif ⁵⁴⁹.

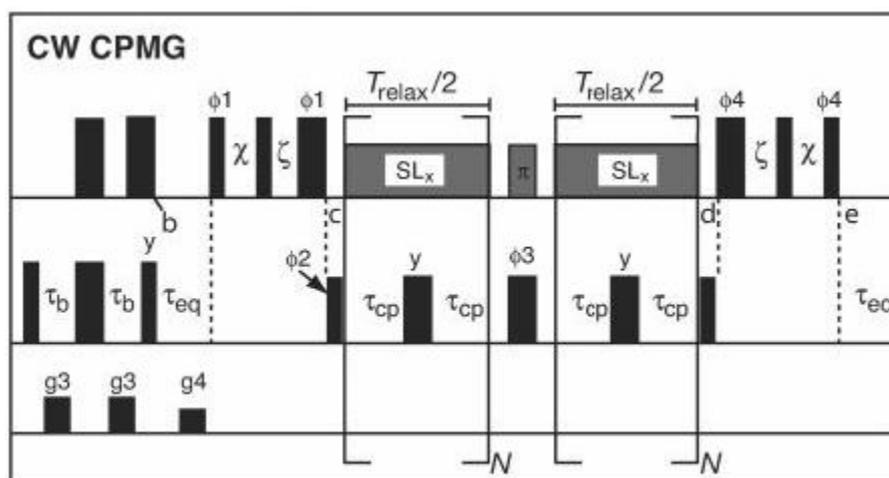


Figure 22B: Extrait de la séquence de dispersion de relaxation découplée ⁵⁵⁰, présentant

l'équivalent de la relaxation de la séquence compensée.

Comme décrit au paragraphe (III.2.2.2.6), nous avons enregistré une série de spectres en faisant varier le nombre d'échos de spin pendant un délai de relaxation T_{relax} fixe (40 ms). Les facteurs $R_{2\text{eff}}$ sont obtenus pour chaque pic de l'HSQC, en comparant pour chaque délai τ l'intensité du pic de l'expérience avec celle d'une mesure de référence (sans CPMG). Nous procédant par la suite au traitement des courbes de dispersion en appliquant l'équation de Meiboom (44) ou de Carver Richards (45), qui correspondent à notre cas (échange rapide ou intermédiaire).

III.2.2.4 L'étude des complexes protéines acides nucléiques par RMN

Les complexes nucléoprotéiques jouent des rôles fondamentaux chez tous les organismes vivants, ce qui souligne l'importance de leur étude. La RMN permet d'obtenir des informations structurales et dynamiques sur ces complexes. Principalement, la procédure d'investigation de ce type d'interaction consiste à déterminer la surface de contact entre les deux biomolécules (CSP ou chemical shift perturbation), de déterminer l'orientation précise des deux espèces au sein du complexe et d'obtenir les contacts intermoléculaires nOe. En général, quand on étudie des complexes macromoléculaires de grande taille, différentes stratégies de marquage sont employées. Dans la majorité des cas, un seul partenaire est marqué ce qui permet de sélectionner un nombre de signaux RMN limités.

Deux types de préparations sont possibles : un premier, qui est le plus fréquent, où la protéine est marquée en ^{15}N et ou en ^{13}C alors que l'acide nucléique n'est pas marqué ; un second, où c'est l'acide nucléique qui est marqué en ^{15}N et/ou en ^{13}C alors que la protéine ne sera pas marquée. Cette technique fournit un moyen pratique de suivre les résonances pendant les expériences de titrage ainsi que d'utiliser des expériences filtrées pour sélectionner ou filtrer l'aimantation des noyaux spécifiques.

Afin de mélanger les deux biomolécules en quantités équimolaires, il est plus facile d'ajouter progressivement un des deux partenaires (protéine) à une solution de l'autre partenaire (acide nucléique), et les spectres RMN sont alors enregistrés après chaque étape d'addition. Les principaux problèmes de préparation des échantillons restent la solubilité et la stabilité du complexe, car, pour identifier les nOe intermoléculaires, l'échantillon doit être fortement concentré et stable à une température raisonnable (entre 25° C et 40° C).

III.2.2.4.1 Les variations de déplacements chimiques

Afin d'étudier les complexes protéines-acides nucléiques, une méthode RMN assez simple consiste à analyser les variations de déplacements chimiques ou CSP "chemical shift perturbations" entre les formes libre et liée de la protéine. Celle-ci se base sur l'analyse des déplacements chimiques du squelette peptidique de la protéine (libre et liée) en enregistrant un spectre hétéronucléaire HSQC ou TROSY afin d'établir la variation des déplacements chimiques de l'azote ^{15}N et du proton amide HN de chaque acide aminé. La variation des déplacements chimiques peut alors être normalisée en utilisant la relation suivante :

$$\Delta_{\omega} = \sqrt{(\Delta_H)^2 + (0,16 \times \Delta_N)^2} \quad (57)$$

Où Δ_N et Δ_{HN} représentent les variations des déplacements chimiques en ppm et la valeur 0,16 est basée sur le rapport des gammes spectrales protons amide/azote peptidique. Cette technique nous permet de détecter les résidus impliqués dans l'interaction avec l'acide nucléique ⁵⁵¹⁻⁵⁵² sachant que les déplacements chimiques sont très sensibles aux perturbations de l'environnement chimique.

III.2.2.4.2 La saturation croisée des protons iminos

Cette méthode est basée sur la saturation sélective des protons d'un partenaire moléculaire qui conduit à une diminution de l'aimantation dans la molécule, qui est principalement contrôlée par la diffusion de spin. La saturation est transférée à l'autre partenaire, entraînant une diminution de l'intensité RMN. Dans le cas des complexes ADN-protéines, les résonances des protons iminos de l'ADN (H3 de la thymine et H1 de la guanine) sont sélectivement saturées en utilisant une impulsion d'inversion sélective centrée sur la gamme de fréquence de résonances des protons iminos ($\delta^1\text{H}$, 12-15 ppm). Pendant la saturation de l'ADN, la relaxation des protons iminos perturbe l'équilibre thermodynamique des spins voisins, ce qui entraîne le transfert de saturation à la protéine. En introduisant la période de saturation avant une séquence d'impulsions classique HSQC ^{15}N , les expériences de saturation croisée des iminos sont enregistrées en utilisant plusieurs temps de saturation (allant de 0,1 à 2 s). Ensuite, l'intensité normalisée des pics HSQC est tracée en fonction du temps de saturation.

III.2.2.4.3 Contraintes de distances intermoléculaires courtes (nOe)

La proximité spatiale des résidus proche de l'interface de contact entre la protéine et l'acide nucléique, peut mener à une interaction intermoléculaire dipôle-dipôle suffisamment large pour créer des nOe mesurables, ce qui crée une évidence directe de leur proximité spatiale. Il est nécessaire d'attribuer les résonances des deux partenaires afin de pouvoir identifier les nOe intermoléculaire, cependant, de forte superposition de résonances ainsi qu'un élargissement des raies peuvent apparaître associés à la taille du complexe et ou à l'échange chimique, ce qui complique l'attribution des nOe. Les progrès récents en RMN ont largement contribué à résoudre ce problème. Dans la plupart des études de complexes protéines acides nucléiques, la protéine est marquée uniformément en $^{13}\text{C}/^{15}\text{N}$ alors que l'acide nucléique reste non marqué. La protéine dans le complexe est alors attribué en utilisant une combinaison classique d'expériences tridimensionnelles (HNCA et HN(CO)CA, HNCO et HN(CA)CO, HNCACB et HN(CO)CACB), TOCSY et des expériences NOESY (3D: ^{15}N HSQC-NOESY, ^{15}N HSQC-TOCSY, HCCH-TOCSY, ^{13}C HSQC-NOESY). De plus, le développement d'expériences filtrées et éditées pour un isotope particulier a particulièrement facilité l'attribution de l'acide nucléique impliqué dans l'interaction. En particulier, il est possible de sélectionner les protons qui sont ou pas impliqués dans une liaison scalaire à un hétéroatome spécifique, ^{15}N ou ^{13}C . La détermination des contraintes de distance intermoléculaires demeure une tâche difficile et nécessite beaucoup de temps, mais elles permettent de définir l'interface des complexes protéine-ADN avec une précision élevées pour permettre des calculs de structure. Le rassemblement d'un nombre significatif de NOE intermoléculaires est certainement le moyen le plus efficace de définir l'interface avec une grande précision.

III.2.2.4.4 Contraintes de distances intermoléculaires longues (Paramagnetic Relaxation Enhancement ou PRE)

Lorsqu'une molécule comporte un (ou plusieurs) électrons non appariés, ces électrons vont interagir par couplage dipolaire avec les noyaux proches dans l'espace. Ces interactions dipolaires magnétiques causent l'effet PRE (Paramagnetic Relaxation Enhancement). Les espèces comportant des électrons non appariés sont en effet paramagnétiques, c'est à dire qu'elles possèdent une susceptibilité magnétique et qu'elles sont capables de s'orienter préférentiellement dans un champ magnétique proche dans l'espace. L'effet PRE mène à une augmentation importante de la relaxation du fait de la valeur importante du g-facteur, l'équivalent pour l'électron du rapport gyromagnétique⁵⁵³.

Du fait du moment magnétique de l'électron (lié à son g-facteur), l'effet PRE est détectable pour des distances (entre la paire électron et noyau impliqué dans le couplage) très importantes : jusqu'à 34 Å pour Mn^{++} ⁵⁵⁴. La RMN paramagnétique représente une source précieuse de longues distances qui peuvent compléter les contraintes NOE pour un complexe situé dans un régime d'échange lent. En effet, plus un noyau se trouve proche des électrons non appariés plus il revient vite à l'équilibre ce qui signifie que le phénomène de relaxation est alors accéléré. Cependant, quand la molécule d'intérêt ne contient pas de centre paramagnétique intrinsèque, des marqueurs extrinsèques doivent être incorporés à la molécule, généralement en utilisant des sondes paramagnétiques ou en générant des mutations ponctuelles de cystéines. Deux classes de sondes paramagnétiques sont utilisées pour la RMN PRE: (i) les radicaux nitroxides (stables); (ii) les chélateurs de métaux (tels que l'EDTA, le DTPA et les peptides se liant aux métaux) qui se lie aux ions paramagnétiques (Mn^{++}) avec une très forte affinité. Ces deux classes de sondes paramagnétiques peuvent être associées de manière covalente aux protéines ainsi qu'aux acides nucléiques utilisés pour l'observation de l'effet PRE intramolécule ou intermoléculaire ^{553, 555}. Cependant, on peut classer la RMN PRE en trois types : intramolécule, intermoléculaire ou provenant du solvant (Figure 23B).

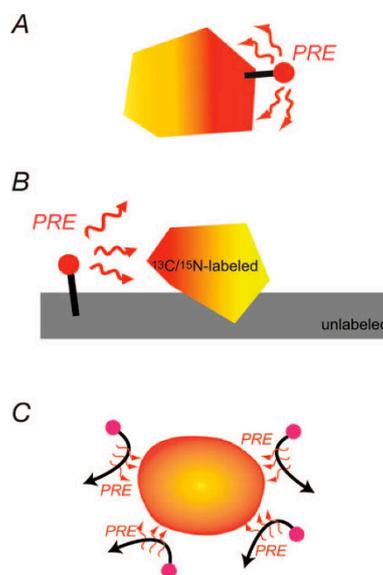


Figure 23B: Types de RMN PRE. (A) Intramolécule où L'effet PRE vient du groupe paramagnétique situé sur la même molécule ; (B) Intermoléculaire où l'effet PRE vient du groupe paramagnétique situé sur le partenaire d'interaction ; (C) Du solvant où l'effet PRE vient des collisions aléatoires entre une macromolécule et les molécules cosolutes paramagnétiques ⁵⁵³.

Dans le cas des complexes protéine-ADN, un agent paramagnétique est communément conjugué à l'ADN. Par exemple, l'utilisation de désoxythymidine dérivée d'EDTA contenant de l'ADN chélatée à Mn^{++} dans le complexe SRY / ADN a permis de

mesurer les effets PRE intermoléculaires et la polarité de la liaison SRY sur l'ADN à définir⁵⁵⁶. Plus récemment, l'incorporation de désoxy-4-thiouracile à l'extrémité de l'oligonucléotide qui réagit avec la sonde 3-(2-iodoacétamido)-proxyl a permis de marquer l'extrémités de l'ADN en 5' avec un agent paramagnétique⁵⁵⁷. L'effet PRE induit par le marquage paramagnétique mène à un élargissement des pics, dont la magnitude dépend de la distance entre le noyau et la sonde paramagnétique. En combinant différents sites de marquage, les données PRE peuvent être extraites de différentes positions du centre paramagnétique, ce qui permet l'extraction de contraintes à longues distances⁵⁵⁸. Enfin, plusieurs méthodes sont possibles pour analyser l'effet PRE : Pour détecter la relaxation électron-proton, on observe le changement d'intensité et la forme des lignes de résonances de la protéine dans le spectre heteronucleaire ; on peut mesurer le temps de relaxation T_2 des formes para- et diamagnétiques, la différence entre les deux représente la seule relaxation provenant de l'interaction entre les électrons et les noyaux.

$$T_2 = R_{2,para} - R_{2,dia} \quad (58)$$

Une autre méthode quantitative consiste à mesurer le ratio :

$$\frac{I_{para}}{I_{dia}} \quad (59)$$

Cependant, ce dernier n'est pas significatif physiquement car il ne dépend pas seulement du T_2 mais également du T_1 , des délais de transfert de cohérences ainsi que du type de traitement de données⁵⁵⁶.

III.2.3 Techniques de Biochimie et de Biologie moléculaire

III.2.3.1 Purification des oligonucléotides

Cette partie concerne les oligonucléotides destinés à l'étude par gel retard.

25 pmol d'oligonucléotides sont repris dans un volume équivalent du tampon de charge bleu urée (B/U) (7 M urée ; 0.05% bleu bromophénol ; 0.05% xylène cyanol), à volume égal. Ils ont été purifiés par électrophorèse sur gel dénaturant 15 % polyacrylamide/7.2 M urée, 1X TBE (90 mM Tris-Borate, 2 mM d'EDTA). À la fin de la migration, les oligonucléotides ont été révélés par *UV-shadowing*, les bandes du gel les contenant ont été découpées, puis l'extraction a été effectuée par élution passive (processus de diffusion), pendant la nuit à température ambiante, dans une solution contenant 0.3 M d'acétate de sodium, 0.1 % de SDS et 1 mM d'EDTA. Au jour J+1, l'éluat est récupéré puis centrifugé sur colonne spinX à 10000 rpm pendant 2 min dans la mini-centrifugeuse de paillasse BECKMAN MicrofugeR Lite. Par la suite, il y a une étape de précipitation avec 2.5 volumes d'éthanol absolu et une centrifugation à 14000 rpm pendant 1 h à 4°C dans la

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

III.2.3.2 Marquage des oligonucléotides en isotope radioactif

Cette partie concerne les ARNs destinés à l'étude par gel retard.

Les oligonucléotides ont été marqués en 5' au ^{32}P en utilisant la T4 polynucléotide kinase (New England Biolabs) et le $[\gamma\text{-}^{32}\text{P}]$ ATP (Perkin Elmer). La réaction a été effectuée dans un volume final de 8 µl. Vingt-cinq pmol d'oligonucléotide dans 3.7 µl d'eau milli-QTM, 0.8 µl de tampon réactionnel (10X) (concentrations finales : 70 mM Tris-HCl pH 7.6 ; 10 mM MgCl_2 et 5 mM DTT), 2.5 µl de $[\gamma\text{-}^{32}\text{P}]$ ATP (25 µCi, 16.7 pmol) et 1 µl de T4 polynucléotide kinase (10 U/µl) constituent le mélange réactionnel. Après une incubation à 37°C pendant 45 min, l'oligodésoxyribonucléotide marqué a été purifié sur gel de polyacrylamide (19:1)/7.2 M urée (15%). Pour vérifier la pureté de l'oligonucléotide marqué en 5', une aliquote correspondant à l'équivalent de 20000 cpm a été analysée par électrophorèse sur gel dénaturant 8% polyacrylamide (19:1)/7.2 M urée dans le tampon d'électrophorèse 1X TBE (90 mM Tris-Borate, 2 mM d'EDTA). Les oligonucléotides ont été révélés par autoradiographie, les bandes du gel les contenant ont été découpées, puis l'extraction a été effectuée par élution passive (processus de diffusion), pendant la nuit à température ambiante, dans une solution contenant 0.3 M d'acétate de sodium, 0.1 % de SDS et 1 mM d'EDTA. Au jour J+1, l'éluat est récupéré puis centrifugé sur colonne spinX à 10000 rpm pendant 2 min dans la mini-centrifugeuse de paillasse. S'en suit une étape de précipitation avec 2.5 volumes d'éthanol absolu et une centrifugation à 14000 rpm pendant 1h à 4°C dans la mini-centrifugeuse de paillasse. Le surnageant est ensuite retiré et un lavage du culot avec 200 µl d'éthanol 70% est effectué pendant 5 min. Les oligonucléotides purifiés ont ensuite été séchés 5 min au speed-vac et remis en solution dans 20 µl d'eau milli-QTM.

III.2.3.3 Marquage des oligonucléotides par des agents paramagnétiques

Cette partie concerne les oligonucléotides destinés à l'étude par RMN PRE.

Le choix du site de marquage est crucial vu qu'il doit être en dehors du site de fixation de la protéine afin de ne pas gêner l'interaction. On a donc choisi de marquer nos oligonucléotides sur la dernière base à son extrémité 5' ou 3'. Dans le cas des ADNs il

contenait en 5' ou 3' une 4-thio-dT (ou S4dT), alors que pour les ARNs, c'est une 4-thio-U (ou S4U), c'est à dire une base qui contient un atome de soufre en position 4 à la place de l'oxygène, c'est par l'intermédiaire de cet atome de soufre que va se lier l'agent paramagnétique. Le couplage se fait donc par réaction chimique d'une sonde Proxyl sur l'oligonucléotide ADN ou ARN modifié. Le protocole du marquage se déroule comme tel : On réalise un mélange pour obtenir au final 0.1 mM d'acide nucléique marqué. On ajoute 100 µl d'ADN ou d'ARN à 1 mM, 200 µl de tampon KPO₄ à 500 mM et à pH 8, 100 µl de proxyl à 100 mM resuspendu à 100% dans du MeTOH, et on complète avec 600 µl d'H₂O pour obtenir un volume final de 1 mL. Cette préparation est protégée de la lumière avec de l'aluminium ou un tube opaque car le proxyl est photosensible et on laisse sous agitation horizontale pendant 24 h à 25°C. Le lendemain du marquage on suit la réaction par dosage UV, si le pic à 335 nm de la base d_{thio} disparaît cela signifie qu'il y a un transfert total du proxyl sur l'oligonucléotide. S'ensuit une étape de purification des oligonucléotides sur colonne d'exclusion NAP-10 en deux fois. La quantité récupérée sera lyophilisée puis resuspendu dans le tampon RMN, avec une mesure de la DO avant de l'utiliser.

III.2.3.4 Test d'hybridation par gel retard

L'hybridation des acides nucléiques est un principe fondamental en biologie moléculaire, il est fondé sur les propriétés d'appariement des bases complémentaires des acides nucléiques. Deux couples ADN/ARN ont été testés pour l'hybridation. Quatre protocoles d'hybridation entre l'ARN marqué au ³²P en 5' et l'ADN ont été testés en absence et en présence de NC(1-55) de l'isolat NL4.3, dans un volume final de 10 µl. Dans chaque protocole, 1 pmol d'ADN a été incubée avec 0.1 pmol d'ARN marquée au ³²P en 5' (20000 cpm). Les échantillons ont été analysés par électrophorèse sur gel polyacrylamide natif 14% (37.5 :1) dans le tampon TBE (1X) thermostaté à 20°C. Le premier protocole d'hybridation a été réalisé sans NC, le deuxième et le troisième en présence de la protéine avec une concentration de 1.5 pmol, 3.3 pmol et 6.6 pmol respectivement.

1^{ère} étape : Dénaturation/Renaturation : L'ADN et l'ARN dans 3.6 µl H₂O et dans deux tubes distincts ont été dénaturés pendant 2 min à 90°C et placés ensuite 2 min dans la glace avant d'ajouter 0.9 µl du tampon d'hybridation 5X (concentrations finales : 25 mM Tris HCl pH 7.5 ; 30 mM NaCl ; 0.2 mM MgCl₂). Ensuite, les échantillons ont été incubés pendant 10 min à 20°C pour permettre la renaturation.

2^{ème} étape : Hybridation des ADN et ARN renaturés : L'ADN et l'ARN renaturés ont été rassemblés dans un même tube et 1 µl du tampon 1X ou 1 µl de NC a été ajouté pour avoir en final 10 µl. Ensuite l'essai a été placé à 20°C pendant 5 min, 15 min, 30 min, 1 h,

2 h, 3 ou 4 h. La réaction a été arrêtée en ajoutant 1 µl de SDS à 2.5% et 1 µl de protéinase K à 3 mg/ml. Après 30 min d'incubation à 20°C, 4 µl du tampon de charge bleu glycérol (BG) (50% glycérol ; 0.05% bleu de bromophénol ; 0.05% xylène cyanol) ont été ajoutés.

III.2.3.5 Production et purification de la protéine de nucléocapside du VIH-1 mature NCp7 marqué en ¹⁵N

III.2.3.5.1 Surexpression de la protéine NCp7

Des souches d'expression bactérienne d'*E. coli* (BL21(DE3)pLysE) ont été transformées par la technique d'électroporation avec un vecteur pET3a (Novagen) portant le gène d'intérêt NCp7. L'expression protéique est sous contrôle du promoteur T7 qui est induit par l'IPTG. Les cellules sont ensuite étalées sur boîte d'agarose. La sélection des cellules se fait par l'ajout d'antibiotique : le chloramphénicol (34 µg/ml) dont le gène de résistance est porté par le plasmide pLysE et l'ampicilline (50 µg/ml) dont le gène de résistance est porté par le plasmide pET3a. Les cellules provenant du clone sélectionné sont cultivées 20 h à 37°C en milieu minimum M9 enrichi par l'isotope azote 15 (¹⁵N) en rajoutant du ¹⁵NH₄CL, supplémenté de 35 µg/ml chloramphénicol et de 50 µg/ml l'ampicilline. Cette préculture servira à ensemercer 1 litre de milieu minimum M9 enrichi en ¹⁵N en présence de chloramphénicol (35 µg/ml) et d'ampicilline (50 µg/ml).

La culture s'effectue à 37°C sous agitation jusqu'à ce qu'elle atteigne une absorbance à 600 nm de 0.5 U.A. L'expression de la protéine est alors induite par ajout d'IPTG à une concentration finale de 0.5 mM. Après 20 h d'induction à 30°C, la culture est centrifugée à 7000 rpm pendant 30 minutes (rotor JLA 8-1000, centrifugeuse Beckman-Coulter Aventi J-26 XP). Les culots bactériens sont repris dans un tampon de lyse (Tris-HCl 50 mM pH 8, glycérol 10%, NaCl 0.1 M, ZnCl₂ 0.1 mM, DTT 5 mM, EDTA 2 mM, PMSF 10 mM, pepstatine A 5 µg/ml). Les cellules sont ensuite lysées à 4°C par sonication pendant 10 minutes à 40 watts (cycles de 20 secondes (sec) entrecoupées de 30 sec de pause). Le lysat bactérien obtenu est ensuite centrifugé à 45000 rpm pendant 25 minutes (min) à 4°C (rotor JA20). Le surnageant est récupéré et les acides nucléiques sont précipités par l'ajout du polyéthylène imine (PEI) à 1% final, goutte à goutte et sous agitation pendant 1 h. Les acides nucléiques précipités et le restant des débris cellulaires sont éliminés par centrifugation à 45000 rpm (rotor JA20) pendant 25 min à 4°C. Le surnageant est filtré sur une membrane de 0.2 µm.

III.2.3.5.2 Purification de la protéine d'intérêt NCp7

Chromatographie échangeuses d'ions

Cette première étape de purification est effectuée à l'aide de deux colonnes montées en série, une Q sépharose (HiLoad 26-40, High Performance HP, GE Healthcare) et une SP sépharose (HiLoad 26-10 (HP), GE Healthcare) à 4°C, sur une FPLC ÄKTApurifier.

La colonne est préalablement équilibrée avec le tampon A (Tris HCl pH 8 50 mM, glycérol 10%, NaCl 0.1 M, ZnCl₂ 0.1 mM, β-mercaptoéthanol 10 mM), puis avec le tampon B à 100% (Tris HCl pH 8 50 mM, glycérol 10%, NaCl 1 M, ZnCl₂ 0.1 mM, β-mercaptoéthanol 10 mM). La colonne est remplacée dans du tampon A puis le surnageant est injecté sur les colonnes.

Le PI de la NCp7 est de 9.9 et nous travaillons ici à pH 8, la protéine est chargée globalement positivement, la protéine se fixe donc sur la colonne SP sepharose. Après le passage de 200 ml de tampon A, la colonne Q-sépharose est détachée puis un gradient de 400 ml entre 20 et 80% de tampon B est effectué sur la SP-sépharose. La protéine NCp7 est éluée à 25% de tampon B.

Tamis moléculaire :

La deuxième étape de purification est effectuée sur un tamis moléculaire (HiLoad 26-60, Superdex 75, Prep Grade, GE Healthcare) préalablement équilibrée avec 300 mL de tampon C (Tris HCl pH 7 50 mM, glycérol 10%, NaCl 0.1 mM, ZnCl₂ 0.1 mM, β-mercaptoéthanol 10 mM), puis l'échantillon récolté de la précédente étape est injecté. La protéine est éluée à 210 ml de tampon C.

Les fractions sont récoltées, concentrées puis dialysées contre le tampon RMN (acétate de sodium deutéré pH 6, 25 mM, NaCl 25 mM, ZnCl₂ 0.1 mM, β-mercaptoéthanol 0.1 mM). La concentration de la protéine est mesurée par son absorbance à 280 nm ($\epsilon = 6410 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$).

IV Résultats

IV.1 Etude du mécanisme d'hybridation des séquences TAR et cTAR impliquées dans le premier transfert de brin du VIH-1 en présence de la NCp7

IV.1.1 Contexte

La protéine de nucléocapside du VIH-1 (NCp7) est impliquée dans le premier et le second transfert de brin, qui se produisent au cours de la transcription inverse, cela, grâce à son activité de protéine chaperon. Le premier transfert de brin consiste sur le plan moléculaire en l'hybridation de séquences complémentaires ARN et ADN. Dans le détail, ce processus passe par deux étapes qui ne sont pas successives mais probablement très imbriquées. La première est la déstabilisation des séquences d'acides nucléiques et la deuxième, l'hybridation des deux partenaires complémentaires. Grâce à son activité de protéine chaperon des acides nucléiques, la NCp7 possède la capacité de faciliter ces deux étapes. Il a été montré que les séquences nucléotidiques impliquées correspondaient aux régions R situées sur l'ADNss (Accepteur) ainsi que sur l'ARNv (Donneur) et que la NCp7 accélérât leur cinétique d'hybridation^{378, 398, 417, 423}. Cependant, il apparaît que les structures qui sont les plus importantes pour le transfert dans la région R sont les séquences TAR ARN et cTAR ADN, cela explique pourquoi la synthèse de la totalité de l'ADNss est nécessaire pour que le transfert de brin se produise, indiquant que les séquences polyA paraissent moins importantes¹⁷⁵. TAR et cTAR forment des tiges boucles hautement structurées dont la déstabilisation constitue une étape limitante pour permettre leur hybridation³⁵⁶. De plus la relation entre la structure et la stabilité des acides nucléique est importante pour l'activité de protéine chaperon de la NCp7.

Des études antérieures, conduites avec la NC(12-55), donc dépourvue de sa partie N-Terminale, ont démontré que la NCp7 déstabilise mieux la séquence ADN cTAR que TAR ARN⁴⁰⁴. Ils ont démontré que la protéine active permettait l'ouverture transitoire des paires de bases situées sur les extrémités stables de la tige, avec un taux de déstabilisation de 7 à 8 pb par oligonucléotide pour cTAR ADN et 1 pb par oligonucléotide pour TAR ARN, avec un ratio NC/nt de 7.5. Ces résultats ont souligné que TAR ARN possède une plus grande stabilité structurale que cTAR ADN et que ce sont les deux doigts de zincs qui sont responsables de l'activité déstabilisatrice de la NC⁴⁰⁴. Cependant, il a également été supposé que la conservation d'une structuration suffisante de TAR ARN, c'est à dire partiellement déstabilisée, était importante pour un transfert de brin efficace²²⁰. D'autres études laissent apparaître que c'est la stabilité thermodynamique

locale et non globale des oligonucléotides qui est la plus importante et que celle-ci constitue le déterminant principal pour l'activité de protéine chaperon de la NCp7⁴⁰²⁻⁴⁰³. Plusieurs études ont également analysé les cinétiques d'hybridation de TAR et cTAR, en présence et en absence de la NCp7, dans le but de déceler les voies d'hybridation empruntées^{406, 424, 426}. Les études avec la forme tronquée de la protéine, NC(12-55) ont démontré que le processus d'hybridation se produit avec une cinétique de second ordre à deux composantes, le processus est plus rapide d'au moins deux ordres de grandeurs en présence de la protéine⁴⁰⁶. Les deux composantes cinétiques ont été assignées à deux voies différentes : (i) une phase rapide qui ne permet pas la formation d'un duplex étendu (ED) mais qui est associée à une hybridation partielle des bases terminales de cTAR et TAR. (ii) une phase plus lente qui permet la formation du ED, mais qui est limitée cinétiquement par la nucléation des résidus situés majoritairement sur les segments centraux double-brin de TAR et cTAR. Pour ces deux phases, le mécanisme de formation du complexe passerait par un intermédiaire de type "zipper" dans lequel les extrémités des molécules ADN et ARN s'hybrideraient^{424-426, 559}. Un mécanisme alternatif a été proposé en présence de la NC(12-55) qui constitue une voie d'hybridation par les boucles « kissing complex », mais qui reste mineur dans les conditions expérimentales choisies⁴⁰⁶. Ce schéma est conservé en présence de la NC(1-55), c'est-à-dire que l'hybridation de TAR et cTAR avec la NCp7 dans des conditions non agrégative, suit une voie de type "zipper", néanmoins ce chemin n'est majoritaire que quand une forte concentration de protéine est utilisée (NC:nt 1:5)^{397, 420, 424-425}.

Par ailleurs, des expériences avec des séquences tronquées de TAR et cTAR (27 nt), ont montré des différences comparées aux séquences complètes (59 nt)⁴²⁶. En absence de NCp7, les mini TAR et cTAR s'hybrident via un mécanisme "kissing complex" et les molécules TAR et cTAR entières sont 10 fois plus lentes à s'hybrider que les séquences courtes en absence de NCp7, avec la formation rapide d'un complexe intermédiaire peu stable, qui interagit par les boucles apicales, suivie par une conversion lente vers le complexe étendu⁴²⁴⁻⁴²⁶. En présence de NCp7, celle-ci facilite l'hybridation de TAR et cTAR entières d'un facteur 10^5 fois plus important qu'en absence de la protéine, stabilisant l'intermédiaire bimoléculaire (10^4 fois plus) et en promouvant l'apparition du duplex étendu (10 fois plus). En résumé, l'hybridation de TAR et cTAR en absence de NCp7 semble se faire par une nucléation à travers les boucles. En présence de NCp7, celle-ci accélère fortement leur cinétique d'hybridation et permet la stabilisation du complexe intermédiaire tout en accélérant l'apparition du complexe étendu⁴²⁴. Ce schéma réactionnel n'a pas été observé avec la NC(12-55)⁴⁰⁶, ce qui impliquerait un rôle de la

partie agrégative dans cette étape. Dans les conditions de forte concentration de NCp7, les travaux de Vo et al., 2006 montrent un basculement du mécanisme d'hybridation vers un mécanisme "zipper", ce qui suggère qu'une plus forte déstabilisation permet l'apparition de ce mécanisme.

Dans notre travail, nous avons voulu tester l'impact du positionnement des boucles internes dans TAR et cTAR; comme il est expliqué dans l'introduction (Figure 20A), le fait que la NC pourrait s'orienter préférentiellement le long des chaînes d'acides nucléiques et le fait que l'un des deux doigt de zinc (le ZF1) est plus important pour la déstabilisation des structures secondaires (Figure 24A) suggèrent un mécanisme dans lequel la position des boucles internes (supposées être importantes pour la fixation des molécules de NC) déterminerait les séquences double brin qui seraient déstabilisées en priorité. Dans un tel modèle l'inversion de la position des boucles internes devrait modifier les cinétiques d'hybridation. Pour réaliser ce travail, nous avons utilisé la technique de FRET à l'aide d'oligonucléotides synthétiques, où l'ADN cTAR est doublement marqué à ces extrémités 5' et 3' par un couple de sonde fluorescentes (Tetramethylrhodamine (TMR) et Fluorescéine (F) respectivement). Nous avons utilisé des séquences sauvages tronquées (cTAR0-TAR0 de 29 nt) ainsi que des couples mutants (cTAR1-TAR1 de 29 nt) où la position des boucles internes a été inversée (Figure 2B). À partir de ces deux couples, d'autres mutants ont été réalisés (loop et stem) dans lesquels les modifications sont les suivantes : (i) deux bases au niveau de la boucle apicale ou (ii) deux bases au niveau des extrémités 5'3' afin d'empêcher l'hybridation par la voie "kissing" et "zipper" respectivement (Figure 2B). Nous avons également étudié l'hybridation de TAR et cTAR en absence et en présence de NCp7 par gels retard. En outre, par des expériences de RMN, nous avons étudié la déstabilisation de la structure secondaire de TAR ARN par la NCp7, à faible ratio de protéine par nucléotides (1:59 nt), afin de déterminer quels sont les premiers sites de nucléation de la déstabilisation.

IV.1.2 Résultats

IV.1.2.1 Analyse de l'hybridation des séquences TAR et cTAR par FRET

L'hybridation est suivie en utilisant un oligonucléotide cTAR doublement marqué par un couple (TMR-F). Dans nos conditions expérimentales (20°C, tampon Acétate I=0.05, pH=7.5), cTAR ADN forme une tige boucle. Les deux chromophores situés aux extrémités 5' et 3' sont alors proches l'un de l'autre et induisent un transfert d'énergie de F vers TMR. Dans le cas du couple TMR, F que nous utilisons, il a été de plus montré qu'un couplage de type excitonique (couplage fort) existe entre les deux chromophores, lorsque ces

derniers sont très proches l'un de l'autre^{404, 559}. Ce couplage est différent du couplage de type FRET, qui est un couplage faible et qui ne peut se produire que quand le couplage fort est perdu, dû à l'ouverture de la tige basse des oligonucléotides. Suite à l'hybridation de l'oligonucléotide cTAR ADN avec TAR ARN, les deux chromophores vont se trouver éloignés l'un de l'autre, le couplage FRET est perdu, ce qui va se traduire par un accroissement important de l'émission de la fluorescence de F.

Nos expériences ont été réalisées avec une concentration d'ADN fixe de 10 ou 50 nM, une concentration d'ARN de 600 nM et différents ratios de protéines (1/5, 1/8 et 1/10 nt). Finalement, nous avons choisi de poursuivre nos travaux avec une concentration fixe d'ADN de 50 nM, vu que les complexes s'avéraient être plus stables et les expériences plus reproductibles dans de telles conditions. Nous avons travaillé dans des conditions de pseudo-premier ordre dans lesquelles l'un des composants est en large excès, ici l'ARN (600 nM). Ces conditions simplifient l'extraction des vitesses d'hybridation du complexe, qui a été effectuée avec le logiciel GraphPrism en utilisant l'équation « non-linear regression (curve fit) two-phase association », ce qui signifie que deux exponentielles sont utilisées pour effectuer les ajustements des courbes expérimentales. Le profil des courbes cinétiques obtenues et les ajustements utilisant des profils à doubles exponentielles ("fits"), montrent que l'hybridation des séquences se déroule en deux phases qui possèdent des constantes de temps significativement différentes. La première phase est rapide, au cours de celle-ci, il se forme un complexe initial partiellement hybridé, cependant la deuxième phase est plus lente, au cours de laquelle le complexe est converti en un duplex étendu parfaitement hybridé (Figure 6C).

Les constantes de vitesses k_1 et k_{-1} sont liées à la formation du complexe initial, alors que la constante k_2 est associée à l'apparition du duplex étendu, si l'on suit le schéma réactionnel proposé par Vo et al. 2006, 2009. En comparant les valeurs de la vitesse (k_2) entre les différents complexes, on détermine si l'hybridation se fait de manière plus rapide.

IV.1.2.1.1 Spectre d'émission de l'ADN

Les courbes obtenues montrent que l'émission de TMR (second pic) est plus intense que celle de F (premier pic) (en cas d'excitation de la fluorescéine à 480 nm), ce qui suggère que le couplage excitonique n'est pas présent car la présence de celui-ci se manifeste par une fluorescence de même niveau pour les deux partenaires F et TMR.

On peut suggérer que la paire de base terminale de type AT est largement ouverte dans nos conditions, empêchant le couplage excitonique. Par contre il semble que le transfert d'énergie de type FRET (par couplage faible) lorsque l'excitation a lieu à 580 nm

(Fluorescéine) entre F et TMR a bien lieu. L'ADN est bien sous une forme repliée en tige-boucle avec ses deux extrémités marquées par F et TMR, proches spatialement, même si la dernière paire de base AT est probablement ouverte. Cependant, en comparant les spectres d'émissions de cTAR0 à différentes températures, on observe que plus la température augmente, plus l'émission de la TMR s'affaiblit. Cela indique que sous l'effet de la température, la structure secondaire est déstabilisée, les fluorochromes s'éloignent de plus en plus l'un de l'autre, ce qui diminue le transfert d'énergie, avec comme résultat une augmentation de l'intensité de la fluorescéine (Figure 1C). Par ailleurs, on remarque un déplacement de la courbe de TMR, cette caractéristique est en fait liée à l'effet de la température sur la sonde fluorescente.

Par la suite, nous avons décidé de travailler à une température de 20°C afin de s'assurer de la conformation en tige boucle de cTAR0. Nous avons également vérifié l'influence de la température sur les spectres d'émissions des autres ADN et nous obtenons le même résultat qu'avec cTAR0 (résultats non présentés). Cependant, dans les mêmes conditions de pH, de température et de concentration, les profils d'émission de fluorescence des différents oligonucléotides ne sont pas identiques. Cela est probablement lié à de légères différences conformationnelles. D'après le logiciel de prédiction de structure secondaire mFold, chaque oligonucléotide présente une structure secondaire spécifique.

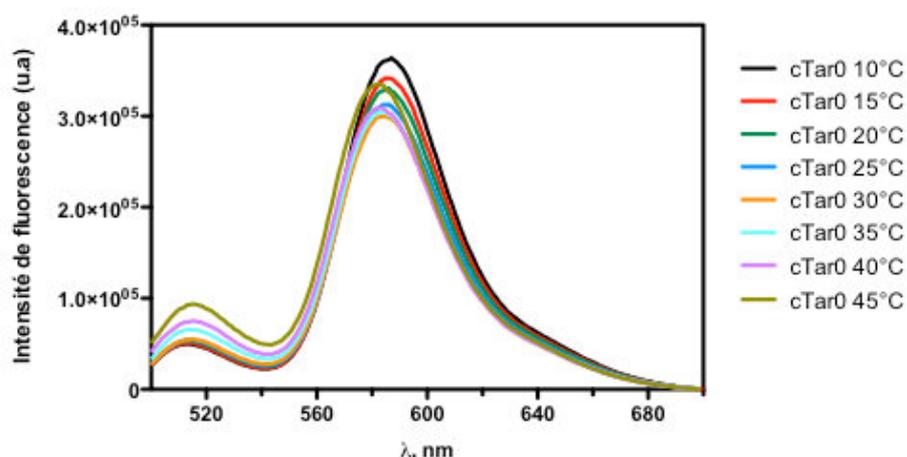


Figure 1C: Spectre d'émission de cTAR0 à différentes températures dans du tampon Acétate $I=0.05$ et $pH=7.5$ avec $\lambda_{ex}=480$.

IV.1.2.1.2 Cinétique d'hybridation TAR et cTAR sauvages en présence de NCp7 à différentes températures

Nous avons tout d'abord, étudié l'hybridation TAR0 et cTAR0 avec un ratio 1/5 de NCp7 à différentes températures (10°C, 15°C, 20°C, 25°C, 30°C, 35°C, 40°C et 45°C) (Figure 2C) et on constate que lorsque la température augmente, la constante de vitesse associée à

l'apparition du duplex étendu k_2 augmente également, ce qui traduit une vitesse d'hybridation des séquences TAR et cTAR plus rapide (Tableau 1C).

Tableau 1C. Constantes cinétiques de l'hybridation du complexe sauvage (TAR0-cTAR0) à différentes températures et un ratio de 1/5 de NC/nt.

| T° | 10°C | 20°C | 25°C | 30°C | 35°C | 40°C | 45°C |
|---------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| k_1 ($M^{-1} s^{-1}$) | 2.5×10^5 | 1.9×10^5 | 2.6×10^5 | 3.9×10^5 | 2.5×10^5 | 2.4×10^5 | 3.1×10^5 |
| k_{-1} (s^{-1}) | 6.7×10^{-2} | 1.2×10^{-2} | 1.7×10^{-2} | 4.7×10^{-2} | 3.7×10^{-2} | 2.9×10^{-2} | 1.9×10^{-2} |
| k_2 (s) | 6.1×10^{-3} | 6.3×10^{-3} | 9.5×10^{-4} | 1.8×10^{-3} | 7×10^{-3} | 2.6×10^{-2} | 4.8×10^{-2} |
| K_d (M) | 2.7×10^{-7} | 5.9×10^{-8} | 6.7×10^{-8} | 1.2×10^{-7} | 1.4×10^{-7} | 1.3×10^{-7} | 6.4×10^{-8} |

Cette vitesse d'hybridation plus rapide, due à l'augmentation de la température, peut s'expliquer par la facilitation de la rupture des liaisons hydrogène entre les bases nucléiques, ce qui permet à la protéine de déstabiliser plus facilement les structures secondaires des acides nucléiques.

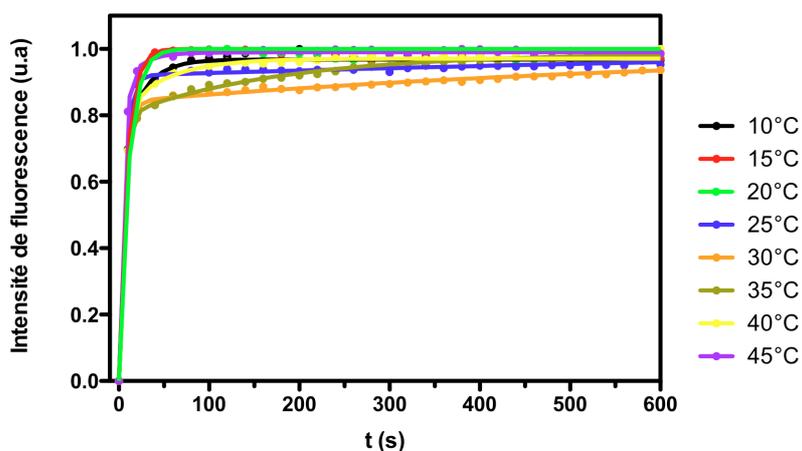


Figure 2C: Influence de la température sur l'hybridation TAR0-cTAR0. Courbes cinétiques normalisées de l'hybridation de cTAR0 (50 nM) et TAR0 (600 nM) en présence de NCp7 (ratio 1/5), dans du tampon Acétate ($I=0.05$ et $pH=7.5$) et à différentes températures, avec $\lambda_{ex}=480$ nm et $\lambda_{em}=516$ nm.

IV.1.2.1.3 Cinétique d'hybridation TAR et cTAR sauvages et mutantes en présence de différentes concentrations de NCp7 à 20°C

Afin de déterminer l'influence de la concentration en protéine sur l'efficacité de l'hybridation, nous avons réalisé des cinétiques en modifiant le rapport de NCp7 par nucléotide (1/5, 1/8, 1/10).

Tableau 2C. Constantes cinétiques de l'hybridation des complexes sauvage (TAR0-cTAR0) et mutant TAR1-cTAR1) à différents ratios de NC/nt et à 20°C.

| Oligonucléotide | cTAR0/TAR0 | | | cTAR1/TAR1 | | |
|---------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | 1/5 | 1/8 | 1/10 | 1/5 | 1/8 | 1/10 |
| k_1 ($M^{-1} s^{-1}$) | 1.9×10^5 | 1.2×10^5 | 1.2×10^5 | 2.7×10^5 | 1.9×10^5 | 9.4×10^5 |
| k_{-1} (s^{-1}) | 1.2×10^{-2} | 1.6×10^{-2} | 1.7×10^{-2} | 1.7×10^{-2} | 1.1×10^{-2} | 1×10^{-2} |
| k_2 (s) | 6.3×10^{-3} | 5.3×10^{-3} | 2.6×10^{-3} | 3.9×10^{-3} | 2.9×10^{-3} | 3.4×10^{-3} |
| K_d (M) | 5.9×10^{-8} | 1.4×10^{-7} | 1.4×10^{-7} | 6.3×10^{-8} | 5.9×10^{-8} | 1.1×10^{-7} |

D'après le profil cinétique et les constantes du tableau ci-dessus (Tableau 2C), on constate que lorsque la concentration en NCp7 augmente, k_2 la constante de vitesse associée à l'apparition du duplex étendu augmente également, ce qui se traduit par une vitesse d'hybridation des séquences TAR et cTAR plus importante. On en déduit que dans des conditions saturantes de NCp7, l'hybridation se fait de plus en plus rapidement ce qui indique une action déstabilisatrice plus efficace. En comparant les deux couples sauvages et mutants (Figure 3C), on observe que l'hybridation du couple sauvage s'avère être plus rapide.

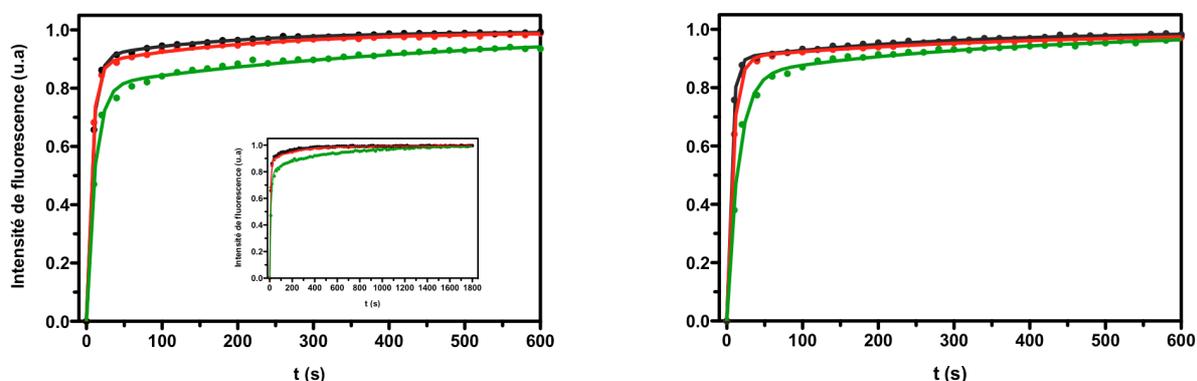


Figure 3C: Influence de la concentration en NCp7 et de la position des boucles internes des acides nucléiques sur l'hybridation TAR-cTAR. Courbes cinétiques normalisées de l'hybridation du couple sauvage (à gauche) et mutant (à droite) en présence de NCp7 (à un ratio de 1/5 (noir), 1/8 (rouge) et 1/10 (vert)) dans du tampon Acétate ($I=0.05$ et $pH=7.5$) et à $20^\circ C$, avec $\lambda_{exc}=480$ nm et $\lambda_{em}=516$ nm. L'insert dans la figure de gauche inclut les courbes cinétiques du couple sauvage jusqu'à $t= 30$ minutes, on arrive à un plateau dès 600 secondes pour tous les complexes.

IV.1.2.1.4 Cinétique d'hybridation des séquences sauvages "loop0" et "stem0" en présence de NCp7

D'après le profil cinétique et les constantes du tableau (3C), on observe que l'hybridation du couple "loop0" s'avère être plus rapide. Rappelons qu'avec ces mutants, l'hybridation

par les boucles est défavorisée, alors que le couple "stem0" s'avère posséder des cinétiques plus lentes par rapport au couple TAR0-cTAR0. Les courbes cinétiques de ces deux couples sont représentées Figure (4C).

Tableau 3C. Constantes cinétiques de l'hybridation des complexes sauvages (TAR0-cTAR0 loop) et mutant TAR0-cTAR0 stem) à différents ratios de NC/nt et à 20°C.

| Oligonucléotide | cTAR0/TAR0loop | | | cTAR0/TAR0stem | | |
|---------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Ratio | 1/5 | 1/8 | 1/10 | 1/5 | 1/8 | 1/10 |
| k_1 ($M^{-1} s^{-1}$) | 1.8×10^5 | 1.3×10^5 | 1.3×10^5 | 1.7×10^5 | 1.4×10^5 | 1.9×10^5 |
| k_{-1} (s^{-1}) | 1.2×10^{-2} | 1.1×10^{-2} | 1.4×10^{-2} | 6.4×10^{-3} | 7.9×10^{-3} | 1.7×10^{-2} |
| k_2 (s) | 7.2×10^{-3} | 4.5×10^{-3} | 3.5×10^{-3} | 4.3×10^{-3} | 2.6×10^{-3} | 2.2×10^{-3} |
| Kd (M) | 6.4×10^{-8} | 8.6×10^{-8} | 1.2×10^{-7} | 3.7×10^{-8} | 5.7×10^{-8} | 9.3×10^{-8} |

Le fait que le k_2 de cTAR0 est plus grand que pour stem0, suggère que l'hybridation pour le couple sauvage se fait par la voie "zipper" et que l'inhibition de cette voie d'hybridation ralentit la vitesse de formation du duplex étendu, ceci même à forte concentration de protéine. Par contre la modification des boucles apicales entraîne une accélération de l'hybridation. Ces résultats pourraient s'expliquer par le fait que les deux chemins d'hybridation possibles sont présents avec les séquences sauvages, mais que le chemin "zipper" est majoritaire. Lorsque les mutations précédentes défavorisent le passage par la voie "kissing complex" (c'est à dire les mutants "loop"), de ce fait la voie "zipper" se trouve favorisée ce qui mène à une hybridation plus rapide. Par contre avec les mutants "stem" qui dévaforisent la formation d'intermédiaires "zipper", le chemin majoritaire devient donc probablement le chemin "kissing", dont la conversion vers le duplex étendu se révèle plus lente avec ces molécules.

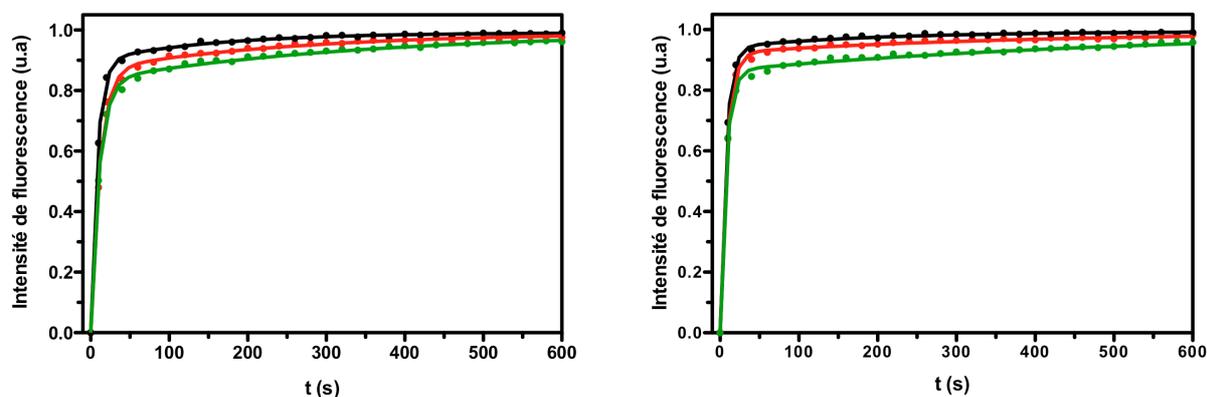


Figure 4C: Influence de la mutation de la boucle apicale ou des extrémités 5' ou 3' de l'ADN

sur l'hybridation TAR0-cTAR0 loop et stem en présence de NCp7. Courbes cinétiques normalisées de l'hybridation des couples sauvages loop0 (à gauche) et stem0 (à droite) en présence de NCp7 (à différents ratios), dans du tampon Acétate (I=0.05 et pH=7.5) et à 20°C, avec $\lambda_{ex}=480$ nm et $\lambda_{em}=516$ nm.

IV.1.2.1.5 Cinétique d'hybridation des séquences mutantes "loop1" et "stem1" en présence de NCp7

Contrairement aux couples sauvages, on observe que l'hybridation du couple "stem1" s'avère être plus rapide, rappelons que dans ces molécules l'hybridation par les extrémités 5' et 3' est défavorisée due aux mutations. Par contre, la constante de vitesse k_2 du couple mutant TAR1-cTAR1 est plus grande que pour le couple mutant "loop1" (Tableau 4C). Cette situation est donc exactement l'inverse de celle du couple précédent cTAR0-TAR0. Les courbes cinétiques de ces deux couples sont représentées Figure (5C).

Tableau 4C. Constantes cinétiques de l'hybridation des complexes mutants (TAR1-cTAR1 loop) et mutant TAR1-cTAR1 stem) à différents ratios de NC/nt et à 20°C.

| Oligonucleotide | cTAR1/TAR1loop | | | cTAR1/TAR1stem | | |
|---------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | 1/5 | 1/8 | 1/10 | 1/5 | 1/8 | 1/10 |
| k_1 ($M^{-1} s^{-1}$) | 12.1×10^5 | 1×10^5 | 1.1×10^5 | 1.6×10^5 | 1.4×10^5 | 1.5×10^5 |
| k_{-1} (s^{-1}) | 4.6×10^{-2} | 7.6×10^{-3} | 8.8×10^{-3} | 8.4×10^{-3} | 6.7×10^{-3} | 1.4×10^{-2} |
| k_2 (s) | 2.2×10^{-3} | 2.1×10^{-3} | 1.9×10^{-3} | 5.4×10^{-3} | 2.8×10^{-3} | 2.5×10^{-3} |
| Kd (M) | 3.8×10^{-8} | 7.4×10^{-8} | 8.4×10^{-8} | 5.4×10^{-8} | 4.9×10^{-8} | 9×10^{-8} |

Ces résultats suggèrent que l'hybridation majoritaire pour le couple mutant (TAR1-cTAR1 dans lequel les boucles internes ont été inversées), se fait par une voie de type "kissing complex", même si comme pour le couple précédent, les deux chemins "zipper" et "kissing" sont probablement simultanément présents, la perturbation de cette voie avec les mutants "loop1" et l'hybridation favorisée quand on a les mutants "stem1" sont cohérentes avec cette hypothèse et indiquent donc que les chemins d'hybridations sont fortement perturbés avec l'inversion de la position des boucles internes.

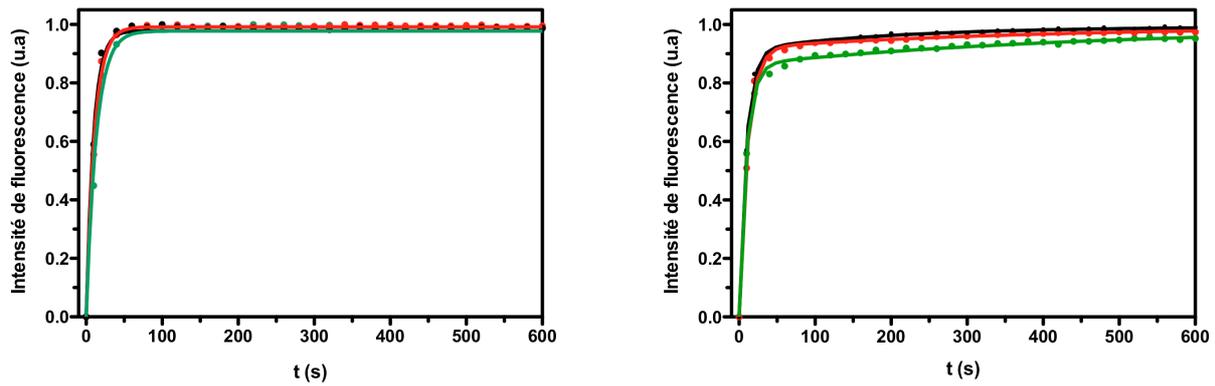


Figure 5C: Influence de la mutation de la boucle apicale ou des extrémités 5' ou 3' de l'ADN sur l'hybridation TAR1-cTAR1 loop et stem en présence de NCp7. Courbes cinétiques normalisées de l'hybridation des couples mutants loop1 (à gauche) et stem1 (à droite) en présence de NCp7 (à un ratio de 1/5 (noir), 1/8 (rouge) et 1/10 (vert)), dans du tampon Acétate ($I=0.05$ et $pH=7.5$) et à $20^{\circ}C$, avec $\lambda_{ex}=480$ nm et $\lambda_{em}=516$ nm.

En conclusion de ces expériences nous proposons sous forme schématique une explication possible des résultats obtenus. Cette explication est basée sur les observations sur les différences de polarité de fixation de la NCp7 aux ADNs et aux ARNs exposées en Introduction^{362, 384}. Puisqu'un certain nombre de données³⁶⁵ indique que ZF1 est davantage impliqué dans l'activité de déstabilisation que ZF2, ces différences de fixation peuvent entraîner des préférences dans les zones simple brin qui vont être déstabilisées. Dans le cas des séquences sauvages (à gauche dans la figure 6C), du fait des modes de fixation proposés, ce sont les zones double brin "supérieures" qui vont être déstabilisées, à la fois dans le cas de l'ADN et de l'ARN. Par contre, avec les séquences mutées, ce seront les zones double brin "inférieures" situées aux extrémités qui seront déstabilisées. Ce schéma peut également expliquer les voies d'hybridation différentes empruntées dans le cas du couple sauvage (mécanisme zipper) et mutant (mécanisme kissing). Notons que nous tentons aussi d'analyser ces données avec d'autres schémas réactionnels, incluant le fait que le mécanisme réactionnel comporte deux étapes avec des cinétiques assez dissemblables. Des schémas réactionnels à plusieurs branches peuvent ainsi être possibles comme montré dans des études récentes⁴⁰⁶. Quelque soit le mécanisme exact, ces études soulignent l'influence importante de la position des boucle internes sur le chemin d'hybridation emprunté.

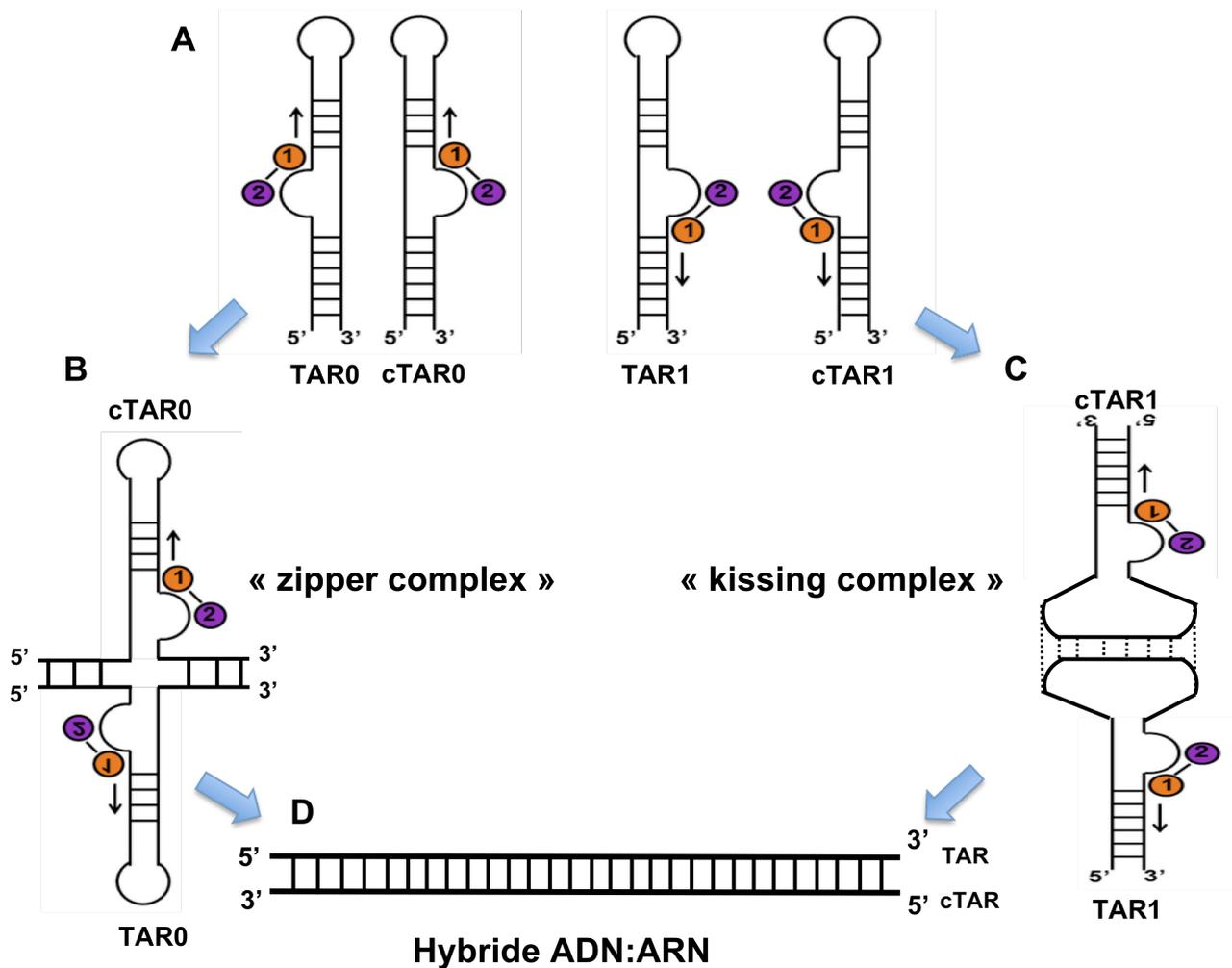


Figure 6C: Schéma hypothétique du mécanisme de fixation, de destabilisation et d'hybridation des couples cTAR0-TAR0 sauvages (à gauche) et cTAR1-TAR1 mutées (à droite) par la NCP7. La protéine est représentée avec ses deux doigts de zinc ZF1 (orange) et ZF2 (violet). **A)** Le mode de fixation sur les ADNs et les ARNs est celui décrit dans Introduction (Figure 20A). ZF1 est le doigt spécialisé dans la destabilisation des structures secondaires. La flèche noire indique la destabilisation des zones double-brin. **B)** La formation du complexe initial (partiellement hybridé) se fait via une voie « zipper mechanism » pour le couple sauvage TAR0-cTAR0. **C)** La formation du complexe initial (partiellement hybridé) se fait via une voie « kissing mechanism » pour le couple mutant TAR1-cTAR1. **D)** La formation du complexe étendu résulte en un hybride ADN:ARN parfaitement hybridé.

IV.1.2.2 Analyse de l'hybridation de TAR et cTAR en absence et en présence de NCP7 par gel retard

Les résultats que j'ai obtenus avec ces expériences ont donné lieu à une publication dans le journal **RNA**. Les principaux résultats obtenus sont résumés ci-dessous.

Afin d'analyser l'efficacité d'hybridation des séquences TAR et cTAR complètes (59 nt) en absence et en présence de NCP7, nous avons eu recours à des tests d'hybridation par gels retard. Pour effectuer ces expériences, l'ARN a été marqué au ^{32}P en 5' dont le protocole de marquage est décrit dans matériels et méthodes. Par rapport aux travaux précédents effectués par d'autres groupes, la particularité de ces expériences est de

travailler en se mettant dans des conditions proches de celles des expériences RMN (voir plus loin) : c'est à dire avec des rapports NC:nt très faibles et des rapports relatifs ARN et ADN équivalents. Cette dernière condition signifie que l'on n'est plus dans des conditions de pseudo-premier ordre. En utilisant de telles conditions nous avons pu étudier le fonctionnement de la NC à de très bas rapports NC:nt pour l'hybridation TAR-cTAR.

IV.1.2.2.1 Hybridation en absence de NCp7

Les premiers tests ont été réalisés sans NCp7 (Article1, Figure 2.A gel du haut, piste 3 à 7). L'hybridation a été étudiée en fonction du temps : 10 min, 1 h, 3 h, 6 h et 24 h. La piste (1) de l'ARN dénaturé permet d'identifier la bande qui correspond à l'ARN sous forme de monomère. La piste (2) correspond au contrôle ARN incubé seul pendant 10 min, afin de vérifier que l'ARN ne se dégrade pas au cours du temps et qu'il ne forme pas de duplex qui pourrait entrer en compétition avec la formation de l'hybride ADN-ARN. Quant aux essais d'hybridation, l'hybride TAR-cTAR correspond à la bande qui apparaît dès 1 h mais qui n'augmente que faiblement au cours du temps. Par conséquent, l'hybridation des tige-boucles en absence de la NCp7 n'est pas efficace.

IV.1.2.2.2 Hybridation en présence de NCp7

Après avoir évalué l'hybridation des tige-boucles complémentaires en absence de NCp7, nous avons voulu comprendre le rôle de la protéine dans l'appariement de ces tige-boucles. Les expériences ont été réalisées avec différentes concentrations de NCp7 : 1.5 μM ; 3.3 μM et 6 μM de NCp7 soit l'équivalent de ratios 1 NC/59 nt ; 1 NC/29 nt et 1 NC/14 nt respectivement. Avec un ratio faible 1 NC/59 nt (Figure 2A gel du bas, piste 3 à 7), on observe que l'hybridation se produit dès 10 min d'incubation et qu'elle augmente à partir de 3h. Il est intéressant de noter que l'hybridation TAR-cTAR en présence de ce faible ratio de protéine par nucléotide est plus efficace que l'hybridation sans NCp7.

Par ailleurs en augmentant la concentration de la protéine à 3.3 μM soit 1 NC/29 nt (Figure 2A gel du bas, piste 8 à 12), on observe que l'hybridation est de plus en plus efficace et qu'elle augmente significativement en fonction du temps. Cependant, afin d'accélérer encore plus le processus d'hybridation, nous avons augmenté la concentration de la protéine pour avoir un ratio de 1 NC/14 nt (Figure 2A gel du haut, piste 8 à 12), l'hybridation est significativement importante dès 10 min d'incubation et augmente significativement en fonction du temps.

IV.1.2.2.3 Analyse de la cinétique d'hybridation

A partir des expériences de gels retard, nous avons réalisé des courbes cinétiques pour les différents complexes qui nous permettent de déterminer le pourcentage de séquences TAR ARN hybridées en fonction du temps (Figure 2B). Selon les modèles de Vo et al., 2009, la cinétique d'hybridation se déroule en deux phases, lente et rapide qui correspondent à la formation d'un duplex partiellement hybridé et étendu respectivement :



Comme nos conditions ne sont pas des conditions de pseudo-premier ordre, puisque les concentrations d'ADN et d'ARN sont du même ordre (rapport ADN:ARN, 1:1), un certain nombre de simplifications qui sont adoptées dans de telles conditions, ne peuvent être faites ici, il est nécessaire d'analyser la première réaction (celle de la formation du complexe initial) dans le cadre d'une réaction biomoléculaire et réversible. La description de l'analyse des données est présentée dans le Matériels et Méthodes de l'article RNA. Les données cinétiques obtenues sont présentées dans le tableau 1 de l'article 1.

La constante de vitesse k_1 affiche une variation importante quand on compare l'hybridation sans NCp7 et en présence de protéine (k_1 augmente d'un facteur 30), on voit donc ici l'effet important de la protéine sur la phase d'hybridation proprement dite du complexe. On remarque que même à des rapports NC:nt très faibles (1 NC/59 nt), l'effet sur la première phase (k_1) est significatif. La présence de NC permet une augmentation de k_1 d'un facteur 5 (comparer les colonnes NC(0:59) et NC(1:59)). Cela signifie donc que dans les conditions où les expériences de RMN ont été effectuées, l'effet de la NC sur l'hybridation est important. En revanche, dans les mêmes conditions, la constante de dissociation k_{-1} n'affiche pas de variation. La constante k_2 augmente d'un facteur 10 entre NC(0:59) et NC(1:14). Nous montrons donc ici clairement le rôle important de la présence de la nucléocapside sur la transition vers le complexe étendu. Il est intéressant de noter que dans le travail précédent de Vo et al, 2009 par le groupe de Musier-Forsyth, les conditions utilisées n'avaient pas permis à ces auteurs de mesurer l'impact de la NC sur la vitesse k_2 dans le cas de l'hybridation TAR-cTAR (molécules entières) (Voir Figure 9(d) de Vo et al, 2009). Dans le Matériels et Méthodes du RNA, au chapitre "Choice of nucleic acids concentrations", il est expliqué qu'une correcte détermination de k_2 dépend de façon essentielle de la fraction f (la fraction du composant rapide), il est important que celle-ci ne soit pas trop grande, car une fraction f trop peuplée, rend la phase lente qui nous intéresse pour étudier k_2 trop peu peuplée. Aux faibles rapports en NC utilisées, il faut utiliser des concentrations en ADN et en ARN importantes (ici 750 micromolaires pour pouvoir

déterminer correctement les vitesses k_2 . On peut alors montrer que, comme on pouvait s'y attendre intuitivement, la quantité de NC est très importante pour favoriser la transition vers le duplex étendu. Encore une fois, il faut noter que ce dernier point n'avait pas pu être mis en évidence par le groupe de Musier-Forsyth⁴²⁴⁻⁴²⁶ du fait du choix de leurs conditions expérimentales.

Ce travail souligne le rôle important de l'activité déstabilisatrice de la NCp7 pour permettre la transition du complexe initial vers le complexe final. On doit noter toutefois qu'aux rapports NC:nt très faibles (1:29) et (1:59), si l'effet sur la formation du complexe initial est bien perceptible, l'effet sur la conversion vers le duplex étendu (qui implique sans doute l'activité déstabilisation des acides nucléiques) est très faible, ce qui signifie que des quantités plus importantes de protéine sont nécessaires pour cette dernière activité.

IV.1.2.3 Analyse de la déstabilisation de la structure secondaire de TAR ARN par la NCp7

Les résultats que j'ai obtenus avec ces expériences ont donné lieu à une publication dans le journal **RNA** et qui est présentée à la fin de ce mémoire. Les résultats obtenus sont résumés ci-après.

Nous avons étudié l'interaction entre la séquence TAR ARN (59 nt), enrichie en ^{15}N et ^{13}C , avec la NCp7 (non enrichie), à un faible ratio (1 NC:59 nt) par RMN. L'attribution des résonances iminos de TAR a été réalisée en enregistrant des spectres 2D ^1H - ^{15}N TROSY et SOFAST-HMQC. Le fait que seul un des deux partenaires soit marqué nous permet d'observer sélectivement les résonances de celui-ci. Par ailleurs, le choix de ce faible ratio de protéine par nt, nous permet de prévenir les problèmes d'agrégations et l'élargissement excessif des résonances, qui mène rapidement à la perte d'une quantité importante d'informations. La formation du complexe TAR:NCp7 induit un élargissement très significatif des résonances des protons iminos (Figure 3 et 4 de l'article 1). Cependant, nous avons pu remarquer que certaines résonances étaient peu affectées alors que d'autres disparaissaient complètement.

Afin d'analyser ces effets différentiels, les résonances des protons iminos ont été classées en trois catégories : (i) celles dont l'élargissement ne permettait plus leur observation (G26, G28 et U38); (ii) celles qui se sont élargies de façon importantes mais qui restent observables (G16, G21, U42, and G43); (iii) celles qui ont montré, de façon relative, un élargissement modéré (G3, U10, G11, U13, U14, U46, G52, G53, and G54) (Figure 3 et 4 de l'article 1). Les protons iminos qui n'ont pas été mentionnés ci-dessus, dans aucune des trois catégories, n'ont pas pu être attribués. Il apparaît immédiatement que les trois

différentes classes de protons iminos décrites plus haut sont localisées dans des parties différentes de la molécule TAR ARN (Figure 1A, de l'article 1). On remarque que le premier groupe, le plus touché, implique les 4 pb de la tige haute et qui vont de G26-C39 à C29-G36 (Figure 3 et 4 de l'article 1). Le deuxième groupe appartient à la courte tige située au centre de la molécule (bp C18–G44 à G21–C41) tandis que les résonances du troisième groupe correspondent à des pb C7–G54 à G16–C45 de la grande tige centrale ainsi qu'au deux paires de bases situées à l'extrémité finale de TAR. Cet élargissement spécifique des résonances nous permet de conclure que les paires de bases les plus touchées appartiennent à un site préférentiel de déstabilisation de TAR ARN. Il apparaît que la NCp7 déstabilise préférentiellement la boucle apicale de la tige boucle TAR. On observe également, qu'il y a un effet de déstabilisation assez faible sur la tige longue centrale. Par ailleurs, la courte tige située aux extrémités 5'3' n'a été que légèrement affectée par la déstabilisation de la NCp7. Finalement, nous concluons à partir de ces données RMN qu'il existe une déstabilisation sélective de TAR ARN par la NCp7 et que la région initiatrice est située au-dessus du bulge UCU avec le résidu G26 comme point possible de nucléation.

IV.1.3 Discussion

En conclusion, les expériences de FRET ont démontré que les cinétiques d'hybridation du couple sauvage cTAR0-TAR0 s'avère être plus rapide que celles du couple mutant cTAR1-TAR1. Au-delà de ces différences quantitatives, ce qui nous apparaît le plus intéressant est que l'inversion de la position des boucles entraîne un changement dans la voie d'hybridation majoritaire. Pour les deux couples de molécules (le couple sauvage et le couple mutant), les hybridations se produisent par les deux chemins (c'est ce que montrent les expériences avec les mutants du type "loop" et "stem").

Pour le couple sauvage le chemin d'hybridation par les extrémités ("zipper") apparaît être le chemin majoritaire tandis que pour le couple mutant, c'est le chemin d'hybridation initié par les complexes boucle-boucle ("kissing complex") qui est majoritaire.

L'utilisation pour chaque couple des molécules appelées "stem" et "loop" permet d'identifier les chemins majoritaires utilisés dans chaque cas. De telles conclusions ne peuvent être tirées, évidemment, que si les résultats des expériences réalisées sur chacun des couples sont bien cohérents entre eux. Par exemple pour le couple sauvage, on constate que l'altération de la tige (molécule "stem" avec laquelle on perturbe l'association par les extrémités sans perturber la formation des complexes boucles-boucles) ralentit les cinétiques d'hybridation (ou plus précisément les valeurs k_2), tandis que l'altération de la boucle (molécule "loop0", cette modification de deux paires de bases dans les boucles

apicales perturbe l'association boucle-boucle sans modifier les possibilités d'association par les extrémités) augmente les cinétiques d'hybridation. Ce dernier comportement est possible si on considère que du fait de la perturbation du chemin "loop", un plus grand nombre de molécules (que dans le couple sauvage) emprunte le chemin d'hybridation "zipper" favorisé thermodynamiquement dans le cas du couple sauvage. L'ensemble de ces trois expériences (couple sauvage et expériences avec "stem0" et "loop0") est donc compatible avec l'hypothèse de l'existence de deux chemins coexistants mais avec un chemin privilégié qui est le chemin "zipper" dans le cas des molécules sauvages.

De façon frappante, les molécules mutantes (c'est à dire TAR1 et cTAR1 avec les boucles internes inversées) montrent des comportements opposés. Le couple mutant "stem1" favorise les hybridations par rapport au couple initial TAR1-cTAR1, tandis que le couple mutant "loop1" montre que l'hybridation avec TAR1-cTAR1 est très ralentie. L'ensemble des résultats des trois types d'expériences sur le couple mutant est donc cohérent mais montre que si, de la même façon que pour les molécules sauvages, les deux chemins d'hybridation sont présents, c'est ici le chemin "kissing complex" qui est majoritaire.

En présence de concentration saturante de NCp7 (à partir de 1:15 NC d'après Vo et al, 2009) avec les molécules TAR-cTAR entières (59nt), il est connu que TAR-cTAR emprunte une voie 'zipper', ce qui est en accord avec nos résultats pour notre couple sauvage. Cependant, il apparaît que le changement de position des boucles internes fait basculer le mécanisme d'hybridation vers un mécanisme de type "kissing complex" et cela même en présence de concentration saturante de NCp7 (puisque l'on travaille à des rapports 1:5). De tels résultats sont compatibles avec les hypothèses suivantes : (i) nous avons remarqué, comme mentionné dans l'introduction (Figure 20A), des modes de fixation différents de la NCp7 sur les ADNs et les ARNs (ceci a été établi sur la base de la comparaison des complexes NC:acides nucléiques existants ³⁶². (ii) le mode de fixation précédent induit un positionnement de ZF1 (le domaine de la NCp7 le plus impliqué dans la déstabilisation des structures secondaires) particulier par rapport aux paires de bases d'une structure tige-boucle (Figure 5 de l'article 1, diagrammes B et C).

En partant des hypothèses faites ci-dessus, avec les séquences sauvages, la NCp7 déstabilisera d'abord les tiges apicales tandis qu'avec les séquences mutantes ce sont les tiges basses qui seront déstabilisées en priorité. Les tiges apicales sont bien celles qui sont majoritairement présentes dans les complexes de type "zipper" tandis que les tiges basses sont majoritairement présentes dans le cas des complexes de type "kissing complex".

L'analyse de l'hybridation des séquences complètes TAR et cTAR en absence et en

présence de la NCp7 par gel retard a confirmé que la protéine est essentielle pour une hybridation efficace des deux partenaires et qu'en absence de la NCp7, la réaction s'avère être très lente. Cela indique que la déstabilisation des séquences TAR et cTAR est une étape limitante pour permettre leur hybridation et que la présence d'une protéine avec une activité chaperon est nécessaire pour leur permettre d'aboutir à un duplex étendu et stable. De plus, on observe que la NCp7 améliore l'efficacité de l'hybridation et ceci même à des concentrations aussi faibles de 1 NCp7 par oligonucléotide (1:59), ce qui n'avait pas été démontré auparavant.

Finalement, nos données soulignent l'importance de l'activité déstabilisatrice de la NCp7 pour l'hybridation de séquences complémentaires qui requiert leur déstabilisation. La NCp7 s'avère être importante pour la formation du duplex partiellement hybridé, ce qui est cohérent avec les résultats de Vo et al., 2009, qui ont déterminé que la NCp7 stabilise ce duplex initial, dont la conversion en duplex étendu est également très dépendante de la NCp7, étape qui requiert des concentrations plus importantes de protéine^{406, 420, 424-425}. Par ailleurs, des études antérieures ont démontré que la déstabilisation de TAR ARN par la NC(12-55) est moins efficace que pour cTAR ADN^{383, 404, 415}. Nous avons entrepris l'étude de la déstabilisation de la structure secondaire de TAR ARN à un faible ratio de protéine par nucléotide (1 NCp7 :59 nt), ces bas rapports étant choisis principalement pour ne pas avoir des élargissements de résonances trop importants. Dans nos conditions, il en ressort qu'une déstabilisation efficace sur la tige boucle s'est avérée que pour seulement trois paires de bases consécutives. En comparant avec les résultats de Bernacchi et al., 2002, bien que leur ratio fût plus élevé (1 NC :7.5 nt) une seule pb a été déstabilisé sur TAR ARN, cependant, la NC(1-55) utilisée dans notre étude a une plus grande activité déstabilisatrice que la NC(12-55)^{177, 424, 560}. En revanche, nos résultats sont très similaires à ceux obtenus avec une étude réalisée avec des expériences de "single-molecule stretching", où les auteurs ont analysé la déstabilisation de TAR ARN par la NCp7³¹². Cette dernière a démontré, comme nous, que les sites spécifiques ciblés par la NCp7 correspondent aux guanines adjacentes aux régions simple brin de la tige, tel que les "mésappariements", les boucles et les "bulges". Tout comme dans notre étude, la base G26 a été identifiée comme un site clé pour l'initiation de la déstabilisation de la tige boucle TAR ARN. Dans notre étude nous avons choisi d'étudier le complexe à faible ratio dans le but de détecter le site initial par lequel la déstabilisation commence. Nous avons vérifié que l'hybridation TAR-cTAR était accélérée par la NCp7 même dans des conditions de faible ratio. Bien que ces conditions soient évidemment très différentes de celles observées *in vivo*³⁵³, elles nous ont permis d'identifier les sites "initiaux" de la

déstabilisation de TAR ARN et de vérifier que même à faible ratio la NCp7 accélère l'hybridation des partenaires TAR-cTAR. Enfin ces conditions très particulières nous ont permis de mesurer correctement les constantes de vitesses k_2 , ce qui n'était pas possible ou très difficile lorsqu'on travaille à des rapports NC:nt trop élevés ⁴²⁴.

Article 1

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

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

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

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

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

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

ABSTRACT

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

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

INTRODUCTION

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

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

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

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

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

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

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

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

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

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

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

RESULTS

NMR analysis of the TAR RNA (59 nt)

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

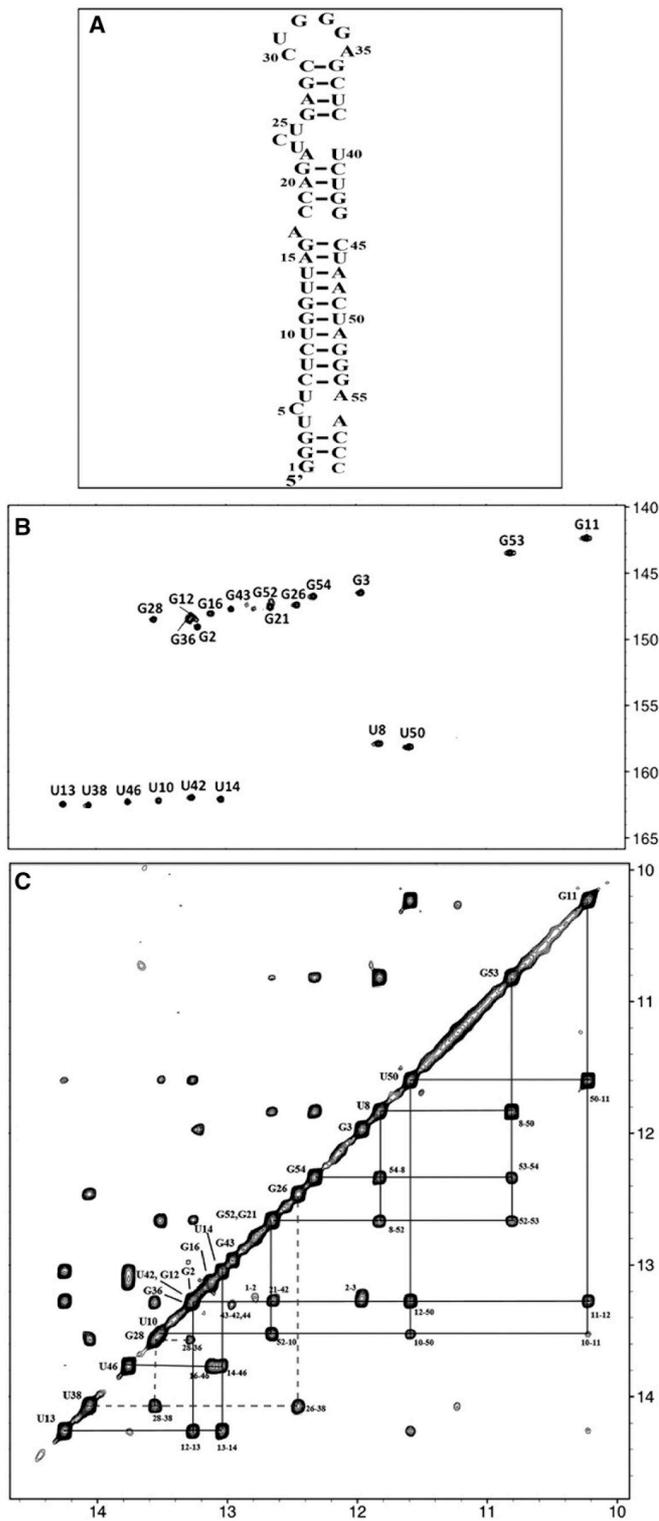


FIGURE 1. (Legend on next page)

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

TAR–cTAR annealing at a low NC:nt ratio

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

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

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

NMR analysis of NC-induced TAR secondary structure destabilization

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

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

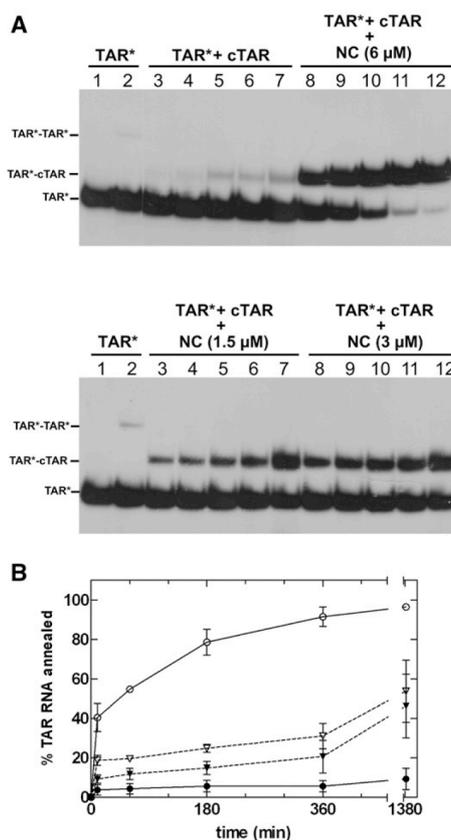


FIGURE 2. (A) Time course of TAR RNA annealing with cTAR DNA in the presence or absence of various concentrations of NC. Lane 1, heat-denatured TAR ^{32}P -RNA. TAR ^{32}P -RNA alone (lanes 2) or mixed with cTAR DNA (lanes 3–12) was incubated at 20°C for 10 min (lanes 2,3,8), 1 h (lanes 4,9), 3 h (lanes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

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

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

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

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

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

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

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

DISCUSSION

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

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

molecule. The assignments were similar, except for the G43 residue, which has a different sequence context due to the truncation of the TAR sequence in the previous studies (Hennig and Williamson 2000; Dethoff et al. 2008). The assigned resonances and the inter-base pair NOE networks were consistent with the self-folding of TAR proposed in Figure 1A. Interestingly, the imino protons of 3 bp were not observable. All these base pairs were U.A base pairs adjacent to a bulge (in Fig. 1A, no imino protons are indicated for C18–G44 due to overlap, resulting in ambiguity, and G1–C59 in the terminal base pair). The A22–U40 imino proton which was not observed here, was also unobservable in previous studies dealing with the upper half of the TAR hairpin (Puglisi et al. 1992; Aboul-ela et al. 1996; Zhang et al. 2006).

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

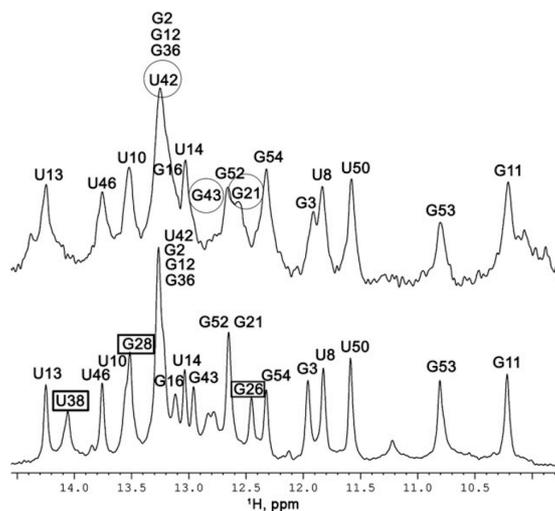


FIGURE 3. Imino proton region of ^{15}N labeled TAR at 10°C . (*Bottom*) Free TAR, (*top*) TAR with NC in a 1:1 ratio (1:59 NC:nt). The resonances that disappear totally are indicated by rectangles in spectrum at the *bottom*; the resonances showing significant broadening and/or chemical shifts are indicated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

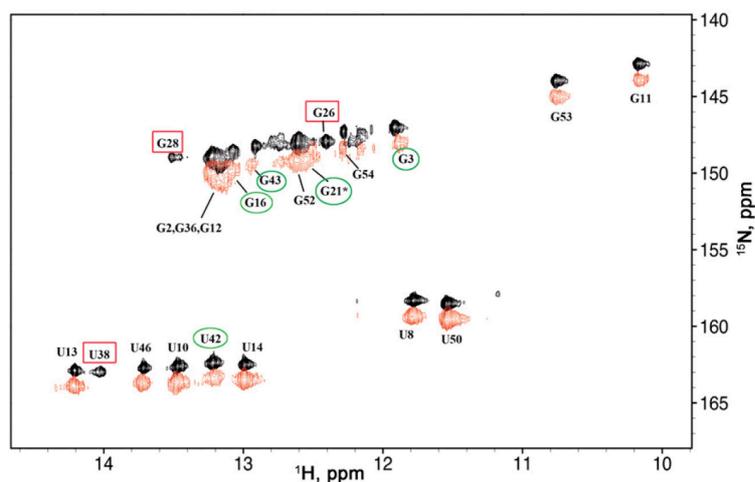


FIGURE 4. 2D ^1H - ^{15}N SOFAST-HMQC spectra at 10°C and 950 MHz of free TAR (black) and TAR-NC in a 1:1 ratio (1:59 NC:nt) (red) at 10°C and 950 MHz. The imino groups are labeled with the name of the residue. The resonances that disappear totally are indicated by red rectangles and those that display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

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

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

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

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

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

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

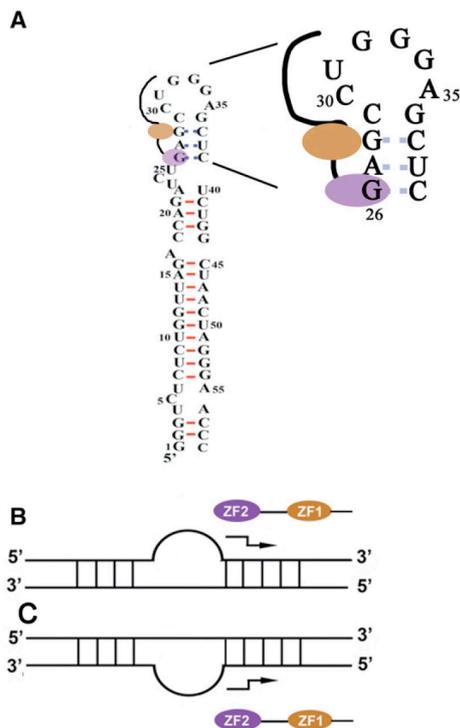


FIGURE 5. Hypothetical diagrams of RNA-DNA binding and destabilization by NC. (A) View of the TAR hairpin with a report of the observed imino protons (red) observed in the NC-TAR complex and those observable in the free RNA but not in the NC-RNA complex (dashed blue lines). NC is depicted with its two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

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

Samples

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

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

NMR

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

Gel retardation assays

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

Kinetic analysis of the gel-retardation assays

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

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

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



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

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

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

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

and

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

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

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

where

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

and

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

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

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

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

Choice of nucleic acid concentrations

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

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

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

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

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

ACKNOWLEDGMENTS

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

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

REFERENCES

- Aboul-ela F, Karn J, Varani G. 1995. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol* **253**: 313–332.
- Aboul-ela F, Karn J, Varani G. 1996. Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res* **24**: 3974–3981.
- Al-Hashimi HM, Gosser Y, Gorin A, Hu W, Majumdar A, Patel DJ. 2002. Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol* **315**: 95–102.
- Amarasinghe GK, De Guzman RN, Turner RB, Chancellor KJ, Wu ZR, Summers MF. 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ -RNA packaging signal. Implications for genome recognition. *J Mol Biol* **301**: 491–511.
- Andersen ES, Contera SA, Knudsen B, Damgaard CK, Bsenbacher F, Kjems J. 2004. Role of the trans-activation response element in dimerization of HIV-1 RNA. *J Biol Chem* **279**: 22243–22249.
- Azoulay J, Clamme JP, Darlix JL, Roques BP, Mély Y. 2003. Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J Mol Biol* **326**: 691–700.
- Bailor MH, Sun XY, Al-Hashimi HM. 2010. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**: 202–206.
- Baker GM, Weng LC. 1992. Apparent first-order behavior under second-order kinetic conditions: a general concept illustrated by the reversible binding of hydrogen peroxide to cytochrome C oxidase. *J Theor Biol* **158**: 221–229.
- Barraud P, Gaudin C, Dardel F, Tisné C. 2007. New insights into the formation of HIV-1 reverse transcription initiation complex. *Biochimie* **89**: 1204–1210.
- Barraud P, Dardel F, Tisné C. 2008. Optimizing HSQC experiment for the observation of exchange broadened signals in RNA–protein complexes. *Cr Chim* **11**: 474–479.
- Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. 2008. Strand transfer events during HIV-1 reverse transcription. *Virus Res* **134**: 19–38.
- Bazzi A, Zargarian L, Chaminade F, Boudier C, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O. 2011. Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyriboses in the binding polarity of NC. *Nucleic Acids Res* **39**: 3903–3916.
- Bazzi A, Zargarian L, Chaminade F, De Rocquigny H, René B, Mély Y, Fossé P, Mauffret O. 2012. Intrinsic nucleic acid dynamics modulates HIV-1 nucleocapsid protein binding to its targets. *PLoS One* **7**: e38905.
- Beltz H, Azoulay J, Bernacchi S, Clamme JP, Ficheux D, Roques B, Darlix JL, Mély Y. 2003. Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* **328**: 95–108.
- Beltz H, Piémont E, Schaub E, Ficheux D, Roques B, Darlix JL, Mély Y. 2004. Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* **338**: 711–723.
- Beltz H, Clauss C, Piémont E, Ficheux D, Gorelick RJ, Roques B, Gabus C, Darlix JL, de Rocquigny H, Mély Y. 2005. Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J Mol Biol* **348**: 1113–1126.
- Berkhout B. 1992. Structural features in TAR RNA of human and simian immunodeficiency viruses: a phylogenetic analysis. *Nucleic Acids Res* **20**: 27–31.
- Berkhout B, Vastenhout NL, Klasens BI, Huthoff H. 2001. Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **7**: 1097–1114.
- Bernacchi S, Stoylov S, Piémont E, Ficheux D, Roques BP, Darlix JL, Mély Y. 2002. HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol Biol* **317**: 385–399.
- Bevington PR, Robinson DK. 2003. *Data reduction and error analysis for the physical sciences*, 3rd ed. McGraw-Hill, NY.
- Bourbigot S, Ramalanjaona N, Boudier C, Salgado GF, Roques BP, Mély Y, Bouaziz S, Morellet N. 2008. How the HIV-1 nucleocapsid protein binds and destabilises the (–) primer binding site during reverse transcription. *J Mol Biol* **383**: 1112–1128.
- Cruceanu M, Gorelick RJ, Musier-Forsyth K, Rouzina I, Williams MC. 2006a. Rapid kinetics of protein–nucleic acid interaction is a major

- component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol* **363**: 867–877.
- Cruceanu M, Urbaneja MA, Hixson CV, Johnson DG, Datta SA, Fivash MJ, Stephen AG, Fisher RJ, Gorelick RJ, Casas-Finet JR, et al. 2006b. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **34**: 593–605.
- Darlix JL, Garrido JL, Morellet N, Mély Y, de Rocquigny H. 2007. Properties, functions, and drug targeting of the multifunctional nucleocapsid protein of the human immunodeficiency virus. *Adv Pharmacol* **55**: 299–346.
- Darlix JL, Godet J, Ivanyi-Nagy R, Fossé P, Mauffret O, Mély Y. 2011. Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J Mol Biol* **410**: 565–581.
- Darlix JL, de Rocquigny H, Mauffret O, Mély Y. 2014. Retrospective on the all-in-one retroviral nucleocapsid protein. *Virus Res* **193**: 2–15.
- De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**: 384–388.
- Dethoff EA, Al-Hashimi HM. 2011. Characterization of ps-Ms dynamics in the TAR apical loop by NMR. *Biophys J* **100**: 238a.
- Dethoff EA, Hansen AL, Musselman C, Watt ED, Andricioaei I, Al-Hashimi HM. 2008. Characterizing complex dynamics in the trans-activation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* **95**: 3906–3915.
- Farjon J, Boisbouvier J, Schanda P, Pardi A, Simorre JP, Brutscher B. 2009. Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *J Am Chem Soc* **131**: 8571–8577.
- Godet J, Mély Y. 2010. Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol* **7**: 687–699.
- Godet J, de Rocquigny H, Raja C, Glasser N, Ficheux D, Darlix JL, Mély Y. 2006. During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol* **356**: 1180–1192.
- Godet J, Kenfack C, Przybilla F, Richert L, Duportail G, Mély Y. 2013. Site-selective probing of cTAR destabilization highlights the necessary plasticity of the HIV-1 nucleocapsid protein to chaperone the first strand transfer. *Nucleic Acids Res* **41**: 5036–5048.
- Golinelli MP, Hughes SH. 2003. Secondary structure in the nucleic acid affects the rate of HIV-1 nucleocapsid-mediated strand annealing. *Biochemistry* **42**: 8153–8162.
- Guo J, Wu T, Kane BF, Johnson DG, Henderson LE, Gorelick RJ, Levin JG. 2002. Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **76**: 4370–4378.
- Heath MJ, Derebail SS, Gorelick RJ, DeStefano JJ. 2003. Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing. *J Biol Chem* **278**: 30755–30763.
- Heilman-Miller SL, Wu T, Levin JG. 2004. Alteration of nucleic acid structure and stability modulates the efficiency of minus-strand transfer mediated by the HIV-1 nucleocapsid protein. *J Biol Chem* **279**: 44154–44165.
- Heng X, Kharytonchik S, Garcia EL, Lu K, Divakaruni SS, LaCotti C, Edme K, Telesnitsky A, Summers MF. 2012. Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J Mol Biol* **417**: 224–239.
- Hennig M, Williamson JR. 2000. Detection of N-H...N hydrogen bonding in RNA via scalar couplings in the absence of observable imino proton resonances. *Nucleic Acids Res* **28**: 1585–1593.
- Hergott CB, Mitra M, Guo J, Wu T, Miller JT, Iwatani Y, Gorelick RJ, Levin JG. 2013. Zinc finger function of HIV-1 nucleocapsid protein is required for removal of 5'-terminal genomic RNA fragments: a paradigm for RNA removal reactions in HIV-1 reverse transcription. *Virus Res* **171**: 346–355.
- Huthoff H, Girard F, Wijmenga SS, Berkhout B. 2004. Evidence for a base triple in the free HIV-1 TAR RNA. *RNA* **10**: 412–423.
- Jankowsky E. 2011. RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36**: 19–29.
- Johnson PE, Turner RB, Wu ZR, Hairston L, Guo J, Levin JG, Summers MF. 2000. A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription. *Biochemistry* **39**: 9084–9091.
- Kanevsky I, Chaminade F, Ficheux D, Moumen A, Gorelick R, Negroni M, Darlix JL, Fossé P. 2005. Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J Mol Biol* **348**: 1059–1077.
- Klaver B, Berkhout B. 1994. Evolution of a disrupted TAR RNA hairpin structure in the HIV-1 virus. *EMBO J* **13**: 2650–2659.
- Krishnamoorthy G, Roques B, Darlix JL, Mély Y. 2003. DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection. *Nucleic Acids Res* **31**: 5425–5432.
- Lee J, Dethoff EA, Al-Hashimi HM. 2014. Invisible RNA state dynamically couples distant motifs. *Proc Natl Acad Sci* **111**: 9485–9490.
- Levin JG, Guo J, Rouzina I, Musier-Forsyth K. 2005. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **80**: 217–286.
- Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **7**: 754–774.
- Liu HW, Zeng Y, Landes CF, Kim YJ, Zhu Y, Ma X, Vo MN, Musier-Forsyth K, Barbara PF. 2007. Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc Natl Acad Sci* **104**: 5261–5267.
- McCaughey MJ, Rouzina I, Manthei KA, Gorelick RJ, Musier-Forsyth K, Williams MC. 2015. Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc Natl Acad Sci* **112**: 13555–13560.
- Mirambeau G, Lyonnais S, Coulaud D, Hameau L, Lafosse S, Jeusset J, Justome A, Delain E, Gorelick RJ, Le Cam E. 2006. Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J Mol Biol* **364**: 496–511.
- Mirambeau G, Lyonnais S, Gorelick RJ. 2010. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol* **7**: 724–734.
- Mitra M, Wang W, Vo MN, Rouzina I, Barany G, Musier-Forsyth K. 2013. The N-terminal zinc finger and flanking basic domains represent the minimal region of the human immunodeficiency virus type-1 nucleocapsid protein for targeting chaperone function. *Biochemistry* **52**: 8226–8236.
- Musselman C, Al-Hashimi HM, Andricioaei I. 2007. IRED analysis of TAR RNA reveals motional coupling, long-range correlations, and a dynamical hinge. *Biophys J* **93**: 411–422.
- Mustoe AM, Brooks CL, Al-Hashimi HM. 2014. Hierarchy of RNA functional dynamics. *Annu Rev Biochem* **83**: 441–466.
- Narayanan N, Gorelick RJ, DeStefano JJ. 2006. Structure/function mapping of amino acids in the N-terminal zinc finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity. *Biochemistry* **45**: 12617–12628.
- Nietispach D. 2005. Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *J Biomol NMR* **31**: 161–166.
- Pervushin K, Riek R, Wider G, Wuthrich K. 1997. Attenuated T₂ relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci* **94**: 12366–12371.
- Pitt SW, Majumdar A, Serganov A, Patel DJ, Al-Hashimi HM. 2004. Argininamide binding arrests global motions in HIV-1 TAR RNA: comparison with Mg²⁺-induced conformational stabilization. *J Mol Biol* **338**: 7–16.

- Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**: 76–80.
- Rein A. 2010. Nucleic acid chaperone activity of retroviral Gag proteins. *RNA Biol* **7**: 700–705.
- Salmon L, Bascom G, Andricioaei I, Al-Hashimi HM. 2013. A general method for constructing atomic-resolution RNA ensembles using NMR residual dipolar couplings: the basis for inter-helical motions revealed. *J Am Chem Soc* **135**: 5457–5466.
- Schanda P, Brutscher B. 2005. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* **127**: 8014–8015.
- Schanda P, Kupce E, Brutscher B. 2005. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**: 199–211.
- Sleiman D, Goldschmidt V, Barraud P, Marquet R, Paillart JC, Tisné C. 2012. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res* **169**: 324–339.
- Sleiman D, Barraud P, Brachet F, Tisné C. 2013. The interaction between tRNA^{Lys} and the primer activation signal deciphered by NMR spectroscopy. *PLoS One* **8**: e64700.
- Spriggs S, Garyu L, Connor R, Summers MF. 2008. Potential intra- and intermolecular interactions involving the unique-5' region of the HIV-1 5'-UTR. *Biochemistry* **47**: 13064–13073.
- Stewart-Maynard KM, Cruceanu M, Wang F, Vo MN, Gorelick RJ, Williams MC, Rouzina I, Musier-Forsyth K. 2008. Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J Virol* **82**: 10129–10142.
- Sun XY, Zhang Q, Al-Hashimi HM. 2007. Resolving fast and slow motions in the internal loop containing stem-loop 1 of HIV-1 that are modulated by Mg²⁺ binding: role in the kissing-duplex structural transition. *Nucleic Acids Res* **35**: 1698–1713.
- Tisné C, Roques BP, Dardel F. 2001. Heteronuclear NMR studies of the interaction of tRNA^{Lys} with HIV-1 nucleocapsid protein. *J Mol Biol* **306**: 443–454.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2006. Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol* **363**: 244–261.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2009a. Effect of Mg²⁺ and Na⁺ on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J Mol Biol* **386**: 773–788.
- Vo MN, Barany G, Rouzina I, Musier-Forsyth K. 2009b. HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop “kissing” to “zipper”. *J Mol Biol* **386**: 789–801.
- Vuilleumier C, Bombarda E, Morellet N, Gérard D, Roques BP, Mély Y. 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **38**: 16816–16825.
- Wang W, Naiyer N, Mitra M, Li J, Williams MC, Rouzina I, Gorelick RJ, Wu Z, Musier-Forsyth K. 2014. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res* **42**: 7145–7159.
- Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **6**: e96.
- Woodson SA. 2010. Taming free energy landscapes with RNA chaperones. *RNA Biol* **7**: 677–686.
- Wu T, Datta SA, Mitra M, Gorelick RJ, Rein A, Levin JG. 2010. Fundamental differences between the nucleic acid chaperone activities of HIV-1 nucleocapsid protein and Gag or Gag-derived proteins: biological implications. *Virology* **405**: 556–567.
- Wu H, Mitra M, McCauley MJ, Thomas JA, Rouzina I, Musier-Forsyth K, Williams MC, Gorelick RJ. 2013. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res* **42**: 2525–2537.
- Wu H, Mitra M, Nauffer MN, McCauley MJ, Gorelick RJ, Rouzina I, Musier-Forsyth K, Williams MC. 2014. Differential contribution of basic residues to HIV-1 nucleocapsid protein's nucleic acid chaperone function and retroviral replication. *Nucleic Acids Res* **42**: 2525–2537.
- You JC, McHenry CS. 1994. Human immunodeficiency virus nucleocapsid protein accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem* **269**: 31491–31495.
- Zarudyana MI, Potyahaylo AL, Kolomiets IM, Hovorun DM. 2014. Phylogenetic study on structural elements of HIV-1 poly(A) region. 2. USE domain and TAR hairpin. *Biopolym Cell* **30**: 29–36.
- Zhang Q, Sun X, Watt ED, Al-Hashimi HM. 2006. Resolving the motional modes that code for RNA adaptation. *Science* **311**: 653–656.
- Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM. 2007. Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **450**: 1263–1267.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

IV.2 Etude des propriétés d'interaction de la NCp7 avec des acides nucléiques

IV.2.1 Contexte

Lors du cycle de réplication, l'affinité de la NCp7 pour ses acides nucléiques substrats varie considérablement, tout comme les rapports NC/nt. Bien que le mode de fixation prédominant soit un mode électrostatique non spécifique et de faible affinité³⁵⁴⁻³⁵⁵ dû au caractère basique de la protéine (fixation très dépendante de la force ionique), il existe une composante de l'interaction, d'origine principalement hydrophobe, qui est plus spécifique. Parmi ces spécificités de liaison, il est connu que la NCp7 a une préférence pour les acides nucléiques simple-brin^{357-358, 561} avec une préférence marquée pour les séquences contenant des guanines et en particulier les séquences TG, TXG, GXG (avec X représentant une base variable)^{175, 177, 357, 362, 364}. Ces propriétés de spécificité de liaison sont principalement associées à la reconnaissance sélective du génome lors de l'étape de l'encapsidation de l'ARNg^{63, 278}. Cependant, les modalités de liaisons de la NCp7 aux acides nucléiques restent encore peu claires.

Nos objectifs dans ce travail, sont d'obtenir des complexes NC:Acide nucléique (ADN et ARN de séquences homologues), comprenant des motifs de fixation préférentiels de la NCp7 avec une forte affinité de liaison, conférant aux complexes une haute stabilité. Ces dernières conditions permettant leur étude et comparaison par RMN. En particulier, nous souhaitons établir l'orientation des deux domaines structurés en doigt de zinc (ZF1 et ZF2) le long de la chaîne d'acide nucléique et ceci selon la nature du substrat acide nucléique (ADN ou ARN). Les techniques de résonance magnétique nucléaire (RMN) peuvent nous permettre de réaliser ces objectifs. Néanmoins, afin d'obtenir des spectres de très bonne qualité, il est préférable de travailler avec des complexes de haute affinité. Dans un premier temps, nous mesurerons donc la stabilité de ces complexes à l'aide des méthodes d'anisotropie de fluorescence, puis nous entreprendrons l'étude des complexes par RMN.

IV.2.2 Résultats

IV.2.2.1 Etude de l'affinité des complexes par anisotropie de fluorescence

Pour les expériences d'anisotropie de fluorescence, nous avons étudié l'interaction de la forme tronquée NC(11-55) de la protéine, afin d'éviter les problèmes d'agrégation, avec les différents oligonucléotides marqués en 5' par une sonde fluorescente (F). Par la suite nous avons également utilisé la NC(1-55) pour vérifier l'interaction avec certains acides nucléiques. Nous avons fixé la concentration de l'oligonucléotide à 20 nM et on a réalisé une titration avec la protéine, ce qui nous permet d'obtenir des courbes dont on peut extraire les constantes d'affinité. Nous avons pu déterminer les constantes de dissociation (K_d) qui nous informent sur l'affinité entre les deux partenaires. Les résultats des expériences effectuées avec les différentes séquences sont indiqués dans les Tableaux (5C, 6C 7C, 8C et 9C). Les courbes de saturations des oligonucléotides fluorescés avec des concentrations croissantes de la protéine sont représentées dans les Figures (6C, 7C, 8C et 9C).

IV.2.2.1.1 Etude des complexes NC-acides nucléiques avec des séquences en tiges boucles

Dans la première partie nous avons étudié cTAR ADN tronquée (cTAR: 26 résidus). La NC(11-55) a une assez bonne affinité avec cTar26 ($K_d = 210$ nM) qui correspond presque exactement à la valeur obtenue dans une étude précédente³⁶² qui utilisait des méthodes de calorimétrie (ITC, Isothermal Titration Calorimetry) pour mesurer les affinités. Ces données valident l'utilisation de la méthode d'anisotropie pour déterminer l'affinité de la NC(11-55) pour ses oligonucléotides substrats. Afin d'examiner l'impact de l'allongement de l'ADN cTAR26 sur la fixation par la NC, nous avons étudié l'interaction entre la NC(11-55) et cTAR33 (33 résidus, une version "allongée" de cTAR26). Nos résultats (Tableau 5C) ont montré qu'il y avait une moins bonne affinité entre la NC(11-55) et cTar33 ($K_d = 1.9$ μ M).

Tableau 5C. Valeurs des constantes de dissociation (K_d) des complexes NC(11-55)-acide nucléique, étudiés par anisotropie de fluorescence, avec des séquences en tige boucles, à température de 5°C et dans du tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ (pH=6.5, I=0.04).

| Oligonucléotide | K_d (nM) avec NC(11-55) |
|-----------------|---------------------------|
| SL3 ADN | 1100 |
| SL3 ARN | 66 |
| SL3 ADN20G | 13 |
| cTAR26 | 210 |
| cTAR33 | 1900 |

Dans un second temps nous avons étudié les propriétés de fixation de la protéine de nucléocapside et de la séquence SL3 ARN (20 nt). Il est connu que la NCp7 présente une forte affinité ($K_d \sim 100$ nM) pour la tige boucle SL3^{54, 352, 360, 562} qui est située dans la région 5' non codante de l'ARN génomique et qui possède le signal d'encapsidation Ψ ^{264, 563}. Nos résultats (Tableau 5C) confirment la forte affinité de NC(11-55) pour SL3 ($K_d = 66$ nM) ce qui est cohérent avec les résultats publiés précédemment par Athavale et al., 2010. Par la suite nous avons étudié l'affinité de la NCp7 pour l'homologue ADN de SL3 (20 nt), nous avons voulu évaluer l'impact de la substitution ARN/ADN pour une même séquence (SL3). Les données (Tableau 5C) indiquent que l'affinité de NC(11-55) pour la séquence SL3-ADN est très faible ($K_d = 1.1 \mu\text{M}$) comparée avec la séquence de SL3 ARN ($K_d = 66$ nM). Nous avons également réalisé une étude avec un mutant de SL3 ADN20G qui comporte un G supplémentaire dans la boucle apicale, qui contient donc 21 résidus (SL3 ADN20G) et qui montre, lui, une très forte affinité pour la NC(11-55).

Le complexe NC(11-55):SL3 ADN, est de faible affinité ($K_d = 1.1 \mu\text{M}$), on observe donc un fort contraste entre les versions ADN et ARN de SL3, il apparaît donc que SL3 ADN n'est pas une séquence favorable pour former un complexe de forte affinité et donc pour entreprendre une étude de RMN avec cette séquence. Afin d'interpréter correctement ces résultats, nous avons considéré les données structurales qui existent sur les complexes NC:Acides nucléiques et en analysant l'interaction avec l'ADN20G par RMN.

Les courbes de titration de SL3 ARN, SL3 ADN et cTAR26 avec la NC(11-55) sont représentées sur la figure ci-dessous :

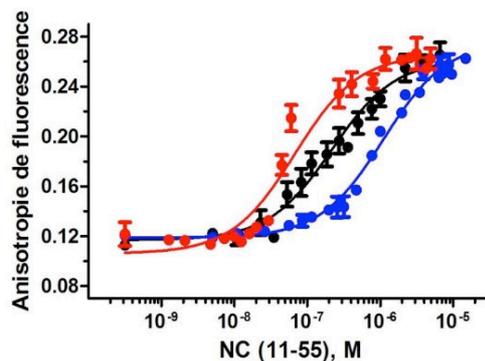


Figure 7C: Analyse quantitative de l'interaction de la NC(11-55) avec différents oligonucleotides en tige boucle. Titration d'anisotropie de fluorescence de SL3 ARN (rouge), SL3 ADN (bleu) et cTAR26 (noir) à 20 nM et avec des concentrations croissantes de protéine (de 10^{-9} à 10^{-6} M). Les échantillons sont dans tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$, force ionique $I=0.04$, $\text{pH}=6.5$ et à 5°C .

IV.2.2.1.2 Etude des complexes NC-acides nucléiques avec des séquences ADN courtes et simple brin

Après l'obtention de ces résultats, nous avons recherché si nous pouvions obtenir des complexes de meilleure affinité avec des séquences linéaires, courtes et dépourvues de la capacité de se replier. Certaines contiennent la séquence GXG (avec X un résidu variable), TXG, et TG, ceci afin d'évaluer la spécificité de la NCp7 et la stabilité des complexes. La longueur de la séquence (7 nucléotides) est un compromis entre une longueur suffisante pour avoir une bonne affinité (d'après les données de la littérature) et la nécessité de n'obtenir qu'un seul site par molécule afin d'empêcher les réactions d'équilibre dynamique entre plusieurs sites. Néanmoins, nous avons aussi testé deux séquences plus longues de 11 et 13 nt respectivement. Les valeurs des constantes de dissociation des ces oligonucléotides avec la NC(11-55) sont représentées sur le Tableau ci-dessous :

Tableau 6C. Valeurs des constantes de dissociation (K_d) des complexes NC(11-55)-acide nucléique étudiés par anisotropie de fluorescence avec des séquences ADN courtes, à température de 5°C et dans du tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ (pH=6.5, I=0.04). Voir Tableau 1B pour lire l'intégralité des séquences.

| Oligonucléotide | K_d (nM) avec NC(11-55) |
|-----------------|---------------------------|
| GCG | 148 |
| GCG11 | 63 |
| GCG13 | 38 |
| GGC | 85 |
| TCG | 250 |
| AGC | 1150 |
| TGC | 170 |
| TAC | 994 |
| ACG | 1800 |
| CCG | 1800 |
| GAG | 160 |
| atGAG | 90 |
| caGAG | 1950 |
| caGCT | 143 |
| caGGC | 440 |
| caTCG | 52 |
| caGTC | 130 |
| caTGC | 14 |

Les oligonucléotides ADN contenant GAG, GCG, TCG et TGC ont tous montré de bonnes affinités d'ordre similaire ($K_d \sim 200$ nM), ils contiennent les séquences GXG, TG ou TXG ; l'ADN avec la séquence GGC a démontré une affinité encore supérieure pour la protéine avec un K_d de 85 nM, ce qui n'était pas connu auparavant, puisque deux guanines sont présentes, mais ne sont pas séparées par un autre résidu, comme dans les séquences GXG précédemment identifiées dans la littérature. Par contre, les séquences TAC, AGC et ACG forment des complexes de très faible affinité avec des constantes de dissociations de l'ordre du μM . Ces dernières séquences ne contenaient pas les séquences préférées mentionnées plus haut. La forte affinité de la NCp7 pour les séquences TG et TXG est un résultat cohérent avec les études antérieures^{177, 384, 564}. Par contre, il n'avait jamais encore été reporté une préférence de la NCp7 pour des séquences avec deux guanines consécutives (voir la très forte affinité pour GGC). On peut faire l'hypothèse qu'un G serait inséré dans le deuxième doigt de zinc et que l'autre G établirait un réseau de liaisons hydrogène, des interactions hydrophobes ainsi qu'électrostatiques. Pour avoir plus d'informations et confirmer ces hypothèses, nous envisageons dans le futur l'étude RMN de ce complexe de forte affinité. Par ailleurs, en analysant l'affinité des oligonucléotides contenant GCG mais de longueur supérieure à 7 nt (11 et 13 nt) nous

avons observé que l'affinité de liaison de la NCp7 augmente pour ces derniers. Cela pourrait s'expliquer par le fait que d'autres molécules de NCp7 arrivent à se lier, vu que c'est un ADN simple brin contrairement au résultat obtenu avec cTAR33.

Quant aux oligonucléotides contenant les séquences TAC, AGC et ACG, il s'avèrent être de très mauvais substrats pour la NC(11-55). Ces observations confirment que les séquences de type GXG, TXG, TG et GGC sont nécessaires pour l'obtention d'une forte affinité. Les courbes de titration de GGC, GAG et ACG avec la NC(11-55) sont représentées sur la figure ci-dessous :

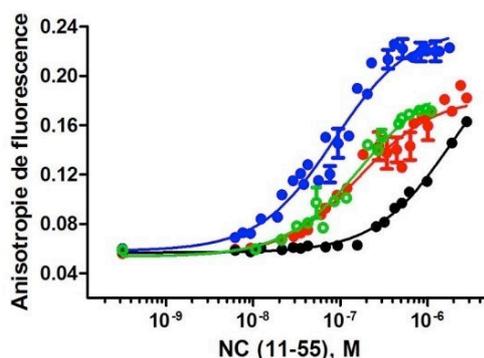


Figure 8C: Analyse quantitative de l'interaction de la NC (11-55) avec des ADNs courts et simple brin. Titration d'anisotropie de fluorescence des quatre oligonucléotides GGC (bleu), GAG (rouge), TGC (vert) et ACG (noir), à 20 nM et par des concentrations croissantes de protéine (de 10⁻⁹ à 10⁻⁶ M). Les échantillons sont dans du tampon NaH₂PO₄/Na₂HPO₄ (force ionique I=0.04, pH=6.5) et à 5°C.

Par la suite, nous avons étudié les séquences caTGC et caGTC afin de comparer l'affinité de la NC selon la polarité 5'-3' de la séquence (TG ou GT). Avec cette expérience, on relève un K_d beaucoup plus faible pour la séquence comportant TG (14 nM) par rapport à GT (130 nM). Ceci reflète donc une meilleure affinité de la NC(11-55) pour TG plutôt que GT (Tableau 6C et Figure 9C). La polarité de la séquence de l'ADN influencerait sur l'interaction de la NC aux acides nucléiques, ces résultats sont similaires à ceux obtenus dans une étude antérieure³⁸⁴.

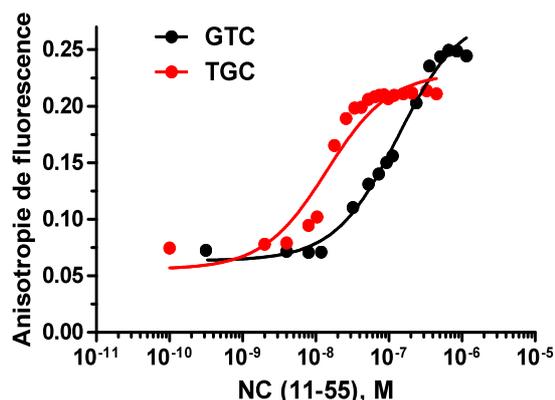


Figure 9C: Analyse quantitative de l'interaction de la NC (11-55) avec les ADNs courts et simple brin : caGTC et caTGC. Titration d'anisotropie de fluorescence des oligonucléotides, caTGC (rouge) et caGTC (noir), à 20 nM et par des concentrations croissantes de protéine (de 10^{-9} à 10^{-6} M). Les échantillons sont dans tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ (force ionique $I=0.04$, $\text{pH}=6.5$) et à 5°C .

IV.2.2.1.3 Etude des complexes NC-acides nucléiques avec des séquences ARN courtes et simple brin

Tous les oligonucléotides ARN courts ont montré de bonnes affinités d'ordre similaire. Cependant, en comparant avec leurs homologues ADN, on remarque que pour les oligonucléotides contenant les séquences GCT et GCU, il n'y a pas de différence significative du K_d , et donc de l'affinité. De même pour les oligonucléotides contenant les séquences TGC et UGC (Tableau 7C). On en conclut que pour ces séquences courtes d'oligonucléotides, l'affinité de la NC ne varie pas considérablement selon la nature de l'acide nucléique (ADN ou ARN).

Tableau 7C. Valeurs des constantes de dissociation (K_d) des complexes NC(11-55)-acide nucléique étudiés par anisotropie de fluorescence avec des séquences ARN courtes et leurs homologues ADN, à température de 5°C et dans du tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$, $\text{pH}=6.5$, $I=0.04$. Voir Tableau 1B pour lire l'intégralité des séquences.

| ARN | K_d (nM) avec NC(11-55) | ADN | K_d (nM) avec NC(11-55) |
|-------|---------------------------|-------|---------------------------|
| caGCU | 120 | caGCT | 440 |
| caUGC | 24 | caTGC | 14 |
| caGUC | 48 | caGTC | 130 |
| caUCG | 25 | caTCG | 52 |

IV.2.2.1.4 Etude des complexes NC-acides nucléiques à différentes concentrations de sels

Une des particularités de la NC est qu'elle est capable de se lier aux acides nucléiques via des interactions spécifiques ou non spécifiques. Les interactions spécifiques dépendent des motifs contenus dans la séquence de l'oligonucléotide, alors que les interactions non spécifiques, sont de types électrostatiques et sont donc souvent insensibles à la séquence, à la structure de l'acide nucléique et ont une forte dépendance aux sels. Nous avons donc réalisé des expériences d'anisotropie de fluorescence à différentes concentrations de sels ainsi qu'avec différentes séquences d'oligonucléotides. Le but de ces expériences est d'évaluer à la fois l'affinité et la spécificité de l'interaction entre les deux partenaires (NC:AN). En effet selon la concentration de NaCl dans le tampon, les affinités peuvent varier de façon significative ³⁸⁴. Les résultats sont représentés sur le Tableau ci dessous :

Tableau 8C. Valeurs des constantes de dissociation (K_d) des complexes NC(1-55)-acide nucléique étudiés par anisotropie de fluorescence à différentes concentrations de sels, à température de 5°C et dans du tampon $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$ (pH=6.5). Voir Tableau 1B pour lire l'intégralité des séquences.

| Oligonucléotide | [NaCl] (mM) | K_d (nM) |
|-----------------|-------------|------------|
| caTGC | 25 | 32 |
| | 75 | 58 |
| | 125 | 11 |
| caGTC | 25 | 22 |
| | 75 | 54 |
| | 125 | 55 |
| caGAG | 25 | 120 |
| | 75 | 160 |
| | 125 | 170 |

Les K_d ainsi obtenus ne fluctuent pas énormément pour caGAG et caTGC, ce qui implique que leur mode de reconnaissance est probablement très spécifique. Cependant, l'affinité de liaison pour caGTC a diminué de façon notable en présence de concentration élevée de sels, ce qui signifie que ces hautes concentrations en sels ont supprimé ou diminué la force des interactions de type ionique.

IV.2.2.1.5 Etude des complexes NC-acides nucléiques avec les deux formes de protéines (complète et tronquée)

Pour pouvoir mesurer l'anisotropie de fluorescence de la NC entière en présence d'oligonucléotide, nous n'avons pas pu continuer les expériences avec le tampon phosphate, à cause des problèmes de précipitations, dus à la partie N-terminale fortement basique. C'est pourquoi, dans la suite de nos expériences, nous avons utilisé le tampon acétate de sodium plus communément utilisé avec cette protéine dans la littérature. Les résultats obtenus (Tableau 9C) démontrent une plus haute affinité de liaison pour la forme complète NC(1-55) comparée à la forme tronquée NC(11-55), particulièrement pour la séquence caGAG et SL3 ADN. Cependant, pour les autres séquences testées, les différences ne sont pas très significatives.

Tableau 9C. Valeurs des constantes de dissociation (K_d) des complexes NC-acide nucléique, avec la protéine tronquée ou complète, étudiés par anisotropie de fluorescence à température de 5°C et dans du tampon Acétate (pH=6.5, I=0.04). Voir Tableau 1B pour lire l'intégralité des séquences.

| Oligonucléotide | Kd (nM) avec NC(11-55) | Kd (nM) avec NC(1-55) |
|-----------------|------------------------|-----------------------|
| SL3 ADN | 1100 | 140 |
| SL3 ARN | 66 | 180 |
| SL3 ADN20G | 13 | 41 |
| cTAR26 | 210 | 100 |
| caGTC | 130 | 58 |
| caTGC | 14 | 32 |
| caGAG | 1950 | 120 |

En conclusion, ces études par anisotropie de fluorescence, nous ont permis de connaître l'affinité de la protéine, sous ses différentes formes (tronquée et entière), avec différentes séquences d'acides nucléiques. L'ensemble des données disponibles dans la littérature et les résultats obtenus dans cette étude montre que la NCp7 peut montrer des affinités élevées pour des oligonucléotides ADN. En fait ce qui paraît déterminant pour l'obtention d'une forte affinité entre la NCp7 et des séquences ADN ou ARN, c'est non seulement la présence de certaines séquences (GXG, TXG, etc.), mais aussi la disponibilité des bases pour interagir avec la protéine. Une étude RMN effectuée dans le laboratoire a démontré que les guanines de la boucle apicale de SL3 ADN ne sont pas correctement accessibles, du fait d'interactions entre ces guanines et d'autres résidus de la boucle, alors que ce n'est pas le cas dans SL3 ARN. Pour cette dernière molécule, les

comparaisons entre la tige-boucle seule et complexée à l'ARN, montrent de façon très nette que les deux guanines de la boucle apicale, qui sont liées à la protéine dans le complexe, sont très accessibles et probablement très mobiles dans la structure de la tige-boucle seule^{360, 565}. Signalons aussi, pour appuyer ces hypothèses, que l'affinité de la NCp7 pour la structure tige-boucle SL4 ARN (qui possède également la séquence GAG dans la boucle apicale) est très faible ($K_d = 14 \mu\text{M}$) et qu'il a été montré que la boucle seule (sans NCp7) était très structurée et qu'aucune guanine n'était disponible³⁵².

A l'issue de ces expériences d'anisotropie de fluorescence, nous avons donc décidé de poursuivre l'étude par RMN du complexe NC(11-55) :d(TAGAGAT), car celui-ci possède une assez forte affinité. Bien que cette dernière fût du même ordre que GCG, mais comme des séquences ARN contenant la séquence GAG ont déjà été obtenus³⁶⁰, cela nous permettra d'effectuer des comparaisons. Par ailleurs, nous avons également procédé à l'étude de ce complexe avec son homologue ARN (UAGAGAU), avec comme objectif d'obtenir deux complexes réalisés dans les mêmes conditions pour pouvoir les comparer.

IV.2.2.2 Analyse par RMN des complexes

IV.2.2.2.1 Attribution de l'acide nucléique d(TAGAGAT)

Les attributions des résonances ^1H de l'ADN libre et complexé ont pu être réalisées en utilisant les spectres 2D TOCSY et NOESY ainsi que des spectres ^1H - ^{13}C de type HSQC. L'absence de résonance de protons échangeables pour l'oligonucléotide libre indique que celui-ci n'est pas replié, ce qui rend à priori relativement accessibles les différents résidus qui le compose. Néanmoins, il est possible d'observer quelques connectivités inter-résiduelles dans l'oligonucléotide libre, ce qui indique quelques éléments de structuration. Ces connectivités sont peu nombreuses, elles indiquent une structuration significative de la partie terminale G5A6T7 de l'oligonucléotide. Les résonances de d(TAGAGAT) dans le complexe ont été attribuées à partir des spectres NOESY en déduisant les signaux liés à la protéine.

Nous avons également eu recours au TOCSY pour confirmer les attributions des méthyles des Thymines T1 et T7. Aussi le concours des HSQC (Figure 10C, à droite), était nécessaire pour distinguer les protons H6 et H8 des bases, en effet le déplacement chimique des carbones aromatiques est fortement dépendant de la nature chimique de la base.

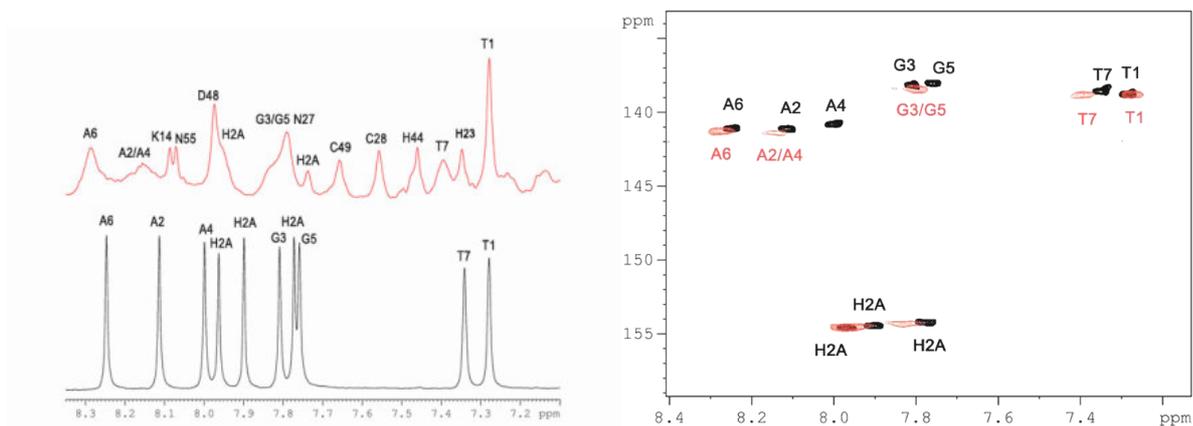


Figure 10C: Spectres 1D (à gauche) et 2D HSQC (à droite) de d(TAGAGAT) sous forme libre (noir) et dans le complexe (rouge), désignant la régions des protons aromatiques, à pH 6.5, dans H₂O, à 20°C et à 500 MHz.

La Figure 10 présente les spectres comparés du d(TAGAGAT) libre et du mélange équimolaire NC(11-55):d(TAGAGAT) avec les attributions réalisés avec les spectres 2D.

Les observations effectuées sur l'ensemble des spectres sont :

- les résidus T1 et T7 montrent peu de variations (surtout T1), l'ensemble des caractéristiques spectrales concernant ces deux résidus montre qu'ils sont peu affectés par l'interaction et ne sont donc probablement pas directement contactés. Les différences d'intensité importantes des protons aromatiques T1 et T7 sont associées à des différences de dynamiques des deux résidus à l'intérieur du complexe ;
- Les résonances de G3 et G5 sont significativement élargies dans notre température d'étude (20°C) et sont de plus superposés ce qui n'était pas le cas dans l'oligonucléotide libre, ce qui empêche à l'heure actuelle leurs attributions respectives. Celle-ci devra être effectuée en travaillant à d'autres températures et en utilisant un champ magnétique plus élevé afin d'augmenter la séparation des signaux. La similitude des déplacements chimiques suggère des environnements similaires pour les deux guanines ;
- La situation est similaire pour A2 et A4 dont l'attribution nécessitera également d'autres conditions. Le résidu A6 est légèrement décalé dans le complexe.

Globalement les données sont compatibles avec l'implication des résidus A2, G3, A4 et G5 dans des interactions avec la protéine, les résidus T1, A6, T7 sont moins affectés. Ceci est compatible avec l'existence d'un site de fixation unique constitué par la séquence AGAG, la bonne qualité des spectres est compatible avec cette hypothèse.

IV.2.2.2.2 Attribution de la protéine NC(11-55) non marquée

Les études des signaux de l'ADN suggèrent une complexation spécifique avec la protéine, l'analyse des résonances de la protéine doit nous confirmer cette interaction. Les attributions ont été faites dans la protéine libre et dans le complexe, ce qui nous permet d'observer les variations de déplacements chimiques subis par la protéine lorsqu'elle est complexée. Les régions NH-CαH qui sont parmi les plus caractéristiques et les plus informatives sont présentées (Figure 11C).

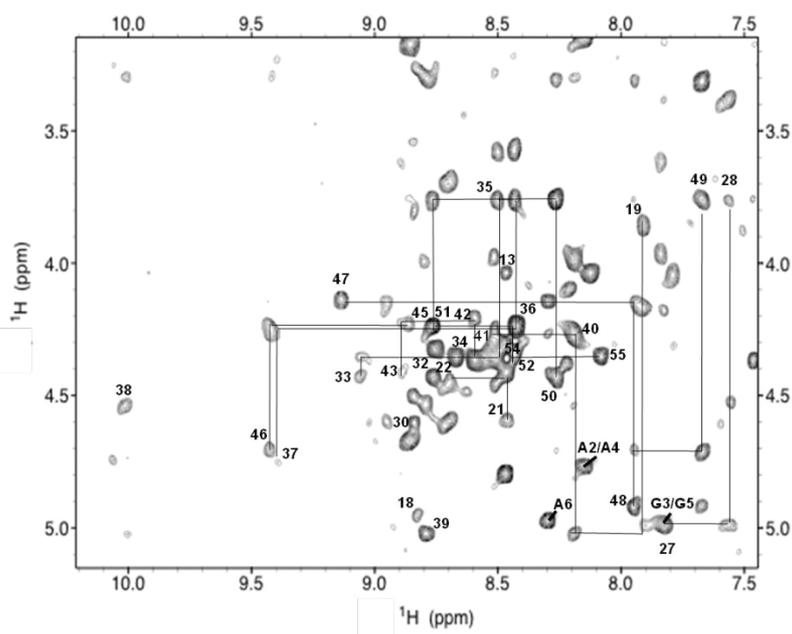


Figure 11C : Région des corrélations NH-CαH du spectre RMN 2D NOESY du complexe NC(11-55):d(TAGAGAT) dans H₂O à 20°C.

L'addition de la protéine à l'oligonucléotide d(TAGAGAT) a induit des changements conformationnels pour les deux partenaires. Pour la protéine, les modifications les plus importantes dans la protéine se produisent pour les résidus : Arg32, Lys33, Lys34, Gly35, Cys36, Trp37, Lys38, Cys39, Gly40, Gly43, His44, Met46 et Lys47. On peut donc identifier ces acides aminés comme constituant le site d'interaction de la NC(11-55) avec l'oligonucléotide. Nous avons comparé nos résultats avec ceux d'une étude effectuée entre un pentanucléotide d(ACGCC) et la NC(12-53) par Morellet et al., 1998. Les changements de déplacements chimiques les plus importants suite à la formation du complexe sont observés pour Val13, Phe16, Thr24, Ala35, Asn17, Arg32, Trp37, Gln45 et Met46. La comparaison des déplacements chimiques de la NC(11-55) et la NC(12-53) sont illustrés dans l'histogramme (Figure 12C). De façon remarquable les variations de déplacement chimique sont presque identiques dans les deux complexes. Ces données montrent sans ambiguïté, que la NC(11-55) s'est liée à notre oligonucléotide, en utilisant majoritairement

ses doigts de zinc. Les détails additionnels sur ce complexe seront obtenus dans les phases ultérieures de l'étude.

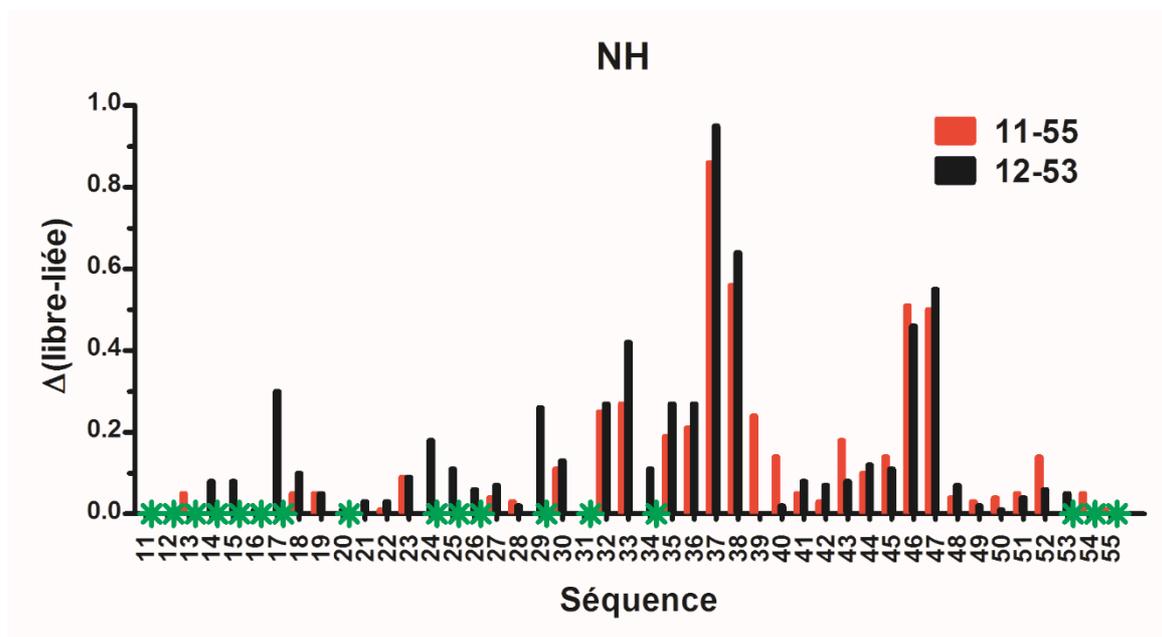


Figure 12C: Histogramme des différences de déplacements chimiques des protons de NC(11-55)-d(TAGAGAT) (rouge) et NC(12-55)-d(ACGCC) (Morellet et al., 1998) (noir) entre la forme libre et la forme complexée. Les étoiles en vert indiquent les données manquantes.

Les résultats RMN ont permis de montrer que c'est bien la séquence GAG qui interagit avec la protéine. Nos données sur le complexe indiquent que les résidus T1, A6 et T7 sont peu ou pas affectés lors de l'interaction, les résidus G3, G5, A2 et A4 sont les résidus principalement affectés. Les acides aminés du ZF2 ont subi des modifications importantes, tout particulièrement le Trp37, ce qui suggère l'empilement des bases avec le noyau indole de cette acide aminé. Les changements de déplacement chimique sont presque identiques avec ceux obtenus dans d'autres complexes, ce qui permet de confirmer la bonne interaction des deux partenaires. Cependant, Il reste encore à différencier les deux guanines et d'identifier les contacts intermoléculaires qui nous permettront de déterminer l'orientation de la NCp7 sur cet oligonucléotide. Nous avons donc poursuivi l'étude de ce complexe à champs magnétique plus fort (950 MHz) sauf que nous avons eu des problèmes de précipitations. C'est pour cela que nous avons décidé de poursuivre notre étude directement avec la méthode de RMN PRE, où l'oligonucléotide (ADN et son homologue ARN) sera marqué une fois à son extrémité 5' et une autre fois en 3' dans le but de voir l'effet paramagnétique sur la protéine et donc de déterminer dans quelle position chaque doigt de zinc est le plus touché.

IV.2.2.3 Analyse par RMN PRE

Les expériences ont été réalisées avec la NCp7 (1-55), marquée en ^{15}N , à 25°C, dans un tampon Acétate (pH 6.5), en utilisant un spectromètre 500 MHz et 600 MHz (équipé d'une cryosonde). Comme décrits dans la partie Matériels et Méthodes, les oligonucléotides sont modifiés à l'une des positions (5' ou 3') par une dThio_T(ADN) ou une Thio_U (ARN), qui par la suite sera couplé à l'agent paramagnétique (Proxyl).

Dans un premier temps, nous avons réalisé l'attribution de l'ADN 5'S4-TAGAGAT en réalisant des 2D NOESY et TOCSY, afin de vérifier que celui-ci n'a pas changé dû à la d_thio_T en 5'. Dans un second temps, nous avons réalisé des spectres de la protéine marquée (^{15}N) seule, afin de procéder à son attribution. Nous avons donc enregistré des spectres 2D ^1H - ^{15}N HSQC, TROSY, NOESY découplée et 3D NOESY-HSQC. La figure 14C représente le spectre ^1H - ^{15}N TROSY de la NCp7 seule et j'ai réussi à attribuer la totalité de la protéine.

Par la suite, nous avons réalisé des spectres des complexes NCp7:ADN ou ARN modifiés sans proxyl ; pour procéder à l'attribution de la protéine dans le complexe et ainsi vérifier que le complexe NC :5'S4-TAGAGAT n'a pas changé due à la base modifié. On observe que les spectres des deux complexes ARN (5'S4U et 3'S4U) sont exactement pareils. Cependant, les deux complexes ADN (5'S4T et 3'S4T) ont affiché des différences sur certains résidus avec une disparition de la partie N-terminale de la NCp7 dans le complexe 5'S4T_D (dû à un échange avec l'eau). L'attribution de ces complexes est toujours en cours.

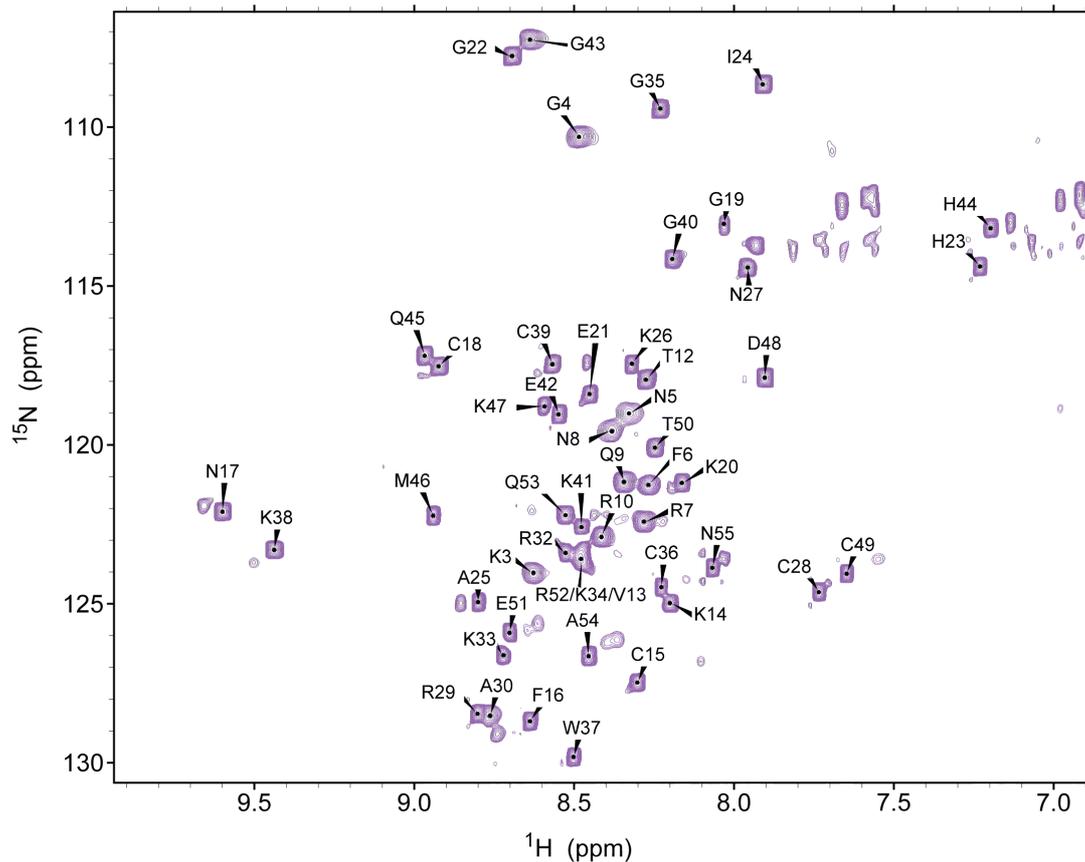


Figure 13C: Spectres 2D ^1H - ^{15}N TROSY de la protéine sous forme libre, à une concentration de 1 mM, dans du tampon Acétate deutéré (pH 6.5), à 25°C et à 600 MHz.

Par la suite, nous avons réalisé les complexes NCP7:ADN-proxyl, avec un ratio 1:1.2 où la concentration de la protéine était fixée à 0.2 mM. Pour commencer, on enregistre des spectres 2D ^1H - ^{15}N HSQC ou TROSY avec l'agent paramagnétique dans son état non réduit (paramagnétique) et on observe les acides aminés qui ont disparus où dont l'intensité a fortement diminué. Ensuite, on réduit l'agent paramagnétique par ajout d'Acide Ascorbique, avec une concentration dix fois supérieure à celle de l'acide nucléique. Les spectres des complexes dans les deux états (paramagnétique et diamagnétique) sont enregistrés dans les mêmes conditions pour pouvoir les comparer.

En résultat, nous observons certains pics qui disparaissent ou dont l'intensité diminue fortement, due à la proximité spatiale avec la sonde paramagnétique. En principe, après la réduction suite à l'ajout de l'acide ascorbique, les pics réapparaissent.

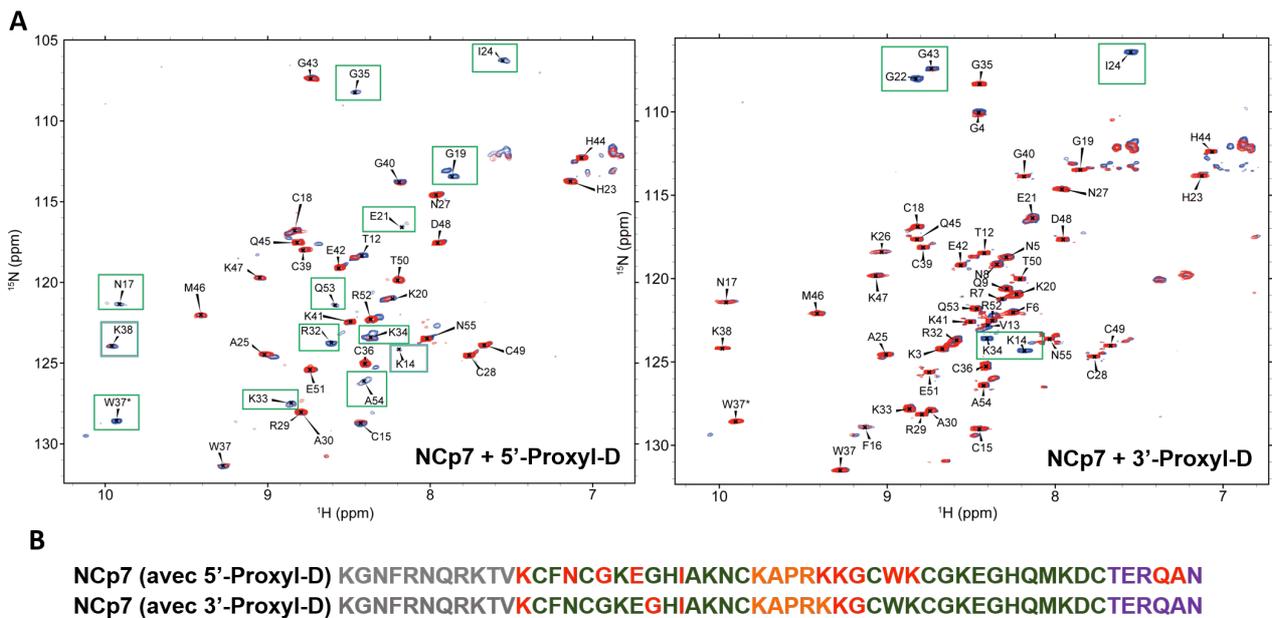


Figure 14C: Spectres 2D ^1H - ^{15}N TROSY de la protéine sous forme complexée à l'ADN 5' et 3' Proxyl-D. L'échantillon est à une concentration de 1 mM ratio 1 :1, dans tampon Acétate deutéré, à pH 6.5, à 25°C et à 600 MHz. **A)** Complexes sous forme para- et dia- magnétique (en rouge et bleu respectivement) sont superposés. Les pics encadrés en vert désignent les résidus les plus affectés par l'agent paramagnétique. **B)** Séquence de la protéine NCp7, où la partie N-terminale est désignée en gris, les doigts de zinc en vert, le linker en orange et la partie C-Terminale en mauve. Les résidus désignés en rouge sont ceux les plus touchés par l'effet paramagnétique dans le complexe.

Les résidus qui sont touchés dans les deux complexes sont : K14 (N-Terminale), I24 (ZF1), ainsi que K34 et G35 (Linker). Cependant, qualitativement, on voit bien que le complexe avec l'ADN 5'Proxyl-D est plus touché par l'agent paramagnétique, avec en plus des résidus cités précédemment : N17, G19 et E21 (ZF1), K33 (Linker), W37 et K38 (ZF2), Q53 et A54 (C-Terminale) qui sont affectés. Ces résultats indiquent que l'extrémité 5' est proche spatialement aussi bien du premier que du deuxième doigt de zinc et qu'elle arrive même à être en contact avec l'extrémité C-Terminale de la protéine. Globalement, l'image qui en ressort, c'est que ce heptanucléotide d(TAGAGAT), est orienté de manière centrée entre les deux doigts de zinc, de la même manière que le pentanucléotide étudié par Morellet et al, 1998. En s'appuyant sur la structure de ce dernier complexe, on peut très bien imaginer que notre oligonucléotide étant plus lent (deux bases en plus), ce qui rendrait l'extrémité 3' éloignée du deuxième doigt de zinc (Figure 15C). En analysant, les distance entre le H2 de la base en 5' du pentanucléotide d(ACGCC) avec les protons amides (HN) de tous les résidus de la NC(12-53) dans le complexe (Figure 16C), on s'aperçoit qu'on retrouve bien de faibles distances avec le doigt de zinc 1 ainsi que le K33 situé sur le Linker, W37 et K38 sur le deuxième doigt de zinc.

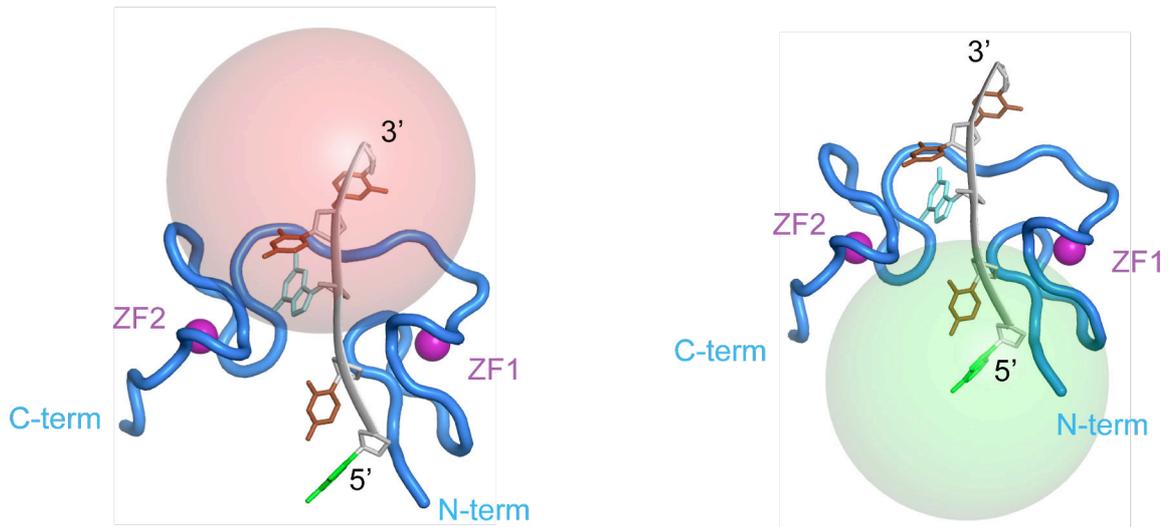


Figure 15C: Structure tridimensionnelle de la NCp7 avec le pentanucléotide ADN d(ACGCC) (code PDB : 1BJ6). Les bases de l'ADN sont indiquées en vert pour l'adénine, marron pour la cytosine et cyan pour la guanine. Les sphères en rose et vert (extrémité 3' et 5' respectivement) indiquent l'espace qui pourrait être affecté par l'effet paramagnétique (12 Å).

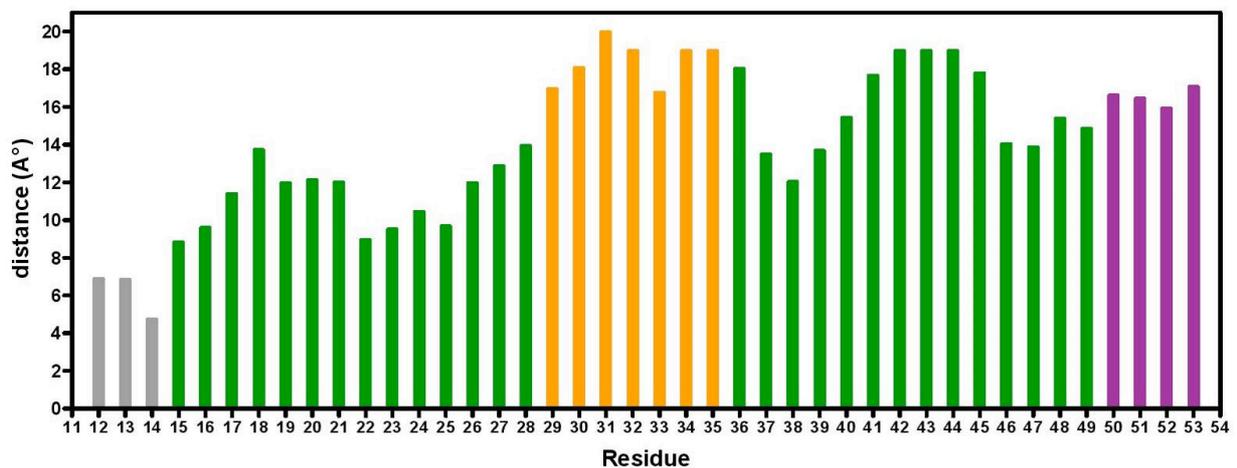


Figure 16C: Histogramme des distances entre le H2 de la base en 5' du pentanucléotide ADN d(ACGCC) avec les HN de chaque acide aminé de la NC(12-53) (code PDB : 1BJ6). Le code couleur est comme tel : la partie N-terminale est désignée en gris, les doigts de zinc en vert, le linker en orange et la partie C-Terminale en mauve.

En perspectives, pour poursuivre notre analyse, il nous faudra traiter quantitativement ces résultats, et analyser les complexes NCp7:Proxyl-ARN. Ce travail est toujours en cours au laboratoire.

IV.3 Etude structurale et dynamique des différentes formes de protéines de nucléocapsides du VIH-1

Les résultats que nous avons obtenus avec cette étude ont donné lieu à deux publications. Les travaux ont été effectués en coopération étroite avec le groupe du Dr Carine Tisné de l'IBPC et je suis co-premier auteur de l'une des publications. Je me suis plus particulièrement impliquée dans l'attribution des signaux de la NCp9 et de la NCp15 ainsi que dans l'obtention des paramètres dynamiques par RMN (T1, T2, NOE, dispersion de relaxation CPMG). Les résultats obtenus sont résumés ci-après.

La polyprotéine Pr55 Gag orchestre l'assemblage et le bourgeonnement des virions^{133, 224} grâce à ses multidomains (la matrice (MA), la capsid (CA), p2, nucléocapside (NC), p1 et le domaine p6) où chacun joue un rôle précis au cours de cette étape du cycle viral.

Le domaine NC de Gag constitue un site majeur d'interaction avec l'ARN, ce qui permet l'encapsidation de l'ARNg, ceci grâce à une fixation spécifique de Gag sur le site d'encapsidation psi (SL1 à SL4). Par ailleurs, au cours du bourgeonnement des particules virales, Gag interagit avec des facteurs cellulaires et des protéines impliquées dans l'ESCRT^{284-285, 287-290} via son domaine p6, qui grâce à son motif PTAP se lie à la protéine Tsg101²⁹⁶⁻³⁰⁰, par le biais de son motif LYPLASL à la protéine Alix²⁹⁹⁻³⁰¹ et avec la protéine virale²⁸⁶ Vpr via son motif LRSFLG³⁰²⁻³⁰³.

D'autre part, la maturation de la protéine de nucléocapside est d'une importance majeure au cours du cycle viral^{211, 310-313, 427}, avec des conséquences dramatiques sur les propriétés de l'activité de protéine chaperon de la NC, qui diffèrent entre les trois formes de protéines de nucléocapside^{237, 311}.

La NCp15 est la protéine qui présente l'activité de protéine chaperon la plus faible. Il a été proposé que son domaine p6 module les fonctions de son domaine NC, par un mécanisme de repliement de p6, qui est riche en acides aminés chargés négativement, sur le domaine NC, qui est riche en acides aminés basique ("fold-back conformation")³¹³. Il faut noter aussi que la plupart des études précédentes menées sur Gag *in vitro* ont été réalisées sans le domaine p6 du fait de la difficulté de produire et de purifier Gag sous sa forme entière.

Au cours de nos travaux, nous avons étudié le paysage conformationnel du domaine NC au sein des formes précurseurs (NCp15 et NCp9) ainsi que sous sa forme mature (NCp7). Pour cela, les trois formes de protéines de nucléocapside ont été

surexprimées et doublement marquées en ^{15}N et ^{13}C afin de procéder à leur attribution. Par la suite nous avons étudié leur dynamique globale et interne en utilisant des expériences de relaxation ^{15}N sous forme de protéine libre ou liée à la séquence SL3 ARN (contenant le signal d'encapsidation psi).

La protéine a été surexprimée et doublement marquée en ^{15}N et ^{13}C pour permettre d'effectuer des expériences en deux dimensions (2D ^1H - ^{15}N TROSY, HSQC, 2D ^1H - ^{13}C HSQC, NOESY découplé en ^{15}N) et une série d'expériences tridimensionnelles qui permettent de procéder à l'attribution du squelette peptidique (HNCO, HNCACO, HNCA, HNCOCA, HNCACB, HNCOCACB) et des chaînes latérales (HCCH TOCSY, NOESYHSQC) des protéines. Nos expériences ont été réalisées à 500, 600 et 950 MHz dans du tampon Acétate deutéré à pH=6.5 et à température de 10°C.

En observant le spectre 2D ^{15}N - ^1H TROSY, on remarque une faible dispersion des résonances des pics des domaines p1-p6, ce qui est caractéristique d'une région intrinsèquement désordonnée (IDD) (Figure 2, article 2). Par ailleurs, en utilisant les déplacements chimiques du squelette peptidique, nous avons calculé le score SSP (secondary structure propensity), qui indique que plusieurs régions dans p1-p6 possèdent une propension significative à former des structures en hélices α .

L'analyse de la relaxation de l'azote ^{15}N (T1, T2 et HetNOE) nous informe sur la dynamique conformationnelle globale de la protéine. De façon très schématique, on peut dire, concernant les valeurs de T1 et HetNOE, que plus elles sont élevées plus cela signifie que la région est structurée et inversement pour les T2. D'après les résultats obtenus on retrouve les zones de structuration déjà identifiées par les méthodes SSP. Globalement, les domaines p1 et p6 apparaissent désordonnés tout en montrant quelques régions de structuration partielle. La conversion des données T1, T2 et HetNOE en densité spectrale à la fréquence zéro "J(0)" nous permet d'obtenir des informations sur la structuration de la protéine. Des valeurs maximales ont été observées pour les résidus L57, I60, W61, N70, F71, E83, E84, S85, F86, L109 et A11. Nos données expérimentales sont en accord avec les prédictions structurales SSP, la NCp15 est globalement désordonnée, avec p1-p6 qui présente certains éléments de structurations.

Nous avons également réalisé des expériences de dispersion de relaxation ^{15}N CPMG (pour les trois formes de protéines NCp7, NCp9 et NCp15) à 950 MHz et 600 MHz, à une température de 10°C, avec un délai de relaxation de 40 ms où durant cette période sont appliqués les pulses CPMG. 20 expériences ont été acquises, dans chaque expérience, le délai de relaxation est constant d'une valeur de 40 ms mais la fréquence des pulses CPMG est modifiée (de 25 à 1000 Hz). Les intensités de pics ont été converties en vitesse

de relaxation R_{2eff} comme il est décrit dans la partie "matériels et méthodes" de l'article 3. Les résultats indiquent que la seule forme protéique qui montre des vitesses d'échanges significatives ($R_{ex} > 10 \text{ s}^{-1}$) est la forme mature NCp7, particulièrement pour ses résidus N17, K33, G35, W37, K38, C39, G43 et K47 (Figure 5A et Figure supplémentaire 9-11 de l'article 3), il s'agit donc une majorité de résidus située sur le ZF2. Ces données démontraient que la NCp7 présente un échange conformationnel entre un état fondamental majoritairement peuplé et un état "excité" faiblement peuplé. Les autres formes NCp9 et NCp15 ne montrent pas de tels échanges.

L'étude de l'interaction des protéines avec la séquences SL3 ARN démontre que les sites de fixation sont les mêmes pour les trois protéines, que l'affinité reste identique mais que les paramètres thermodynamiques de l'interaction changent de façon importante.

Pour la première fois nous avons réalisé l'étude de la protéine NCp15 complète. Nous avons démontré que p1-p6 au sein de la NCp15 est désordonnée en solution (faibles dispersion des résonances H-N sur la 2D TROSY ou HSQC, séquence en aa riche en Prolines et en résidus chargés négativement, analyse des paramètres dynamique notamment $J(0)$ qui démontre que p6 est globalement désordonnée) et que la protéine n'adopte pas une conformation repliée ("fold-back") sur le domaine NC, cependant, des hélices α se forment de manière transitoire suite au contact avec le domaine NC, impliquant des interactions de nature électrostatique, vu que p6 est chargé négativement alors que le domaine NC est chargé positivement. Grâce aux expériences de dispersion de relaxation CPMG nous avons identifié les résidus impliqués dans la sélection conformationnelle de la NCp7 pour se lier aux acides nucléiques, aptitude exclusive à la forme mature de nucléocapside.

En conclusion, le processus de maturation de la protéine de nucléocapside module les propriétés de liaison de la NC aux acides nucléiques. Notre travail apporte une meilleure compréhension des propriétés dynamiques et structurales de la nucléocapside au cours de l'étape d'assemblage, de bourgeonnement et de maturation des nouveaux virions (Figure 7 de l'article 3).

Article 2

^1H , ^{13}C and ^{15}N backbone and partial side-chain resonance assignments of the C-terminal domain of HIV-1 Pr55^{Gag} encompassed in NCp15.

Valéry Larue^{1*}, Marjorie Catala^{1,2}, Anissa Belfetmi³, Lousiné Zargarian³, Olivier Mauffret³, Carine Tisné^{1,2*}.

(1) Laboratoire de Cristallographie et RMN biologiques, CNRS, Paris Sorbonne Cité, Unité mixte de recherche 8015, 4 avenue de l'Observatoire, 75006 Paris

(2) Present address: Laboratoire d'Expression génétique microbienne, IBPC, CNRS, Paris Sorbonne Cité, Unité mixte de recherche 8261, 13 rue Pierre et Marie Curie, 75005 Paris.

(3) Laboratoire de Biologie et Pharmacologie Appliquée, Ecole Normale Supérieure de Cachan, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8113, Cachan, France.

* corresponding authors: valery.larue@parisdescartes.fr ; carine.tisne@cns.fr

Abstract

During HIV-1 assembly, the Pr55^{Gag} polyprotein precursor (Gag) interacts with the genomic RNA, with lipids of the plasma membrane, with host proteins (ALIX, TSG101) through the ESCRT complex, with the viral protein Vpr and are involved in intermolecular interactions with other Pr55^{Gag} proteins. This network of interactions is responsible for the formation of the viral particle, the selection of genomic RNA and the packaging of Vpr. The C-terminal domain of Gag encompassed in NCp15 is involved in the majority of these interactions, either by its nucleocapsid or its p6 domains. We study the NCp15 protein as a model of the C-terminal domain of Gag to better understand the role of this domain in the assembly and budding of HIV-1. Here, we report the ^1H , ^{13}C and ^{15}N chemical shift assignments of NCp15 obtained by heteronuclear multidimensional NMR spectroscopy as well as the analysis of its secondary structure in solution. These assignments of NCp15 pave the way for interaction studies with its numerous partners.

Keywords: HIV-1, NCp15, NMR, Nucleocapsid, Resonance assignment.

Biological context

Gag polyprotein precursor is composed of the following structural protein domains: the matrix (MA), the capsid (CA), a spacer peptide p2, the nucleocapsid domain (NC_d), a hydrophobic peptide p1, and an acidic p6 domain. The C-terminal domain of Pr55^{Gag}

containing the NC_d, p1 and p6 are responsible for numerous interactions, either by NC_d or p6 during assembly and budding of virions. The NC_d within Gag is the major site of interaction with gRNA, even if MA is also able to interact with nucleic acids^{24, 566-568}. Besides, p6 harbours motifs responsible for the interactions with proteins belonging to the human ESCRT machinery: the P₇₈TAP₈₁ sequence is involved in the binding of TSG101⁵⁶⁹ and the L₁₀₆YPLASL₁₁₂ sequence in the binding of ALIX⁵⁷⁰. ALIX also interacts with the NC_d, however the exact function of this interaction is not clearly understood^{284, 571-572}. p6 is also involved *via* its conserved L₁₁₂RSLFG₁₁₇ motif in the binding of the retroviral protein Vpr⁵⁷³⁻⁵⁷⁵.

During HIV-1 maturation, three different forms of NC_d appear successively. The first cleavage by the HIV-1 protease occurs between MA-CA-p2 and NC_d-p1-p6, thereby liberating NCp15 (NC_d-p1-p6, 123 amino acids). The second cleavage frees NCp9 (NC_d-p1) whereas the last one releases the NC_d itself, which constitutes the final maturation form of NC_d, also called NCp7. NCp7 and p6 were previously studied separately by NMR. In the HIV-1 life cycle, the nucleic-acid chaperone activity of the mature NC_d is essential during viral replication and notably during reverse transcription^{42, 427}. NCp7 is a small basic protein of 55 residues containing two zinc-knuckles (ZKs) of the consensus sequence CX₂CX₄HX₄C separated by a small linker of 7 residues. Our aim is to study the NCp15 protein as a model of the C-terminal domain of Pr55^{Gag} to better understand the role of this domain in the assembly, budding and maturation of HIV-1. We report here the chemical shift assignments of HIV-1 NCp15 protein and analyze its secondary structure.

Methods and experiments

Protein expression and purification

¹⁵N and ¹³C-labeled NCp15 protein was expressed and purified in *E. coli* as previously described (Mouhand *et al.*, manuscript in preparation). NMR samples of NCp15 were prepared at a concentration of ~0.8 mM, ¹⁵N-labeled or ¹⁵N/¹³C-labeled, in a 25 mM deuterated acetate buffer pH 6.5 containing 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM β-mercaptoethanol and 10% ²H₂O.

NMR experiments

All NMR experiments were recorded at 10°C on Bruker AVANCE III HD 600 MHz and 950 MHz equipped with a 5 mm TCI cryoprobe with z-gradients. Internal TSP standard was used for direct referencing of the ¹H chemical shifts, and indirect referencing of ¹⁵N and ¹³C shifts. Backbone assignments were obtained using the standard and/or rapid BEST-TROSY versions ⁵⁷⁶ of the following 3D NMR experiments : HNCA, HNCACB, HNCO, HNCACO and BEST-TROSY HNCANH NMR experiment (Solyom *et al.*, 2013). [¹H-¹³C] and [¹H-¹⁵N] 3D NOESY-HSQC, 3D TOCSY-HSQC and 3D HCCH-TOCSY were recorded to validate and complete the assignments. NMRpipe ⁵⁷⁷ and Sparky ⁵⁷⁸ softwares were used to process 2D and 3D experiments. CcpNmr Analysis program (<http://www.ccpn.ac.uk/software/analysis>) was used to compare NMR data like Nuclear Overhauser Effects (NOEs) and secondary chemical shifts of H α , C α , C β and CO. Dangle software ⁵⁷⁹ included in CCPN software ⁵⁸⁰ was used to predict secondary structure of the NCp15 protein.

Resonance assignments and data deposition

NMR assignments were based on 3D heteronuclear NMR experiments performed on $^{15}\text{N}/^{13}\text{C}$ -labeled NCp15. Backbone amide ^1H - ^{15}N resonance assignment of NCp15 at 10°C was achieved for 105 over 112 non-proline residues. In addition, 84.5% of $\text{C}\alpha$, 75.6% of $\text{C}\beta$ and 80.5% of CO were assigned. Amide groups from residues F6, R7, K11, Q53, S63, H64, S74 could not be assigned. Side chains ^1H and ^{13}C were assigned for which connections were only observed up to the β position. Partial assignments of side chains were possible for carbons of the 11 prolines. No assignments could be achieved for $\text{C}\alpha$, $\text{C}\beta$, CO and $\text{H}\alpha$ of the first ten residues. The ^1H - ^{15}N HSQC spectrum of NCp15 (Fig. 1) shows a low dispersion for most of the amide groups of the protein (1 ppm) indicating that one part of the protein could be disordered.

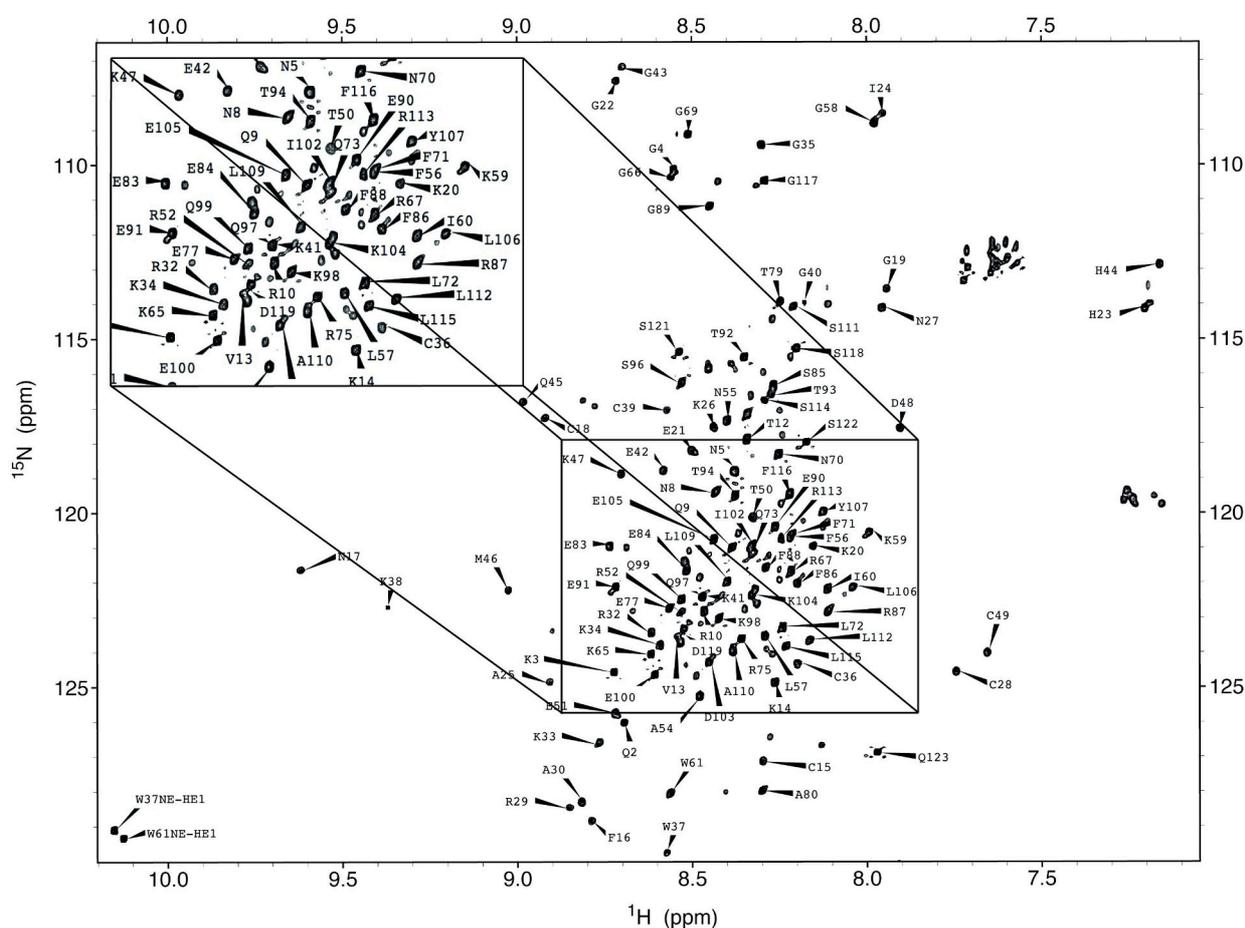


Fig. 1 ^1H - ^{15}N HSQC spectrum of HIV-1 NCp15 recorded at 10°C and 950 MHz. Backbone resonance assignments of amide groups are indicated by one letter amino acid code.

NOEs connectivities $d(\text{H}^{\text{N}}-\text{H}\alpha)$, $d(\text{H}^{\text{N}}_n-\text{H}^{\text{N}}_{n+/-1})$, $d(\text{H}^{\text{N}}-\text{H}\beta)$, $d(\text{H}^{\text{N}}_n-\text{H}^{\text{N}}_{n+/-2})$ and $d(\text{H}\alpha-\text{H}^{\text{N}}_{n+/-2})$ were deduced from $[^1\text{H}-^{15}\text{N}]$ 3D NOESY-HSQC experiment using two mixing time (80ms and 150ms) (Fig 2). With 343 NOE peaks found, 16% were strong, 56% were medium, 24% were weak and 4% with a very weak intensity. Some NOE cross peaks $d(\text{H}^{\text{N}}_n-\text{H}^{\text{N}}_{n+2})$ and $d(\text{H}\alpha-\text{H}^{\text{N}}_{n+2})$ have been found in the following short regions (F16-G19, E21-G22, K26-

C28, K38-G40, D48-C49, D103-E105 and E105-Y107) of NCp15 tending to a α -helical structuration.

Secondary chemical shifts for $H\alpha$ ⁵⁸¹ and for $C\alpha$, $C\beta$, CO ⁵⁸² were calculated as the difference between measured chemical shifts for NCp15 and random-coil chemical shifts ⁵⁸³ for each amino acid and were corrected for next-neighbor effects ⁵⁸⁴. The maximum variation of chemical shift for $C\alpha$, $C\beta$, CO are +4.5 ppm for positives values and -4.8 ppm for negatives values while for the $H\alpha$ proton variation, the maximum values are between 0.5 and -0.9 for positives and negatives values respectively. The results are presented in the chart table in Fig. 2. Positive values of secondary chemical shifts are indicative of an α -helical conformation whereas negative values corresponds to a β -strand conformation. Combining both 1H and ^{13}C chemical shifts allowed to calculate the Chemical Shift Index (CSI) for each residue (Wishart, et al., 1992) (Fig. 2). Globally, all these data converge to show that besides the two zinc fingers (residues 13 to 51), p1-p6 (residues 55 to123) is globally disordered. This is consistent with the DANGLE prediction (Fig. 2), that confirmed NCp15, is globally disordered, except for residues involved in the zinc fingers and for a short parts (E84-F86, D103-E105, S111-L112) of p6 region which could be the beginning of an helicoidal organisation. Previous NMR studies, when p6 was observed separately, in various media or in hydrophobic environment, identified helices in same regions of C terminal domain. (Solbak et al., 2013^{574, 585}. The chemical shift values for 1H , ^{13}C and ^{15}N resonances of NCp15 were deposited in the BioMagResBank (www.bmrb.wisc.edu) under accession number 26843.

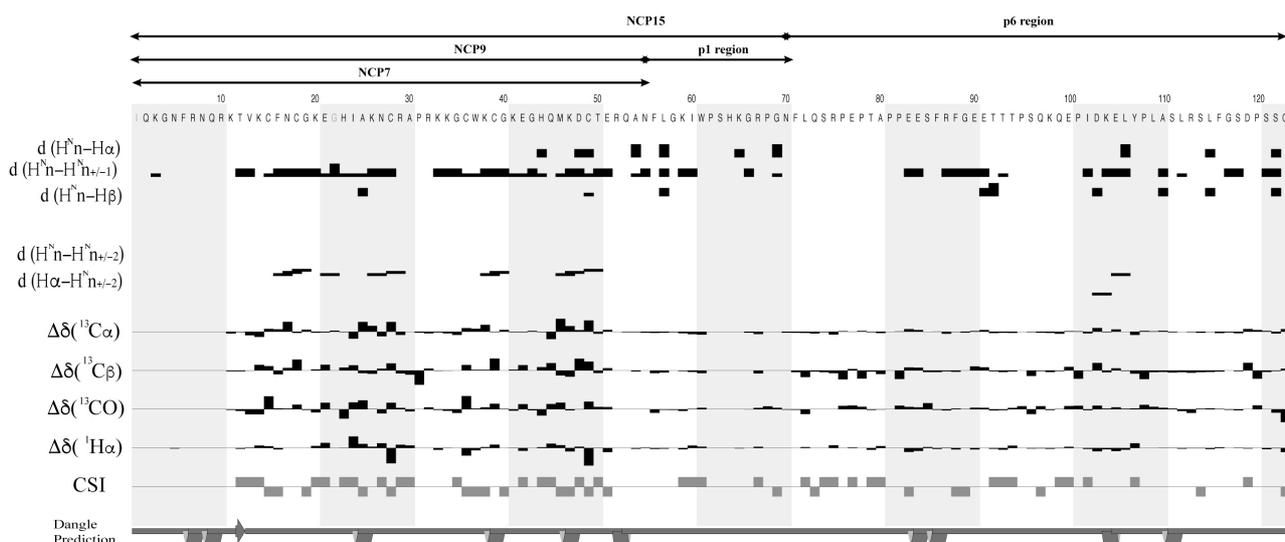


Fig. 2 : Chart table of NCp15 NOE connectivities $d(H^N-H\alpha)$, $d(H^N_n-H^{N_{n+1}})$, $d(H^N-H\beta)$, $d(H^N_n-H^{N_{n+2}})$, $d(H\alpha-H^{N_{n+2}})$ and secondary structure from analysis of secondary chemical shifts $\Delta\delta(^1H\alpha)$, $\Delta\delta(^{13}C\alpha)$, $\Delta\delta(^{13}C\beta)$ and $\Delta\delta(^{13}CO)$. Chemical Shift Index (CSI) indicate a global value for the secondary structure (-1 value for alpha helix and +1 value for beta strand). DANGLE (Dihedral Angles from

Global Likelihood Estimates) program predict protein backbone angles and secondary structure of the protein. This chart table has been generated using the CcpNmr Analysis program (<http://www.ccpn.ac.uk/software/analysis>).

Acknowledgments

This work was supported by grants from SIDACTION and ANRS. We thank Nelly Morellet and Ewen Lescop for technical and helpful discussion. Financial support from the TGIR-RMN-THC FR3050 CNRS for conducting the research is gratefully acknowledged.

References

- Alfadhli, A., Still, A., and Barklis, E. (2009). Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *J Virol* 83, 12196-12203.
- Bachand, F., Yao, X.J., Hrimech, M., Rougeau, N., and Cohen, E.A. (1999). Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *J Biol Chem* 274, 9083-9091.
- Cheung, M.S., Maguire, M.L., Stevens, T.J., and Broadhurst, R.W. (2010). DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* 202, 223-233.
- Chukkapalli, V., Oh, S.J., and Ono, A. (2010). Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc Natl Acad Sci U S A* 107, 1600-1605.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293.
- Demirov, D.G., Ono, A., Orenstein, J.M., and Freed, E.O. (2002). Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proc Natl Acad Sci U S A* 99, 955-960.
- Deshmukh, L., Ghirlando, R., and Clore, G.M. (2015). Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proc Natl Acad Sci U S A* 112, 3374-3379.
- Dussupt, V., Javid, M.P., Abou-Jaoude, G., Jadwin, J.A., de La Cruz, J., Nagashima, K., and Bouamr, F. (2009). The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding. *PLoS Pathog* 5, e1000339.

- Dussupt, V., Sette, P., Bello, N.F., Javid, M.P., Nagashima, K., and Bouamr, F. (2011). Basic residues in the nucleocapsid domain of Gag are critical for late events of HIV-1 budding. *J Virol* 85, 2304-2315.
- Fossen, T., Wray, V., Bruns, K., Rachmat, J., Henklein, P., Tessmer, U., Maczurek, A., Klinger, P., and Schubert, U. (2005). Solution structure of the human immunodeficiency virus type 1 p6 protein. *J Biol Chem* 280, 42515-42527.
- Goddard, T.D., and Kneller, D.G. SPARKY 3. University of California, San Francisco.
- Lee, S., Joshi, A., Nagashima, K., Freed, E.O., and Hurley, J.H. (2007). Structural basis for viral late-domain binding to Alix. *Nat Struct Mol Biol* 14, 194-199.
- Lescop, E., Schanda, P., and Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187, 163-169.
- Salgado, G.F., Marquant, R., Vogel, A., Alves, I.D., Feller, S.E., Morellet, N., and Bouaziz, S. (2009a). Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. *Biochemistry* 48, 2355-2367.
- Salgado, G.F., Vogel, A., Marquant, R., Feller, S.E., Bouaziz, S., and Alves, I.D. (2009b). The role of membranes in the organization of HIV-1 Gag p6 and Vpr: p6 shows high affinity for membrane bilayers which substantially increases the interaction between p6 and Vpr. *J Med Chem* 52, 7157-7162.
- Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 123, 2970-2978.
- Sette, P., Dussupt, V., and Bouamr, F. (2012). Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA. *J Virol* 86, 11608-11615.
- Shkriabai, N., Datta, S.A., Zhao, Z., Hess, S., Rein, A., and Kvaratskhelia, M. (2006). Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry* 45, 4077-4083.
- Sleiman, D., Goldschmidt, V., Barraud, P., Marquet, R., Paillart, J.C., and Tisne, C. (2012). Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res* 169, 324-339.
- Solbak, SM., Reksten, TR., Hahn, F., Wray, V., Henklein, P., Henklein, P., Halskau, Ø., Schubert, U., Fossen, T. T., (2013). HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein. *Biochim Biophys Acta*. 1828(2), 816-823.
- Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., Brutscher, B. (2013). BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 55(4):311-321.

- Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687-696.
- Wishart, D.S., and Sykes, B.D. (1994). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* 4, 171-180.
- Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31, 1647-1651.
- Wu, T., Gorelick, R.J., and Levin, J.G. (2014). Selection of fully processed HIV-1 nucleocapsid protein is required for optimal nucleic acid chaperone activity in reverse transcription. *Virus Res* 193, 52-64.
- Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* 25, 173-195.

Article 3

Modulation of the HIV-1 nucleocapsid conformational landscape along its maturation pathway

Assia Mouhand^{1,2§}, Anissa Belfetmi^{3§}, Marjorie Catala^{1,2}, Valéry Larue¹, Loussiné Zargarian³, Franck Brachet¹, Robert J. Gorelick⁴, Serge Bouaziz¹, Carine Van Heijenoort⁵, Gilles Mirambeau⁶, Pierre Barraud^{1,2}, Olivier Mauffret³ & Carine Tisné^{1,2*}

¹ Laboratoire de cristallographie et RMN biologiques, CNRS, Université Paris Descartes, USPC, 4 avenue de l'Observatoire, 75006 Paris

² Laboratoire d'Expression génétique microbienne, IBPC, CNRS, Université Paris Diderot, USPC, 13 rue Pierre et Marie Curie, 75005 Paris.

³ LBPA, CNRS UMR 8113, ENS Paris-Saclay, Université Paris-Saclay, 61 Avenue du Pdt Wilson, F-94235, Cachan

⁴ AIDS and Cancer Virus Program, SAIC-Frederick, Inc., Frederick National Laboratory, Maryland 21702-1201, USA.

⁵ Institut de Chimie des Substances Naturelles, CNRS UPR2301, Univ. Paris Sud, Université Paris-Saclay, Avenue de la Terrasse, 91190 Gif-sur-Yvette.

⁶ IDIBAPS, University of Barcelona, Barcelona, Spain.

§ These authors contributed equally to the work presented in the paper.

* To whom correspondence should be addressed:

Carine Tisné, carine.tisne@cnrs.fr

Phone: + 33 1 53 73 15 72 ; Fax: + 33 1 53 73 99 25

ABSTRACT

During HIV-1 assembly and budding, the Gag protein, and in particular its C-terminal domain composed of the nucleocapsid domain (NC_d), p1 and p6, is the location of numerous interactions with viral and cellular factors (genomic RNA, Vpr, TSG101, ALIX). To date, most of the studies regarding Gag have been performed with Gag deleted from its p6 domain. Using NMR, we show that the p1-p6 domain, within the C-terminal part of Gag (NCp15), is largely disordered, and transiently interacts with NC_d. These interactions modify the dynamic properties of NC_d in several timescales. Notably, in the mature form of NC_d, a minor state pre-organized for the interaction with nucleic acids (NA) is present. This state does not exist in the precursor forms of NC. Altogether, our study demonstrates that Gag C-terminus processing by the viral protease finely tunes the dynamic properties of NC_d and thus controls its NA binding properties.

INTRODUCTION

In retroviruses, the Pr55^{Gag} polyprotein precursor (Gag) is the only protein required for particle formation. It is composed of the structural protein domains: the matrix (MA), the capsid (CA), a spacer peptide p2, the nucleocapsid domain (NC_d), a hydrophobic-rich peptide p1, and an acidic p6 domain. As part of Gag, the MA domain targets Gag to the plasma membrane; CA drives Gag multimerization during assembly; NC_d recruits the viral RNA genome (gRNA) into virions and facilitates the assembly process; and p6 recruits the endosomal sorting complex (ESCRT), which catalyses the membrane fission step to complete the budding process¹³³. The C-terminal domain of Gag encompassed within NCp15 (NC_d-p1-p6) is thus responsible for numerous interactions, either by NC_d or p6 during assembly and budding of virions.

The NC_d within Gag is the major site of interaction with gRNA, even though MA is also able to interact with nucleic acids^{24, 566-568}. It was recently shown that global changes in the RNA binding specificity of Gag regulate virion synthesis¹⁵¹. Prior to, and after virion assembly and maturation, the NC_d of Gag preferentially binds to Ψ region at the level of SL3 and to Rev Response elements of the viral genome¹⁵¹. The Ψ region of gRNA contains several stem-loops (SL1 to SL4), SL3 being one of the major determinants of viral encapsidation^{36, 254, 267}. During virion genesis, this specificity transiently changes in a manner that facilitates genome packaging, *i.e.* the NC_d of Gag non-specifically binds to many sites on the gRNA. Then, the RNA-binding properties of NC_d in mature virions resemble that of unassembled Gag in the cytosol¹⁵¹. Regarding p6, two major interaction motifs for proteins belonging to the human ESCRT machinery are present along its

sequence: P₇₈TAP₈₁ involved in the binding of TSG101⁵⁶⁹ and L₁₀₆YPLASL₁₁₂ involved in the binding of ALIX⁵⁷⁰. ALIX also interacts with the NC_d, however the exact function of this interaction is not clearly understood^{284, 571-572}. p6 is also engaged *via* its conserved L₁₁₂RSLFG₁₁₇ motif in the binding of the retroviral protein Vpr⁵⁷³⁻⁵⁷⁵.

During HIV-1 maturation, three different forms of NC_d appear successively. The first cleavage by the HIV-1 protease occurs between MA-CA-p2 and NC_d-p1-p6, thereby liberating NCp15 (Fig. 1). The second cleavage frees NCp9 (NC_d-p1) whereas the last one releases the NC_d itself, which constitutes the final maturation form of NC_d, also called NCp7. NCp15 and NCp9 intermediates appear transiently but their lifetime is specifically programmed^{309, 334}. The correct processing of the cleavage site between p1 and p6 is essential and NCp15-containing virions are non-infectious^{336, 338, 586}. The coordination between protease activation, viral assembly and budding is critical³⁰⁹. The involvement of TSG101 and ALIX during HIV-1 budding is required for speeding up the closure of budding virions and the release of fully formed particles before the activation of the HIV-1 protease, this timing of events ensuring the infectivity of progeny virions³⁰⁹.

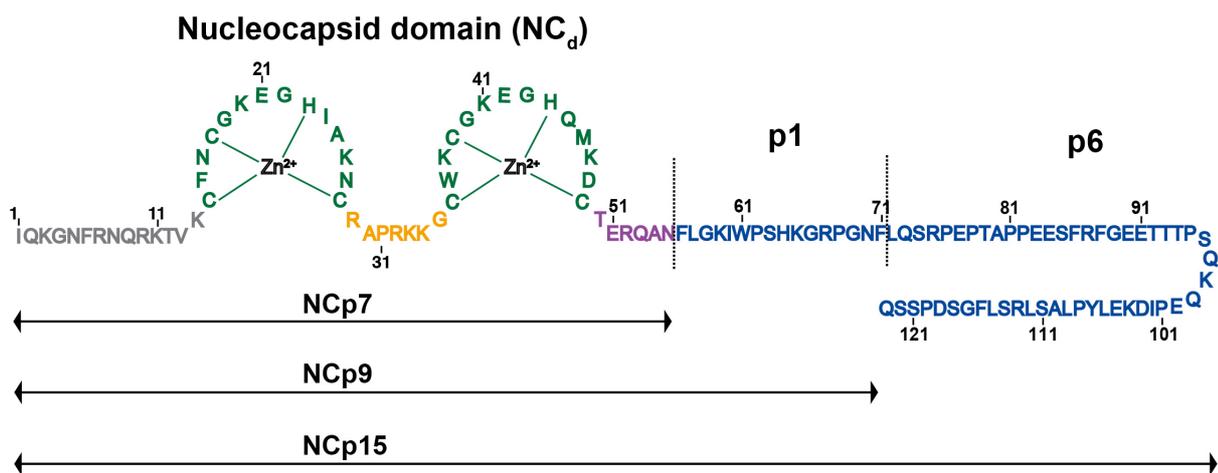


Figure 1 Sequence of the HIV-1 C-terminal domain of Gag (NCp15). The first cleavage by the HIV-1 protease first liberates NCp15 protein, then NCp9 and finally the mature form of the nucleocapsid domain NCp7. The dashed line represents the two sites of protease cleavage present in NCp15. Residues are coloured in grey for residues in the N-terminal part of NCd, green for those in the zinc fingers and orange for residues in the linker between the two zinc knuckles of NCd and purple for the C-terminal domain of NCd.

In the HIV-1 life cycle, the nucleic-acid chaperone activity of NCp7 is essential during viral replication and notably during reverse transcription (RT)^{42, 427}. Previous studies compared the nucleic-acid binding and chaperone properties of mature NCp7 to those of the precursor proteins NCp9 and NCp15^{311, 313, 427}. Even if NCp9 and NCp15 have strong annealing activities, they are unable to sustain efficient RT-catalyzed DNA elongation⁴²⁷. It was proposed that p6 regulates the function of NC through interactions of its positively-charged residues with negatively-charged residues of NC_d. NCp15 would be structured in

a “fold-back” conformation^{313, 427}.

NCp7 is a small basic protein of 55 residues containing two zinc-knuckles (ZKs) of the consensus sequence CX₂CX₄HX₄C (Green and Berg, 1990) separated by a small linker of 7 residues. The structures of NCp7 alone^{344, 349}, in complex with RNA stem-loops (SL2³⁶¹ and SL3³⁶⁰) or in complex with DNA oligonucleotides^{362, 365-366} have been solved by NMR. Overall, these structures show that the folding of each individual ZK is the same in all complexes. Analysis of these structures revealed that each ZK is specialized in one function: ZK2 (C-terminal ZK) is responsible for binding of one accessible and flexible guanine, while ZK1 (N-terminal ZK) either binds a second G or destabilizes the stem through its hydrophobic residues³⁶⁹. The major residues involved in the hydrophobic contacts with nucleic acids are the aromatic residues present at the second position of each ZK (*i.e.* F16 and W37, Fig. 1) together with other hydrophobic residues (*i.e.* V13, I24, A25 and M46). The functional versatility of NCp7 relies upon plasticity in the spatial orientation of the two ZKs. Notably, differential flexibility within the linker results in asymmetric motion of the ZKs which may explain their functionally distinct roles despite high sequence identity^{369, 587}. The conformation and dynamics of a large HIV-1 Gag fragment comprising the MA, the N- and C-terminal domains of CA, p2 and NC_d was also studied by NMR⁵⁶⁸. In solution, the five structured domains of this Gag construct retain their fold and reorient semi-independently from one another.

In the present study, we investigated by NMR the conformational landscape of the NC_d in its precursor forms (NCp15 and NCp9) as a model of the C-terminal portion of Gag and in its mature form (NCp7). We examined the dynamics occurring in several timescales using ¹⁵N relaxation measurements of the NC_d in its precursor and mature forms alone or in complex with SL3. For the first time, we show that p1-p6 within NCp15 is disordered in solution and that NCp15 does not exhibit a ground “fold-back” conformation. Short-lived α -helices are formed in p1-p6 upon long-range and transient contacts with the NC_d, likely promoting the binding of p6 to its partners. We also demonstrate by NMR and microcalorimetry that this dynamical behavior changes the thermodynamics of the NC_d interaction to SL3. Additional ¹⁵N relaxation dispersion NMR experiments led to the identification of residues involved in nucleic-acid conformational selection in NCp7, but not in NCp9 and NCp15. The maturation process of the C-terminal part of Gag by the protease thus finely tunes the NC_d binding properties to nucleic acid (NA).

RESULTS

p1-p6 within NCp15 is largely disordered, but presents regions that transiently adopt α -helical structures.

We first explore the hypothesis of a “fold-back” conformation by a comprehensive NMR dynamic study of NC_d in its different stages of maturation. We expressed, purified and assigned the NMR backbone resonances of NCp7, NCp9 and NCp15 proteins that correspond to the three states of maturation of NC_d by the HIV-1 protease.

The reduced frequency dispersion observed for the peaks originating from p1-p6 (Supplementary Fig. 1A), especially in the ¹H dimension, is an evidence of a high degree of conformational flexibility characteristic of an intrinsically disordered domain (IDD) (for recent reviews ⁵⁸⁸⁻⁵⁸⁹). As expected for IDDs, only very limited structural information for p1-p6 could be obtained from ¹H-¹H NOE measurements involving amide groups and observations of these NOE cross-peaks were mostly limited to correlations within intra-residue or sequential correlations (Supplementary Fig. 1B). NMR chemical shifts and ¹⁵N backbone relaxation (T1, T2 and heteronuclear ¹H-¹⁵N NOE (HetNOE)) were used to characterize the structural propensities and the conformational dynamics of p1-p6 within NCp15 (Fig. 2 and supplementary Fig. 2).

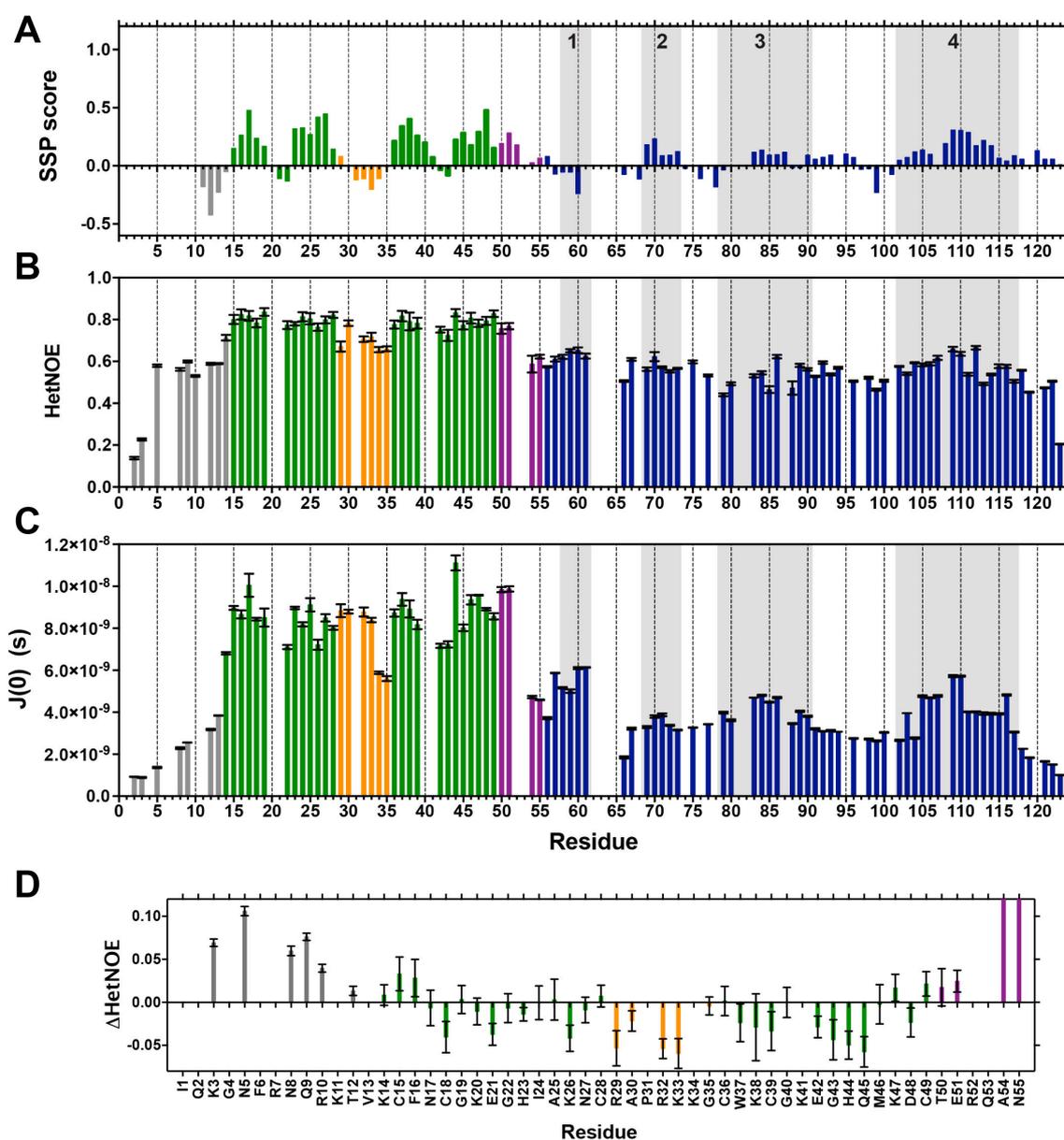


Figure 2 Structural and dynamic characterization of NCp15 from NMR chemical shifts and ^{15}N relaxation data recorded at 10°C and 950 MHz. (A) Secondary structural propensities calculated by the SSP program⁵⁹⁰. $\text{C}\alpha$, $\text{C}\beta$, CO and $\text{H}\alpha$ have been used as input data for the calculation of the SSP score. Positive score value indicate the amount of α -helical conformation present along the sequence whereas negative values indicate extended or β -strand conformations. **(B)** ^1H - ^{15}N NOE (HetNOE) values are indicative of the magnitude of local subnanosecond motions (high values: restricted motions; low values: high-amplitude motions). **(C)** Spectral densities $J(0)$ extracted by spectral density mapping from ^{15}N relaxation data (T_1 , T_2 , HetNOE) of NCp15. $J(0)$ values are indicative of slow overall and segmental tumbling motions present in NCp15. The boxes in grey indicated the four regions of p1-p6 domains showing significant secondary structure propensities. **(D)** Differences of HetNOE within the NCd between NCp15 and NCp7 measured at 10°C and 950 MHz. The color code for residues is described in figure 1.

Based on the backbone NMR chemical shift assignments of the NCp15 protein (BMRB entry number 26843), the secondary structure propensity (SSP) score⁵⁹⁰ highlights several regions in p1-p6 with a significant propensity to form α -helical structures (Fig. 2A). Helical propensities of 17%, 14% and 11%, were estimated for regions 4, 2 and 3

respectively. The measurements of HetNOE at 950 MHz (Fig. 2B) and 600 MHz (Supplementary Fig. 2B) confirm this trend, *ie* p1-p6 is disordered (most of the residues displaying HetNOE smaller than 0.6 at 950 MHz and smaller than 0.3 at 600 MHz), but some regions present a higher propensity to fold. The conversion of the measured ^{15}N T1, T2 and HetNOE data into spectral densities at zero-frequency $J(0)$ gives information on the folding propensity at a single residue resolution. Basically, higher $J(0)$ values are observed for regions displaying well-folded or transient secondary structures. These latter regions are also characterized by the occurrence of $J(0)$ maxima in their center⁵⁹¹⁻⁵⁹². In p1-p6, $J(0)$ maxima are observed for regions 1, 2, 3 and 4 (Fig. 2C). Interestingly, these maxima found at residues L57, I60, W61, N70, F71, E83, E84, S85, F86, L109 and A110 involved mainly hydrophobic residues. For regions 1 and 2 that contain the sites of protease cleavage, there is also some evidence of transient structuration, potentially in β -strand for region 1 and in α -helix for region 2.

Regarding the NC_d within NCp15, the NMR data (Fig. 2 and Supplementary Fig. 2) show that the two ZKs define two well-structured regions (high values of HetNOE and $J(0)$) separated by a more flexible linker, similarly to data for NCp7 and NCp9 (Supplementary Fig. 3), showing that the global folding of the NC_d is conserved. Fine analysis of the dynamics of the NC_d in the mature state NCp7 or within the precursor NCp15 (Fig. 2D) show that the HetNOE magnitudes present differences. The precursor is more rigid for residues of the N-terminal part and globally more flexible for residues of the ZKs and the linker. In addition, the magnitude of $J(0)$ values increase for residues of the NC_d by a 2.5 fold when comparing NCp7 to NCp15, then by adding the disordered p1-p6 tail. This shows that the motions of the NC_d and p1-p6 are dynamically coupled and could be an indicator of transient contacts between NC_d and p1-p6.

Altogether, the picture that emerges from the analysis of the NCp15 dynamics is that besides a well-folded part corresponding to the two ZKs, p1-p6 within NCp15 is largely disordered but, contains regions that transiently adopt secondary structures. This is notably the case for the regions known to interact with protein partners, namely TSG101, interacting with region 2, and ALIX and Vpr, interacting with region 4. These NMR data definitely show that the p1-p6 domain is disordered and flexible within NCp15 and that therefore a static (ground) “fold-back” conformation of NCp15 does not exist. As suggested by the dynamics of NCp15, we next investigated if the NMR data could provide evidence of transient interactions between p1-p6 and NC_d as sometimes observed between a structured domain and an unfolded tail of a protein (ref ets1 HMMG).

Long-range transient interactions occur between p1 and p1-p6 with the NC_d.

Fig. 3A displays zooms on amide group resonances extracted from ¹H-¹⁵N BEST-TROSY NMR experiments for the state of maturation of the NC_d. Large chemical shift variations are observed in the NC_d notably for residues F16, N17, C18, A25, K33 and K38 (Fig. 2AB) indicative of structural perturbations. It appears that amide group chemical shifts of F16, N17, C18, K33 are the residues that are the most affected by the presence of p1, whereas the chemical shifts of the amide groups of A25 and K38 are impacted by the presence of both p1 and p6 (Fig. 3). For these two latter residues, it seems that the phenomenon responsible for their chemical shift perturbations when p1 is present is reinforced by the presence of p6. This is also observed for residues of p1, L57, I60 and W61, that show chemical shift variations in presence of p6 (Fig 3A and Supplementary Fig. 4A). This comparison of chemical shifts between different NC_d constructs under identical conditions provides the information that transient long-range interactions occur between p1 or p1-p6 with the NC_d in both precursor forms of NC_d. Indeed, since p1-p6 (respectively p1) is disordered within NCp15 (respectively NCp9), when one compares the chemical shifts of the NC_d within NCp7, NCp9 and NCp15, the chemical shift changes observed for residues that are not sequentially close to the cleavage sites are indicative of transient long-range interactions with the protein part that is absent in the shorter of the two constructs. Such long-range transient interactions are clearly visible in data presented in Fig. 3B. For instance, the cleavage of p1 in NCp9 leads to large chemical shift variations for F16, N17, C18, A25, K33 and K38 in NC_d, whereas the cleavage of p6 in NCp15 causes additional chemical shift perturbations mainly for A25 and K38. These data strongly support that p1 makes long-range transient interactions with NC_d and that p6 reinforces these interactions.

3D ¹⁵N- and ¹³C-edited NOESY-HSQC spectra were recorded to determine if any persistent inter-residue contact was detectable between p1 or p1-p6 and NC_d. Only intra-residue and nearest-neighbor inter-residue ¹H-¹H NOE were observed, and no long-range NOE was detected between p1 and NC_d or p1-p6 and NC_d. This data confirms that there is no evidence of any persistent structural element within p1 or p1-p6, and show that there does not exist any long-lived interaction between p1-p6 and NC_d. Only weak ¹H-¹H NOEs could be detected and assigned in 2D NOESY spectra of NCp9 recorded in D₂O with mixing times of 150 or 300 ms in order to obtain enhanced sensitivity for the weak NOEs (Fig. 3C). Indeed, the Hε3 proton of W37 exhibits small cross-relaxation with both methyl groups of L57 and methyl group of I60. The NOEs are observable only in 2D spectra for NCp9 and NCp15. Altogether, these data show that p1 interacts with the NC_d at the level

of a region defined by F16, N17, A25, C18 and K33, both in NCp9 and NCp15. For p6, the data show that it reinforces the interaction of p1, by increasing the population of the interacting species (Noe avec NCp15?). In NCp15, the N-terminal residues are less flexible than in NCp7 and NCp9, these motion restraints can be indicative of transient interactions between p6 and the N-terminal part of NCp15. Indeed, p1-p6 rapidly interconverts between an ensemble of conformations that are partially restrained by transient interactions with the NCd. This ensemble may involve binding of the flexible p1-p6 at one predominant site or at multiple sites on the NCd. The NMR data collected for the three proteins are in agreement with p1 transiently interacting with the NCd at one site described by the chemical shift perturbations, whereas the acidic p6 domain could have multiple sites of interactions in the N-terminal domain that is unfolded and positively charged (Fig 3D).

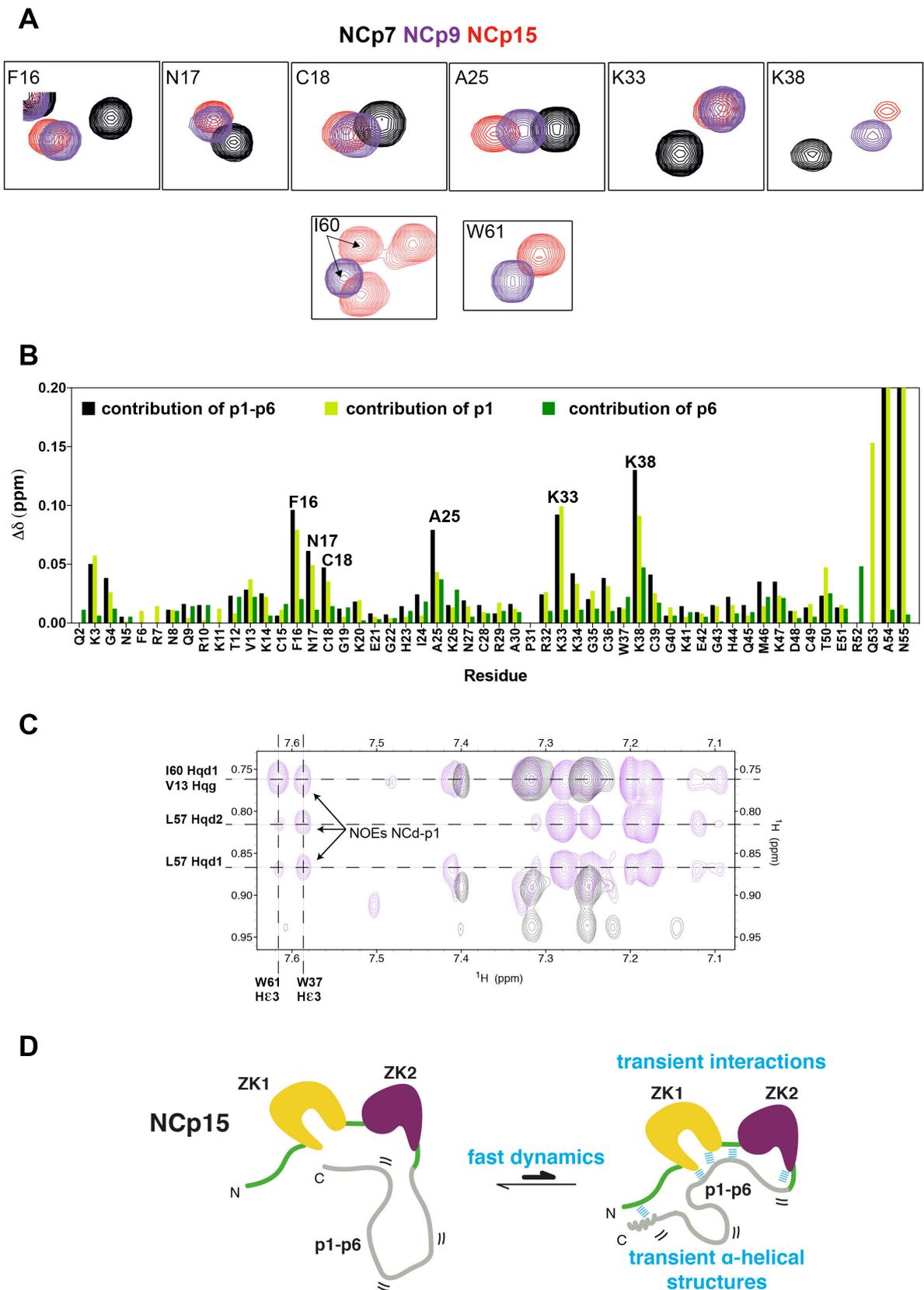


Figure 3 Transient interactions between p1-p6 and NC_d domains are evidenced by NMR data measured on the precursor and mature forms of the NC_d. (A) Peaks of F16, N17, C18, A25 and K33 amide groups in NCp7 (in black), NCp9 (in purple) and NCp15 (in red) extracted from ¹H-¹⁵N BEST-TROSY recorded at 950 MHz. (B) Contributions of p1-p6 (black bars), p1(light green bars) and p6 (dark green bars) to NC_d chemical shift perturbations measured on the amide groups. The p1-p6 (respectively p1) contributions were calculated as the difference in chemical shifts of NC_d amide groups (combined ¹H/¹⁵N shifts) between NCp15 (respectively NCp9) and NCp7, the p6 contributions were calculated as the difference in chemical shifts of NC_d amide groups (combined

¹H/¹⁵N shifts) between NCp15 and NCp9. **(C)** 2D-NOESY (150 ms) measured in D₂O on NCp9 (in black) showing NOEs between proton He3 of W37 and protons of methyl groups of L57 and I60, the same region of a 2D-NOESY recorded in the same conditions on NCp7 is displayed in red. **(D)** Drawing summarizing the interactions between NC_d, p1 and p6 evidenced by the NMR chemical shift and the dynamic analysis of the NC_d in NCp7, NCp9 and NCp15.

Interestingly, previous NMR studies of p6 or parts of p6 in various media^{574, 585} identified helices in the regions 3 and 4 where we uncovered transient helical propensities, but no significant appearance of secondary structures was found for p6 in water. This strongly suggests that the transient folding of p6 occurs concomitantly to the transient contacts between p1-p6 and NC_d. This behavior resembles those of multifunctional, intrinsically disordered proteins (IDPs) of HCV virus such as NS5A in which a network of electrostatic long-range interactions channels the transient folding of charged elements⁵⁹³. The influence of the distribution of residues with opposite-charged residues on the conformations of IDPs appears to be a general feature of this class of proteins⁵⁹⁴. Note also that similarly to NS5A protein, p6 contains a large number of phosphorylation sites that could bring further negative charges⁵⁹⁵⁻⁵⁹⁷.

Beside these transient interactions, weak interdomain cross-peaks between the Phe16 aromatic protons and the H β protons of W37, previously observed in NCp7, are also observed in NCp9 (Supplementary Fig. 5) and NCp15. The appearance of these weak interknuckle NOEs were explained by a transient interaction between the two zinc knuckles, a small fraction of the protein existing in multiple “close-approach” conformations, the equilibrium being shifted toward the non-interacting species. These connections are still observable for NCp9 and NCp15.

In the next paragraph, we investigate if these transient interactions can impact the interaction of NC with the major component of the gRNA encapsidation signal SL3 by NMR and ITC.

Changes in fast dynamics observed for the different states of maturation of the nucleocapsid domain correlate with changes in conformational entropy of SL3 binding.

The chemical shift perturbations measured on amide groups of the NC_d upon binding to SL3 are very similar whatever the state of maturation of this domain (Fig. 4A and Supplementary Fig. 6). Therefore, SL3 is very likely positioned with the same specific interactions into the hydrophobic plateau of the NC_d whatever its state of maturation. Moreover, p1-p6 does not present large chemical shift perturbations upon SL3 binding (Fig. 4B) and p6 alone does not interact with SL3 (Supplementary Fig. 7). Altogether, this

shows that only the NC_d is involved in SL3 binding either in the precursor or mature forms of this domain and that the SL3-binding site is globally identical in the three complexes.

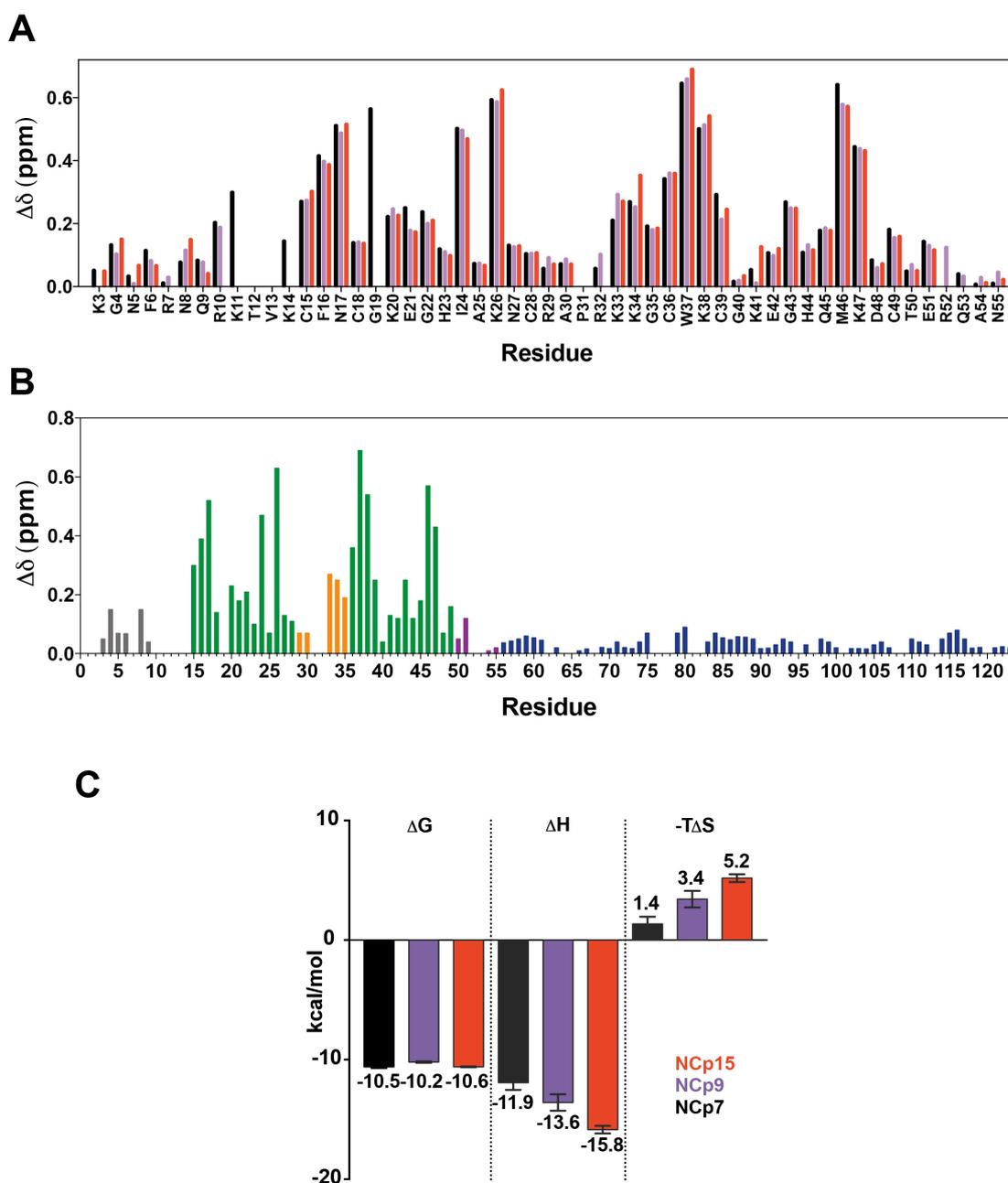


Figure 4 NMR and ITC analysis of NC/SL3 complexes. (A) Chemical shift perturbations upon SL3 binding for NCp7 (black bars), NCp9 (purple bars) and NCp15 (red bars) observed at the level of the NC domain. (B) Chemical shift perturbations upon SL3 binding measured for NCp15 as a function of the sequence of NCp15. The color code for residues is that of figure 1. (C) Thermodynamic components (ΔG : free enthalpy, ΔH : enthalpy and $-T\Delta S$: entropy) displayed as bars extracted from the calorimetric titrations of SL3 RNA with NCp7 (in black), NCp9 (in purple) and NCp15 (in red). Each titration was conducted at two different concentrations of NaCl 25 mM (dark bars) and 150 mM (light bars).

We used ITC to gain insight into the thermodynamics of the SL3/NC interaction (Fig. 4C and Supplementary Fig. 8). The free energy of binding (ΔG) for NCp7, NCp9 and NCp15 to SL3 complexes is similar (~ -10 kcal/mol) and SL3 binding to the NC_d is enthalpically

driven for these three proteins. However, the enthalpy and entropy differed significantly (Fig. 4C). Notably, in the constructs containing p1 or p1-p6, a higher entropic penalty is measured upon SL3 binding compensated by a higher enthalpy of interaction. An entropy change could arise from a number of factors, including differences in solvent or conformational entropy. If we assume that the entropy from solvent upon ligand binding is similar in the three complexes, which is reasonable given that both the ligand and the ligand-binding site are identical, then the differences between the three complexes in the entropic contribution to binding are due to differences in the conformational entropy as previously observed in other complexes⁵⁹⁸⁻⁵⁹⁹. Therefore, compared to NCp7, there is a penalty for SL3-binding arising from the conformational entropy, which amounts to 2 kcal/mol and 4 kcal/mol, for NCp9 and NCp15, respectively. Moreover, NCp9 calorimetric analysis is intermediate between NCp7 and NCp15 showing that both p1 and p6 are responsible for the changes in the conformational entropy (Fig. 4C and Supplementary Fig. 8). It was recently shown that changes in conformational entropy originate from enhanced pico to nanosecond internal motions and that fast timescale dynamics tune the entropic contribution of ligand affinity^{598, 599, Liu, 2016 #205, 600-601}. In our case, we have demonstrated that NCp15 and NCp9 present fast motions corresponding to transient interactions between the NC_d and p1 in NCp9 or p1-p6 in NCp15. Moreover, in NCp15, these transient interactions are responsible for a higher flexibility of the ZKs and the linker. Therefore, the enhanced flexibility encountered in the precursor forms of NCd are very likely responsible for the larger conformational entropic penalties observed upon their binding to SL3.

A minor state pre-organized to interact with nucleic acids exists in the mature form of NC_d.

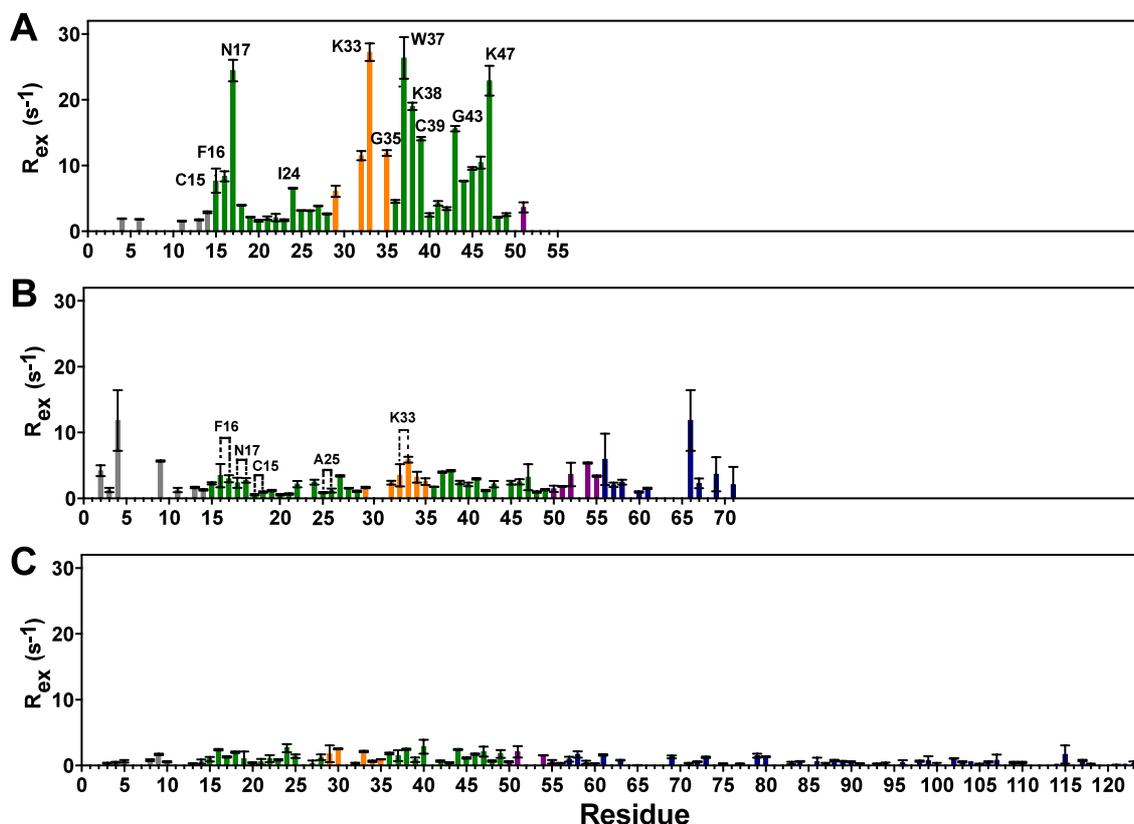


Figure 5. A minor conformation is present in the mature NC_d. Exchange contributions to the transverse ¹⁵N relaxation (R_{ex} values) extracted from CPMG relaxation dispersion experiments recorded at various frequencies of CPMG pulses train for **(A)** NCp7 **(B)** NCp9 and **(C)** NCp15. The color code for residues is that of figure 1.

To investigate dynamics in the microsecond to millisecond timescale, ¹⁵N CPMG relaxation dispersion experiments were measured at 950 MHz and 600 MHz for the three maturation forms of the NC_d, namely NCp15, NCp9 and NCp7. Interestingly, only in NCp7, residues N17, K33, G35, W37, K38, C39, G43 and K47 exhibit particularly large exchange rates (R_{ex} > 10 s⁻¹) (Fig 5A and Supplementary Fig. 9-11). These residues are more numerous in ZK2 and in the linker than in ZK1. These dispersion data therefore show that NCp7 undergoes conformational exchange between a ground state, highly populated, and an excited state, lowly populated. Using a two-state analysis and two different methodologies (see Materials and Methods), the dispersion relaxation data recorded at two different magnetic fields were fitted with good agreement to a model in which the minor conformer populated at 5.5 % interconverts with the ground state at a rate of 470 s⁻¹ at 10°C (Supplementary Table 1). The ground state corresponds to the conformation of NCp7 previously solved by NMR^{349, 360}. The residues showing high exchange contributions to relaxation define the specific guanine-binding pocket in ZK2 (Fig. 6C) suggesting that the

higher energy conformation sampled by NCp7 plays an important role in capturing this guanine. Indeed, there are strong linear correlations between the CPMG-derived values of $|\Delta\omega|$ (the chemical shift differences between the ground and excited states) and the absolute differences in ^{15}N chemical shift ($|\Delta\delta|$) between the NCp7 free and bound to SL3³⁶⁰ (Fig. 6A). Interestingly, the residues displaying significant exchange contribution to relaxation can be sorted into two classes: (i) class 1 (red, Fig. 6) for which $|\Delta\omega|$ is larger than $|\Delta\delta|$ but linearly related to it and (ii) class 2 (purple, Fig. 6) for which the two values are linearly correlated (slope 1.10, $R^2=0.80$). Therefore, residues of class 2 sample a higher energy substate in which they adopt a conformation similar to the one they exhibit in the SL3-bound structure of NCp7. A similar observation has been reported for ribonuclease A or DHFR⁶⁰²⁻⁶⁰³. For residues of class 1, their excited conformation is actually unknown since significant reorganization of the protein occurs upon SL3-binding as shown by the large differences between $|\Delta\delta|$ and $|\Delta\omega|$. However, residues of class 1 are distant in the structure of NCp7 but close in the structure of the complex NCp7/SL3 (Fig. 6B and C) where they contribute with some residues of class 2 to fashion the guanine-binding pocket in ZK2. Interestingly, class 2 residues are rather involved in structuring ZK2 whereas class 1 residues are involved in RNA binding. Only two residues of ZK1, namely F16 and N17, display significant exchange contributions to relaxation. Given their position in the free and bound structure of NCp7, they probably contribute to make contact between ZK1 and ZK2. Our NMR data are in agreement with an equilibrium between a major conformer in which N17 and W37 residues are far from each other (Fig. 6B) and a minor conformer in which these two latter residues are in close contact inducing large chemical shift changes such as in the SL3-bound protein (Fig. 6C). Altogether, our data show that a pre-organized form (the excited conformer) at the level of the guanine-binding pocket in ZK2 exists in solution, and a reorganization of the selected conformer after RNA binding is needed. Conversely, no significant R_{ex} value (*i.e.* larger than 3 Hz) could be observed in NCp15 and NCp9 neither in the NC nor p1-p6 domains. Indeed, the slow motions occurring in NCp7 are no more detectable in the NC_d within NCp15 and NCp9 demonstrating a major impact of p1 and p6 on the dynamics of the NC_d in the slow timescale. The cleavage of p6 liberating NCp9 is not sufficient to recover the large R_{ex} values observed for some residues in NCp7 (Fig. 5). The removal of p6 from NCp15 and then of p1 from NCp9, progressively reveals the presence of this minor conformer along the maturation pathway of the NC_d.

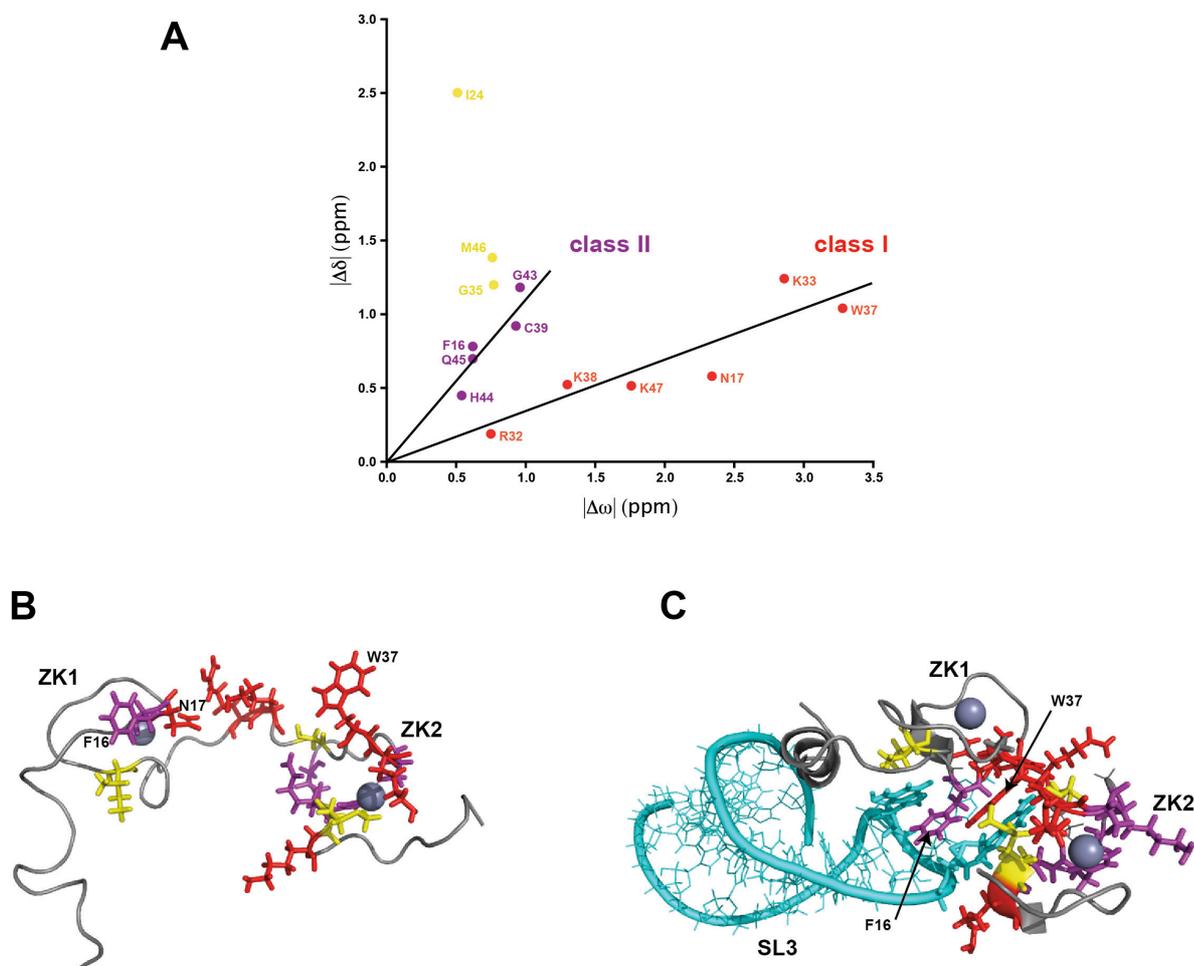


Figure 6 The minor conformation of the mature NCd (NCp7) organizes NC ZK2 for RNA binding. (A) Correlation between the $|\Delta\omega|$ values obtained from relaxation dispersion experiments and differences $|\Delta\delta|$ between chemical shifts of NCp7 free and bound to SL3 RNA. The different residues of NCp7 could be classified in two categories depending on the slope along which they align. Class 1 in red corresponds to residues that are involved in the SL3-binding and that will experience further chemical shift variation upon SL3-binding, class 2 (in purple) corresponds to residues that present the same chemical shifts in the minor conformer and in the SL3-bound structure of NCp7 and others residues showing significant exchange contribution to relaxation are in yellow. Visualization of the positions of NCp7 residues belonging to the different classes in **(B)** the structure of the free protein (PDB code: 1MFS) and **(C)** in the structure of the SL3/NCp7 complex (PDB code: 1A1T).

DISCUSSION

The C-terminal domain of Pr55^{Gag} is a platform for molecular interactions involved in the binding of gRNA, ALIX, TSG101 and Vpr. In this work, we have studied the C-terminal part of Pr55^{Gag} encompassed in NCp15, and we report the first comprehensive study that analyses the conformational landscape of each step of maturation of the NC_d (NCp7, NCp9 and NCp15) together with their properties of interactions with SL3. We can anticipate that the dynamic properties of NCp15 should be conserved in the context of Pr55^{Gag}, since it was previously shown that the structured domains of a Gag construct lacking p1-p6 retain their fold and reorient semi-independently from each other⁵⁶⁸. Our

work thus provides a new picture of the structural and dynamic properties of the NC_d during the assembly, budding and maturation of the virions (Fig. 7). Indeed, we have demonstrated that p1-p6 within NCp15 is an IDD that experiences transient α -helical structuration through long-range transient interactions with the NC_d (Fig. 7A). This leads to a dramatic change in the NC dynamics on several timescales. These transient interactions between p1-p6 and NC together with the resulting global higher flexibility of the ZKs and the linker occurring in the fast timescale (ns-ps) do not change the affinity of NC_d to SL3 but result in a conformational entropic penalty to SL3 binding. In the slow timescale (ms- μ s), these transient interactions prevent the formation of a minor conformer in the NC_d that partially mimics the nucleic-acid binding structure of NCp7 by forming the guanine-binding pocket in ZK2.

These results can have several implications for the function of Gag and NC throughout the viral replication cycle (Fig. 7). Prior to virion assembly, Pr55^{Gag} exists as a diffuse pool of monomers or low-order multimers in the cytoplasm of infected cells with its NC_d bound to G-U rich RNA sequences including sequences in the Ψ domain of the HIV-1 genome, consistent with the specificity of the mature NC_d (NCp7) ¹⁵¹. This result is in agreement with our study showing that NCp15 is able to bind SL3 with the same affinity as NCp7. During assembly, as Gag molecules become more tightly packed, coincident with higher-order multimerization at the plasma membrane, a dramatic change in Gag-RNA binding specificity to A-rich sequences contributes to selective packaging of viral genome ¹⁵¹. We speculate that the feature to constrain thousands of C-terminal parts of Gag (NCp15) into a pseudo-curved array induced a dramatic increase of local NCp15 concentration. This could induce an enhancement of the transient interactions between p1-p6 and NC_d leading to two consequences: 1) a higher level of p1-p6 structurations and 2) a higher penalty of conformational entropy for NC_d binding to RNA. Regarding the first point, the regions of p6 exhibiting larger propensities to form transient helical conformations are located in the binding sites of TSG101, ALIX and Vpr and p6 forms helices in contact with its partners. Therefore, the enhancement of these transient foldings during assembly and budding could help the recruitment of p6 partners (Fig. 7C). For point 2, the conformational entropic penalty could lead to inhibition of NC binding to its “traditional” specific targets (Fig. 7CD). Further studies will be needed to explain the shift of NC-RNA binding specificity to A-rich sequences.

During maturation, a minor state that exists in the mature form of NC_d (NCp7) will progressively emerge upon removal of p6 and then p1 from the precursor forms (NCp15 and NCp9) (Fig 7EF). In mature virions (Fig. 7F), the NCp7 ground state is in slow

equilibrium with a minor excited conformer being poised for interaction with RNA (Fig. 7E). This conformer formed by the interaction between F16, N17 of ZK1 and W37, K38, K47 of ZK2 could involve interactions between aromatic residues F16 and W37 present in each ZK. This minor state consists of a pre-organization of the guanine-binding pocket in ZK2. This minor conformation explains the weak interknuckle NOEs that were previously described in earlier NMR studies of NCp7 between F16 and W37^{349, 371} and that were proposed to correspond to a small fraction of the protein in which the knuckles exist in “close approach” conformations. Our results are also in agreement with the conformational landscape of NCp7 depicted from ensemble calculations based on RDC data⁵⁸⁷. While the presence of a minor conformer *per se* is not a formal proof of its involvement in the binding process, the close relation between the chemical shifts of the minor conformer and those of the SL3-bound form is in favor of such a role. Our data thus strongly suggest that a conformational selection mechanism could exist for the mature NC_d but not for its precursor forms, and give direct information on the conformation of the minor form that could be selected by RNA.

This conformational selection property of NCp7 to capture free guanine in NA is likely to be related to its NA chaperone activity that culminates in mature virions. Interestingly, NA chaperone activity improves greatly as the Gag precursor is progressively processed to create NCp15, NCp9 and NCp7³⁷⁴ and HIV-1 NCp7 presents the optimal NA chaperone activity that involves NA aggregation, destabilization and rapid protein-NA interaction kinetics. Therefore, the NA chaperone activity of NC_d is optimal when the minor conformation appears in its mature form. We propose that the minor conformer uncovered in this work is essential for the optimal NA chaperone activity of NCp7. This minor conformer is formed by the contact between F16/N17 and W37 (Fig. 6C). It was previously shown that mutations of either F16 or W37 to alanine in NCp7 significantly slows down NCp7's DNA interaction kinetics, while retaining some helix-destabilization capability³⁷⁷. A double F16A/W37A substitution in NCp7 further reduces the latter activity³⁷⁷. Mutating any of these aromatic residues probably prevents the occurrence of the minor conformer.

Therefore, the maturation by the protease of the C-terminal part of Gag will finely and progressively tune the interaction and the function of NC_d by cleavage of the IDD. We have thus shed light on the fine regulation of the NC_d dynamics and binding properties by p1-p6 and by the maturation process orchestrated by the protease. The use of a highly disordered domain in the regulation of the RNA recognition modes for the various steps of maturation of the NC_d adds to the known repertoire of the unstructured protein's roles.

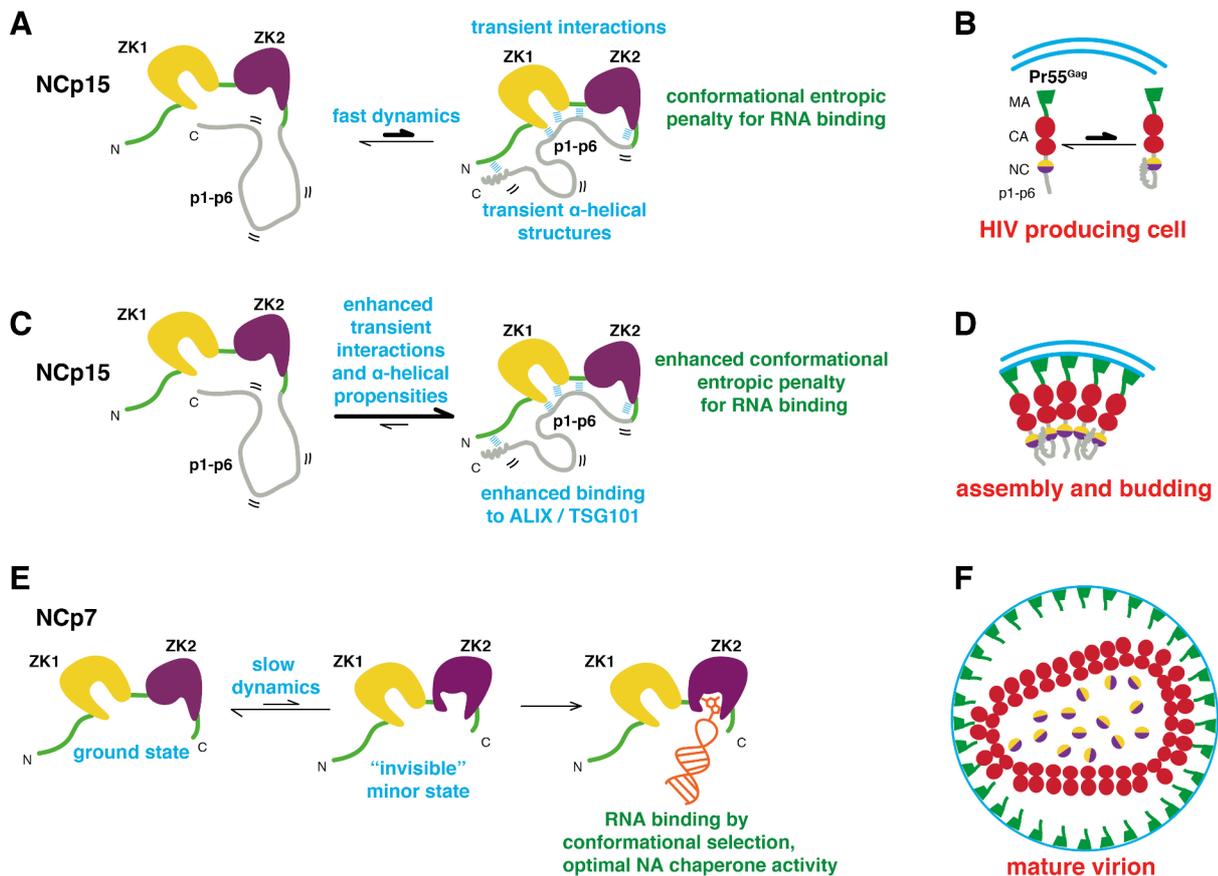


Figure 7 Model summarizing the fast and slow dynamics behaviour observed for the C-terminal part of Gag (NCp15) along its maturation pathway and the impact for its interactions to nucleic acids and ESCRT proteins during virion genesis. (A) Fast dynamics motions observed in NCp15 likely occurring at the C-terminal of Gag in the cytosol **(B)**. **(C)** Transient interactions could be amplified during assembly and budding of virions **(D)** leading to enhanced binding to ESCRT proteins and enhanced conformational entropic penalty for NA binding. **(E, F)** In mature virion, slow dynamics motions occurring in NCp7 takes place revealing a minor conformation prone to capture unpaired guanine by a conformational selection mechanism involving ZK2 and interactions between F16, N17 residues in ZK1 and W37 in ZK2. The minor conformation present in NCp7 is likely to optimize its NA chaperone activity. For sake of clarity, the gRNA is not represented.

METHODS

Expression and purification of recombinant HIV-1 NCp7, NCp9 and NCp15. NCp7, NCp9 and NCp15 (HIV-1 strain NL4-3) were expressed in *E. coli* from plasmids built into a pET-3a vector (Novagen). NCp7 and NCp9 were overexpressed in *E. coli* BL21(DE3)pLysE strain while NCp15 was overexpressed in *E. coli* BL21(DE3)star. NCp7 was overexpressed isotopically ^{15}N -labeled and ^{13}C - ^{15}N -labeled as previously described^{360, 371}. The same protocols were used for NCp9. For NCp15, the protocols of purification were modified since NCp15 presented a lower expression level and was more sensitive to proteolysis. Briefly, the cells were resuspended in a lysis buffer containing 50 mM Tris.Cl (pH 8.5), 10% glycerol, 100 mM NaCl, 10 mM EDTA, 10 mM β -mercaptoethanol, 1 mM

PMSF, 5 µg/ml pepstatin A and one protease inhibitor tablet at 4°C. The cells were sonicated and then centrifuged at 45000 g for 30 min at 4°C. The supernatant was collected and the nucleic acids were precipitated by adding 5% polyethyleneimine dropwise to a final concentration of 1%, and the mixtures were stirred for 45 min before centrifugation at 45 000g for 30 min at 4°C. The supernatant was collected and loaded onto a 53 ml Q-Sepharose (HiLoad 26/40, HP, GE Healthcare) equilibrated in buffer A (50 mM Tris.Cl pH 8.5, 10% glycerol, 100 mM NaCl, 10 mM EDTA, 10 mM β-mercaptoethanol, 1 mM PMSF). The fractions containing NCp15 were pooled and the sample was supplemented with ammonium sulfate to a final concentration of 1 M before loading it onto a 53 ml phenyl-Sepharose (HiLoad 26/10, HP, GE Healthcare). The protein was eluted with a 500 ml gradient of 30 to 70% buffer B (50 mM Tris.Cl pH 8.5, 10% glycerol, 1 M ammonium sulfate, 5 mM EDTA, 10 mM β-mercaptoethanol). The fractions containing NCp15 were pooled and loaded onto a 320 ml Superdex75 (HiLoad 26/60, Prep Grade, GE Healthcare) equilibrated in buffer C (50 mM Tris.Cl pH 8.5, 10% glycerol, 100 mM NaCl, 10 mM β-mercaptoethanol, 0.1 mM ZnCl₂). The eluted protein was pooled and loaded onto a column of 5 ml Heparin Sepharose (HiTrap Heparin, HP, GE Healthcare) equilibrated with buffer D (50 mM Tris.Cl pH 8.5, 100 mM NaCl, 10 mM β-mercaptoethanol, 0.1 mM ZnCl₂). The protein was eluted with a 200 ml gradient of 15% to 30% buffer E (50 mM Tris.Cl pH 8.5, 1 M NaCl, 10 mM β-mercaptoethanol, 0.1 mM ZnCl₂). The fractions containing NCp15 were pooled, concentrated and dialyzed against the NMR buffer (25 mM deuterated sodium acetate pH 6.0, 25 mM NaCl, 0.1 mM ZnCl₂ and 0.1 mM β-mercaptoethanol) using an amicon unit of 3 kDa (Millipore).

RNA sample preparation. The SL3 RNA stem-loop, 5'-GGACUAGCGGAGGCUAGUCC-3', deprotected and desalted, was purchased from Dharmacon Research[®]. The RNA sample was heating at 90°C for 5 minutes and cool on ice quickly for 10 minutes. It was then lyophilized and dissolved in the NMR buffer.

Analysis of NMR chemical shifts. All NMR spectra were recorded at 10°C or 25°C on Bruker 600, 700 or 950 MHz spectrometers equipped with cryogenic probes. Backbone chemical shift assignments of NCp7, NCp9 and NCp15 were performed using BEST-TROSY and standard 3D NMR experiments⁶⁰⁴. The backbone NMR chemical shifts of NCp15 were deposited to BMRB (entry code 26843). Combined chemical shift perturbations of amide groups $\Delta\delta$ (in ppm) were derived from ¹H and ¹⁵N chemical shift differences : $\Delta\delta(H,N) = \sqrt{[(\Delta^{15}N W_N)^2 + (\Delta^1H W_H)^2]}$, where $\Delta = \delta_{state1} - \delta_{state2}$ (difference of chemical shifts between states 1 and 2 of the protein) and $W_H=1$ and $W_N=1/6$.

SSP scores were calculated using C^α, C^β, CO and H^α chemical shifts, five-residue

weighted averaging, and all other default parameters⁵⁹⁰.

¹⁵N NMR relaxation experiments. The ¹⁵N backbone relaxation experiments were performed on 0.9-1 mM samples of ¹⁵N-uniformly labeled NCp15, NCp9 and NCp7. The T1, T2 and hetNOE ¹⁵N experiments have been performed using the standard Bruker pulse sequence libraries. All the experiments were recorded with a relaxation delay of 4s between two successive scans. The T1 data were collected using 20, 50 (repeat), 80, 100, 150, 200, 250, 300, 400 (repeat), 500, 700, 800, and 3000 ms for the values (13) of recovery delay. The T2-CPMG experiments were collected with 10 values of 16, 32, 64, 80, 128 (repeat), 160, 208 (repeat), 256, 320 and 400 ms. CPMG pulse trains were used with a 0.9 ms delay between successive ¹⁵N 180° pulses. The (¹H^N → ¹⁵N) HetNOE values were taken as the ratio between the intensities recorded with and without saturation of the amide protons. In all experiments, the points corresponding to the different relaxation delays were acquired in an interleaved manner.

The experiments were processed using NMRPipe⁵⁷⁷ and SPARKY⁵⁷⁸ was used to measure the intensities values of cross-peaks at the different relaxation delays. CurveFit (AG. Palmer Lab) and PYTHON(www.python.org), R (www.r-project.org) scripts were used to determine the R₁ and R₂ relaxation rates as well as the associated uncertainties from the single-exponential decay. Additionally, reduced power spectral density mapping for the determination of J(0) were used using the strategy designed by Farrow et al.⁵⁴⁸ and PYTHON scripts to achieve the procedure. The uncertainties in the calculated spectral density values were obtained from Monte Carlo simulations using the errors deduced from the preceding fits of R1 and R2 values and the spectral noise errors for hetNOE experiments. Using these errors 500 synthetic files were generated for each observable (T1, T2, hetNOE) and after extraction of the 500 J(0) values, these later were examined and the means and standard deviations extracted. The whole calculations were achieved using PYTHON and R scripts.

¹⁵N relaxation dispersion experiments. Spectra were recorded at 950 and 600 MHz at 10°C and in the same conditions using the pseudo-three-dimensional (interleaved) constant time CPMG (Carr-Purcell-Meiboom-Gill) optimized as previously described^(550, 605). The constant CPMG delay (T_{CPMG}) was established to 40 ms. 20 experiments were acquired with ¹⁵N 180° pulse repetition frequencies varied between 25 and 1000 Hz. A ¹H spin-lock is applied at a field strength of approximately 13 kHz during the CPMG sequence; the exact field strength slightly varies (~10%) for different values of CPMG frequency as previously described⁵⁵⁰. Peak intensities were converted into relaxation rates according procedures described in previous works^{550, 606}. The evaluation of

uncertainty of the $R_{2,eff}$ measurements was made following the methodology described in ⁶⁰⁷. CPMG relaxation data were analyzed using the program CATIA (<http://www.biochem.ucl.ac.uk/hansen/>), which numerically integrates the Bloch-McConnell equations ⁶⁰⁸ and permits the extraction of $k_{ex}=k_A+k_B$, p_B (population of minor conformer) and $|\Delta\omega|$, the chemical shift difference between the major and minor conformer in the case of a two state equilibrium ($A \xrightleftharpoons[k_B]{k_A} B$) between a major state A and a minor state B ($p_A \gg p_B$). Alternatively to this approach the relaxation dispersion profiles at the two fields were fitted with the Carver-Richards equation to extract the same parameters as described above ⁶⁰⁹⁻⁶¹¹. To fit the Carver-Richards equation, we use a suite of PYTHON and R scripts. The results from the two procedures were compared (Supplementary Fig. 11). The error for each parameter (k_{ex} , p_B , $\Delta\omega$) was estimated from Monte Carlo simulations using the uncertainty determined for each $R_{2,eff}$ measurement).

Isothermal titration calorimetry (ITC). Dissociation constants for NCp7, NCp15 and NCp9 binding to SL3 were recorded on a MicroCal[®] ITC₂₀₀ system. SL3 RNA and the protein samples were extensively dialyzed against the ITC buffer (25 mM Na-acetate pH 6.0, 25 mM NaCl, 0.1 mM TCEP, 0.1% Tween, 0.1 mM ZnCl₂) prior to titration. Heats of reaction were measured at 30°C for 15 injections of 2.54 μ L of one NC protein (in the syringe at a concentration between 60 to 70 μ M) into 280 μ L of SL3 RNA (in the cell at a concentration of 5 μ M). The heats of dilution were obtained by titrating the identical protein sample into a cell containing the ITC buffer and was subtracted from the raw data. The ITC data were analyzed with the software ORIGIN[®] using a one set of site model. The standard errors were estimated from the data spread and from the uncertainty of the titrant concentration determination as previously described ⁶¹².

Acknowledgements:

Financial support from the TGIR-RMN-THC FR3050 CNRS for conducting the research is gratefully acknowledged. This work was supported by grants from SIDACTION, ANRS and in part with federal funds from National Cancer Institute, NIH, Contract HHSN261200800001E (to R.G.).

Author contributions: GM, RG, OM and CT conceived the project ; AM, MC and CT designed the protocols and/or produced the samples for NMR and ITC ; AM, AB, MC, VL, LZ, PB, OM and CT designed, performed and analyzed the NMR experiments recorded for the three proteins and for the three complexes with SL3; AM, FB, PB and CT performed and analyzed the ITC experiments ; SB provides p6 peptide ; CVH provides CPMG

dispersion relaxation pulse program and helps in fitting the data; AM, AB, VL, OM, PB and CT wrote the paper.

Competing financial interests: The authors declare no competing financial interests.

REFERENCES

1. Freed, E.O. HIV-1 assembly, release and maturation. *Nature reviews. Microbiology* 13, 484-96 (2015).
2. Alfadhli, A., Still, A. & Barklis, E. Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *Journal of virology* 83, 12196-203 (2009).
3. Chukkapalli, V., Oh, S.J. & Ono, A. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proceedings of the National Academy of Sciences of the United States of America* 107, 1600-5 (2010).
4. Shkriabai, N. et al. Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry* 45, 4077-83 (2006).
5. Deshmukh, L., Ghirlando, R. & Clore, G.M. Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* 112, 3374-9 (2015).
6. Kutluay, S.B. et al. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* 159, 1096-109 (2014).
7. Paillart, J.C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nature reviews. Microbiology* 2, 461-72 (2004).
8. Lu, K., Heng, X. & Summers, M.F. Structural determinants and mechanism of HIV-1 genome packaging. *Journal of molecular biology* 410, 609-33 (2011).
9. Abd El-Wahad, E.W. et al. Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nature communications* 5, 4304 (2014).
10. Demirov, D.G., Ono, A., Orenstein, J.M. & Freed, E.O. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proceedings of the National Academy of Sciences of the United States of America* 99, 955-60 (2002).
11. Lee, S., Joshi, A., Nagashima, K., Freed, E.O. & Hurley, J.H. Structural basis for viral late-domain binding to Alix. *Nature structural & molecular biology* 14, 194-9

- (2007).
12. Dussupt, V. et al. The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding. *PLoS pathogens* 5, e1000339 (2009).
 13. Dussupt, V. et al. Basic residues in the nucleocapsid domain of Gag are critical for late events of HIV-1 budding. *Journal of virology* 85, 2304-15 (2011).
 14. Sette, P., Dussupt, V. & Bouamr, F. Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA. *Journal of virology* 86, 11608-15 (2012).
 15. Bachand, F., Yao, X.J., Hrimech, M., Rougeau, N. & Cohen, E.A. Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *The Journal of biological chemistry* 274, 9083-91 (1999).
 16. Salgado, G.F. et al. Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. *Biochemistry* 48, 2355-67 (2009).
 17. Salgado, G.F. et al. The role of membranes in the organization of HIV-1 Gag p6 and Vpr: p6 shows high affinity for membrane bilayers which substantially increases the interaction between p6 and Vpr. *Journal of medicinal chemistry* 52, 7157-62 (2009).
 18. Mirambeau, G., Lyonnais, S. & Gorelick, R.J. Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA biology* 7, 724-34 (2010).
 19. Bendjennat, M. & Saffarian, S. The Race against Protease Activation Defines the Role of ESCRTs in HIV Budding. *PLoS pathogens* 12, e1005657 (2016).
 20. Coren, L.V. et al. Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *Journal of virology* 81, 10047-54 (2007).
 21. Kafaie, J. et al. Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization. *Virology* 385, 233-44 (2009).
 22. Lee, S.K., Harris, J. & Swanstrom, R. A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *Journal of virology* 83, 8536-43 (2009).
 23. Sleiman, D. et al. Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus research* 169, 324-39 (2012).
 24. Wu, T., Gorelick, R.J. & Levin, J.G. Selection of fully processed HIV-1 nucleocapsid

- protein is required for optimal nucleic acid chaperone activity in reverse transcription. *Virus research* 193, 52-64 (2014).
25. Cruceanu, M. et al. Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic acids research* 34, 593-605 (2006).
 26. Wang, W. et al. Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic acids research* 42, 7145-59 (2014).
 27. Morellet, N. et al. Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *The EMBO journal* 11, 3059-65 (1992).
 28. Summers, M.F. et al. Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein science : a publication of the Protein Society* 1, 563-74 (1992).
 29. Amarasinghe, G.K. et al. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *Journal of molecular biology* 301, 491-511 (2000).
 30. De Guzman, R.N. et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* 279, 384-8 (1998).
 31. Bourbigot, S. et al. How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *Journal of molecular biology* 383, 1112-28 (2008).
 32. Bazzi, A. et al. 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 research* 39, 3903-16 (2011).
 33. Morellet, N. et al. Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *Journal of molecular biology* 283, 419-34 (1998).
 34. Zargarian, I. et al. Dynamics of linker residues modulate the nucleic acid binding properties of the HIV-1 nucleocapsid protein fingers. *PloS one* 9, e102150 (2014).
 35. Deshmukh, L., Schwieters, C.D., Grishaev, A. & Clore, G.M. Quantitative Characterization of Configurational Space Sampled by HIV-1 Nucleocapsid Using Solution NMR, X-ray Scattering and Protein Engineering. *Chemphyschem : a European journal of chemical physics and physical chemistry* 17, 1548-52 (2016).
 36. Jensen, M.R., Ruigrok, R.W. & Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Current opinion in structural biology* 23, 426-

- 35 (2013).
37. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of magnetic resonance* 241, 74-85 (2014).
 38. Marsh, J.A., Singh, V.K., Jia, Z. & Forman-Kay, J.D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein science : a publication of the Protein Society* 15, 2795-804 (2006).
 39. Farrow, N.A., Zhang, O., Szabo, A., Torchia, D.A. & Kay, L.E. Spectral density function mapping using ¹⁵N relaxation data exclusively. *Journal of biomolecular NMR* 6, 153-62 (1995).
 40. Brutscher, B., Bruschweiler, R. & Ernst, R.R. Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ¹H, ¹³C, and ¹⁵N nuclear magnetic resonance spectroscopy. *Biochemistry* 36, 13043-53 (1997).
 41. Feuerstein, S. et al. Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *Journal of molecular biology* 420, 310-23 (2012).
 42. Fossen, T. et al. Solution structure of the human immunodeficiency virus type 1 p6 protein. *The Journal of biological chemistry* 280, 42515-27 (2005).
 43. Solyom, Z. et al. The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophysical journal* 109, 1483-96 (2015).
 44. Das, R.K. & Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America* 110, 13392-7 (2013).
 45. Muller, B., Patschinsky, T. & Krausslich, H.G. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *Journal of virology* 76, 1015-24 (2002).
 46. Hemonnot, B. et al. The host cell MAP kinase ERK-2 regulates viral assembly and release by phosphorylating the p6gag protein of HIV-1. *The Journal of biological chemistry* 279, 32426-34 (2004).
 47. Kudoh, A. et al. The phosphorylation of HIV-1 Gag by atypical protein kinase C facilitates viral infectivity by promoting Vpr incorporation into virions. *Retrovirology* 11, 9 (2014).
 48. Frederick, K.K., Marlow, M.S., Valentine, K.G. & Wand, A.J. Conformational entropy

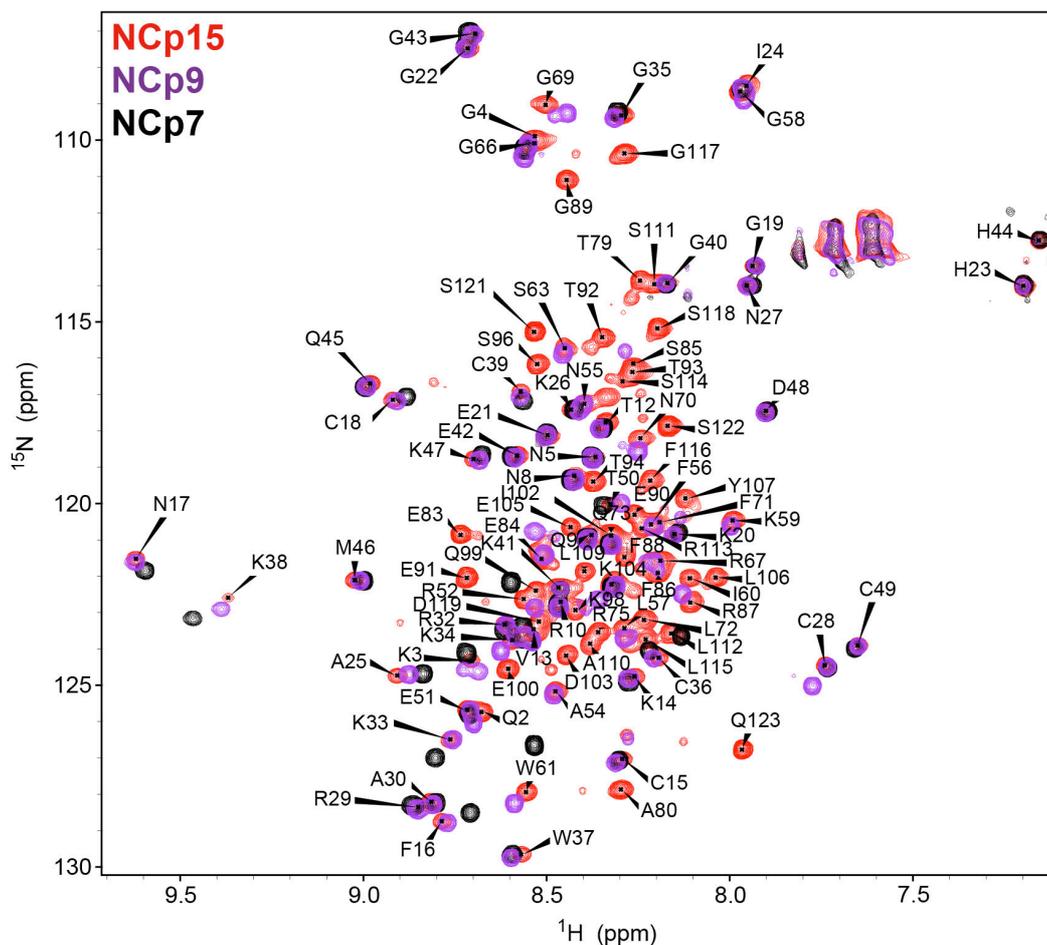
- in molecular recognition by proteins. *Nature* 448, 325-9 (2007).
49. Tzeng, S.R. & Kalodimos, C.G. Protein activity regulation by conformational entropy. *Nature* 488, 236-40 (2012).
 50. Liu, X. et al. Distinct Roles for Conformational Dynamics in Protein-Ligand Interactions. *Structure* 24, 2053-2066 (2016).
 51. Chakrabarti, K.S. et al. Conformational Selection in a Protein-Protein Interaction Revealed by Dynamic Pathway Analysis. *Cell reports* 14, 32-42 (2016).
 52. Beach, H., Cole, R., Gill, M.L. & Loria, J.P. Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *Journal of the American Chemical Society* 127, 9167-76 (2005).
 53. Boehr, D.D., McElheny, D., Dyson, H.J. & Wright, P.E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313, 1638-42 (2006).
 54. Lee, B.M., De Guzman, R.N., Turner, B.G., Tjandra, N. & Summers, M.F. Dynamical behavior of the HIV-1 nucleocapsid protein. *Journal of molecular biology* 279, 633-49 (1998).
 55. Cruceanu, M., Gorelick, R.J., Musier-Forsyth, K., Rouzina, I. & Williams, M.C. Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Journal of molecular biology* 363, 867-77 (2006).
 56. Wu, H. et al. Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus research* 171, 263-77 (2013).
 57. Salzman, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* 95, 13585-90 (1998).
 58. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* 6, 277-93 (1995).
 59. Goddard, T.D. & Kneller, D.G. SPARKY 3. *University of California, San Francisco*.
 60. Hansen, D.F., Vallurupalli, P. & Kay, L.E. An improved ¹⁵N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *The journal of physical chemistry. B* 112, 5898-904 (2008).
 61. Loria, J.P. & Nowak, T. Conformational changes in yeast pyruvate kinase studied by 205Ti+ NMR. *Biochemistry* 37, 6967-74 (1998).
 62. Mulder, F.A., Skrynnikov, N.R., Hon, B., Dahlquist, F.W. & Kay, L.E. Measurement

of slow (micro-seconds) time scale dynamics in protein side chains by ^{15}N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *Journal of the American Chemical Society* 123, 967-75 (2001).

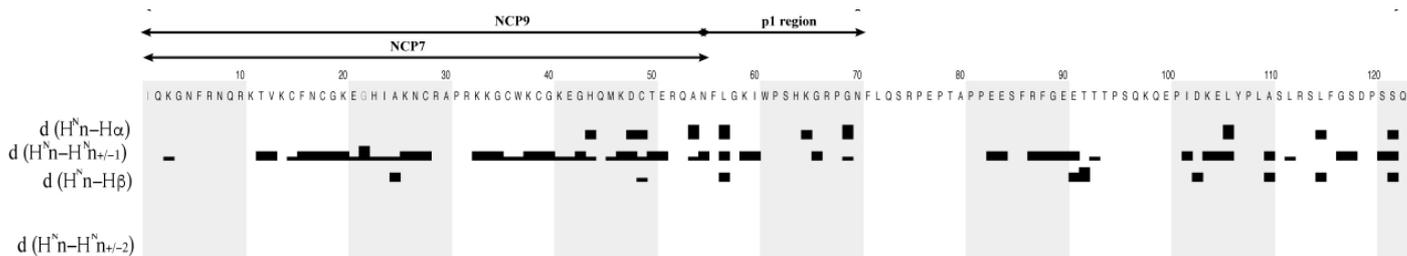
63. Ishima, R. & Torchia, D.A. Error estimation and global fitting in transverse-relaxation dispersion experiments to determine chemical-exchange parameters. *Journal of biomolecular NMR* 32, 41-54 (2005).
64. McConnell, H.M. Reaction Rates by nuclear magnetic resonance. *The Journal of Chemical Physics* 28, 430-431 (1958).
65. Davis, D.G., Perlman, M.E. & London, R.E. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *Journal of magnetic resonance. Series B* 104, 266-75 (1994).
66. Millet, O., Loria, J.P., Kroenke, C.D., Pons, M. & Palmer, A.G., 3rd. The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale. *J. Am. Chem. Soc.* 122, 2867-2877 (2000).
67. Mulder, F.A., Mittermaier, A., Hon, B., Dahlquist, F.W. & Kay, L.E. Studying excited states of proteins by NMR spectroscopy. *Nature structural biology* 8, 932-5 (2001).
68. Tellinghuisen, J. & Chodera, J.D. Systematic errors in isothermal titration calorimetry: concentrations and baselines. *Analytical biochemistry* 414, 297-9 (2011).

Supplementary Fig. 1 Comparison of NMR data recorded on nucleocapsid proteins at 10°C and 950 MHz. The three samples at a concentration of 0.9 mM were dialysed against the NMR buffer in the same container before recording the NMR experiments. **(A)** ^1H - ^{15}N BEST-TROSY {Solyom, 2013 #171} of NCp7 in black, NCp9 in purple and NCp15 in red. **(B)** NCp15 protein Chart table generated from the CcpNmr Analysis program (<http://www.ccpn.ac.uk/software/analysis>) showing a graphical representation of observed NOEs between protons H^{N} and $\text{H}\alpha$, $\text{H}^{\text{N}}_{\text{n}}$ and $\text{H}^{\text{N}}_{\text{n}+/-1}$, H^{N} and $\text{H}\beta$, $\text{H}^{\text{N}}_{\text{n}}$ and $\text{H}^{\text{N}}_{\text{n}+/-2}$, $\text{H}\alpha$ and $\text{H}^{\text{N}}_{\text{n}+/-2}$.

A

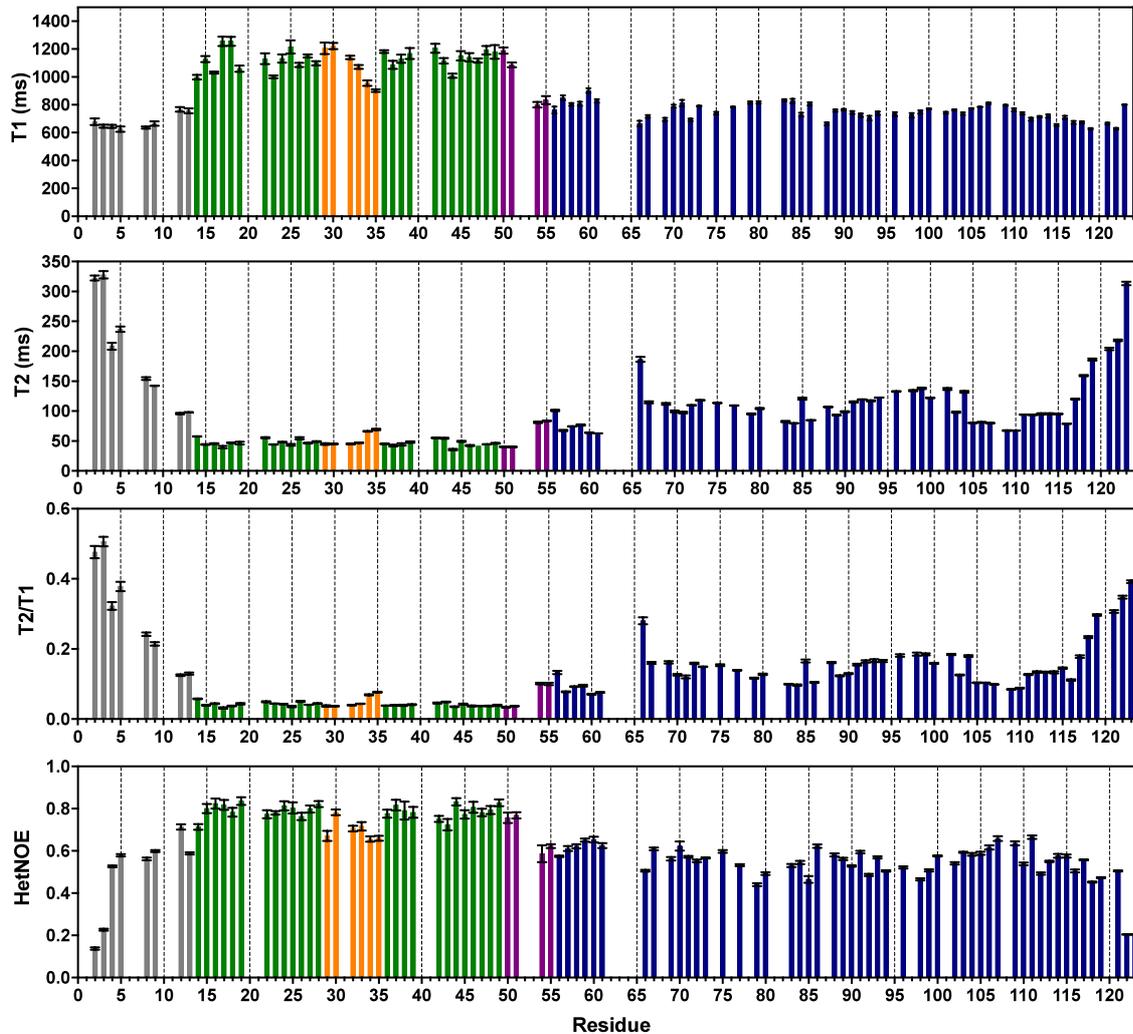


B

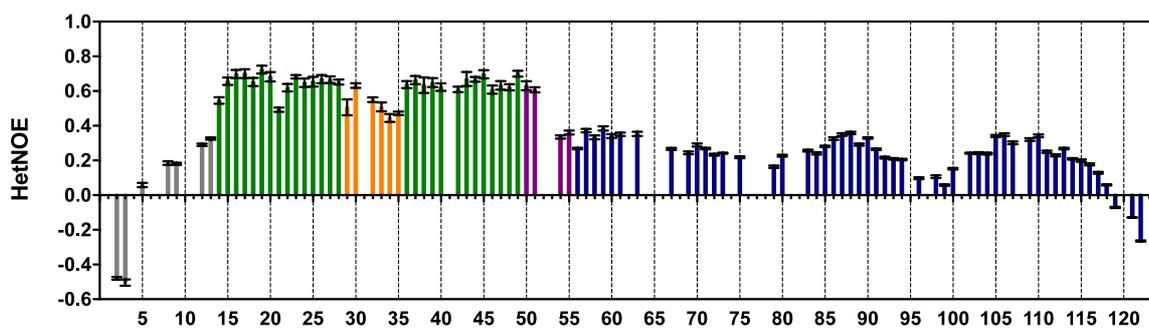


Supplementary Fig. 2 Dynamic characterization of NCp15 at 10°C (A) from ^{15}N NMR relaxation data recorded at 950 MHz (B) ^1H - ^{15}N HetNOE measured at 600 MHz. The color code for residues is described on figure 1.

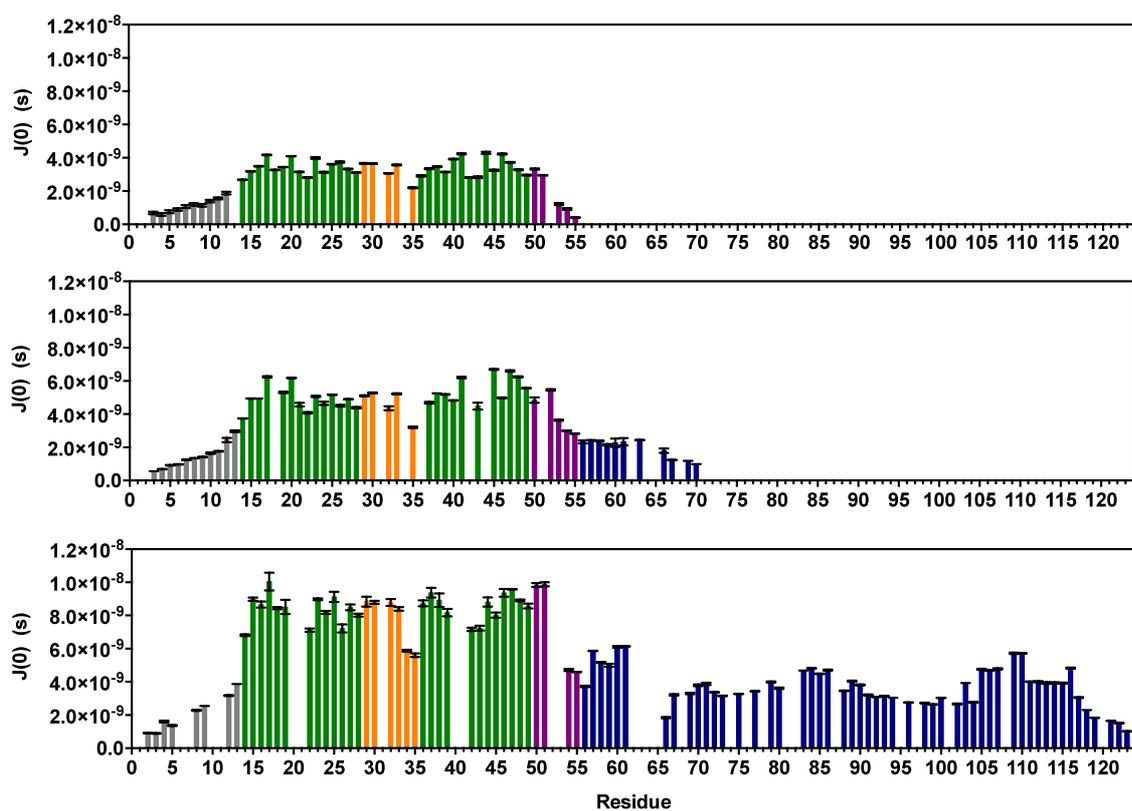
A



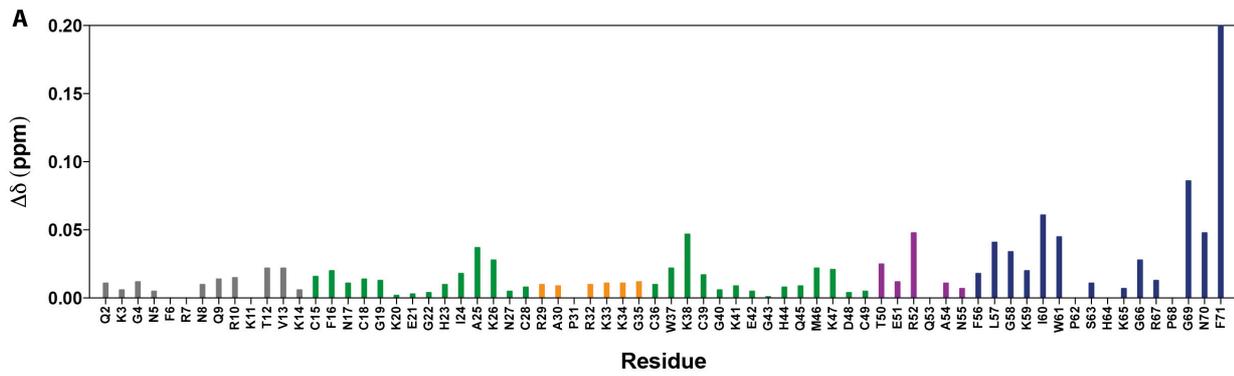
B



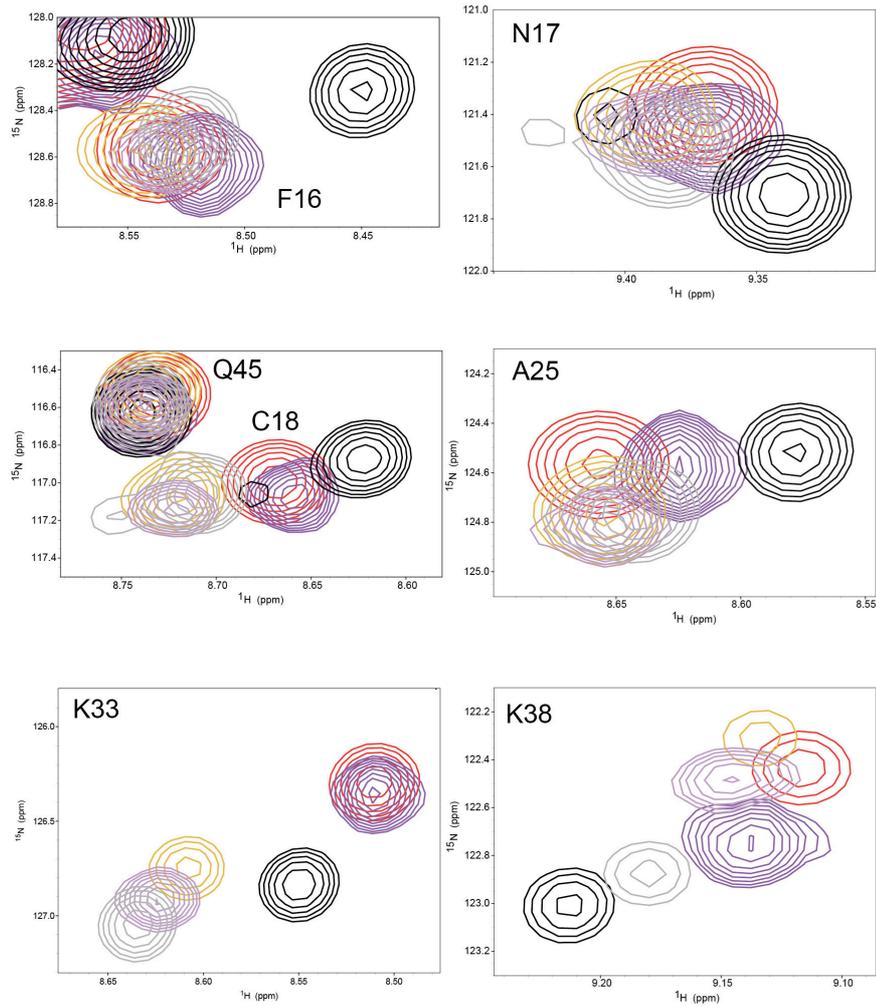
Supplementary Fig. 3 Spectral densities $J(0)$ extracted by spectral density mapping from ^{15}N relaxation data (T1, T2, HetNOE) of (A) NCp7, (B) NCp9 and (C) NCp15 at 950 MHz and 10°C. The color code for residues is described on figure 1.



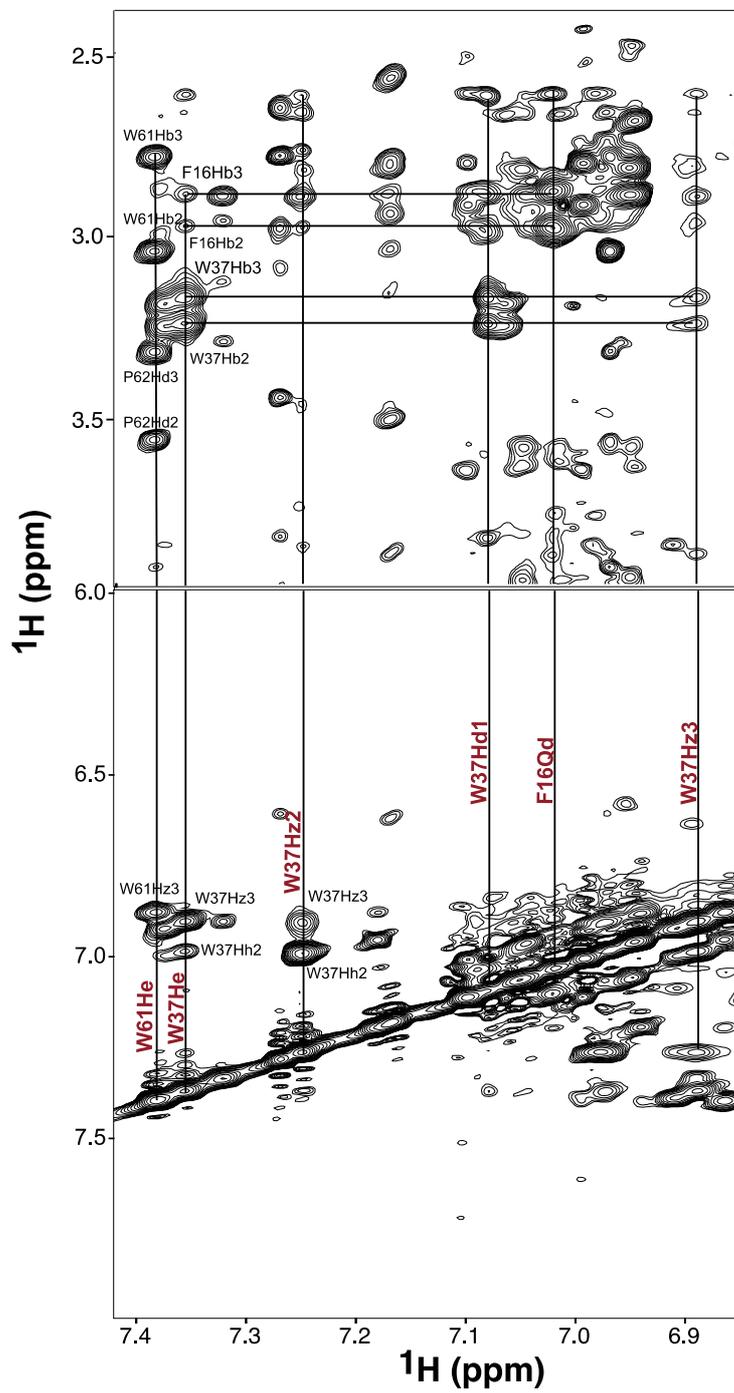
Supplementary Fig. 4 Analysis of NMR data recorded on nucleocapsid proteins recorded at 10°C. (A) Contributions of p6 to NC_d-p1 chemical shift perturbations measured on the amide groups. The p6 contribution was calculated as the difference in chemical shifts of NC_d-p1 amide groups (combined ¹H/¹⁵N shifts) between NCp15 and NCp9. The color code for residues is described on figure 1. (B) Comparison of NMR data recorded on nucleocapsid proteins recorded at 10°C and 700 MHz at two salt concentrations (25 mM and 500 mM NaCl). Zooms showing amide group resonances of F16, N17, C18, A25, K33 and K38 of NCp7 (Black and grey), NCp9 (purple and light purple), NCp15 (red and orange).



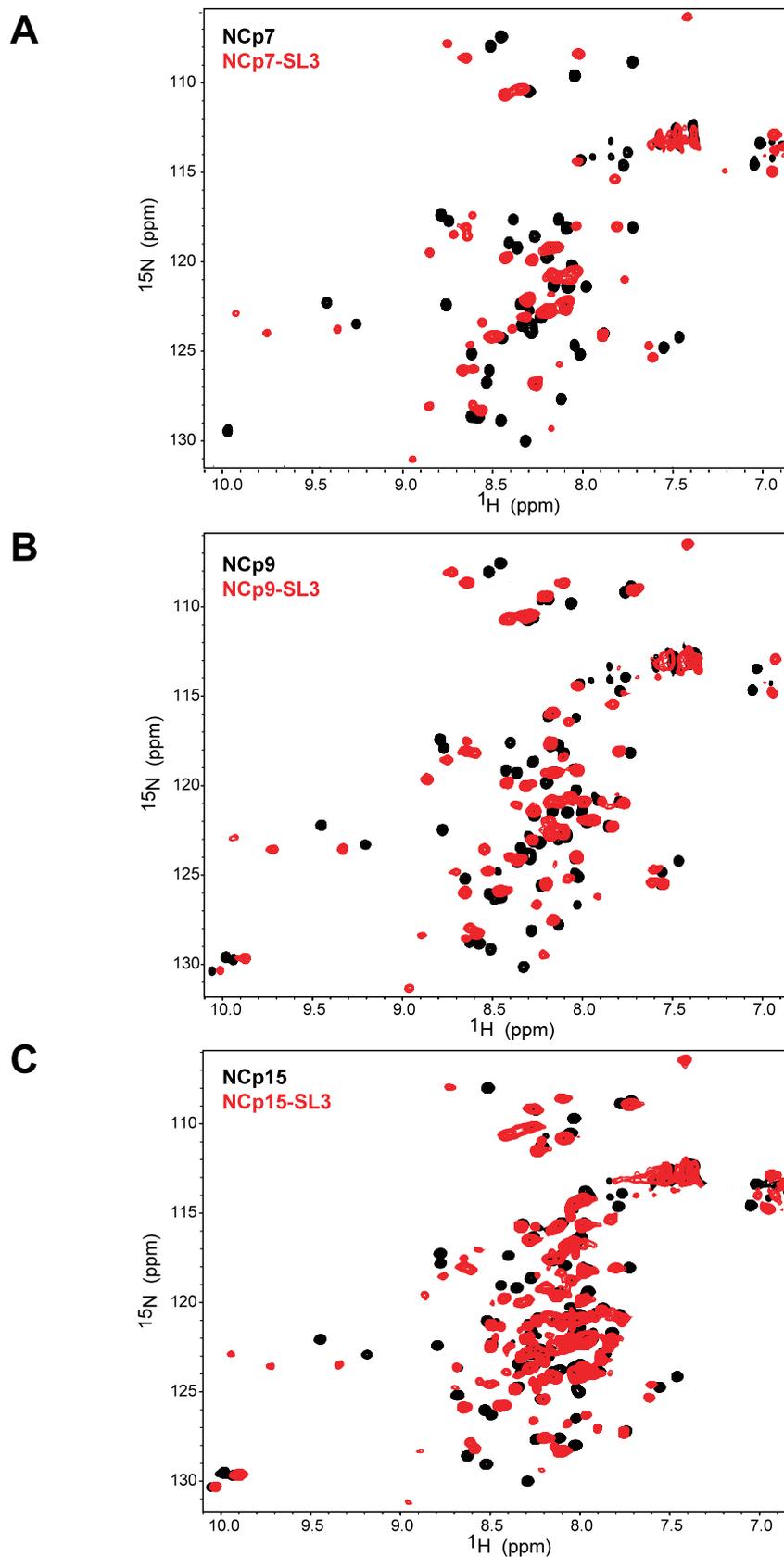
B **NCp7 (25 mM, 500 mM NaCl)** **NCp9 (25 mM, 500 mM NaCl)** **NCp15 (25 mM, 500 mM NaCl)**



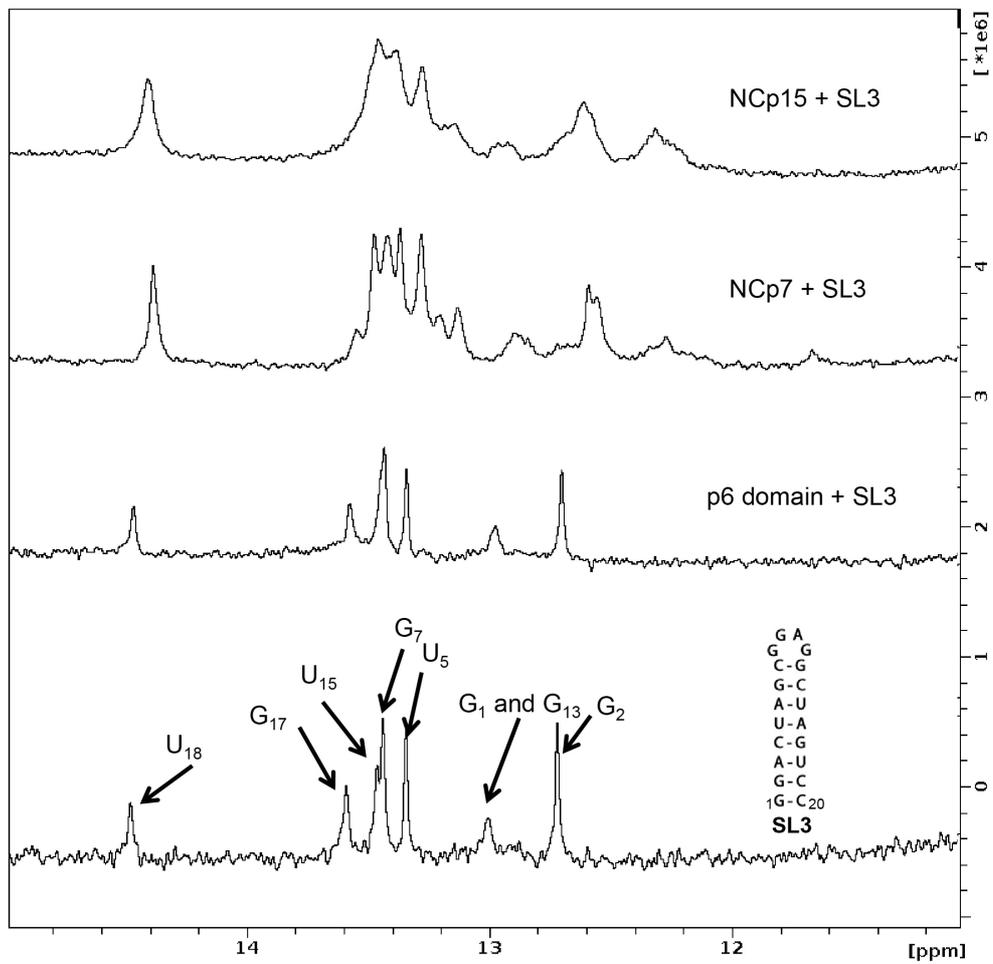
Supplementary Fig. 5 NOESY D₂O 950 MHz 150 ms NCp9 showing inter knuckles NOEs between F16 and W37.



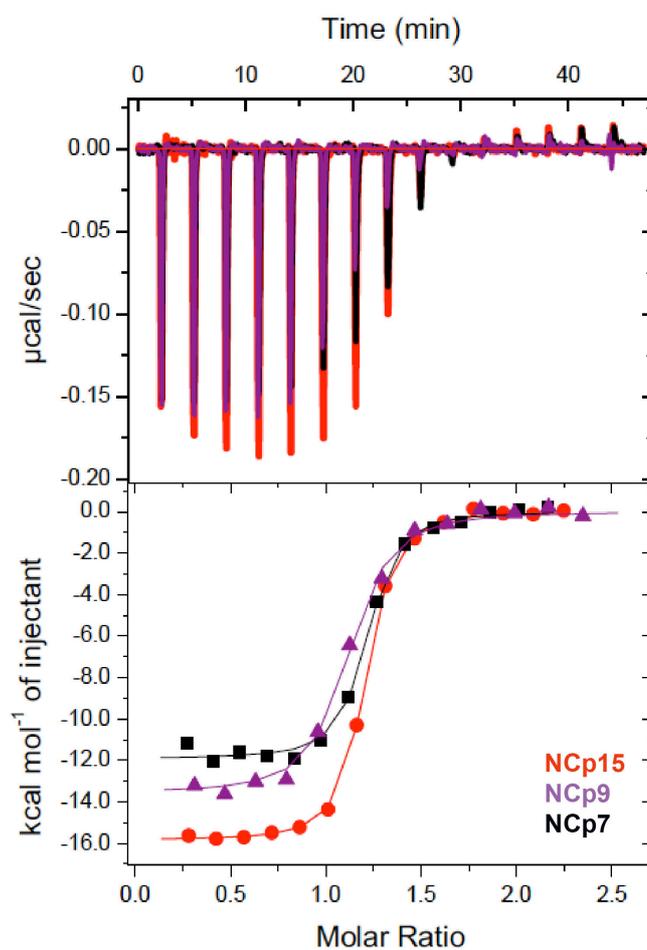
Supplementary Fig. 6 NMR chemical shift mappings of SL3 binding to the NC_d according to its maturation state. Superimposition of ¹H-¹⁵N BEST-TROSY NMR experiments recorded at 25°C and 950 Mhz for the protein alone (in black) and bound to SL3 (in red) at 1 : 1 ratio **(A)** NCp7, **(B)** NCp9 and **(C)** NCp15.



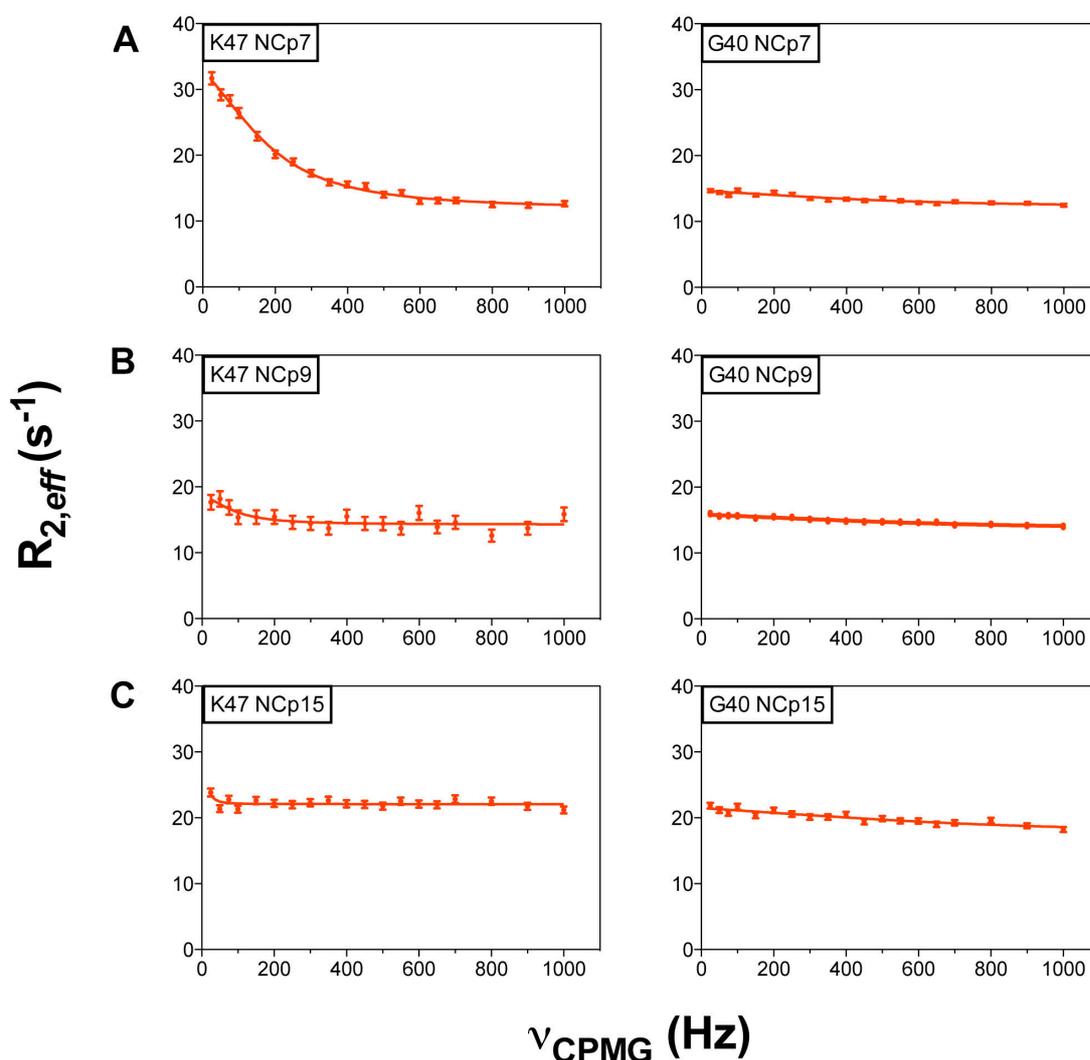
Supplementary Fig. 7 Interaction between NCp15, NCp7 and p6 with SL3 studied by 1D ^1H NMR experiment. 1D ^1H NMR spectra recorded at 25°C and 600 MHz showing the imino protons of SL3 upon addition of NC proteins. From bottom to top : free SL3, SL3 upon addition of p6, SL3 upon binding of NCp7 protein and SL3 upon binding of NCp15 protein at 1:1 ratio.



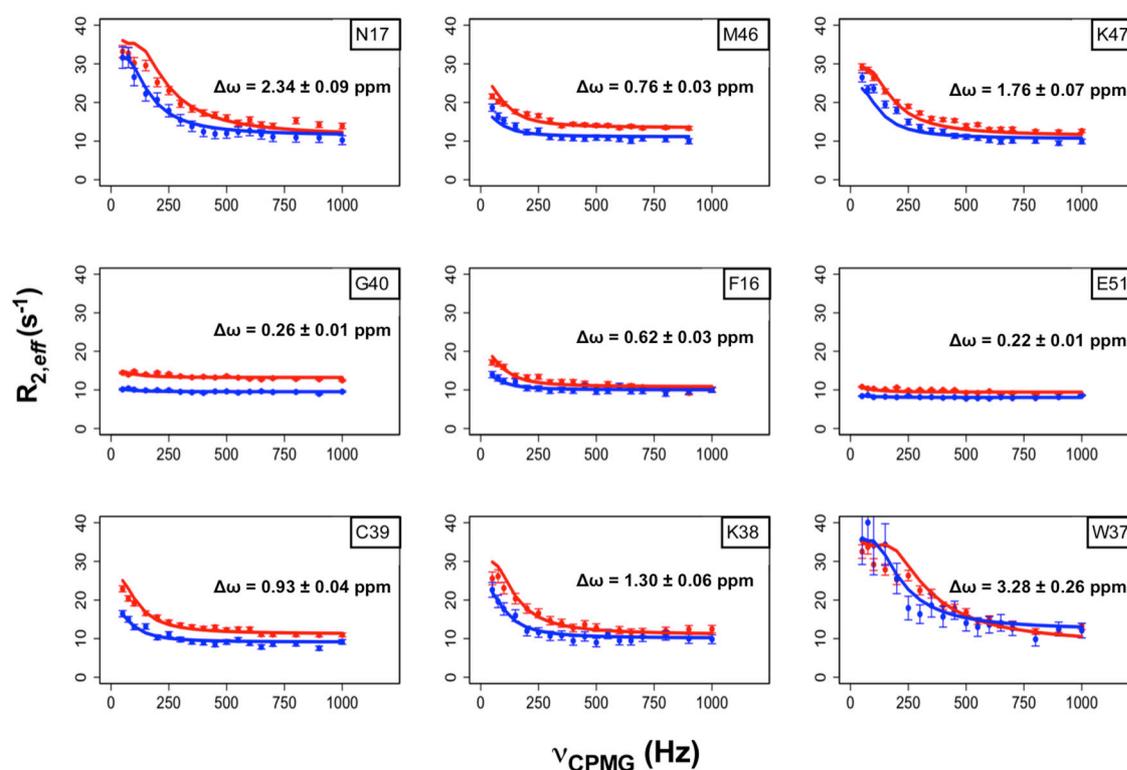
Supplementary Fig. 8 Representative ITC binding isotherms of the calorimetric titration of SL3 by NCp7 (in black), NCp9 (in purple) and NCp15 (in red) at 30°C and 25 mM NaCl. On the top: Raw data with the integration baseline. On the bottom: Data after peak integration and subtraction of blank titration. The line represents the fit to a 1:1 binding model.



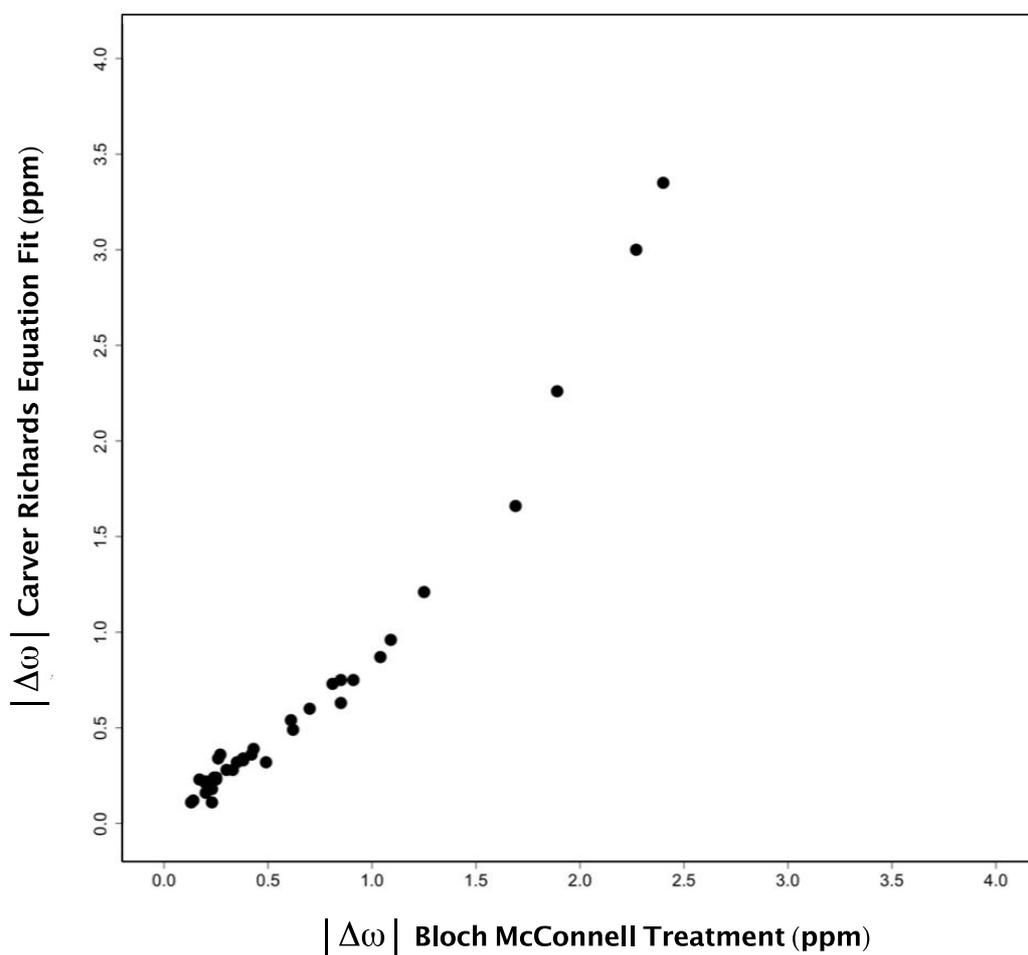
Supplementary Fig. 9 Examples of ^{15}N relaxation dispersion profiles and fits obtained for K47 and G40 in (A) NCp7, (B) NCp9 and (C) NCp15. The plain line for (A) was calculated using the parameters extracted from the global fit of the Carver-Richards equation for data obtained for all residues and at two magnetic fields (950 MHz and 600 MHz) and as described in Materials and Methods. For (B) and (C) the plain lines result from individual (per residue) fits of the Meiboom equation for data recorded at 950 MHz and at 10°C. The fit to the Meiboom equation (fast exchange) is better adapted for NCp9 and NCp15 where the observed R_{ex} are small. Error bars are \pm one standard deviation and were determined as described in Materials & Methods.



Supplementary Fig. 10 ^{15}N -CPMG relaxation dispersion experiments show some exchanging residues for NCp7. Data recorded at 22.3 T ($\nu_{\text{H}} = 950$ MHz) and 14.1 T ($\nu_{\text{H}} = 600$ MHz) are represented in red and blue respectively. The plain lines correspond to the calculated $R_{2,\text{eff}}$ using the mean parameters (k_{ex} , p_{B} and $\Delta\omega$) obtained from global fit using the data for all residues and at the two fields. The relaxation dispersion data were fit with the general Carver-Richards equation for two site-exchange as described in Materials & Methods. The mean parameters, were obtained from 100 simulated Monte Carlo data using the experimental $R_{2,\text{eff}}$ uncertainty. The mean parameters are reported in Supplementary Table 1.



Supplementary Fig. 11 Correlation between $|\Delta\omega|$ values of the different residues for NCp7 obtained from relaxation dispersion measurements and extracted using the two different methods described in Material & Methods. Each point corresponds to one residue.



Supplementary Table 1 Fitted Parameters of ^{15}N -CPMG Relaxation Dispersion

Curves for NCp7. The relaxation dispersion data were fit with the general Carver-Richards equation for two site-exchange as described in Materials & Methods

| Residue | $\Delta\omega$ (ppm) | R_2 (s^{-1}) 600 MHz | R_2 (s^{-1}) 950 MHz |
|---------|----------------------|-----------------------------------|-----------------------------------|
| K14 | 0.32 ± 0.02 | 7.31 ± 0.10 | 8.70 ± 0.09 |
| C15 | 0.41 ± 0.02 | 9.60 ± 0.12 | 11.33 ± 0.11 |
| F16 | 0.62 ± 0.03 | 10.01 ± 0.22 | 10.86 ± 0.19 |
| N17 | 2.34 ± 0.09 | 11.37 ± 0.46 | 11.43 ± 0.39 |
| C18 | 0.33 ± 0.02 | 9.34 ± 0.11 | 11.28 ± 0.10 |
| G19 | 0.23 ± 0.02 | 8.86 ± 0.10 | 11.04 ± 0.11 |
| K20 | 0.23 ± 0.01 | 10.01 ± 0.07 | 13.16 ± 0.07 |
| E21 | 0.26 ± 0.01 | 8.81 ± 0.05 | 10.75 ± 0.06 |
| H23 | 0.24 ± 0.01 | 3.44 ± 0.04 | 11.24 ± 0.05 |
| I24 | 0.51 ± 0.02 | 8.91 ± 0.08 | 10.78 ± 0.09 |
| A25 | 0.30 ± 0.02 | 9.15 ± 0.06 | 11.53 ± 0.09 |
| K26 | 0.33 ± 0.02 | 10.48 ± 0.08 | 13.24 ± 0.08 |
| N27 | 0.38 ± 0.02 | 9.33 ± 0.08 | 11.70 ± 0.08 |
| C28 | 0.29 ± 0.01 | 8.37 ± 0.05 | 10.42 ± 0.04 |
| R29 | 0.25 ± 0.01 | 9.66 ± 0.09 | 11.52 ± 0.09 |
| A30 | 0.18 ± 0.01 | 9.09 ± 0.05 | 11.01 ± 0.06 |
| K33 | 2.86 ± 0.08 | 10.61 ± 0.42 | 9.29 ± 0.40 |
| G35 | 0.77 ± 0.03 | 5.59 ± 0.34 | 6.79 ± 0.13 |
| C36 | 0.42 ± 0.02 | 7.94 ± 0.10 | 10.67 ± 0.11 |
| W37 | 3.28 ± 0.26 | 12.23 ± 0.85 | 8.49 ± 0.87 |
| K38 | 1.30 ± 0.06 | 10.13 ± 0.36 | 11.03 ± 0.37 |
| C39 | 0.93 ± 0.04 | 9.11 ± 0.21 | 11.20 ± 0.19 |
| G40 | 0.26 ± 0.01 | 9.53 ± 0.07 | 13.20 ± 0.07 |
| E42 | 0.35 ± 0.01 | 7.87 ± 0.05 | 9.46 ± 0.05 |
| H44 | 0.54 ± 0.02 | 10.00 ± 0.10 | 13.88 ± 0.11 |
| Q45 | 0.62 ± 0.02 | 9.19 ± 0.07 | 10.89 ± 0.09 |
| M46 | 0.76 ± 0.03 | 11.13 ± 0.24 | 13.47 ± 0.18 |
| K47 | 1.76 ± 0.07 | 10.75 ± 0.22 | 11.62 ± 0.22 |
| D48 | 0.27 ± 0.02 | 9.32 ± 0.07 | 12.01 ± 0.09 |
| C49 | 0.24 ± 0.01 | 8.21 ± 0.05 | 10.57 ± 0.06 |
| E51 | 0.22 ± 0.01 | 8.04 ± 0.04 | 9.47 ± 0.05 |
| Q53 | 0.18 ± 0.01 | 4.09 ± 0.03 | 4.26 ± 0.02 |

| | | | |
|------------|-----------------|-----------------|-----------------|
| A54 | 0.19 ± 0.01 | 3.24 ± 0.02 | 3.01 ± 0.02 |
|------------|-----------------|-----------------|-----------------|

Results from global fitting of data with $k_{\text{ex}} = 472 \pm 22 \text{ s}^{-1}$ and $p_{\text{B}} = 0.055 \pm 0.002$. The error for each parameter was estimated from Monte Carlo simulations.

V Conclusions et perspectives

Dans un premier temps, nous avons mené une série d'études sur les propriétés de fixation de la nucléocapside aux acides nucléiques ainsi que sur la capacité de la NC à déstabiliser leur structure secondaire. Les études étaient dirigées principalement selon deux axes :

1) mieux comprendre le processus de déstabilisation par la NC et notamment la relation entre le processus de fixation "initial" aux séquences simple-brin et la déstabilisation des séquences double-brin adjacentes ;

2) identifier une éventuelle "polarité de fixation", de la protéine aux acides nucléiques ADN et ARN ; cette dernière hypothèse nous étant suggéré par les analyses effectuées sur les complexes NC-acides nucléiques existants ^{353, 362}.

Notons que nous avons aussi tenté de comprendre si ces deux aspects ne pouvaient être liés, c'est à dire si le mode de fixation de la NC n'était pas en mesure d'influencer le choix des zones déstabilisées.

La première étude, présentée sous forme de publication dans **RNA**, montre que les "premières" paires de bases déstabilisées par la NC dans TAR ARN, sont toutes localisées à un endroit particulier : la tige "apicale". Ces résultats suggèrent une fixation à la boucle interne et une déstabilisation "à partir" de ce site. De façon intéressante, plusieurs sites du génome du VIH-1 qui sont déstabilisés par la NC, présentent une structure semblable à celle identifiée dans cette étude ^{264, 412}. Ces travaux nous ont permis aussi de déterminer l'effet de la NC sur les associations des acides nucléiques, lorsque les rapports NC:nt sont très faibles. Dans de telles circonstances l'effet de la NC sur les vitesses d'hybridation est toujours perceptible. Nous avons également pu montrer que la vitesse de conversion vers le duplexe étendu est fortement dépendant du rapport NC:nt.

Nos études par anisotropie de fluorescence, nous ont permis de connaître l'affinité de la protéine avec différentes séquences d'acides nucléiques. Nous avons retrouvé les principales préférences connues de la NCp7 : les séquences GXG, TG et TXG. Une séquence qui n'avait pas été identifiée auparavant et qui apparaît comme un substrat de haute affinité : la séquence GGC.

La comparaison des interactions de la NC(11-55) avec la même séquence (SL3) sous forme ARN et sous forme ADN montre des différences importantes, avec une plus forte affinité de liaison pour les séquences ARN comparé aux ADN. L'association de nos données avec celles provenant de notre étude RMN sur l'interaction de la NC avec SL3

ADN et SL3 ADN20G, a permis de montrer que la présence de guanines non appariées n'est pas suffisante pour garantir une haute affinité. Il faut qu'en plus ces guanines soient bien accessibles, être exemptes à la fois d'interactions d'appariement et d'interactions d'empilement. Si ces conditions sont respectées, les substrats contenant deux guanines avec la séquence GXG ont une très forte affinité pour la NC(11-55). Ce résultat nous a ouvert la voie vers l'étude du complexes NC-d(TAGAGAT) par RMN ainsi que son homologue ARN. L'analyse complète de ces deux complexes par RMN PRE est en cours, devrait nous permettre de réaliser une étude comparative, qui a pour but de confirmer et de préciser la polarité de fixation de la protéine le long des chaînes d'acides nucléiques, selon leur nature ARN/ADN.

En outre, l'étude de la capacité de la NCp7 à favoriser les associations de formes repliées d'ADN et d'ARN en utilisant la méthode FRET, a démontré que la position des boucles internes dans ces séquences repliées influait sur les vitesses d'association. Ces données sont en accord avec une fixation et une déstabilisation ordonnée de la protéine sur les acides nucléiques. Par ailleurs, les couples mutants "stem" et "loop" nous ont permis de déceler les voies d'hybridation empruntées par les séquences TAR-cTAR en présence de la NCp7. Il est évident que la position des boucles internes influence les voies d'hybridation empruntées. Nous retrouvons pour les séquences sauvages TAR0-cTAR0 un mécanisme majoritaire de type "zipper", ce qui est en accord avec ce qui a été démontré ultérieurement dans la littérature ⁴²⁴ et pour les séquences mutantes TAR1-cTAR1, celles-ci s'hybrideraient plutôt selon un mécanisme "kissing".

Dans la dernière partie de la thèse nous nous sommes intéressés à la dynamique interne de la protéine NCp7 et des formes immatures NCp9 et NCp15. Lors d'un précédent travail, nous avons, en utilisant les études de dynamique interne de la protéine, proposé un modèle de fonctionnement "séquentiel" de la NC qui tentait d'expliquer certaines propriétés de fonctionnement de la protéine. Les deux doigts de zinc apparaissent en effet occupés des fonctions assez différentes. Ainsi, un grand nombre d'études soulignent que les guanines reconnues par la NC interagissent majoritairement avec le Tryptophane 37 (situé dans le doigt de zinc 2), et d'autre part de nombreux travaux indiquent que ce sont les mutations dans le doigt de zinc 1 qui ont le plus d'impact sur l'activité et l'infectiosité du virus. Ce dernier doigt de zinc est aussi clairement plus important pour l'activité de protéine chaperon. De plus, il semble que non seulement la nature, mais aussi la position des deux doigts de zinc dans la séquence, soient importantes. Nous avons donc ici poursuivi ces études dynamiques sur plusieurs échelles de temps et en nous attachant à la comparaison avec la dynamique des protéines

immatures NCp9 et NCp15. Ces deux protéines se distinguent de la forme mature par la présence de parties désordonnées (p1 dans NCp9 et p1-p6 dans NCp15) qui interagissent avec la partie NCp7, sachant qu'on a une composante hydrophobe pour p1 et une électrostatique pour p6.

Les domaines p1 et p6 apparaissent capables de moduler de façon importante la dynamique du domaine NC, sur différentes échelles de temps, cela du fait de contacts transitoires entre les domaines p1 et p6, qui sont largement désordonnés. La nature transitoire de ces contacts est déduite notamment du fait que les parties p1 et p6 restent principalement sous une forme désordonnée.

Les paramètres de relaxation sont précieux pour identifier le rôle des domaines additionnels p1 et p6. Notamment, on observe un doublement des valeurs $J(0)$ pour les domaines de doigt de zinc repliés entre La NCp7 et La NCp15. Ce comportement révèle que les deux doigts de zinc se comportent davantage comme une entité unique, probablement du fait des contacts transitoires entre p1-p6 et p7. Les domaines p1 et p6 non structurés jouent aussi probablement un rôle de freinage dans la libre rotation des protéines. On doit noter que la partie "linker" est toujours animée d'une dynamique importante dans NCp9 et NCp15.

Par ailleurs nous avons également mis en évidence un équilibre entre la conformation majeure de la NCp7 et une forme mineure (5%), en utilisant les expériences de dispersion de relaxation. De tels équilibres ne semblent pas être présents dans NCp9 et dans NCp15. En effectuant les mesures à deux champs magnétiques, nous avons pu extraire les paramètres associés à cet échange conformationnel dans la NCp7. De façon intéressante, la forme mineure semble être proche de la forme liée à l'ARN SL3, ce qui suggère un mécanisme de reconnaissance du type sélection coformationnel.

Les propriétés d'interaction avec les acides nucléiques (et notamment avec la séquence SL3) des trois protéines semblent impliquer les mêmes résidus. L'affinité de liaison apparaît également similaire, mais de larges différences sont observées au niveau des paramètres thermodynamiques de l'interaction.

Ces propriétés influent sur l'activité de protéine chaperon de la NC au cours des différentes étapes du cycle viral, nous amène à reconsidérer le rôle des différentes parties de la polyprotéine Gag sur les propriétés du domaine NC, au cours de l'étape de reconnaissance et de la sélection du génome viral ARN ainsi qu'au cours de l'assemblage des nouvelles particules virales. Les interactions domaine p7-domaine p6, notamment à l'intérieur de Gag, peuvent en effet varier significativement selon que Gag est sous forme monomérique ou multimérique. Il est clair que les contacts domaine NC-p6 seront plus

importants dans un contexte oligomérique que dans un contexte monomérique. Ces contacts moduleraient à leur tour les propriétés d'interaction avec les ARNs. Rappelons que les propriétés d'interaction du domaine NC de Gag avec les ARN varient de façon importante au cours du cycle de replication et qu'un changement important de ces propriétés se produit ¹⁵¹.

En perspective, de nouvelles données sur les complexes devront être obtenus en travaillant à d'autres températures et à des champs plus élevés en RMN. L'étude des équilibres conformationnels et leurs relations avec les propriétés de protéine chaperon de la NC sera également poursuivie.

Ce travail pourrait appuyer des études de développement d'une classe d'inhibiteurs spécifique de la NCp7. En effet, le rôle de la protéine de nucléocapside à différentes étapes du cycle, couplé au fait que sa séquence est très conservée, la désigne comme une cible thérapeutique attrayante.

VI Références bibliographiques

1. Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L., Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220* (4599), 868-71.
2. Gallo, R. C.; Sliski, A.; Wong-Staal, F., Origin of human T-cell leukaemia-lymphoma virus. *Lancet* **1983**, *2* (8356), 962-3.
3. Blattner, W.; Gallo, R. C.; Temin, H. M., HIV causes AIDS. *Science* **1988**, *241* (4865), 515-6.
4. de Silva, T. I.; Cotten, M.; Rowland-Jones, S. L., HIV-2: the forgotten AIDS virus. *Trends Microbiol* **2008**, *16* (12), 588-95.
5. Mourez, T.; Simon, F.; Plantier, J. C., Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev* **2013**, *26* (3), 448-61.
6. Hahn, B. H.; Shaw, G. M.; De Cock, K. M.; Sharp, P. M., AIDS as a zoonosis: scientific and public health implications. *Science* **2000**, *287* (5453), 607-14.
7. Simon, J. H.; Gaddis, N. C.; Fouchier, R. A.; Malim, M. H., Evidence for a newly discovered cellular anti-HIV-1 phenotype. *Nat Med* **1998**, *4* (12), 1397-400.
8. Hirsch, V. M.; Olmsted, R. A.; Murphey-Corb, M.; Purcell, R. H.; Johnson, P. R., An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **1989**, *339* (6223), 389-92.
9. Muller-Trutwin, M. C.; Corbet, S.; Souquiere, S.; Roques, P.; Versmisse, P.; Ayouba, A.; Delarue, S.; Nerrienet, E.; Lewis, J.; Martin, P.; Simon, F.; Barre-Sinoussi, F.; Mauclore, P., SIVcpz from a naturally infected Cameroonian chimpanzee: biological and genetic comparison with HIV-1 N. *J Med Primatol* **2000**, *29* (3-4), 166-72.
10. Peeters, M.; Gershy-Damet, G. M.; Franssen, K.; Koffi, K.; Coulibaly, M.; Delaporte, E.; Piot, P.; van der Groen, G., Virological and polymerase chain reaction studies of HIV-1/HIV-2 dual infection in Cote d'Ivoire. *Lancet* **1992**, *340* (8815), 339-40.
11. Peeters, M.; Honore, C.; Huet, T.; Bedjabaga, L.; Ossari, S.; Bussi, P.; Cooper, R. W.; Delaporte, E., Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* **1989**, *3* (10), 625-30.
12. Gao, F.; Bailes, E.; Robertson, D. L.; Chen, Y.; Rodenburg, C. M.; Michael, S. F.; Cummins, L. B.; Arthur, L. O.; Peeters, M.; Shaw, G. M.; Sharp, P. M.; Hahn, B. H., Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **1999**, *397* (6718), 436-41.
13. Gao, F.; Yue, L.; White, A. T.; Pappas, P. G.; Barchue, J.; Hanson, A. P.; Greene, B. M.; Sharp, P. M.; Shaw, G. M.; Hahn, B. H., Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* **1992**, *358* (6386), 495-9.
14. Keele, B. F.; Van Heuverswyn, F.; Li, Y.; Bailes, E.; Takehisa, J.; Santiago, M. L.;

Bibollet-Ruche, F.; Chen, Y.; Wain, L. V.; Liegeois, F.; Loul, S.; Ngole, E. M.; Bienvenue, Y.; Delaporte, E.; Brookfield, J. F.; Sharp, P. M.; Shaw, G. M.; Peeters, M.; Hahn, B. H., Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* **2006**, *313* (5786), 523-6.

15. Peeters, M.; Gueye, A.; Mboup, S.; Bibollet-Ruche, F.; Ekaza, E.; Mulanga, C.; Ouedrago, R.; Gandji, R.; Mpele, P.; Dibanga, G.; Koumare, B.; Saidou, M.; Esu-Williams, E.; Lombart, J. P.; Badombena, W.; Luo, N.; Vanden Haesevelde, M.; Delaporte, E., Geographical distribution of HIV-1 group O viruses in Africa. *AIDS* **1997**, *11* (4), 493-8.

16. McCutchan, F. E., Understanding the genetic diversity of HIV-1. *AIDS* **2000**, *14 Suppl 3*, S31-44.

17. Jonassen, T. O.; Stene-Johansen, K.; Berg, E. S.; Hungnes, O.; Lindboe, C. F.; Froland, S. S.; Grinde, B., Sequence analysis of HIV-1 group O from Norwegian patients infected in the 1960s. *Virology* **1997**, *231* (1), 43-7.

18. Chen, Z.; Telfier, P.; Gettie, A.; Reed, P.; Zhang, L.; Ho, D. D.; Marx, P. A., Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J Virol* **1996**, *70* (6), 3617-27.

19. Kannangai, R.; David, S.; Sridharan, G., Human immunodeficiency virus type-2-A milder, kinder virus: an update. *Indian J Med Microbiol* **2012**, *30* (1), 6-15.

20. Gelderblom, H. R.; Hausmann, E. H.; Ozel, M.; Pauli, G.; Koch, M. A., Fine structure of human immunodeficiency virus (HIV) and immunolocalization of structural proteins. *Virology* **1987**, *156* (1), 171-6.

21. Corbin, A.; Grigorov, B.; Roingeard, P.; Darlix, J. L.; Muriaux, D., [Revisiting HIV-1 assembly]. *Med Sci (Paris)* **2008**, *24* (1), 49-55.

22. Abbas, W.; Herbein, G., Plasma membrane signaling in HIV-1 infection. *Biochim Biophys Acta* **2014**, *1838* (4), 1132-42.

23. Sierra, S.; Kupfer, B.; Kaiser, R., Basics of the virology of HIV-1 and its replication. *J Clin Virol* **2005**, *34* (4), 233-44.

24. Alfadhli, A.; Still, A.; Barklis, E., Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *Journal of virology* **2009**, *83* (23), 12196-203.

25. Datta, S. A.; Zhao, Z.; Clark, P. K.; Tarasov, S.; Alexandratos, J. N.; Campbell, S. J.; Kvaratskhelia, M.; Lebowitz, J.; Rein, A., Interactions between HIV-1 Gag molecules in solution: an inositol phosphate-mediated switch. *J Mol Biol* **2007**, *365* (3), 799-811.

26. Fiorentini, S.; Marini, E.; Caracciolo, S.; Caruso, A., Functions of the HIV-1 matrix protein p17. *New Microbiol* **2006**, *29* (1), 1-10.

27. Ghanam, R. H.; Samal, A. B.; Fernandez, T. F.; Saad, J. S., Role of the HIV-1 Matrix Protein in Gag Intracellular Trafficking and Targeting to the Plasma Membrane for Virus Assembly. *Front Microbiol* **2012**, *3*, 55.
28. Ganser-Pornillos, B. K.; Cheng, A.; Yeager, M., Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* **2007**, *131* (1), 70-9.
29. Pornillos, O.; Ganser-Pornillos, B. K.; Kelly, B. N.; Hua, Y.; Whitby, F. G.; Stout, C. D.; Sundquist, W. I.; Hill, C. P.; Yeager, M., X-ray structures of the hexameric building block of the HIV capsid. *Cell* **2009**, *137* (7), 1282-92.
30. Bampi, C.; Jacquenet, S.; Lener, D.; Decimo, D.; Darlix, J. L., The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance. *Int J Biochem Cell Biol* **2004**, *36* (9), 1668-86.
31. Greene, W. C.; Peterlin, B. M., Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nat Med* **2002**, *8* (7), 673-80.
32. Balasubramaniam, M.; Freed, E. O., New insights into HIV assembly and trafficking. *Physiology (Bethesda)* **2011**, *26* (4), 236-51.
33. Basu, V. P.; Song, M.; Gao, L.; Rigby, S. T.; Hanson, M. N.; Bambara, R. A., Strand transfer events during HIV-1 reverse transcription. *Virus Res* **2008**, *134* (1-2), 19-38.
34. Bender, W.; Davidson, N., Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* **1976**, *7* (4), 595-607.
35. Paillart, J. C.; Berthoux, L.; Ottmann, M.; Darlix, J. L.; Marquet, R.; Ehresmann, B.; Ehresmann, C., A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J Virol* **1996**, *70* (12), 8348-54.
36. Paillart, J. C.; Shehu-Xhilaga, M.; Marquet, R.; Mak, J., Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol* **2004**, *2* (6), 461-72.
37. Paillart, J. C.; Marquet, R.; Skripkin, E.; Ehresmann, C.; Ehresmann, B., Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* **1996**, *78* (7), 639-53.
38. Paillart, J. C.; Skripkin, E.; Ehresmann, B.; Ehresmann, C.; Marquet, R., A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* **1996**, *93* (11), 5572-7.
39. Buckman, J. S.; Bosche, W. J.; Gorelick, R. J., Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *J Virol* **2003**, *77* (2), 1469-80.

40. Lapadat-Tapolsky, M.; De Rocquigny, H.; Van Gent, D.; Roques, B.; Plasterk, R.; Darlix, J. L., Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res* **1993**, *21* (4), 831-9.
41. Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess, J. W., Jr.; Swanstrom, R.; Burch, C. L.; Weeks, K. M., Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **2009**, *460* (7256), 711-6.
42. Sleiman, D.; Goldschmidt, V.; Barraud, P.; Marquet, R.; Paillart, J. C.; Tisne, C., Initiation of HIV-1 reverse transcription and functional role of nucleocapsid-mediated tRNA/viral genome interactions. *Virus Res* **2012**, *169* (2), 324-39.
43. Harrich, D.; Ulich, C.; Gaynor, R. B., A critical role for the TAR element in promoting efficient human immunodeficiency virus type 1 reverse transcription. *J Virol* **1996**, *70* (6), 4017-27.
44. Ronsard, L.; Ganguli, N.; Singh, V. K.; Mohankumar, K.; Rai, T.; Sridharan, S.; Pajaniradje, S.; Kumar, B.; Rai, D.; Chaudhuri, S.; Coumar, M. S.; Ramachandran, V. G.; Banerjea, A. C., Impact of Genetic Variations in HIV-1 Tat on LTR-Mediated Transcription via TAR RNA Interaction. *Front Microbiol* **2017**, *8*, 706.
45. Ronsard, L.; Rai, T.; Rai, D.; Ramachandran, V. G.; Banerjea, A. C., In silico Analyses of Subtype Specific HIV-1 Tat-TAR RNA Interaction Reveals the Structural Determinants for Viral Activity. *Front Microbiol* **2017**, *8*, 1467.
46. Krawczyk, K.; Demharter, S.; Knapp, B.; Deane, C. M.; Minary, P., In silico structural modelling of multiple epigenetic marks on DNA. *Bioinformatics* **2017**.
47. Marcello, A.; Zoppe, M.; Giacca, M., Multiple modes of transcriptional regulation by the HIV-1 Tat transactivator. *IUBMB Life* **2001**, *51* (3), 175-81.
48. Darlix, J. L.; Cristofari, G.; Rau, M.; Pechoux, C.; Berthoux, L.; Roques, B., Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs. *Adv Pharmacol* **2000**, *48*, 345-72.
49. Helga-Maria, C.; Hammarskjold, M. L.; Rekosh, D., An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. *J Virol* **1999**, *73* (5), 4127-35.
50. Baudin, F.; Marquet, R.; Isel, C.; Darlix, J. L.; Ehresmann, B.; Ehresmann, C., Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J Mol Biol* **1993**, *229* (2), 382-97.
51. Kleiman, L., tRNA(Lys3): the primer tRNA for reverse transcription in HIV-1. *IUBMB Life* **2002**, *53* (2), 107-14.
52. Huthoff, H.; Bugala, K.; Barciszewski, J.; Berkhout, B., On the importance of the primer activation signal for initiation of tRNA(lys3)-primed reverse transcription of the HIV-1 RNA genome. *Nucleic Acids Res* **2003**, *31* (17), 5186-94.

53. Tisne, C.; Roques, B. P.; Dardel, F., The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome. *J Biol Chem* **2004**, *279* (5), 3588-95.
54. Clever, J.; Sasseti, C.; Parslow, T. G., RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J Virol* **1995**, *69* (4), 2101-9.
55. Harrison, G. P.; Lever, A. M., The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J Virol* **1992**, *66* (7), 4144-53.
56. Abd El-Wahab, E. W.; Smyth, R. P.; Mailler, E.; Bernacchi, S.; Vivet-Boudou, V.; Hijnen, M.; Jossinet, F.; Mak, J.; Paillart, J. C.; Marquet, R., Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nat Commun* **2014**, *5*, 4304.
57. Houzet, L.; Paillart, J. C.; Smagulova, F.; Maurel, S.; Morichaud, Z.; Marquet, R.; Mougel, M., HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res* **2007**, *35* (8), 2695-704.
58. O'Reilly, M. M.; McNally, M. T.; Beemon, K. L., Two strong 5' splice sites and competing, suboptimal 3' splice sites involved in alternative splicing of human immunodeficiency virus type 1 RNA. *Virology* **1995**, *213* (2), 373-85.
59. Ashe, M. P.; Pearson, L. H.; Proudfoot, N. J., The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J* **1997**, *16* (18), 5752-63.
60. Lever, A.; Gottlinger, H.; Haseltine, W.; Sodroski, J., Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J Virol* **1989**, *63* (9), 4085-7.
61. Clavel, F.; Orenstein, J. M., A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J Virol* **1990**, *64* (10), 5230-4.
62. Harrison, G. P.; Miele, G.; Hunter, E.; Lever, A. M., Functional analysis of the core human immunodeficiency virus type 1 packaging signal in a permissive cell line. *J Virol* **1998**, *72* (7), 5886-96.
63. Aldovini, A.; Young, R. A., Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* **1990**, *64* (5), 1920-6.
64. Abbink, T. E.; Berkhout, B., A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J Biol Chem* **2003**, *278* (13), 11601-11.
65. Keane, S. C.; Heng, X.; Lu, K.; Kharytonchyk, S.; Ramakrishnan, V.; Carter, G.; Barton, S.; Hoscic, A.; Florwick, A.; Santos, J.; Bolden, N. C.; McCowin, S.; Case, D. A.;

- Johnson, B. A.; Salemi, M.; Telesnitsky, A.; Summers, M. F., RNA structure. Structure of the HIV-1 RNA packaging signal. *Science* **2015**, *348* (6237), 917-21.
66. Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E. L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A.; Divakaruni, S. S.; LaCotti, C.; Barton, S.; Tummillo, D.; Hosic, A.; Edme, K.; Albrecht, S.; Telesnitsky, A.; Summers, M. F., NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **2011**, *334* (6053), 242-5.
67. Malim, M. H.; Bohnlein, S.; Hauber, J.; Cullen, B. R., Functional dissection of the HIV-1 Rev trans-activator--derivation of a trans-dominant repressor of Rev function. *Cell* **1989**, *58* (1), 205-14.
68. Malim, M. H.; Hauber, J.; Le, S. Y.; Maizel, J. V.; Cullen, B. R., The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **1989**, *338* (6212), 254-7.
69. DiMattia, M. A.; Watts, N. R.; Stahl, S. J.; Rader, C.; Wingfield, P. T.; Stuart, D. I.; Steven, A. C.; Grimes, J. M., Implications of the HIV-1 Rev dimer structure at 3.2 Å resolution for multimeric binding to the Rev response element. *Proc Natl Acad Sci U S A* **2010**, *107* (13), 5810-4.
70. Reicin, A. S.; Kalpana, G.; Paik, S.; Marmon, S.; Goff, S., Sequences in the human immunodeficiency virus type 1 U3 region required for in vivo and in vitro integration. *J Virol* **1995**, *69* (9), 5904-7.
71. Reicin, A. S.; Paik, S.; Berkowitz, R. D.; Luban, J.; Lowy, I.; Goff, S. P., Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity. *J Virol* **1995**, *69* (2), 642-50.
72. Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M., Nucleotide sequence of the AIDS virus, LAV. *Cell* **1985**, *40* (1), 9-17.
73. Chukkapalli, V.; Ono, A., Molecular determinants that regulate plasma membrane association of HIV-1 Gag. *J Mol Biol* **2011**, *410* (4), 512-24.
74. Hogue, I. B.; Llewellyn, G. N.; Ono, A., Dynamic Association between HIV-1 Gag and Membrane Domains. *Mol Biol Int* **2012**, *2012*, 979765.
75. Perilla, J. R.; Gronenborn, A. M., Molecular Architecture of the Retroviral Capsid. *Trends Biochem Sci* **2016**, *41* (5), 410-20.
76. Luban, J., TRIM5 and the Regulation of HIV-1 Infectivity. *Mol Biol Int* **2012**, *2012*, 426840.
77. Zhou, Y.; Cahan, S. H., A novel family of terminal-repeat retrotransposon in miniature (TRIM) in the genome of the red harvester ant, *Pogonomyrmex barbatus*. *PloS one* **2012**, *7* (12), e53401.

78. Kaplan, A. H.; Krogstad, P.; Kempf, D. J.; Norbeck, D. W.; Swanstrom, R., Human immunodeficiency virus type 1 virions composed of unprocessed Gag and Gag-Pol precursors are capable of reverse transcribing viral genomic RNA. *Antimicrob Agents Chemother* **1994**, *38* (12), 2929-33.
79. Berkhout, B.; Silverman, R. H.; Jeang, K. T., Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* **1989**, *59* (2), 273-82.
80. Feinberg, M. B.; Jarrett, R. F.; Aldovini, A.; Gallo, R. C.; Wong-Staal, F., HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* **1986**, *46* (6), 807-17.
81. Romani, B.; Engelbrecht, S.; Glashoff, R. H., Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *J Gen Virol* **2010**, *91* (Pt 1), 1-12.
82. Pollard, V. W.; Malim, M. H., The HIV-1 Rev protein. *Annu Rev Microbiol* **1998**, *52*, 491-532.
83. Liu, H.; Wu, X.; Newman, M.; Shaw, G. M.; Hahn, B. H.; Kappes, J. C., The Vif protein of human and simian immunodeficiency viruses is packaged into virions and associates with viral core structures. *J Virol* **1995**, *69* (12), 7630-8.
84. Marin, M.; Rose, K. M.; Kozak, S. L.; Kabat, D., HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. *Nat Med* **2003**, *9* (11), 1398-403.
85. Mehle, A.; Strack, B.; Ancuta, P.; Zhang, C.; McPike, M.; Gabuzda, D., Vif overcomes the innate antiviral activity of APOBEC3G by promoting its degradation in the ubiquitin-proteasome pathway. *J Biol Chem* **2004**, *279* (9), 7792-8.
86. Romani, B.; Engelbrecht, S.; Glashoff, R. H., Antiviral roles of APOBEC proteins against HIV-1 and suppression by Vif. *Arch Virol* **2009**, *154* (10), 1579-88.
87. Simon, V.; Zennou, V.; Murray, D.; Huang, Y.; Ho, D. D.; Bieniasz, P. D., Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. *PLoS Pathog* **2005**, *1* (1), e6.
88. Sheehy, A. M.; Gaddis, N. C.; Malim, M. H., The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat Med* **2003**, *9* (11), 1404-7.
89. Stopak, K.; de Noronha, C.; Yonemoto, W.; Greene, W. C., HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. *Mol Cell* **2003**, *12* (3), 591-601.
90. Strebel, K., HIV accessory genes Vif and Vpu. *Adv Pharmacol* **2007**, *55*, 199-232.
91. Yu, X.; Yu, Y.; Liu, B.; Luo, K.; Kong, W.; Mao, P.; Yu, X. F., Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **2003**, *302* (5647), 1056-60.

92. Agostini, I.; Popov, S.; Hao, T.; Li, J. H.; Dubrovsky, L.; Chaika, O.; Chaika, N.; Lewis, R.; Bukrinsky, M., Phosphorylation of Vpr regulates HIV type 1 nuclear import and macrophage infection. *AIDS Res Hum Retroviruses* **2002**, *18* (4), 283-8.
93. Hogan, T. H.; Nonnemacher, M. R.; Krebs, F. C.; Henderson, A.; Wigdahl, B., HIV-1 Vpr binding to HIV-1 LTR C/EBP cis-acting elements and adjacent regions is sequence-specific. *Biomed Pharmacother* **2003**, *57* (1), 41-8.
94. Sherman, M. P.; de Noronha, C. M.; Eckstein, L. A.; Hataye, J.; Mundt, P.; Williams, S. A.; Neidleman, J. A.; Goldsmith, M. A.; Greene, W. C., Nuclear export of Vpr is required for efficient replication of human immunodeficiency virus type 1 in tissue macrophages. *J Virol* **2003**, *77* (13), 7582-9.
95. Romani, B.; Engelbrecht, S., Human immunodeficiency virus type 1 Vpr: functions and molecular interactions. *J Gen Virol* **2009**, *90* (Pt 8), 1795-805.
96. Hrimech, M.; Yao, X. J.; Bachand, F.; Rougeau, N.; Cohen, E. A., Human immunodeficiency virus type 1 (HIV-1) Vpr functions as an immediate-early protein during HIV-1 infection. *J Virol* **1999**, *73* (5), 4101-9.
97. Jowett, J. B.; Planelles, V.; Poon, B.; Shah, N. P.; Chen, M. L.; Chen, I. S., The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. *J Virol* **1995**, *69* (10), 6304-13.
98. Popov, S.; Rexach, M.; Ratner, L.; Blobel, G.; Bukrinsky, M., Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex. *J Biol Chem* **1998**, *273* (21), 13347-52.
99. Bukrinsky, M.; Adzhubei, A., Viral protein R of HIV-1. *Rev Med Virol* **1999**, *9* (1), 39-49.
100. Cohen, G. B.; Gandhi, R. T.; Davis, D. M.; Mandelboim, O.; Chen, B. K.; Strominger, J. L.; Baltimore, D., The selective downregulation of class I major histocompatibility complex proteins by HIV-1 protects HIV-infected cells from NK cells. *Immunity* **1999**, *10* (6), 661-71.
101. Collins, K. L.; Chen, B. K.; Kalams, S. A.; Walker, B. D.; Baltimore, D., HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **1998**, *391* (6665), 397-401.
102. Goldsmith, M. A.; Warmerdam, M. T.; Atchison, R. E.; Miller, M. D.; Greene, W. C., Dissociation of the CD4 downregulation and viral infectivity enhancement functions of human immunodeficiency virus type 1 Nef. *J Virol* **1995**, *69* (7), 4112-21.
103. Luo, T.; Garcia, J. V., The association of Nef with a cellular serine/threonine kinase and its enhancement of infectivity are viral isolate dependent. *J Virol* **1996**, *70* (9), 6493-6.
104. Lama, J.; Mangasarian, A.; Trono, D., Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner. *Curr Biol* **1999**, *9* (12), 622-31.

105. Bour, S.; Perrin, C.; Strebel, K., Cell surface CD4 inhibits HIV-1 particle release by interfering with Vpu activity. *J Biol Chem* **1999**, *274* (47), 33800-6.
106. Willey, R. L.; Maldarelli, F.; Martin, M. A.; Strebel, K., Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. *J Virol* **1992**, *66* (1), 226-34.
107. Strebel, K.; Klimkait, T.; Maldarelli, F.; Martin, M. A., Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. *J Virol* **1989**, *63* (9), 3784-91.
108. Neil, S. J.; Eastman, S. W.; Jouvenet, N.; Bieniasz, P. D., HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. *PLoS Pathog* **2006**, *2* (5), e39.
109. Varthakavi, V.; Smith, R. M.; Martin, K. L.; Derdowski, A.; Lapierre, L. A.; Goldenring, J. R.; Spearman, P., The pericentriolar recycling endosome plays a key role in Vpu-mediated enhancement of HIV-1 particle release. *Traffic* **2006**, *7* (3), 298-307.
110. Dube, M.; Bego, M. G.; Paquay, C.; Cohen, E. A., Modulation of HIV-1-host interaction: role of the Vpu accessory protein. *Retrovirology* **2010**, *7*, 114.
111. Dube, M.; Roy, B. B.; Guiot-Guillain, P.; Binette, J.; Mercier, J.; Chiasson, A.; Cohen, E. A., Antagonism of tetherin restriction of HIV-1 release by Vpu involves binding and sequestration of the restriction factor in a perinuclear compartment. *PLoS Pathog* **2010**, *6* (4), e1000856.
112. Huang, C. C.; Tang, M.; Zhang, M. Y.; Majeed, S.; Montabana, E.; Stanfield, R. L.; Dimitrov, D. S.; Korber, B.; Sodroski, J.; Wilson, I. A.; Wyatt, R.; Kwong, P. D., Structure of a V3-containing HIV-1 gp120 core. *Science* **2005**, *310* (5750), 1025-8.
113. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A., Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998**, *393* (6686), 648-59.
114. Lasky, L. A.; Nakamura, G.; Smith, D. H.; Fennie, C.; Shimasaki, C.; Patzer, E.; Berman, P.; Gregory, T.; Capon, D. J., Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* **1987**, *50* (6), 975-85.
115. McDougal, J. S.; Maddon, P. J.; Dalgleish, A. G.; Clapham, P. R.; Littman, D. R.; Godfrey, M.; Maddon, D. E.; Chess, L.; Weiss, R. A.; Axel, R., The T4 glycoprotein is a cell-surface receptor for the AIDS virus. *Cold Spring Harb Symp Quant Biol* **1986**, *51 Pt 2*, 703-11.
116. Dimitrov, A. S.; Louis, J. M.; Bewley, C. A.; Clore, G. M.; Blumenthal, R., Conformational changes in HIV-1 gp41 in the course of HIV-1 envelope glycoprotein-mediated fusion and inactivation. *Biochemistry* **2005**, *44* (37), 12471-9.

117. Keele, B. F.; Giorgi, E. E.; Salazar-Gonzalez, J. F.; Decker, J. M.; Pham, K. T.; Salazar, M. G.; Sun, C.; Grayson, T.; Wang, S.; Li, H.; Wei, X.; Jiang, C.; Kirchherr, J. L.; Gao, F.; Anderson, J. A.; Ping, L. H.; Swanstrom, R.; Tomaras, G. D.; Blattner, W. A.; Goepfert, P. A.; Kilby, J. M.; Saag, M. S.; Delwart, E. L.; Busch, M. P.; Cohen, M. S.; Montefiori, D. C.; Haynes, B. F.; Gaschen, B.; Athreya, G. S.; Lee, H. Y.; Wood, N.; Seoighe, C.; Perelson, A. S.; Bhattacharya, T.; Korber, B. T.; Hahn, B. H.; Shaw, G. M., Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A* **2008**, *105* (21), 7552-7.
118. Wilen, C. B.; Wang, J.; Tilton, J. C.; Miller, J. C.; Kim, K. A.; Rebar, E. J.; Sherrill-Mix, S. A.; Patro, S. C.; Secreto, A. J.; Jordan, A. P.; Lee, G.; Kahn, J.; Aye, P. P.; Bunnell, B. A.; Lackner, A. A.; Hoxie, J. A.; Danet-Desnoyers, G. A.; Bushman, F. D.; Riley, J. L.; Gregory, P. D.; June, C. H.; Holmes, M. C.; Doms, R. W., Engineering HIV-resistant human CD4+ T cells with CXCR4-specific zinc-finger nucleases. *PLoS Pathog* **2011**, *7* (4), e1002020.
119. Arhel, N. J.; Souquere-Besse, S.; Munier, S.; Souque, P.; Guadagnini, S.; Rutherford, S.; Prevost, M. C.; Allen, T. D.; Charneau, P., HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *EMBO J* **2007**, *26* (12), 3025-37.
120. Stevenson, M., Portals of entry: uncovering HIV nuclear transport pathways. *Trends Cell Biol* **1996**, *6* (1), 9-15.
121. Krishnan, L.; Engelman, A., Retroviral integrase proteins and HIV-1 DNA integration. *J Biol Chem* **2012**, *287* (49), 40858-66.
122. Farnet, C. M.; Bushman, F. D., HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. *Cell* **1997**, *88* (4), 483-92.
123. Freed, E. O., HIV-1 replication. *Somat Cell Mol Genet* **2001**, *26* (1-6), 13-33.
124. Purcell, D. F.; Martin, M. A., Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J Virol* **1993**, *67* (11), 6365-78.
125. van Bel, N.; Das, A. T.; Cornelissen, M.; Abbink, T. E.; Berkhout, B., A short sequence motif in the 5' leader of the HIV-1 genome modulates extended RNA dimer formation and virus replication. *J Biol Chem* **2014**, *289* (51), 35061-74.
126. Felber, B. K.; Hadzopoulou-Cladaras, M.; Cladaras, C.; Copeland, T.; Pavlakis, G. N., rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc Natl Acad Sci U S A* **1989**, *86* (5), 1495-9.
127. Daly, T. J.; Cook, K. S.; Gray, G. S.; Maione, T. E.; Rusche, J. R., Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature* **1989**, *342* (6251), 816-9.
128. Fornerod, M.; Ohno, M.; Yoshida, M.; Mattaj, I. W., CRM1 is an export receptor for

leucine-rich nuclear export signals. *Cell* **1997**, *90* (6), 1051-60.

129. Kim, S. Y.; Byrn, R.; Groopman, J.; Baltimore, D., Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol* **1989**, *63* (9), 3708-13.

130. Dorman, N.; Lever, A., Comparison of viral genomic RNA sorting mechanisms in human immunodeficiency virus type 1 (HIV-1), HIV-2, and Moloney murine leukemia virus. *J Virol* **2000**, *74* (23), 11413-7.

131. Levin, J. G.; Rosenak, M. J., Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: evidence for persistence of viral messenger RNA. *Proc Natl Acad Sci U S A* **1976**, *73* (4), 1154-8.

132. Messer, L. I.; Levin, J. G.; Chattopadhyay, S. K., Metabolism of viral RNA in murine leukemia virus-infected cells; evidence for differential stability of viral message and virion precursor RNA. *J Virol* **1981**, *40* (3), 683-90.

133. Freed, E. O., HIV-1 assembly, release and maturation. *Nat Rev Microbiol* **2015**, *13* (8), 484-96.

134. Adamson, C. S.; Freed, E. O., Human immunodeficiency virus type 1 assembly, release, and maturation. *Adv Pharmacol* **2007**, *55*, 347-87.

135. D'Souza, V.; Summers, M. F., How retroviruses select their genomes. *Nat Rev Microbiol* **2005**, *3* (8), 643-55.

136. Jewell, N. A.; Mansky, L. M., In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *J Gen Virol* **2000**, *81* (Pt 8), 1889-99.

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

138. Thomas, J. A.; Gorelick, R. J., Nucleocapsid protein function in early infection processes. *Virus Res* **2008**, *134* (1-2), 39-63.

139. Hu, W. S.; Temin, H. M., Retroviral recombination and reverse transcription. *Science* **1990**, *250* (4985), 1227-33.

140. Levin, J. G.; Mitra, M.; Mascarenhas, A.; Musier-Forsyth, K., Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **2010**, *7* (6), 754-74.

141. Telesnitsky, A.; Goff, S. P., Reverse Transcriptase and the Generation of Retroviral DNA. **1997**.

142. Driscoll, M. D.; Golinelli, M. P.; Hughes, S. H., In vitro analysis of human immunodeficiency virus type 1 minus-strand strong-stop DNA synthesis and genomic RNA processing. *J Virol* **2001**, *75* (2), 672-86.

143. Purohit, V.; Roques, B. P.; Kim, B.; Bambara, R. A., Mechanisms that prevent template inactivation by HIV-1 reverse transcriptase RNase H cleavages. *J Biol Chem* **2007**, *282* (17), 12598-609.
144. Mak, J.; Kleiman, L., Primer tRNAs for reverse transcription. *J Virol* **1997**, *71* (11), 8087-95.
145. Hsu, M.; Rong, L.; de Rocquigny, H.; Roques, B. P.; Wainberg, M. A., The effect of mutations in the HIV-1 nucleocapsid protein on strand transfer in cell-free reverse transcription reactions. *Nucleic Acids Res* **2000**, *28* (8), 1724-9.
146. Isel, C.; Marquet, R.; Keith, G.; Ehresmann, C.; Ehresmann, B., Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J Biol Chem* **1993**, *268* (34), 25269-72.
147. Yusupova, G.; Lanchy, J. M.; Yusupov, M.; Keith, G.; Le Grice, S. F.; Ehresmann, C.; Ehresmann, B.; Marquet, R., Primer selection by HIV-1 reverse transcriptase on RNA-tRNA(3Lys) and DNA-tRNA(3Lys) hybrids. *J Mol Biol* **1996**, *261* (3), 315-21.
148. Mougel, M.; Houzet, L.; Darlix, J. L., When is it time for reverse transcription to start and go? *Retrovirology* **2009**, *6*, 24.
149. Puglisi, E. V.; Puglisi, J. D., Secondary structure of the HIV reverse transcription initiation complex by NMR. *J Mol Biol* **2011**, *410* (5), 863-74.
150. Isel, C.; Ehresmann, C.; Marquet, R., Initiation of HIV Reverse Transcription. *Viruses* **2010**, *2* (1), 213-43.
151. Kutluay, S. B.; Zang, T.; Blanco-Melo, D.; Powell, C.; Jannain, D.; Errando, M.; Bieniasz, P. D., Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* **2014**, *159* (5), 1096-109.
152. Liu, S.; Harada, B. T.; Miller, J. T.; Le Grice, S. F.; Zhuang, X., Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat Struct Mol Biol* **2010**, *17* (12), 1453-60.
153. Hu, W. S.; Hughes, S. H., HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2012**, *2* (10).
154. Wakefield, J. K.; Morrow, C. D., Mutations within the primer binding site of the human immunodeficiency virus type 1 define sequence requirements essential for reverse transcription. *Virology* **1996**, *220* (2), 290-8.
155. Aiyar, A.; Cobrinik, D.; Ge, Z.; Kung, H. J.; Leis, J., Interaction between retroviral U5 RNA and the T psi C loop of the tRNA(Trp) primer is required for efficient initiation of reverse transcription. *J Virol* **1992**, *66* (4), 2464-72.
156. Puglisi, E. V.; Puglisi, J. D., HIV-1 A-rich RNA loop mimics the tRNA anticodon structure. *Nat Struct Biol* **1998**, *5* (12), 1033-6.

157. Beerens, N.; Berkhout, B., Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA* **2002**, *8* (3), 357-69.
158. Rigourd, M.; Bec, G.; Benas, P.; Le Grice, S. F.; Ehresmann, B.; Ehresmann, C.; Marquet, R., Effects of tRNA 3 Lys aminoacylation on the initiation of HIV-1 reverse transcription. *Biochimie* **2003**, *85* (5), 521-5.
159. Abbink, T. E.; Beerens, N.; Berkhout, B., Forced selection of a human immunodeficiency virus type 1 variant that uses a non-self tRNA primer for reverse transcription: involvement of viral RNA sequences and the reverse transcriptase enzyme. *J Virol* **2004**, *78* (19), 10706-14.
160. Ooms, M.; Cupac, D.; Abbink, T. E.; Huthoff, H.; Berkhout, B., The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. *Nucleic Acids Res* **2007**, *35* (5), 1649-59.
161. Jones, C. P.; Saadatmand, J.; Kleiman, L.; Musier-Forsyth, K., Molecular mimicry of human tRNA^{Lys} anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. *RNA* **2013**, *19* (2), 219-29.
162. Sleiman, D.; Barraud, P.; Brachet, F.; Tisne, C., The Interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. *PloS one* **2013**, *8* (6), e64700.
163. Seif, E.; Niu, M.; Kleiman, L., In virio SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology* **2015**, *12*, 40.
164. Coey, A.; Larsen, K.; Puglisi, J. D.; Viani Puglisi, E., Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *RNA* **2016**, *22* (11), 1689-1698.
165. Beerens, N.; Jepsen, M. D.; Nechyporuk-Zloy, V.; Kruger, A. C.; Darlix, J. L.; Kjems, J.; Birkedal, V., Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA* **2013**, *19* (4), 517-26.
166. Liu, S.; Abbondanzieri, E. A.; Rausch, J. W.; Le Grice, S. F.; Zhuang, X., Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science* **2008**, *322* (5904), 1092-7.
167. Zhang, H.; Bagasra, O.; Niikura, M.; Poiesz, B. J.; Pomerantz, R. J., Intravirion reverse transcripts in the peripheral blood plasma on human immunodeficiency virus type 1-infected individuals. *J Virol* **1994**, *68* (11), 7591-7.
168. Coffin, J. M.; Haseltine, W. A., Nucleotide sequence of Rous sarcoma virus RNA at the initiation site of DNA synthesis: The 102nd nucleotide is U. *J Mol Biol* **1977**, *117* (3), 805-14.
169. Berkhout, B.; Vastenhouw, N. L.; Klasens, B. I.; Huthoff, H., Structural features in the HIV-1 repeat region facilitate strand transfer during reverse transcription. *RNA* **2001**, *7* (8),

1097-114.

170. Post, K.; Kankia, B.; Gopalakrishnan, S.; Yang, V.; Cramer, E.; Saladores, P.; Gorelick, R. J.; Guo, J.; Musier-Forsyth, K.; Levin, J. G., Fidelity of plus-strand priming requires the nucleic acid chaperone activity of HIV-1 nucleocapsid protein. *Nucleic Acids Res* **2009**, *37* (6), 1755-66.

171. Onafuwa-Nuga, A.; Telesnitsky, A., The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. *Microbiol Mol Biol Rev* **2009**, *73* (3), 451-80, Table of Contents.

172. Godet, J.; Mely, Y., Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol* **2010**, *7* (6), 687-99.

173. Ohi, Y.; Clever, J. L., Sequences in the 5' and 3' R elements of human immunodeficiency virus type 1 critical for efficient reverse transcription. *J Virol* **2000**, *74* (18), 8324-34.

174. Chen, Y.; Balakrishnan, M.; Roques, B. P.; Fay, P. J.; Bambara, R. A., Mechanism of minus strand strong stop transfer in HIV-1 reverse transcription. *J Biol Chem* **2003**, *278* (10), 8006-17.

175. Kanevsky, I.; Chaminade, F.; Ficheux, D.; Moumen, A.; Gorelick, R.; Negroni, M.; Darlix, J. L.; Fosse, P., Specific interactions between HIV-1 nucleocapsid protein and the TAR element. *J Mol Biol* **2005**, *348* (5), 1059-77.

176. Moumen, A.; Polomack, L.; Roques, B.; Buc, H.; Negroni, M., The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* **2001**, *29* (18), 3814-21.

177. Levin, J. G.; Guo, J.; Rouzina, I.; Musier-Forsyth, K., Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog Nucleic Acid Res Mol Biol* **2005**, *80*, 217-86.

178. Godet, J.; Mely, Y., Biophysical studies of the nucleic acid chaperone properties of the HIV-1 nucleocapsid protein. *RNA Biol* **2010**, *7* (6), 687-99.

179. Panganiban, A. T.; Fiore, D., Ordered interstrand and intrastrand DNA transfer during reverse transcription. *Science* **1988**, *241* (4869), 1064-9.

180. van Wamel, J. L.; Berkhout, B., The first strand transfer during HIV-1 reverse transcription can occur either intramolecularly or intermolecularly. *Virology* **1998**, *244* (2), 245-51.

181. Yu, H.; Jetzt, A. E.; Ron, Y.; Preston, B. D.; Dougherty, J. P., The nature of human immunodeficiency virus type 1 strand transfers. *J Biol Chem* **1998**, *273* (43), 28384-91.

182. Beerens, N.; Kjemis, J., Circularization of the HIV-1 genome facilitates strand transfer during reverse transcription. *RNA* **2010**, *16* (6), 1226-35.

183. Brule, F.; Bec, G.; Keith, G.; Le Grice, S. F.; Roques, B. P.; Ehresmann, B.; Ehresmann, C.; Marquet, R., 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* **2000**, *28* (2), 634-40.
184. Piekna-Przybylska, D.; DiChiacchio, L.; Mathews, D. H.; Bambara, R. A., A sequence similar to tRNA 3 Lys gene is embedded in HIV-1 U3-R and promotes minus-strand transfer. *Nat Struct Mol Biol* **2010**, *17* (1), 83-9.
185. Drummond, J. E.; Mounts, P.; Gorelick, R. J.; Casas-Finet, J. R.; Bosche, W. J.; Henderson, L. E.; Waters, D. J.; Arthur, L. O., Wild-type and mutant HIV type 1 nucleocapsid proteins increase the proportion of long cDNA transcripts by viral reverse transcriptase. *AIDS Res Hum Retroviruses* **1997**, *13* (7), 533-43.
186. Anthony, R. M.; Destefano, J. J., In vitro synthesis of long DNA products in reactions with HIV-RT and nucleocapsid protein. *J Mol Biol* **2007**, *365* (2), 310-24.
187. Lanciault, C.; Champoux, J. J., Effects of unpaired nucleotides within HIV-1 genomic secondary structures on pausing and strand transfer. *J Biol Chem* **2005**, *280* (4), 2413-23.
188. Simon-Loriere, E.; Martin, D. P.; Weeks, K. M.; Negroni, M., RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J Virol* **2010**, *84* (24), 12675-82.
189. Zhang, J.; Temin, H. M., Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* **1993**, *259* (5092), 234-8.
190. Coffin, J. M., HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **1995**, *267* (5197), 483-9.
191. Simon-Loriere, E.; Holmes, E. C., Why do RNA viruses recombine? *Nat Rev Microbiol* **2011**, *9* (8), 617-26.
192. Pathak, V. K.; Temin, H. M., Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc Natl Acad Sci U S A* **1990**, *87* (16), 6019-23.
193. Mansky, L. M.; Temin, H. M., Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* **1995**, *69* (8), 5087-94.
194. Abram, M. E.; Ferris, A. L.; Shao, W.; Alvord, W. G.; Hughes, S. H., Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol* **2010**, *84* (19), 9864-78.
195. Smyth, R. P.; Davenport, M. P.; Mak, J., The origin of genetic diversity in HIV-1. *Virus Res* **2012**, *169* (2), 415-29.
196. Anderson, J. A.; Bowman, E. H.; Hu, W. S., Retroviral recombination rates do not

- increase linearly with marker distance and are limited by the size of the recombining subpopulation. *J Virol* **1998**, *72* (2), 1195-202.
197. Jetzt, A. E.; Yu, H.; Klarmann, G. J.; Ron, Y.; Preston, B. D.; Dougherty, J. P., High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol* **2000**, *74* (3), 1234-40.
198. Zhuang, J.; Jetzt, A. E.; Sun, G.; Yu, H.; Klarmann, G.; Ron, Y.; Preston, B. D.; Dougherty, J. P., Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* **2002**, *76* (22), 11273-82.
199. Dykes, C.; Balakrishnan, M.; Planelles, V.; Zhu, Y.; Bambara, R. A.; Demeter, L. M., Identification of a preferred region for recombination and mutation in HIV-1 gag. *Virology* **2004**, *326* (2), 262-79.
200. Levy, D. N.; Aldrovandi, G. M.; Kutsch, O.; Shaw, G. M., Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci U S A* **2004**, *101* (12), 4204-9.
201. Chin, M. P.; Lee, S. K.; Chen, J.; Nikolaitchik, O. A.; Powell, D. A.; Fivash, M. J., Jr.; Hu, W. S., Long-range recombination gradient between HIV-1 subtypes B and C variants caused by sequence differences in the dimerization initiation signal region. *J Mol Biol* **2008**, *377* (5), 1324-33.
202. Galli, A.; Kearney, M.; Nikolaitchik, O. A.; Yu, S.; Chin, M. P.; Maldarelli, F.; Coffin, J. M.; Pathak, V. K.; Hu, W. S., Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. *J Virol* **2010**, *84* (15), 7651-61.
203. Schlub, T. E.; Smyth, R. P.; Grimm, A. J.; Mak, J.; Davenport, M. P., Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput Biol* **2010**, *6* (4), e1000766.
204. Ramirez, B. C.; Simon-Loriere, E.; Galetto, R.; Negroni, M., Implications of recombination for HIV diversity. *Virus Res* **2008**, *134* (1-2), 64-73.
205. Galetto, R.; Moumen, A.; Giacomoni, V.; Veron, M.; Charneau, P.; Negroni, M., The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo. *J Biol Chem* **2004**, *279* (35), 36625-32.
206. Coffin, J. M., Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J Gen Virol* **1979**, *42* (1), 1-26.
207. Hwang, C. K.; Svarovskaia, E. S.; Pathak, V. K., 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* **2001**, *98* (21), 12209-14.
208. Negroni, M.; Ricchetti, M.; Nouvel, P.; Buc, H., Homologous recombination promoted

- by reverse transcriptase during copying of two distinct RNA templates. *Proc Natl Acad Sci U S A* **1995**, 92 (15), 6971-5.
209. Rein, A.; Henderson, L. E.; Levin, J. G., Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends Biochem Sci* **1998**, 23 (8), 297-301.
210. Darlix, J. L.; Lapadat-Tapolsky, M.; de Rocquigny, H.; Roques, B. P., First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* **1995**, 254 (4), 523-37.
211. Rodriguez-Rodriguez, L.; Tsuchihashi, Z.; Fuentes, G. M.; Bambara, R. A.; Fay, P. J., Influence of human immunodeficiency virus nucleocapsid protein on synthesis and strand transfer by the reverse transcriptase in vitro. *J Biol Chem* **1995**, 270 (25), 15005-11.
212. Druillennec, S.; Caneparo, A.; de Rocquigny, H.; Roques, B. P., Evidence of interactions between the nucleocapsid protein NCp7 and the reverse transcriptase of HIV-1. *J Biol Chem* **1999**, 274 (16), 11283-8.
213. Peliska, J. A.; Balasubramanian, S.; Giedroc, D. P.; Benkovic, S. J., Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity. *Biochemistry* **1994**, 33 (46), 13817-23.
214. Cameron, C. E.; Ghosh, M.; Le Grice, S. F.; Benkovic, S. J., 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* **1997**, 94 (13), 6700-5.
215. Delviks-Frankenberry, K.; Galli, A.; Nikolaitchik, O.; Mens, H.; Pathak, V. K.; Hu, W. S., Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* **2011**, 3 (9), 1650-80.
216. Chen, Y.; Balakrishnan, M.; Roques, B. P.; Bambara, R. A., Acceptor RNA cleavage profile supports an invasion mechanism for HIV-1 minus strand transfer. *J Biol Chem* **2005**, 280 (15), 14443-52.
217. Xie, P., A polymerase-site-jumping model for strand transfer during DNA synthesis by reverse transcriptase. *Virus Res* **2009**, 144 (1-2), 65-73.
218. Moumen, A.; Polomack, L.; Unge, T.; Veron, M.; Buc, H.; Negroni, M., Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA. *J Biol Chem* **2003**, 278 (18), 15973-82.
219. Balakrishnan, M.; Fay, P. J.; Bambara, R. A., The kissing hairpin sequence promotes recombination within the HIV-1 5' leader region. *J Biol Chem* **2001**, 276 (39), 36482-92.
220. Heilman-Miller, S. L.; Wu, T.; Levin, J. G., Alteration of nucleic acid structure and stability modulates the efficiency of minus-strand transfer mediated by the HIV-1 nucleocapsid protein. *J Biol Chem* **2004**, 279 (42), 44154-65.

221. Song, R.; Kafaie, J.; Laughrea, M., Role of the 5' TAR stem-loop and the U5-AUG duplex in dimerization of HIV-1 genomic RNA. *Biochemistry* **2008**, *47* (10), 3283-93.
222. Roda, R. H.; Balakrishnan, M.; Hanson, M. N.; Wohrl, B. M.; Le Grice, S. F.; Roques, B. P.; Gorelick, R. J.; Bambara, R. A., Role of the Reverse Transcriptase, Nucleocapsid Protein, and Template Structure in the Two-step Transfer Mechanism in Retroviral Recombination. *J Biol Chem* **2003**, *278* (34), 31536-46.
223. Roda, R. H.; Balakrishnan, M.; Kim, J. K.; Roques, B. P.; Fay, P. J.; Bambara, R. A., Strand transfer occurs in retroviruses by a pause-initiated two-step mechanism. *J Biol Chem* **2002**, *277* (49), 46900-11.
224. Muriaux, D.; Darlix, J. L., Properties and functions of the nucleocapsid protein in virus assembly. *RNA Biol* **2010**, *7* (6), 744-53.
225. Kutluay, S. B.; Bieniasz, P. D., Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog* **2010**, *6* (11), e1001200.
226. Kuzembayeva, M.; Dilley, K.; Sardo, L.; Hu, W. S., Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **2014**, *454-455*, 362-70.
227. Rein, A.; Datta, S. A.; Jones, C. P.; Musier-Forsyth, K., Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem Sci* **2011**, *36* (7), 373-80.
228. Aldovini, A.; Young, R. A., Development of a BCG recombinant vehicle for candidate AIDS vaccines. *Int Rev Immunol* **1990**, *7* (1), 79-83.
229. Berkowitz, R. D.; Luban, J.; Goff, S. P., Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* **1993**, *67* (12), 7190-200.
230. McBride, M. S.; Panganiban, A. T., The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J Virol* **1996**, *70* (5), 2963-73.
231. Luban, J.; Goff, S. P., Binding of human immunodeficiency virus type 1 (HIV-1) RNA to recombinant HIV-1 gag polyprotein. *J Virol* **1991**, *65* (6), 3203-12.
232. Clever, J. L.; Miranda, D., Jr.; Parslow, T. G., RNA structure and packaging signals in the 5' leader region of the human immunodeficiency virus type 1 genome. *J Virol* **2002**, *76* (23), 12381-7.
233. Bernacchi, S.; Abd El-Wahab, E. W.; Dubois, N.; Hijnen, M.; Smyth, R. P.; Mak, J.; Marquet, R.; Paillart, J. C., HIV-1 Pr55Gag binds genomic and spliced RNAs with different affinity and stoichiometry. *RNA Biol* **2017**, *14* (1), 90-103.
234. Comas-Garcia, M.; Davis, S. R.; Rein, A., On the Selective Packaging of Genomic RNA by HIV-1. *Viruses* **2016**, *8* (9).

235. Mailler, E.; Bernacchi, S.; Marquet, R.; Paillart, J. C.; Vivet-Boudou, V.; Smyth, R. P., The Life-Cycle of the HIV-1 Gag-RNA Complex. *Viruses* **2016**, *8* (9).
236. Sakuragi, J.; Shioda, T.; Panganiban, A. T., Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions. *J Virol* **2001**, *75* (6), 2557-65.
237. Jalalirad, M.; Laughrea, M., Formation of immature and mature genomic RNA dimers in wild-type and protease-inactive HIV-1: differential roles of the Gag polyprotein, nucleocapsid proteins NCp15, NCp9, NCp7, and the dimerization initiation site. *Virology* **2010**, *407* (2), 225-36.
238. Bell, N. M.; Lever, A. M., HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol* **2013**, *21* (3), 136-44.
239. Kafaie, J.; Song, R.; Abrahamyan, L.; Mouland, A. J.; Laughrea, M., Mapping of nucleocapsid residues important for HIV-1 genomic RNA dimerization and packaging. *Virology* **2008**, *375* (2), 592-610.
240. Haddrick, M.; Lear, A. L.; Cann, A. J.; Heaphy, S., Evidence that a kissing loop structure facilitates genomic RNA dimerisation in HIV-1. *J Mol Biol* **1996**, *259* (1), 58-68.
241. Laughrea, M.; Jette, L., HIV-1 genome dimerization: formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1-232 and 296-790 regions flanking the kissing-loop domain. *Biochemistry* **1996**, *35* (29), 9366-74.
242. Muriaux, D.; Girard, P. M.; Bonnet-Mathoniere, B.; Paoletti, J., Dimerization of HIV-1Lai RNA at low ionic strength. An autocomplementary sequence in the 5' leader region is evidenced by an antisense oligonucleotide. *J Biol Chem* **1995**, *270* (14), 8209-16.
243. Ennifar, E.; Walter, P.; Dumas, P., Cation-dependent cleavage of the duplex form of the subtype-B HIV-1 RNA dimerization initiation site. *Nucleic Acids Res* **2010**, *38* (17), 5807-16.
244. Mundigala, H.; Michaux, J. B.; Feig, A. L.; Ennifar, E.; Rueda, D., HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. *Nucleic Acids Res* **2014**, *42* (11), 7281-9.
245. Skripkin, E.; Paillart, J. C.; Marquet, R.; Blumenfeld, M.; Ehresmann, B.; Ehresmann, C., Mechanisms of inhibition of in vitro dimerization of HIV type I RNA by sense and antisense oligonucleotides. *J Biol Chem* **1996**, *271* (46), 28812-7.
246. Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C., Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A* **1994**, *91* (11), 4945-9.
247. Rist, M. J.; Marino, J. P., Mechanism of nucleocapsid protein catalyzed structural isomerization of the dimerization initiation site of HIV-1. *Biochemistry* **2002**, *41* (50), 14762-70.

248. Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P., Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* **2001**, *8* (12), 1064-8.
249. Windbichler, N.; Werner, M.; Schroeder, R., Kissing complex-mediated dimerisation of HIV-1 RNA: coupling extended duplex formation to ribozyme cleavage. *Nucleic Acids Res* **2003**, *31* (22), 6419-27.
250. Muriaux, D.; De Rocquigny, H.; Roques, B. P.; Paoletti, J., NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J Biol Chem* **1996**, *271* (52), 33686-92.
251. Takahashi, K.; Baba, S.; Hayashi, Y.; Koyanagi, Y.; Yamamoto, N.; Takaku, H.; Kawai, G., NMR analysis of intra- and inter-molecular stems in the dimerization initiation site of the HIV-1 genome. *J Biochem* **2000**, *127* (4), 681-6.
252. Damgaard, C. K.; Dyhr-Mikkelsen, H.; Kjems, J., Mapping the RNA binding sites for human immunodeficiency virus type-1 gag and NC proteins within the complete HIV-1 and -2 untranslated leader regions. *Nucleic Acids Res* **1998**, *26* (16), 3667-76.
253. Lawrence, D. C.; Stover, C. C.; Noznitsky, J.; Wu, Z.; Summers, M. F., Structure of the intact stem and bulge of HIV-1 Psi-RNA stem-loop SL1. *J Mol Biol* **2003**, *326* (2), 529-42.
254. Abd El-Wahad, E. W.; Smyth, R. P.; Mailler, E.; Bernacchi, S.; Vivet-Boudou, V.; Hijnen, M.; Jossinet, F.; Mak, J.; Paillart, J. C.; Marquet, R., Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nature communications* **2014**, *5*, 4304.
255. Smyth, R. P.; Despons, L.; Huili, G.; Bernacchi, S.; Hijnen, M.; Mak, J.; Jossinet, F.; Weixi, L.; Paillart, J. C.; von Kleist, M.; Marquet, R., Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat Methods* **2015**, *12* (9), 866-72.
256. Berkhout, B.; van Wamel, J. L., Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol* **1996**, *70* (10), 6723-32.
257. Laughrea, M.; Jette, L., HIV-1 genome dimerization: kissing-loop hairpin dictates whether nucleotides downstream of the 5' splice junction contribute to loose and tight dimerization of human immunodeficiency virus RNA. *Biochemistry* **1997**, *36* (31), 9501-8.
258. Chen, J.; Rahman, S. A.; Nikolaitchik, O. A.; Grunwald, D.; Sardo, L.; Burdick, R. C.; Plisov, S.; Liang, E.; Tai, S.; Pathak, V. K.; Hu, W. S., HIV-1 RNA genome dimerizes on the plasma membrane in the presence of Gag protein. *Proc Natl Acad Sci U S A* **2016**, *113* (2), E201-8.
259. Clever, J. L.; Parslow, T. G., Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J Virol* **1997**, *71* (5), 3407-14.
260. Ooms, M.; Huthoff, H.; Russell, R.; Liang, C.; Berkhout, B., A riboswitch regulates

- RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol* **2004**, 78 (19), 10814-9.
261. Huthoff, H.; Berkhout, B., Two alternating structures of the HIV-1 leader RNA. *RNA* **2001**, 7 (1), 143-57.
262. Miele, G.; Mouland, A.; Harrison, G. P.; Cohen, E.; Lever, A. M., The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J Virol* **1996**, 70 (2), 944-51.
263. Paillart, J. C.; Dettenhofer, M.; Yu, X. F.; Ehresmann, C.; Ehresmann, B.; Marquet, R., First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J Biol Chem* **2004**, 279 (46), 48397-403.
264. Wilkinson, K. A.; Gorelick, R. J.; Vasa, S. M.; Guex, N.; Rein, A.; Mathews, D. H.; Giddings, M. C.; Weeks, K. M., High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* **2008**, 6 (4), e96.
265. Damgaard, C. K.; Andersen, E. S.; Knudsen, B.; Gorodkin, J.; Kjems, J., RNA interactions in the 5' region of the HIV-1 genome. *J Mol Biol* **2004**, 336 (2), 369-79.
266. Heng, X.; Kharytonchyk, S.; Garcia, E. L.; Lu, K.; Divakaruni, S. S.; LaCotti, C.; Edme, K.; Telesnitsky, A.; Summers, M. F., Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J Mol Biol* **2012**, 417 (3), 224-39.
267. Lu, K.; Heng, X.; Summers, M. F., Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* **2011**, 410 (4), 609-33.
268. Nikolaitchik, O.; Rhodes, T. D.; Ott, D.; Hu, W. S., Effects of mutations in the human immunodeficiency virus type 1 Gag gene on RNA packaging and recombination. *J Virol* **2006**, 80 (10), 4691-7.
269. Lingappa, J. R.; Dooher, J. E.; Newman, M. A.; Kiser, P. K.; Klein, K. C., Basic residues in the nucleocapsid domain of Gag are required for interaction of HIV-1 gag with ABCE1 (HP68), a cellular protein important for HIV-1 capsid assembly. *J Biol Chem* **2006**, 281 (7), 3773-84.
270. Robinson, B. A.; Reed, J. C.; Geary, C. D.; Swain, J. V.; Lingappa, J. R., A temporospatial map that defines specific steps at which critical surfaces in the Gag MA and CA domains act during immature HIV-1 capsid assembly in cells. *J Virol* **2014**, 88 (10), 5718-41.
271. Mouland, A. J.; Mercier, J.; Luo, M.; Bernier, L.; DesGroseillers, L.; Cohen, E. A., The double-stranded RNA-binding protein Staufen is incorporated in human immunodeficiency virus type 1: evidence for a role in genomic RNA encapsidation. *J Virol* **2000**, 74 (12), 5441-51.

272. Reed, J. C.; Molter, B.; Geary, C. D.; McNevin, J.; McElrath, J.; Giri, S.; Klein, K. C.; Lingappa, J. R., HIV-1 Gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. *J Cell Biol* **2012**, *198* (3), 439-56.
273. Olety, B.; Ono, A., Roles played by acidic lipids in HIV-1 Gag membrane binding. *Virus Res* **2014**, *193*, 108-15.
274. Jouvenet, N.; Simon, S. M.; Bieniasz, P. D., Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc Natl Acad Sci U S A* **2009**, *106* (45), 19114-9.
275. Ono, A.; Waheed, A. A.; Joshi, A.; Freed, E. O., Association of human immunodeficiency virus type 1 gag with membrane does not require highly basic sequences in the nucleocapsid: use of a novel Gag multimerization assay. *J Virol* **2005**, *79* (22), 14131-40.
276. Burniston, M. T.; Cimarelli, A.; Colgan, J.; Curtis, S. P.; Luban, J., Human immunodeficiency virus type 1 Gag polyprotein multimerization requires the nucleocapsid domain and RNA and is promoted by the capsid-dimer interface and the basic region of matrix protein. *J Virol* **1999**, *73* (10), 8527-40.
277. Muriaux, D.; Darlix, J. L.; Cimarelli, A., Targeting the assembly of the human immunodeficiency virus type I. *Curr Pharm Des* **2004**, *10* (30), 3725-39.
278. Cimarelli, A.; Darlix, J. L., Assembling the human immunodeficiency virus type 1. *Cell Mol Life Sci* **2002**, *59* (7), 1166-84.
279. Thomas, J. A.; Gorelick, R. J., Nucleocapsid protein function in early infection processes. *Virus Res* **2008**, *134* (1-2), 39-63.
280. Crist, R. M.; Datta, S. A.; Stephen, A. G.; Soheilian, F.; Mirro, J.; Fisher, R. J.; Nagashima, K.; Rein, A., Assembly properties of human immunodeficiency virus type 1 Gag-leucine zipper chimeras: implications for retrovirus assembly. *J Virol* **2009**, *83* (5), 2216-25.
281. Llewellyn, G. N.; Hogue, I. B.; Grover, J. R.; Ono, A., Nucleocapsid promotes localization of HIV-1 gag to uropods that participate in virological synapses between T cells. *PLoS Pathog* **2010**, *6* (10), e1001167.
282. Sandefur, S.; Smith, R. M.; Varthakavi, V.; Spearman, P., Mapping and characterization of the N-terminal I domain of human immunodeficiency virus type 1 Pr55(Gag). *J Virol* **2000**, *74* (16), 7238-49.
283. El Meshri, S. E.; Dujardin, D.; Godet, J.; Richert, L.; Boudier, C.; Darlix, J. L.; Didier, P.; Mely, Y.; de Rocquigny, H., Role of the nucleocapsid domain in HIV-1 Gag oligomerization and trafficking to the plasma membrane: a fluorescence lifetime imaging microscopy investigation. *J Mol Biol* **2015**, *427* (6 Pt B), 1480-94.
284. Dussupt, V.; Javid, M. P.; Abou-Jaoude, G.; Jadwin, J. A.; de La Cruz, J.;

- Nagashima, K.; Bouamr, F., The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding. *PLoS Pathog* **2009**, *5* (3), e1000339.
285. Sette, P.; Dussupt, V.; Bouamr, F., Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA. *J Virol* **2012**, *86* (21), 11608-15.
286. Fossen, T.; Wray, V.; Bruns, K.; Rachmat, J.; Henklein, P.; Tessmer, U.; Maczurek, A.; Klinger, P.; Schubert, U., Solution structure of the human immunodeficiency virus type 1 p6 protein. *J Biol Chem* **2005**, *280* (52), 42515-27.
287. Garrus, J. E.; von Schwedler, U. K.; Pornillos, O. W.; Morham, S. G.; Zavitz, K. H.; Wang, H. E.; Wettstein, D. A.; Stray, K. M.; Cote, M.; Rich, R. L.; Myszka, D. G.; Sundquist, W. I., Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* **2001**, *107* (1), 55-65.
288. Huang, M.; Orenstein, J. M.; Martin, M. A.; Freed, E. O., p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J Virol* **1995**, *69* (11), 6810-8.
289. Fujii, K.; Hurley, J. H.; Freed, E. O., Beyond Tsg101: the role of Alix in 'ESCRTing' HIV-1. *Nat Rev Microbiol* **2007**, *5* (12), 912-6.
290. Strack, B.; Calistri, A.; Craig, S.; Popova, E.; Gottlinger, H. G., AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* **2003**, *114* (6), 689-99.
291. Kondo, E.; Mammano, F.; Cohen, E. A.; Gottlinger, H. G., The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *J Virol* **1995**, *69* (5), 2759-64.
292. Paxton, W.; Connor, R. I.; Landau, N. R., Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analysis. *J Virol* **1993**, *67* (12), 7229-37.
293. Lu, Y. L.; Bennett, R. P.; Wills, J. W.; Gorelick, R.; Ratner, L., A leucine triplet repeat sequence (LXX)₄ in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. *J Virol* **1995**, *69* (11), 6873-9.
294. Accola, M. A.; Ohagen, A.; Gottlinger, H. G., Isolation of human immunodeficiency virus type 1 cores: retention of Vpr in the absence of p6(gag). *J Virol* **2000**, *74* (13), 6198-202.
295. Welker, R.; Hohenberg, H.; Tessmer, U.; Huckhagel, C.; Krausslich, H. G., Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1. *J Virol* **2000**, *74* (3), 1168-77.
296. Carter, C. A.; Ehrlich, L. S., Cell biology of HIV-1 infection of macrophages. *Annu Rev Microbiol* **2008**, *62*, 425-43.

297. Pornillos, O.; Higginson, D. S.; Stray, K. M.; Fisher, R. D.; Garrus, J. E.; Payne, M.; He, G. P.; Wang, H. E.; Morham, S. G.; Sundquist, W. I., HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. *J Cell Biol* **2003**, *162* (3), 425-34.
298. Martin-Serrano, J.; Neil, S. J., Host factors involved in retroviral budding and release. *Nat Rev Microbiol* **2011**, *9* (7), 519-31.
299. Bieniasz, P. D., Late budding domains and host proteins in enveloped virus release. *Virology* **2006**, *344* (1), 55-63.
300. Morita, E.; Sundquist, W. I., Retrovirus budding. *Annu Rev Cell Dev Biol* **2004**, *20*, 395-425.
301. Carlton, J. G.; Martin-Serrano, J., The ESCRT machinery: new functions in viral and cellular biology. *Biochem Soc Trans* **2009**, *37* (Pt 1), 195-9.
302. Morellet, N.; Roques, B. P.; Bouaziz, S., Structure-function relationship of Vpr: biological implications. *Curr HIV Res* **2009**, *7* (2), 184-210.
303. Kogan, M.; Rappaport, J., HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention. *Retrovirology* **2011**, *8*, 25.
304. Zhu, H.; Jian, H.; Zhao, L. J., Identification of the 15FRFG domain in HIV-1 Gag p6 essential for Vpr packaging into the virion. *Retrovirology* **2004**, *1*, 26.
305. Popov, S.; Popova, E.; Inoue, M.; Gottlinger, H. G., Human immunodeficiency virus type 1 Gag engages the Bro1 domain of ALIX/AIP1 through the nucleocapsid. *J Virol* **2008**, *82* (3), 1389-98.
306. Sundquist, W. I.; Krausslich, H. G., HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med* **2012**, *2* (7), a006924.
307. Ohishi, M.; Nakano, T.; Sakuragi, S.; Shioda, T.; Sano, K.; Sakuragi, J., The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. *Nucleic Acids Res* **2011**, *39* (8), 3404-17.
308. Konnyu, B.; Sadiq, S. K.; Turanyi, T.; Hirmondo, R.; Muller, B.; Krausslich, H. G.; Coveney, P. V.; Muller, V., Gag-Pol processing during HIV-1 virion maturation: a systems biology approach. *PLoS Comput Biol* **2013**, *9* (6), e1003103.
309. Bendjennat, M.; Saffarian, S., The Race against Protease Activation Defines the Role of ESCRTs in HIV Budding. *PLoS Pathog* **2016**, *12* (6), e1005657.
310. de Marco, A.; Muller, B.; Glass, B.; Riches, J. D.; Krausslich, H. G.; Briggs, J. A., Structural analysis of HIV-1 maturation using cryo-electron tomography. *PLoS Pathog* **2010**, *6* (11), e1001215.
311. Cruceanu, M.; Urbaneja, M. A.; Hixson, C. V.; Johnson, D. G.; Datta, S. A.; Fivash,

M. J.; Stephen, A. G.; Fisher, R. J.; Gorelick, R. J.; Casas-Finet, J. R.; Rein, A.; Rouzina, I.; Williams, M. C., Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res* **2006**, *34* (2), 593-605.

312. McCauley, M. J.; Rouzina, I.; Manthei, K. A.; Gorelick, R. J.; Musier-Forsyth, K.; Williams, M. C., Targeted binding of nucleocapsid protein transforms the folding landscape of HIV-1 TAR RNA. *Proc Natl Acad Sci U S A* **2015**, *112* (44), 13555-60.

313. Wang, W.; Naiyer, N.; Mitra, M.; Li, J.; Williams, M. C.; Rouzina, I.; Gorelick, R. J.; Wu, Z.; Musier-Forsyth, K., Distinct nucleic acid interaction properties of HIV-1 nucleocapsid protein precursor NCp15 explain reduced viral infectivity. *Nucleic Acids Res* **2014**, *42* (11), 7145-59.

314. Berthoux, L.; Pechoux, C.; Ottmann, M.; Morel, G.; Darlix, J. L., Mutations in the N-terminal domain of human immunodeficiency virus type 1 nucleocapsid protein affect virion core structure and proviral DNA synthesis. *J Virol* **1997**, *71* (9), 6973-81.

315. Craven, R. C.; Parent, L. J., Dynamic interactions of the Gag polyprotein. *Curr Top Microbiol Immunol* **1996**, *214*, 65-94.

316. Krishnamoorthy, G.; Roques, B.; Darlix, J. L.; Mely, Y., DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection. *Nucleic Acids Res* **2003**, *31* (18), 5425-32.

317. Briggs, J. A.; Simon, M. N.; Gross, I.; Krausslich, H. G.; Fuller, S. D.; Vogt, V. M.; Johnson, M. C., The stoichiometry of Gag protein in HIV-1. *Nat Struct Mol Biol* **2004**, *11* (7), 672-5.

318. Chertova, E.; Chertov, O.; Coren, L. V.; Roser, J. D.; Trubey, C. M.; Bess, J. W., Jr.; Sowder, R. C., 2nd; Barsov, E.; Hood, B. L.; Fisher, R. J.; Nagashima, K.; Conrads, T. P.; Veenstra, T. D.; Lifson, J. D.; Ott, D. E., Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *J Virol* **2006**, *80* (18), 9039-52.

319. Mervis, R. J.; Ahmad, N.; Lillehoj, E. P.; Raum, M. G.; Salazar, F. H.; Chan, H. W.; Venkatesan, S., The gag gene products of human immunodeficiency virus type 1: alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *J Virol* **1988**, *62* (11), 3993-4002.

320. Darlix, J. L.; Gabus, C.; Nugeyre, M. T.; Clavel, F.; Barre-Sinoussi, F., Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J Mol Biol* **1990**, *216* (3), 689-99.

321. Chen, J.; Nikolaitchik, O.; Singh, J.; Wright, A.; Bencsics, C. E.; Coffin, J. M.; Ni, N.; Lockett, S.; Pathak, V. K.; Hu, W. S., High efficiency of HIV-1 genomic RNA packaging and heterozygote formation revealed by single virion analysis. *Proc Natl Acad Sci U S A* **2009**, *106* (32), 13535-40.

322. Zhang, H.; Dornadula, G.; Pomerantz, R. J., Natural endogenous reverse transcription of HIV-1. *J Reprod Immunol* **1998**, *41* (1-2), 255-60.
323. Huseby, D.; Barklis, R. L.; Alfadhli, A.; Barklis, E., Assembly of human immunodeficiency virus precursor gag proteins. *J Biol Chem* **2005**, *280* (18), 17664-70.
324. Carteau, S.; Batson, S. C.; Poljak, L.; Mouscadet, J. F.; de Rocquigny, H.; Darlix, J. L.; Roques, B. P.; Kas, E.; Auclair, C., Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg²⁺-dependent DNA integration in vitro. *J Virol* **1997**, *71* (8), 6225-9.
325. Rajkowitsch, L.; Schroeder, R., Dissecting RNA chaperone activity. *RNA* **2007**, *13* (12), 2053-60.
326. Cen, S.; Huang, Y.; Khorchid, A.; Darlix, J. L.; Wainberg, M. A.; Kleiman, L., The role of Pr55(gag) in the annealing of tRNA³Lys to human immunodeficiency virus type 1 genomic RNA. *J Virol* **1999**, *73* (5), 4485-8.
327. Dorfman, T.; Luban, J.; Goff, S. P.; Haseltine, W. A.; Gottlinger, H. G., Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **1993**, *67* (10), 6159-69.
328. Gorelick, R. J.; Gagliardi, T. D.; Bosche, W. J.; Wiltrout, T. A.; Coren, L. V.; Chabot, D. J.; Lifson, J. D.; Henderson, L. E.; Arthur, L. O., 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* **1999**, *256* (1), 92-104.
329. de Rocquigny, H.; Shvadchak, V.; Avilov, S.; Dong, C. Z.; Dietrich, U.; Darlix, J. L.; Mely, Y., Targeting the viral nucleocapsid protein in anti-HIV-1 therapy. *Mini Rev Med Chem* **2008**, *8* (1), 24-35.
330. Mori, M.; Manetti, F.; Botta, M., Targeting protein-protein and protein-nucleic acid interactions for anti-HIV therapy. *Curr Pharm Des* **2011**, *17* (33), 3713-28.
331. Pettit, S. C.; Simsic, J.; Loeb, D. D.; Everitt, L.; Hutchison, C. A., 3rd; Swanstrom, R., Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. *J Biol Chem* **1991**, *266* (22), 14539-47.
332. Shehu-Xhilaga, M.; Kraeusslich, H. G.; Pettit, S.; Swanstrom, R.; Lee, J. Y.; Marshall, J. A.; Crowe, S. M.; Mak, J., Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol* **2001**, *75* (19), 9156-64.
333. Sheng, N.; Pettit, S. C.; Tritch, R. J.; Ozturk, D. H.; Rayner, M. M.; Swanstrom, R.; Erickson-Viitanen, S., Determinants of the human immunodeficiency virus type 1 p15NC-RNA interaction that affect enhanced cleavage by the viral protease. *J Virol* **1997**, *71* (8), 5723-32.

334. Mirambeau, G.; Lyonnais, S.; Gorelick, R. J., Features, processing states, and heterologous protein interactions in the modulation of the retroviral nucleocapsid protein function. *RNA Biol* **2010**, *7* (6), 724-34.
335. Tanchou, V.; Gabus, C.; Rogemond, V.; Darlix, J. L., Formation of stable and functional HIV-1 nucleoprotein complexes in vitro. *J Mol Biol* **1995**, *252* (5), 563-71.
336. Coren, L. V.; Thomas, J. A.; Chertova, E.; Sowder, R. C., 2nd; Gagliardi, T. D.; Gorelick, R. J.; Ott, D. E., Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1. *Journal of virology* **2007**, *81* (18), 10047-54.
337. de Marco, A.; Heuser, A. M.; Glass, B.; Krausslich, H. G.; Muller, B.; Briggs, J. A., Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J Virol* **2012**, *86* (24), 13708-16.
338. Lee, S. K.; Harris, J.; Swanstrom, R., A strongly transdominant mutation in the human immunodeficiency virus type 1 gag gene defines an Achilles heel in the virus life cycle. *Journal of virology* **2009**, *83* (17), 8536-43.
339. Lee, S. K.; Potempa, M.; Kolli, M.; Ozen, A.; Schiffer, C. A.; Swanstrom, R., Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J Biol Chem* **2012**, *287* (16), 13279-90.
340. Mely, Y.; Jullian, N.; Morellet, N.; De Rocquigny, H.; Dong, C. Z.; Piemont, E.; Roques, B. P.; Gerard, D., Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer. *Biochemistry* **1994**, *33* (40), 12085-91.
341. De Rocquigny, H.; Gabus, C.; Vincent, A.; Fournie-Zaluski, M. C.; Roques, B.; Darlix, J. L., Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci U S A* **1992**, *89* (14), 6472-6.
342. Morellet, N.; de Rocquigny, H.; Mely, Y.; Jullian, N.; Demene, H.; Ottmann, M.; Gerard, D.; Darlix, J. L.; Fournie-Zaluski, M. C.; Roques, B. P., Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by ¹H NMR. *J Mol Biol* **1994**, *235* (1), 287-301.
343. Berg, J. M., Potential metal-binding domains in nucleic acid binding proteins. *Science* **1986**, *232* (4749), 485-7.
344. Summers, M. F.; Henderson, L. E.; Chance, M. R.; Bess, J. W., Jr.; South, T. L.; Blake, P. R.; Sagi, I.; Perez-Alvarado, G.; Sowder, R. C., 3rd; Hare, D. R.; et al., Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. *Protein science : a publication of the Protein Society* **1992**, *1* (5), 563-74.
345. Bess, J. W., Jr.; Powell, P. J.; Issaq, H. J.; Schumack, L. J.; Grimes, M. K.;

- Henderson, L. E.; Arthur, L. O., Tightly bound zinc in human immunodeficiency virus type 1, human T-cell leukemia virus type I, and other retroviruses. *J Virol* **1992**, *66* (2), 840-7.
346. Mely, Y.; De Rocquigny, H.; Morellet, N.; Roques, B. P.; Gerad, D., Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. *Biochemistry* **1996**, *35* (16), 5175-82.
347. Bombarda, E.; Grell, E.; Roques, B. P.; Mely, Y., Molecular mechanism of the Zn²⁺-induced folding of the distal CCHC finger motif of the HIV-1 nucleocapsid protein. *Biophys J* **2007**, *93* (1), 208-17.
348. Bombarda, E.; Morellet, N.; Cherradi, H.; Spiess, B.; Bouaziz, S.; Grell, E.; Roques, B. P.; Mely, Y., Determination of the pK(a) of the four Zn²⁺-coordinating residues of the distal finger motif of the HIV-1 nucleocapsid protein: consequences on the binding of Zn²⁺. *J Mol Biol* **2001**, *310* (3), 659-72.
349. Morellet, N.; Jullian, N.; De Rocquigny, H.; Maigret, B.; Darlix, J. L.; Roques, B. P., Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR. *The EMBO journal* **1992**, *11* (8), 3059-65.
350. Omichinski, J. G.; Clore, G. M.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M., Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy. *FEBS Lett* **1991**, *292* (1-2), 25-30.
351. Demene, H.; Dong, C. Z.; Ottmann, M.; Rouyez, M. C.; Jullian, N.; Morellet, N.; Mely, Y.; Darlix, J. L.; Fournie-Zaluski, M. C.; Saragosti, S.; et al., 1H NMR structure and biological studies of the His23-->Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity. *Biochemistry* **1994**, *33* (39), 11707-16.
352. Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F., NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **2000**, *301* (2), 491-511.
353. Darlix, J. L.; Godet, J.; Ivanyi-Nagy, R.; Fosse, P.; Mauffret, O.; Mely, Y., Flexible nature and specific functions of the HIV-1 nucleocapsid protein. *J Mol Biol* **2011**, *410* (4), 565-81.
354. Athavale, S. S.; Ouyang, W.; McPike, M. P.; Hudson, B. S.; Borer, P. N., Effects of the nature and concentration of salt on the interaction of the HIV-1 nucleocapsid protein with SL3 RNA. *Biochemistry* **2010**, *49* (17), 3525-33.
355. Lapadat-Tapolsky, M.; Pernelle, C.; Borie, C.; Darlix, J. L., Analysis of the nucleic acid annealing activities of nucleocapsid protein from HIV-1. *Nucleic Acids Res* **1995**, *23* (13), 2434-41.
356. You, J. C.; McHenry, C. S., Human immunodeficiency virus nucleocapsid protein

accelerates strand transfer of the terminally redundant sequences involved in reverse transcription. *J Biol Chem* **1994**, *269* (50), 31491-5.

357. Fisher, R. J.; Rein, A.; Fivash, M.; Urbaneja, M. A.; Casas-Finet, J. R.; Medaglia, M.; Henderson, L. E., Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J Virol* **1998**, *72* (3), 1902-9.

358. Williams, M. C.; Rouzina, I.; Wenner, J. R.; Gorelick, R. J.; Musier-Forsyth, K.; Bloomfield, V. A., Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc Natl Acad Sci U S A* **2001**, *98* (11), 6121-6.

359. Wu, J. Q.; Ozarowski, A.; Maki, A. H.; Urbaneja, M. A.; Henderson, L. E.; Casas-Finet, J. R., Binding of the nucleocapsid protein of type 1 human immunodeficiency virus to nucleic acids studied using phosphorescence and optically detected magnetic resonance. *Biochemistry* **1997**, *36* (41), 12506-18.

360. De Guzman, R. N.; Wu, Z. R.; Stalling, C. C.; Pappalardo, L.; Borer, P. N.; Summers, M. F., Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **1998**, *279* (5349), 384-8.

361. Amarasinghe, G. K.; De Guzman, R. N.; Turner, R. B.; Chancellor, K. J.; Wu, Z. R.; Summers, M. F., NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *J Mol Biol* **2000**, *301* (2), 491-511.

362. Bazzi, A.; Zargarian, L.; Chaminade, F.; Boudier, C.; De Rocquigny, H.; Rene, B.; Mely, Y.; Fosse, P.; Mauffret, O., Structural insights into the cTAR DNA recognition by the HIV-1 nucleocapsid protein: role of sugar deoxyriboses in the binding polarity of NC. *Nucleic Acids Res* **2011**, *39* (9), 3903-16.

363. Zargarian, L.; Kanevsky, I.; Bazzi, A.; Boynard, J.; Chaminade, F.; Fosse, P.; Mauffret, O., Structural and dynamic characterization of the upper part of the HIV-1 cTAR DNA hairpin. *Nucleic Acids Res* **2009**, *37* (12), 4043-54.

364. Kanevsky, I.; Chaminade, F.; Chen, Y.; Godet, J.; Rene, B.; Darlix, J. L.; Mely, Y.; Mauffret, O.; Fosse, P., Structural determinants of TAR RNA-DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *Nucleic Acids Res* **2011**, *39* (18), 8148-62.

365. Bourbigot, S.; Ramalanjaona, N.; Boudier, C.; Salgado, G. F.; Roques, B. P.; Mely, Y.; Bouaziz, S.; Morellet, N., How the HIV-1 nucleocapsid protein binds and destabilises the (-)primer binding site during reverse transcription. *J Mol Biol* **2008**, *383* (5), 1112-28.

366. Morellet, N.; Demene, H.; Teilleux, V.; Huynh-Dinh, T.; de Rocquigny, H.; Fournie-Zaluski, M. C.; Roques, B. P., Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J Mol Biol* **1998**, *283* (2), 419-34.

367. Spriggs, S.; Garyu, L.; Connor, R.; Summers, M. F., Potential intra- and intermolecular interactions involving the unique-5' region of the HIV-1 5'-UTR. *Biochemistry* **2008**, *47* (49), 13064-73.
368. Paoletti, A. C.; Shubsda, M. F.; Hudson, B. S.; Borer, P. N., Affinities of the nucleocapsid protein for variants of SL3 RNA in HIV-1. *Biochemistry* **2002**, *41* (51), 15423-8.
369. Zargarian, I.; Tisné, C.; Barraud, P.; Xu, X.; Morellet, N.; René, B.; Mély, Y.; Fossé, P.; Mauffret, O., Dynamics of linker residues modulate the nucleic acid binding properties of the HIV-1 nucleocapsid protein fingers. *PloS one* **2014**, *9* (7), e102150.
370. Ramboarina, S.; Srividya, N.; Atkinson, R. A.; Morellet, N.; Roques, B. P.; Lefevre, J. F.; Mely, Y.; Kieffer, B., Effects of temperature on the dynamic behaviour of the HIV-1 nucleocapsid NCp7 and its DNA complex. *J Mol Biol* **2002**, *316* (3), 611-27.
371. Lee, B. M.; De Guzman, R. N.; Turner, B. G.; Tjandra, N.; Summers, M. F., Dynamical behavior of the HIV-1 nucleocapsid protein. *J Mol Biol* **1998**, *279* (3), 633-49.
372. Abbondanzieri, E. A.; Bokinsky, G.; Rausch, J. W.; Zhang, J. X.; Le Grice, S. F.; Zhuang, X., Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature* **2008**, *453* (7192), 184-9.
373. Gorelick, R. J.; Chabot, D. J.; Rein, A.; Henderson, L. E.; Arthur, L. O., The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent. *J Virol* **1993**, *67* (7), 4027-36.
374. Cruceanu, M.; Gorelick, R. J.; Musier-Forsyth, K.; Rouzina, I.; Williams, M. C., Rapid kinetics of protein-nucleic acid interaction is a major component of HIV-1 nucleocapsid protein's nucleic acid chaperone function. *J Mol Biol* **2006**, *363* (5), 867-77.
375. Barraud, P.; Gaudin, C.; Dardel, F.; Tisne, C., New insights into the formation of HIV-1 reverse transcription initiation complex. *Biochimie* **2007**, *89* (10), 1204-10.
376. Tisne, C.; Roques, B. P.; Dardel, F., Heteronuclear NMR studies of the interaction of tRNA(Lys)₃ with HIV-1 nucleocapsid protein. *J Mol Biol* **2001**, *306* (3), 443-54.
377. Wu, H.; Mitra, M.; McCauley, M. J.; Thomas, J. A.; Rouzina, I.; Musier-Forsyth, K.; Williams, M. C.; Gorelick, R. J., Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication. *Virus Res* **2013**, *171* (2), 263-77.
378. Guo, J.; Wu, T.; Kane, B. F.; Johnson, D. G.; Henderson, L. E.; Gorelick, R. J.; Levin, J. G., Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **2002**, *76* (9), 4370-8.
379. Heath, M. J.; Derebail, S. S.; Gorelick, R. J.; DeStefano, J. J., Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-

enhanced nucleic acid annealing. *J Biol Chem* **2003**, *278* (33), 30755-63.

380. Avilov, S. V.; Piemont, E.; Shvadchak, V.; de Rocquigny, H.; Mely, Y., Probing dynamics of HIV-1 nucleocapsid protein/target hexanucleotide complexes by 2-aminopurine. *Nucleic Acids Res* **2008**, *36* (3), 885-96.

381. Aduri, R.; Briggs, K. T.; Gorelick, R. J.; Marino, J. P., Molecular determinants of HIV-1 NCp7 chaperone activity in maturation of the HIV-1 dimerization initiation site. *Nucleic Acids Res* **2013**, *41* (4), 2565-80.

382. Webb, J. A.; Jones, C. P.; Parent, L. J.; Rouzina, I.; Musier-Forsyth, K., Distinct binding interactions of HIV-1 Gag to Psi and non-Psi RNAs: implications for viral genomic RNA packaging. *RNA* **2013**, *19* (8), 1078-88.

383. Beltz, H.; Clauss, C.; Piemont, E.; Ficheux, D.; Gorelick, R. J.; Roques, B.; Gabus, C.; Darlix, J. L.; de Rocquigny, H.; Mely, Y., Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis. *J Mol Biol* **2005**, *348* (5), 1113-26.

384. Vuilleumier, C.; Bombarda, E.; Morellet, N.; Gerard, D.; Roques, B. P.; Mely, Y., Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **1999**, *38* (51), 16816-25.

385. Mely, Y.; Piemont, E.; Sorinas-Jimeno, M.; de Rocquigny, H.; Jullian, N.; Morellet, N.; Roques, B. P.; Gerard, D., Structural and dynamic characterization of the aromatic amino acids of the human immunodeficiency virus type I nucleocapsid protein zinc fingers and their involvement in heterologous tRNA(Phe) binding: a steady-state and time-resolved fluorescence study. *Biophys J* **1993**, *65* (4), 1513-22.

386. Urbaneja, M. A.; Kane, B. P.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Casas-Finet, J. R., Binding properties of the human immunodeficiency virus type 1 nucleocapsid protein p7 to a model RNA: elucidation of the structural determinants for function. *J Mol Biol* **1999**, *287* (1), 59-75.

387. Draper, D. E., Themes in RNA-protein recognition. *J Mol Biol* **1999**, *293* (2), 255-70.

388. Shajani, Z.; Varani, G., ¹³C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J Mol Biol* **2005**, *349* (4), 699-715.

389. Ravindranathan, S.; Oberstrass, F. C.; Allain, F. H., Increase in backbone mobility of the VTS1p-SAM domain on binding to SRE-RNA. *J Mol Biol* **2010**, *396* (3), 732-46.

390. Oberstrass, F. C.; Allain, F. H.; Ravindranathan, S., Changes in dynamics of SRE-RNA on binding to the VTS1p-SAM domain studied by ¹³C NMR relaxation. *J Am Chem Soc* **2008**, *130* (36), 12007-20.

391. Zhang, Q.; Stelzer, A. C.; Fisher, C. K.; Al-Hashimi, H. M., Visualizing spatially correlated dynamics that directs RNA conformational transitions. *Nature* **2007**, *450* (7173),

1263-7.

392. Mittermaier, A. K.; Kay, L. E., Observing biological dynamics at atomic resolution using NMR. *Trends Biochem Sci* **2009**, *34* (12), 601-11.

393. Palmer, A. G., 3rd; Kroenke, C. D.; Loria, J. P., Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol* **2001**, *339*, 204-38.

394. Palmer, A. G., 3rd, NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **2004**, *104* (8), 3623-40.

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

396. Tsuchihashi, Z.; Brown, P. O., DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol* **1994**, *68* (9), 5863-70.

397. Urbaneja, M. A.; Wu, M.; Casas-Finet, J. R.; Karpel, R. L., 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* **2002**, *318* (3), 749-64.

398. Darlix, J. L.; Vincent, A.; Gabus, C.; de Rocquigny, H.; Roques, B., Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C R Acad Sci III* **1993**, *316* (8), 763-71.

399. Feng, Y. X.; Campbell, S.; Harvin, D.; Ehresmann, B.; Ehresmann, C.; Rein, A., The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J Virol* **1999**, *73* (5), 4251-6.

400. Roldan, A.; Warren, O. U.; Russell, R. S.; Liang, C.; Wainberg, M. A., 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* **2005**, *280* (17), 17488-96.

401. Guo, J.; Wu, T.; Anderson, J.; Kane, B. F.; Johnson, D. G.; Gorelick, R. J.; Henderson, L. E.; Levin, J. G., Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer. *J Virol* **2000**, *74* (19), 8980-8.

402. Beltz, H.; Azoulay, J.; Bernacchi, S.; Clamme, J. P.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mely, Y., Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7. *J Mol Biol* **2003**, *328* (1), 95-108.

403. Beltz, H.; Piemont, E.; Schaub, E.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mely, Y., Role of the structure of the top half of HIV-1 cTAR DNA on the nucleic acid destabilizing

activity of the nucleocapsid protein NCp7. *J Mol Biol* **2004**, 338 (4), 711-23.

404. Bernacchi, S.; Stoylov, S.; Piemont, E.; Ficheux, D.; Roques, B. P.; Darlix, J. L.; Mely, Y., HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence. *J Mol Biol* **2002**, 317 (3), 385-99.

405. Zeng, Y.; Liu, H. W.; Landes, C. F.; Kim, Y. J.; Ma, X.; Zhu, Y.; Musier-Forsyth, K.; Barbara, P. F., Probing nucleation, reverse annealing, and chaperone function along the reaction path of HIV-1 single-strand transfer. *Proc Natl Acad Sci U S A* **2007**, 104 (31), 12651-6.

406. Godet, J.; de Rocquigny, H.; Raja, C.; Glasser, N.; Ficheux, D.; Darlix, J. L.; Mely, Y., During the early phase of HIV-1 DNA synthesis, nucleocapsid protein directs hybridization of the TAR complementary sequences via the ends of their double-stranded stem. *J Mol Biol* **2006**, 356 (5), 1180-92.

407. Godet, J.; Boudier, C.; Humbert, N.; Ivanyi-Nagy, R.; Darlix, J. L.; Mely, Y., Comparative nucleic acid chaperone properties of the nucleocapsid protein NCp7 and Tat protein of HIV-1. *Virus Res* **2012**, 169 (2), 349-60.

408. Rasclé, J. B.; Ficheux, D.; Darlix, J. L., Possible roles of nucleocapsid protein of MoMuLV in the specificity of proviral DNA synthesis and in the genetic variability of the virus. *J Mol Biol* **1998**, 280 (2), 215-25.

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

410. Herschlag, D., RNA chaperones and the RNA folding problem. *J Biol Chem* **1995**, 270 (36), 20871-4.

411. Herschlag, D.; Khosla, M.; Tsuchihashi, Z.; Karpel, R. L., An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO J* **1994**, 13 (12), 2913-24.

412. Belfetmi, A.; Zargarian, L.; Tisne, C.; Sleiman, D.; Morellet, N.; Lescop, E.; Maskri, O.; Rene, B.; Mely, Y.; Fosse, P.; Mauffret, O., Insights into the mechanisms of RNA secondary structure destabilization by the HIV-1 nucleocapsid protein. *RNA* **2016**, 22 (4), 506-17.

413. Mustoe, A. M.; Al-Hashimi, H. M.; Brooks, C. L., 3rd, Coarse grained models reveal essential contributions of topological constraints to the conformational free energy of RNA bulges. *J Phys Chem B* **2014**, 118 (10), 2615-27.

414. Lyonnais, S.; Gorelick, R. J.; Heniche-Boukhalfa, F.; Bouaziz, S.; Parissi, V.; Mouscadet, J. F.; Restle, T.; Gatell, J. M.; Le Cam, E.; Mirambeau, G., A protein ballet around the viral genome orchestrated by HIV-1 reverse transcriptase leads to an architectural switch: from nucleocapsid-condensed RNA to Vpr-bridged DNA. *Virus Res* **2013**, 171 (2), 287-303.

415. Azoulay, J.; Clamme, J. P.; Darlix, J. L.; Roques, B. P.; Mely, Y., Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations. *J Mol Biol* **2003**, *326* (3), 691-700.
416. Hong, M. K.; Harbron, E. J.; O'Connor, D. B.; Guo, J.; Barbara, P. F.; Levin, J. G.; Musier-Forsyth, K., Nucleic acid conformational changes essential for HIV-1 nucleocapsid protein-mediated inhibition of self-priming in minus-strand transfer. *J Mol Biol* **2003**, *325* (1), 1-10.
417. Driscoll, M. D.; Hughes, S. H., 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* **2000**, *74* (19), 8785-92.
418. Egele, C.; Schaub, E.; Ramalanjaona, N.; Piemont, E.; Ficheux, D.; Roques, B.; Darlix, J. L.; Mely, Y., HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions. *J Mol Biol* **2004**, *342* (2), 453-66.
419. Ramalanjaona, N.; de Rocquigny, H.; Millet, A.; Ficheux, D.; Darlix, J. L.; Mely, Y., Investigating the mechanism of the nucleocapsid protein chaperoning of the second strand transfer during HIV-1 DNA synthesis. *J Mol Biol* **2007**, *374* (4), 1041-53.
420. Liu, H. W.; Zeng, Y.; Landes, C. F.; Kim, Y. J.; Zhu, Y.; Ma, X.; Vo, M. N.; Musier-Forsyth, K.; Barbara, P. F., Insights on the role of nucleic acid/protein interactions in chaperoned nucleic acid rearrangements of HIV-1 reverse transcription. *Proc Natl Acad Sci U S A* **2007**, *104* (13), 5261-7.
421. Barat, C.; Schatz, O.; Le Grice, S.; Darlix, J. L., Analysis of the interactions of HIV1 replication primer tRNA(Lys,3) with nucleocapsid protein and reverse transcriptase. *J Mol Biol* **1993**, *231* (2), 185-90.
422. Golinelli, M. P.; Hughes, S. H., Secondary structure in the nucleic acid affects the rate of HIV-1 nucleocapsid-mediated strand annealing. *Biochemistry* **2003**, *42* (27), 8153-62.
423. Guo, J.; Henderson, L. E.; Bess, J.; Kane, B.; Levin, J. G., 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* **1997**, *71* (7), 5178-88.
424. Vo, M. N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K., HIV-1 nucleocapsid protein switches the pathway of transactivation response element RNA/DNA annealing from loop-loop "kissing" to "zipper". *J Mol Biol* **2009**, *386* (3), 789-801.
425. Vo, M. N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K., Effect of Mg(2+) and Na(+) on the nucleic acid chaperone activity of HIV-1 nucleocapsid protein: implications for reverse transcription. *J Mol Biol* **2009**, *386* (3), 773-88.
426. Vo, M. N.; Barany, G.; Rouzina, I.; Musier-Forsyth, K., Mechanistic studies of mini-TAR RNA/DNA annealing in the absence and presence of HIV-1 nucleocapsid protein. *J Mol Biol* **2006**, *363* (1), 244-61.

427. Wu, T.; Gorelick, R. J.; Levin, J. G., Selection of fully processed HIV-1 nucleocapsid protein is required for optimal nucleic acid chaperone activity in reverse transcription. *Virus Res* **2014**, *193*, 52-64.
428. Mirambeau, G.; Lyonnais, S.; Coulaud, D.; Hameau, L.; Lafosse, S.; Jeusset, J.; Borde, I.; Reboud-Ravaux, M.; Restle, T.; Gorelick, R. J.; Le Cam, E., HIV-1 protease and reverse transcriptase control the architecture of their nucleocapsid partner. *PloS one* **2007**, *2* (7), e669.
429. Mirambeau, G.; Lyonnais, S.; Coulaud, D.; Hameau, L.; Lafosse, S.; Jeusset, J.; Justome, A.; Delain, E.; Gorelick, R. J.; Le Cam, E., Transmission electron microscopy reveals an optimal HIV-1 nucleocapsid aggregation with single-stranded nucleic acids and the mature HIV-1 nucleocapsid protein. *J Mol Biol* **2006**, *364* (3), 496-511.
430. Qualley, D. F.; Stewart-Maynard, K. M.; Wang, F.; Mitra, M.; Gorelick, R. J.; Rouzina, I.; Williams, M. C.; Musier-Forsyth, K., C-terminal domain modulates the nucleic acid chaperone activity of human T-cell leukemia virus type 1 nucleocapsid protein via an electrostatic mechanism. *J Biol Chem* **2010**, *285* (1), 295-307.
431. Morcock, D. R.; Kane, B. P.; Casas-Finet, J. R., Fluorescence and nucleic acid binding properties of the human T-cell leukemia virus-type 1 nucleocapsid protein. *Biochim Biophys Acta* **2000**, *1481* (2), 381-94.
432. Morcock, D. R.; Sowder, R. C., 2nd; Casas-Finet, J. R., Role of the histidine residues of visna virus nucleocapsid protein in metal ion and DNA binding. *FEBS Lett* **2000**, *476* (3), 190-3.
433. Stewart-Maynard, K. M.; Cruceanu, M.; Wang, F.; Vo, M. N.; Gorelick, R. J.; Williams, M. C.; Rouzina, I.; Musier-Forsyth, K., Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J Virol* **2008**, *82* (20), 10129-42.
434. Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J., The Protein Data Bank and the challenge of structural genomics. *Nat Struct Biol* **2000**, *7 Suppl*, 957-9.
435. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999**, *293* (2), 321-31.
436. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *J Mol Graph Model* **2001**, *19* (1), 26-59.
437. Uversky, V. N., Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **2002**, *11* (4), 739-56.
438. Uversky, V. N., Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't? *FEBS Lett* **2002**, *514* (2-3), 181-3.

439. Kriwacki, R. W.; Hengst, L.; Tennant, L.; Reed, S. I.; Wright, P. E., Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* **1996**, *93* (21), 11504-9.
440. Ivanyi-Nagy, R.; Davidovic, L.; Khandjian, E. W.; Darlix, J. L., Disordered RNA chaperone proteins: from functions to disease. *Cell Mol Life Sci* **2005**, *62* (13), 1409-17.
441. Tompa, P., Intrinsically unstructured proteins. *Trends Biochem Sci* **2002**, *27* (10), 527-33.
442. Demchenko, A. P., Recognition between flexible protein molecules: induced and assisted folding. *J Mol Recognit* **2001**, *14* (1), 42-61.
443. Iakoucheva, L. M.; Radivojac, P.; Brown, C. J.; O'Connor, T. R.; Sikes, J. G.; Obradovic, Z.; Dunker, A. K., The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* **2004**, *32* (3), 1037-49.
444. Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zamai, M.; Zambonin, M., Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **1986**, *25* (8), 1847-51.
445. Ross, J. L., The Dark Matter of Biology. *Biophys J* **2016**, *111* (5), 909-16.
446. Ward, J. J.; McGuffin, L. J.; Bryson, K.; Buxton, B. F.; Jones, D. T., The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **2004**, *20* (13), 2138-9.
447. Iakoucheva, L. M.; Dunker, A. K., Order, disorder, and flexibility: prediction from protein sequence. *Structure* **2003**, *11* (11), 1316-7.
448. Zuker, M., On finding all suboptimal foldings of an RNA molecule. *Science* **1989**, *244* (4900), 48-52.
449. Uhlenbeck, O. C., Keeping RNA happy. *RNA* **1995**, *1* (1), 4-6.
450. Schroeder, R.; Barta, A.; Semrad, K., Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* **2004**, *5* (11), 908-19.
451. Lorsch, J. R., RNA chaperones exist and DEAD box proteins get a life. *Cell* **2002**, *109* (7), 797-800.
452. Tompa, P.; Csermely, P., The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* **2004**, *18* (11), 1169-75.
453. Woodson, S. A., Recent insights on RNA folding mechanisms from catalytic RNA. *Cell Mol Life Sci* **2000**, *57* (5), 796-808.
454. Semrad, K.; Green, R.; Schroeder, R., RNA chaperone activity of large ribosomal subunit proteins from *Escherichia coli*. *RNA* **2004**, *10* (12), 1855-60.

455. Rajkowitsch, L.; Semrad, K.; Mayer, O.; Schroeder, R., Assays for the RNA chaperone activity of proteins. *Biochem Soc Trans* **2005**, *33* (Pt 3), 450-6.
456. Mayer, O.; Rajkowitsch, L.; Lorenz, C.; Konrat, R.; Schroeder, R., RNA chaperone activity and RNA-binding properties of the E. coli protein StpA. *Nucleic Acids Res* **2007**, *35* (4), 1257-69.
457. Zhang, A.; Derbyshire, V.; Salvo, J. L.; Belfort, M., Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA* **1995**, *1* (8), 783-93.
458. Coetzee, T.; Herschlag, D.; Belfort, M., Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. *Genes Dev* **1994**, *8* (13), 1575-88.
459. Rajkowitsch, L.; Chen, D.; Stampfl, S.; Semrad, K.; Waldsich, C.; Mayer, O.; Jantsch, M. F.; Konrat, R.; Blasi, U.; Schroeder, R., RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **2007**, *4* (3), 118-30.
460. Jiang, W.; Hou, Y.; Inouye, M., CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *J Biol Chem* **1997**, *272* (1), 196-202.
461. Jones, P. G.; Mitta, M.; Kim, Y.; Jiang, W.; Inouye, M., Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in Escherichia coli. *Proc Natl Acad Sci U S A* **1996**, *93* (1), 76-80.
462. Phadtare, S.; Severinov, K., Nucleic acid melting by Escherichia coli CspE. *Nucleic Acids Res* **2005**, *33* (17), 5583-90.
463. Phadtare, S.; Severinov, K., RNA remodeling and gene regulation by cold shock proteins. *RNA Biol* **2010**, *7* (6), 788-95.
464. Max, K. E.; Zeeb, M.; Bienert, R.; Balbach, J.; Heinemann, U., T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB. *J Mol Biol* **2006**, *360* (3), 702-14.
465. Zeeb, M.; Max, K. E.; Weininger, U.; Low, C.; Sticht, H.; Balbach, J., Recognition of T-rich single-stranded DNA by the cold shock protein Bs-CspB in solution. *Nucleic Acids Res* **2006**, *34* (16), 4561-71.
466. Sachs, R.; Max, K. E.; Heinemann, U.; Balbach, J., RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. *RNA* **2012**, *18* (1), 65-76.
467. Phadtare, S.; Inouye, M.; Severinov, K., The mechanism of nucleic acid melting by a CspA family protein. *J Mol Biol* **2004**, *337* (1), 147-55.
468. Phadtare, S.; Tyagi, S.; Inouye, M.; Severinov, K., Three amino acids in Escherichia coli CspE surface-exposed aromatic patch are critical for nucleic acid melting activity leading to transcription antitermination and cold acclimation of cells. *J Biol Chem* **2002**,

277 (48), 46706-11.

469. Phadtare, S.; Inouye, M.; Severinov, K., The nucleic acid melting activity of *Escherichia coli* CspE is critical for transcription antitermination and cold acclimation of cells. *J Biol Chem* **2002**, *277* (9), 7239-45.

470. Prats, A. C.; Sarih, L.; Gabus, C.; Litvak, S.; Keith, G.; Darlix, J. L., Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA. *EMBO J* **1988**, *7* (6), 1777-83.

471. Zuniga, S.; Sola, I.; Cruz, J. L.; Enjuanes, L., Role of RNA chaperones in virus replication. *Virus Res* **2009**, *139* (2), 253-66.

472. Huang, Z. S.; Su, W. H.; Wang, J. L.; Wu, H. N., Selective strand annealing and selective strand exchange promoted by the N-terminal domain of hepatitis delta antigen. *J Biol Chem* **2003**, *278* (8), 5685-93.

473. Zuniga, S.; Sola, I.; Moreno, J. L.; Sabella, P.; Plana-Duran, J.; Enjuanes, L., Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* **2007**, *357* (2), 215-27.

474. Mir, M. A.; Panganiban, A. T., Characterization of the RNA chaperone activity of hantavirus nucleocapsid protein. *J Virol* **2006**, *80* (13), 6276-85.

475. DeStefano, J. J.; Titilope, O., Poliovirus protein 3AB displays nucleic acid chaperone and helix-destabilizing activities. *J Virol* **2006**, *80* (4), 1662-71.

476. Henriet, S.; Sinck, L.; Bec, G.; Gorelick, R. J.; Marquet, R.; Paillart, J. C., Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res* **2007**, *35* (15), 5141-53.

477. Kuciak, M.; Gabus, C.; Ivanyi-Nagy, R.; Semrad, K.; Storchak, R.; Chaloin, O.; Muller, S.; Mely, Y.; Darlix, J. L., The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro. *Nucleic Acids Res* **2008**, *36* (10), 3389-400.

478. Karakasiliotis, I.; Chaudhry, Y.; Roberts, L. O.; Goodfellow, I. G., Feline calicivirus replication: requirement for polypyrimidine tract-binding protein is temperature-dependent. *J Gen Virol* **2006**, *87* (Pt 11), 3339-47.

479. Shi, S. T.; Lai, M. M., Viral and cellular proteins involved in coronavirus replication. *Curr Top Microbiol Immunol* **2005**, *287*, 95-131.

480. Domitrovich, A. M.; Felmler, D. J.; Siddiqui, A., Hepatitis C virus nonstructural proteins inhibit apolipoprotein B100 secretion. *J Biol Chem* **2005**, *280* (48), 39802-8.

481. Anderson, E. C.; Hunt, S. L.; Jackson, R. J., Internal initiation of translation from the human rhinovirus-2 internal ribosome entry site requires the binding of Unr to two distinct sites on the 5' untranslated region. *J Gen Virol* **2007**, *88* (Pt 11), 3043-52.

482. Domitrovich, A. M.; Diebel, K. W.; Ali, N.; Sarker, S.; Siddiqui, A., Role of La autoantigen and polypyrimidine tract-binding protein in HCV replication. *Virology* **2005**, *335* (1), 72-86.
483. Hunt, S. L.; Jackson, R. J., Polypyrimidine-tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. *RNA* **1999**, *5* (3), 344-59.
484. Daros, J. A.; Flores, R., A chloroplast protein binds a viroid RNA in vivo and facilitates its hammerhead-mediated self-cleavage. *EMBO J* **2002**, *21* (4), 749-59.
485. Bertrand, E. L.; Rossi, J. J., Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. *EMBO J* **1994**, *13* (12), 2904-12.
486. Shojania, S.; O'Neil, J. D., HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. *J Biol Chem* **2006**, *281* (13), 8347-56.
487. Dosztanyi, Z.; Chen, J.; Dunker, A. K.; Simon, I.; Tompa, P., Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* **2006**, *5* (11), 2985-95.
488. Long, K. S.; Crothers, D. M., Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry* **1999**, *38* (31), 10059-69.
489. Doetsch, M.; Furtig, B.; Gstrein, T.; Stampfl, S.; Schroeder, R., The RNA annealing mechanism of the HIV-1 Tat peptide: conversion of the RNA into an annealing-competent conformation. *Nucleic Acids Res* **2011**, *39* (10), 4405-18.
490. Boudier, C.; Storchak, R.; Sharma, K. K.; Didier, P.; Follenius-Wund, A.; Muller, S.; Darlix, J. L.; Mely, Y., The mechanism of HIV-1 Tat-directed nucleic acid annealing supports its role in reverse transcription. *J Mol Biol* **2010**, *400* (3), 487-501.
491. Harrich, D.; Ulich, C.; Garcia-Martinez, L. F.; Gaynor, R. B., Tat is required for efficient HIV-1 reverse transcription. *EMBO J* **1997**, *16* (6), 1224-35.
492. Ulich, C.; Dunne, A.; Parry, E.; Hooker, C. W.; Gaynor, R. B.; Harrich, D., Functional domains of Tat required for efficient human immunodeficiency virus type 1 reverse transcription. *J Virol* **1999**, *73* (3), 2499-508.
493. Apolloni, A.; Meredith, L. W.; Suhrbier, A.; Kiernan, R.; Harrich, D., The HIV-1 Tat protein stimulates reverse transcription in vitro. *Curr HIV Res* **2007**, *5* (5), 473-83.
494. Dettenhofer, M.; Cen, S.; Carlson, B. A.; Kleiman, L.; Yu, X. F., Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. *J Virol* **2000**, *74* (19), 8938-45.

495. Suo, Z.; Johnson, K. A., Effect of RNA secondary structure on the kinetics of DNA synthesis catalyzed by HIV-1 reverse transcriptase. *Biochemistry* **1997**, *36* (41), 12459-67.
496. Klasens, B. I.; Huthoff, H. T.; Das, A. T.; Jeeninga, R. E.; Berkhout, B., The effect of template RNA structure on elongation by HIV-1 reverse transcriptase. *Biochim Biophys Acta* **1999**, *1444* (3), 355-70.
497. Henriët, S.; Richer, D.; Bernacchi, S.; Decroly, E.; Vigne, R.; Ehresmann, B.; Ehresmann, C.; Paillart, J. C.; Marquet, R., Cooperative and specific binding of Vif to the 5' region of HIV-1 genomic RNA. *J Mol Biol* **2005**, *354* (1), 55-72.
498. Zhang, J.; Tamilarasu, N.; Hwang, S.; Garber, M. E.; Huq, I.; Jones, K. A.; Rana, T. M., HIV-1 TAR RNA enhances the interaction between Tat and cyclin T1. *J Biol Chem* **2000**, *275* (44), 34314-9.
499. Bernacchi, S.; Henriët, S.; Dumas, P.; Paillart, J. C.; Marquet, R., RNA and DNA binding properties of HIV-1 Vif protein: a fluorescence study. *J Biol Chem* **2007**, *282* (36), 26361-8.
500. Post, K.; Olson, E. D.; Naufer, M. N.; Gorelick, R. J.; Rouzina, I.; Williams, M. C.; Musier-Forsyth, K.; Levin, J. G., Mechanistic differences between HIV-1 and SIV nucleocapsid proteins and cross-species HIV-1 genomic RNA recognition. *Retrovirology* **2016**, *13* (1), 89.
501. Marc, D.; Mercey, R.; Lantier, F., Scavenger, transducer, RNA chaperone? What ligands of the prion protein teach us about its function. *Cell Mol Life Sci* **2007**, *64* (7-8), 815-29.
502. Zanusso, G.; Petersen, R. B.; Jin, T.; Jing, Y.; Kanoush, R.; Ferrari, S.; Gambetti, P.; Singh, N., Proteasomal degradation and N-terminal protease resistance of the codon 145 mutant prion protein. *J Biol Chem* **1999**, *274* (33), 23396-404.
503. Gabus, C.; Derrington, E.; Leblanc, P.; Chnaiderman, J.; Dormont, D.; Swietnicki, W.; Morillas, M.; Surewicz, W. K.; Marc, D.; Nandi, P.; Darlix, J. L., The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. *J Biol Chem* **2001**, *276* (22), 19301-9.
504. Brockes, C.; Vogt, P. R.; Roth, T. B.; Turina, J., Double aortic arch: diagnosis missed for 29 years. *Vasa* **2000**, *29* (1), 77-9.
505. Mouillet-Richard, S.; Ermonval, M.; Chebassier, C.; Laplanche, J. L.; Lehmann, S.; Launay, J. M.; Kellermann, O., Signal transduction through prion protein. *Science* **2000**, *289* (5486), 1925-8.
506. Caughey, B.; Brown, K.; Raymond, G. J.; Katzenstein, G. E.; Thresher, W., Binding of the protease-sensitive form of PrP (prion protein) to sulfated glycosaminoglycan and congo red [corrected]. *J Virol* **1994**, *68* (4), 2135-41.
507. Weiss, S.; Proske, D.; Neumann, M.; Groschup, M. H.; Kretzschmar, H. A.; Famulok,

- M.; Winnacker, E. L., RNA aptamers specifically interact with the prion protein PrP. *J Virol* **1997**, *71* (11), 8790-7.
508. Nandi, P. K.; Leclerc, E., Polymerization of murine recombinant prion protein in nucleic acid solution. *Arch Virol* **1999**, *144* (9), 1751-63.
509. Akowitz, A.; Sklaviadis, T.; Manuelidis, L., Endogenous viral complexes with long RNA cosediment with the agent of Creutzfeldt-Jakob disease. *Nucleic Acids Res* **1994**, *22* (6), 1101-7.
510. Darlix, J. L.; Lapadat-Tapolsky, M.; de Rocquigny, H.; Roques, B. P., First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* **1995**, *254* (4), 523-37.
511. Carp, R. I.; Meeker, H. C.; Caruso, V.; Sersen, E., Scrapie strain-specific interactions with endogenous murine leukaemia virus. *J Gen Virol* **1999**, *80* (Pt 1), 5-10.
512. Heyduk, T.; Lee, J. C., Application of fluorescence energy transfer and polarization to monitor Escherichia coli cAMP receptor protein and lac promoter interaction. *Proc Natl Acad Sci U S A* **1990**, *87* (5), 1744-8.
513. Hill, J. J.; Royer, C. A., Fluorescence approaches to study of protein-nucleic acid complexation. *Methods Enzymol* **1997**, *278*, 390-416.
514. Weber, G., Rotational Brownian motion and polarization of the fluorescence of solutions. *Adv Protein Chem* **1953**, *8*, 415-59.
515. Rossi, A. M.; Taylor, C. W., Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* **2011**, *6* (3), 365-87.
516. Levison, S. A.; Hicks, A. N.; Portmann, A. J.; Dandliker, W. B., Fluorescence polarization and intensity kinetic studies of antiluorescein antibody obtained at different stages of the immune response. *Biochemistry* **1975**, *14* (17), 3778-86.
517. Rawitch, A. B.; Weber, G., The reversible association of lysozyme and thyroglobulin. Cooperative binding by near-neighbor interactions. *J Biol Chem* **1972**, *247* (3), 680-5.
518. Fernando, T.; Royer, C., Role of protein--protein interactions in the regulation of transcription by trp repressor investigated by fluorescence spectroscopy. *Biochemistry* **1992**, *31* (13), 3429-41.
519. Royer, C. A.; Ropp, T.; Scarlata, S. F., Solution studies of the interactions between the histone core proteins and DNA using fluorescence spectroscopy. *Biophys Chem* **1992**, *43* (2), 197-211.
520. Bernacchi, S.; Piemont, E.; Potier, N.; Dorsselaer, A.; Mely, Y., Excitonic heterodimer formation in an HIV-1 oligonucleotide labeled with a donor-acceptor pair used for fluorescence resonance energy transfer. *Biophys J* **2003**, *84* (1), 643-54.

521. Cousin, S. F.; Charlier, C.; Kaderavek, P.; Marquardsen, T.; Tyburn, J. M.; Bovier, P. A.; Ulzega, S.; Speck, T.; Wilhelm, D.; Engelke, F.; Maas, W.; Sakellariou, D.; Bodenhausen, G.; Pelupessy, P.; Ferrage, F., High-resolution two-field nuclear magnetic resonance spectroscopy. *Phys Chem Chem Phys* **2016**, *18* (48), 33187-33194.
522. Kerfah, R.; Plevin, M. J.; Sounier, R.; Gans, P.; Boisbouvier, J., Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr Opin Struct Biol* **2015**, *32*, 113-22.
523. Bell, N. G.; Michalchuk, A. A.; Blackburn, J. W.; Graham, M. C.; Uhrin, D., Isotope-Filtered 4D NMR Spectroscopy for Structure Determination of Humic Substances. *Angew Chem Int Ed Engl* **2015**, *54* (29), 8382-5.
524. Kanelis, V.; Forman-Kay, J. D.; Kay, L. E., Multidimensional NMR methods for protein structure determination. *IUBMB Life* **2001**, *52* (6), 291-302.
525. Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K., Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* **1997**, *94* (23), 12366-71.
526. Riek, R.; Pervushin, K.; Wuthrich, K., TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends Biochem Sci* **2000**, *25* (10), 462-8.
527. Takahashi, H.; Nakanishi, T.; Kami, K.; Arata, Y.; Shimada, I., A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat Struct Biol* **2000**, *7* (3), 220-3.
528. Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H., Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci U S A* **1995**, *92* (20), 9279-83.
529. Tjandra, N.; Bax, A., Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **1997**, *278* (5340), 1111-4.
530. Tugarinov, V.; Hwang, P. M.; Kay, L. E., Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* **2004**, *73*, 107-46.
531. Wuthrich, K.; Wider, G.; Wagner, G.; Braun, W., Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J Mol Biol* **1982**, *155* (3), 311-9.
532. Hartmann, S. R.; Hahn, E. L., Nuclear Double Resonance in Rotating Frame. *Phys Rev* **1962**, *128* (5), 2042-&.
533. Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G.; Rance, M.; Skelton, N. J., Protein NMR Spectroscopy: Principles and Practice, 2nd Edition. *Protein Nmr Spectroscopy: Principles and Practice, 2nd Edition* **2007**, 1-888.

534. Fischer, M. W.; Zeng, L.; Majumdar, A.; Zuiderweg, E. R., Characterizing semilocal motions in proteins by NMR relaxation studies. *Proc Natl Acad Sci U S A* **1998**, *95* (14), 8016-9.
535. Korzhnev, D. M.; Bocharov, E. V.; Zhuravlyova, A. V.; Orekhov, V. Y.; Ovchinnikova, T. V.; Billeter, M.; Arseniev, A. S., Backbone dynamics of the channel-forming antibiotic zervamicin IIB studied by ¹⁵N NMR relaxation. *FEBS Lett* **2001**, *495* (1-2), 52-5.
536. Palmer, A. G., 3rd, Nmr probes of molecular dynamics: overview and comparison with other techniques. *Annu Rev Biophys Biomol Struct* **2001**, *30*, 129-55.
537. Solomon, I., Relaxation Processes in a System of 2 Spins. *Phys Rev* **1955**, *99* (2), 559-565.
538. Luginbuhl, P.; Wuthrich, K., Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Prog Nucl Mag Res Sp* **2002**, *40* (3), 199-247.
539. Jarymowycz, V. A.; Stone, M. J., Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem Rev* **2006**, *106* (5), 1624-71.
540. Hansen, D. F.; Vallurupalli, P.; Kay, L. E., Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR* **2008**, *41* (3), 113-20.
541. Meiboom, S.; Gill, D., Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev Sci Instrum* **1958**, *29* (8), 688-691.
542. Carr, H. Y.; Purcell, E. M., Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys Rev* **1954**, *94* (3), 630-638.
543. Baldwin, A. J.; Kay, L. E., NMR spectroscopy brings invisible protein states into focus. *Nat Chem Biol* **2009**, *5* (11), 808-14.
544. Luz, Z.; Meiboom, S., Nuclear Magnetic Resonance Study of Protolysis of Trimethylammonium Ion in Aqueous Solution - Order of Reaction with Respect to Solvent. *Journal of Chemical Physics* **1963**, *39* (2), 366-&.
545. Carver, J. P.; Richards, R. E., General 2-Site Solution for Chemical Exchange Produced Dependence of T₂ Upon Carr-Purcell Pulse Separation. *J Magn Reson* **1972**, *6* (1), 89-&.
546. Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A. A.; Forman-Kay, J. D.; Kay, L. E., Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* **2001**, *123* (46), 11341-11352.
547. Press, W. H., *Numerical recipes : the art of scientific computing*. Cambridge University Press: Cambridge Cambridgeshire ; New York, 1986; p xx, 818 p.

548. Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E., Spectral density function mapping using ^{15}N relaxation data exclusively. *J Biomol NMR* **1995**, *6* (2), 153-62.
549. Hansen, D. F.; Kay, L. E., Improved magnetization alignment schemes for spin-lock relaxation experiments. *J Biomol NMR* **2007**, *37* (4), 245-55.
550. Hansen, D. F.; Vallurupalli, P.; Kay, L. E., An improved ^{15}N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B* **2008**, *112* (19), 5898-904.
551. Yamasaki, K.; Kigawa, T.; Inoue, M.; Tatenno, M.; Yamasaki, T.; Yabuki, T.; Aoki, M.; Seki, E.; Matsuda, T.; Tomo, Y.; Hayami, N.; Terada, T.; Shirouzu, M.; Osanai, T.; Tanaka, A.; Seki, M.; Shinozaki, K.; Yokoyama, S., Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell* **2004**, *16* (12), 3448-59.
552. Veprintsev, D. B.; Freund, S. M.; Andreeva, A.; Rutledge, S. E.; Tidow, H.; Canadillas, J. M.; Blair, C. M.; Fersht, A. R., Core domain interactions in full-length p53 in solution. *Proc Natl Acad Sci U S A* **2006**, *103* (7), 2115-9.
553. Clore, G. M.; Iwahara, J., Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **2009**, *109* (9), 4108-39.
554. Clore, G. M., Seeing the invisible by paramagnetic and diamagnetic NMR. *Biochem Soc Trans* **2013**, *41* (6), 1343-54.
555. Duss, O.; Konte, N. D. D.; Allain, F. H. T., Cut and Paste RNA for Nuclear Magnetic Resonance, Paramagnetic Resonance Enhancement, and Electron Paramagnetic Resonance Structural Studies. *Method Enzymol* **2015**, *565*, 537-562.
556. Iwahara, J.; Anderson, D. E.; Murphy, E. C.; Clore, G. M., EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J Am Chem Soc* **2003**, *125* (22), 6634-5.
557. Cai, S.; Lichtenberger, D. L.; Walker, F. A., NMR and EPR studies of the bis(pyridine) and bis(tert-butyl isocyanide) complexes of iron(III) octaethylchlorin. *Inorg Chem* **2005**, *44* (6), 1890-903.
558. Clore, G. M.; Tang, C.; Iwahara, J., Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr Opin Struct Biol* **2007**, *17* (5), 603-16.
559. Bernacchi, S.; Ennifar, E.; Toth, K.; Walter, P.; Langowski, J.; Dumas, P., Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J Biol Chem* **2005**, *280* (48), 40112-21.
560. Mitra, M.; Wang, W.; Vo, M. N.; Rouzina, I.; Barany, G.; Musier-Forsyth, K., The N-

terminal zinc finger and flanking basic domains represent the minimal region of the human immunodeficiency virus type-1 nucleocapsid protein for targeting chaperone function. *Biochemistry* **2013**, 52 (46), 8226-36.

561. You, J. C.; McHenry, C. S., HIV nucleocapsid protein. Expression in *Escherichia coli*, purification, and characterization. *J Biol Chem* **1993**, 268 (22), 16519-27.

562. Berkowitz, R. D.; Goff, S. P., Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology* **1994**, 202 (1), 233-46.

563. Geigenmuller, U.; Linial, M. L., Specific binding of human immunodeficiency virus type 1 (HIV-1) Gag-derived proteins to a 5' HIV-1 genomic RNA sequence. *J Virol* **1996**, 70 (1), 667-71.

564. Fisher, R. J.; Fivash, M. J.; Stephen, A. G.; Hagan, N. A.; Shenoy, S. R.; Medaglia, M. V.; Smith, L. R.; Worthy, K. M.; Simpson, J. T.; Shoemaker, R.; McNitt, K. L.; Johnson, D. G.; Hixson, C. V.; Gorelick, R. J.; Fabris, D.; Henderson, L. E.; Rein, A., Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. *Nucleic Acids Res* **2006**, 34 (2), 472-84.

565. Pappalardo, L.; Kerwood, D. J.; Pelczer, I.; Borer, P. N., Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J Mol Biol* **1998**, 282 (4), 801-18.

566. Chukkapalli, V.; Oh, S. J.; Ono, A., Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, 107 (4), 1600-5.

567. Shkriabai, N.; Datta, S. A.; Zhao, Z.; Hess, S.; Rein, A.; Kvaratskhelia, M., Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry* **2006**, 45 (13), 4077-83.

568. Deshmukh, L.; Ghirlando, R.; Clore, G. M., Conformation and dynamics of the Gag polyprotein of the human immunodeficiency virus 1 studied by NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, 112 (11), 3374-9.

569. Demirov, D. G.; Ono, A.; Orenstein, J. M.; Freed, E. O., Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99 (2), 955-60.

570. Lee, S.; Joshi, A.; Nagashima, K.; Freed, E. O.; Hurley, J. H., Structural basis for viral late-domain binding to Alix. *Nat Struct Mol Biol* **2007**, 14 (3), 194-9.

571. Dussupt, V.; Sette, P.; Bello, N. F.; Javid, M. P.; Nagashima, K.; Bouamr, F., Basic residues in the nucleocapsid domain of Gag are critical for late events of HIV-1 budding. *Journal of virology* **2011**, 85 (5), 2304-15.

572. Sette, P.; Dussupt, V.; Bouamr, F., Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA. *Journal of virology* **2012**, *86* (21), 11608-15.
573. Bachand, F.; Yao, X. J.; Hrimech, M.; Rougeau, N.; Cohen, E. A., Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *The Journal of biological chemistry* **1999**, *274* (13), 9083-91.
574. Salgado, G. F.; Marquant, R.; Vogel, A.; Alves, I. D.; Feller, S. E.; Morellet, N.; Bouaziz, S., Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. *Biochemistry* **2009**, *48* (11), 2355-67.
575. Salgado, G. F.; Vogel, A.; Marquant, R.; Feller, S. E.; Bouaziz, S.; Alves, I. D., The role of membranes in the organization of HIV-1 Gag p6 and Vpr: p6 shows high affinity for membrane bilayers which substantially increases the interaction between p6 and Vpr. *J Med Chem* **2009**, *52* (22), 7157-62.
576. Lescop, E.; Schanda, P.; Brutscher, B., A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* **2007**, *187* (1), 163-9.
577. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995**, *6* (3), 277-93.
578. Goddard, T. D.; Kneller, D. G., SPARKY 3. *University of California, San Francisco*.
579. Cheung, M. S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W., DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J Magn Reson* **2010**, *202* (2), 223-33.
580. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **2005**, *59* (4), 687-96.
581. Wishart, D. S.; Sykes, B. D.; Richards, F. M., The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31* (6), 1647-51.
582. Wishart, D. S.; Sykes, B. D., The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **1994**, *4* (2), 171-80.
583. Zhang, H.; Neal, S.; Wishart, D. S., RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* **2003**, *25* (3), 173-95.
584. Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J., Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **2001**, *123* (13), 2970-8.

585. Fossen, T.; Wray, V.; Bruns, K.; Rachmat, J.; Henklein, P.; Tessmer, U.; Maczurek, A.; Klinger, P.; Schubert, U., Solution structure of the human immunodeficiency virus type 1 p6 protein. *The Journal of biological chemistry* **2005**, *280* (52), 42515-27.
586. Kafaie, J.; Dolatshahi, M.; Ajamian, L.; Song, R.; Mouland, A. J.; Rouiller, I.; Laughrea, M., Role of capsid sequence and immature nucleocapsid proteins p9 and p15 in Human Immunodeficiency Virus type 1 genomic RNA dimerization. *Virology* **2009**, *385* (1), 233-44.
587. Deshmukh, L.; Schwieters, C. D.; Grishaev, A.; Clore, G. M., Quantitative Characterization of Configurational Space Sampled by HIV-1 Nucleocapsid Using Solution NMR, X-ray Scattering and Protein Engineering. *Chemphyschem* **2016**, *17* (11), 1548-52.
588. Jensen, M. R.; Ruigrok, R. W.; Blackledge, M., Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **2013**, *23* (3), 426-35.
589. Konrat, R., NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* **2014**, *241*, 74-85.
590. Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D., Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein science : a publication of the Protein Society* **2006**, *15* (12), 2795-804.
591. Brutscher, B.; Bruschiweiler, R.; Ernst, R. R., Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by ¹H, ¹³C, and ¹⁵N nuclear magnetic resonance spectroscopy. *Biochemistry* **1997**, *36* (42), 13043-53.
592. Feuerstein, S.; Solyom, Z.; Aladag, A.; Favier, A.; Schwarten, M.; Hoffmann, S.; Willbold, D.; Brutscher, B., Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J Mol Biol* **2012**, *420* (4-5), 310-23.
593. Solyom, Z.; Ma, P.; Schwarten, M.; Bosco, M.; Polidori, A.; Durand, G.; Willbold, D.; Brutscher, B., The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophys J* **2015**, *109* (7), 1483-96.
594. Das, R. K.; Pappu, R. V., Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (33), 13392-7.
595. Muller, B.; Patschinsky, T.; Krausslich, H. G., The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *Journal of virology* **2002**, *76* (3), 1015-24.
596. Hemonnot, B.; Cartier, C.; Gay, B.; Rebuffat, S.; Bardy, M.; Devaux, C.; Boyer, V.; Briant, L., The host cell MAP kinase ERK-2 regulates viral assembly and release by phosphorylating the p6gag protein of HIV-1. *The Journal of biological chemistry* **2004**, *279*

(31), 32426-34.

597. Kudoh, A.; Takahama, S.; Sawasaki, T.; Ode, H.; Yokoyama, M.; Okayama, A.; Ishikawa, A.; Miyakawa, K.; Matsunaga, S.; Kimura, H.; Sugiura, W.; Sato, H.; Hirano, H.; Ohno, S.; Yamamoto, N.; Ryo, A., The phosphorylation of HIV-1 Gag by atypical protein kinase C facilitates viral infectivity by promoting Vpr incorporation into virions. *Retrovirology* **2014**, *11*, 9.

598. Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J., Conformational entropy in molecular recognition by proteins. *Nature* **2007**, *448* (7151), 325-9.

599. Tzeng, S. R.; Kalodimos, C. G., Protein activity regulation by conformational entropy. *Nature* **2012**, *488* (7410), 236-40.

600. Liu, X.; Speckhard, D. C.; Shepherd, T. R.; Sun, Y. J.; Hengel, S. R.; Yu, L.; Fowler, C. A.; Gakhar, L.; Fuentes, E. J., Distinct Roles for Conformational Dynamics in Protein-Ligand Interactions. *Structure* **2016**, *24* (12), 2053-2066.

601. Chakrabarti, K. S.; Agafonov, R. V.; Pontiggia, F.; Otten, R.; Higgins, M. K.; Schertler, G. F.; Oprian, D. D.; Kern, D., Conformational Selection in a Protein-Protein Interaction Revealed by Dynamic Pathway Analysis. *Cell Rep* **2016**, *14* (1), 32-42.

602. Beach, H.; Cole, R.; Gill, M. L.; Loria, J. P., Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *J Am Chem Soc* **2005**, *127* (25), 9167-76.

603. Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E., The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **2006**, *313* (5793), 1638-42.

604. Salzman, M.; Pervushin, K.; Wider, G.; Senn, H.; Wuthrich, K., TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (23), 13585-90.

605. Loria, J. P.; Nowak, T., Conformational changes in yeast pyruvate kinase studied by 205TI+ NMR. *Biochemistry* **1998**, *37* (19), 6967-74.

606. Mulder, F. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E., Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. *J Am Chem Soc* **2001**, *123* (5), 967-75.

607. Ishima, R.; Torchia, D. A., Error estimation and global fitting in transverse-relaxation dispersion experiments to determine chemical-exchange parameters. *J Biomol NMR* **2005**, *32* (1), 41-54.

608. McConnell, H. M., Reaction Rates by nuclear magnetic resonance. *The Journal of Chemical Physics* **1958**, *28*, 430-431.

609. Davis, D. G.; Perlman, M. E.; London, R. E., Direct measurements of the dissociation-

rate constant for inhibitor-enzyme complexes via the T1 rho and T2 (CPMG) methods. *J Magn Reson B* **1994**, *104* (3), 266-75.

610. Millet, O.; Loria, J. P.; Kroenke, C. D.; Pons, M.; Palmer, A. G., 3rd, The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale. *J. Am. Chem. Soc.* **2000**, *122*, 2867-2877.

611. Mulder, F. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E., Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* **2001**, *8* (11), 932-5.

612. Tellinghuisen, J.; Chodera, J. D., Systematic errors in isothermal titration calorimetry: concentrations and baselines. *Anal Biochem* **2011**, *414* (2), 297-9.

Titre : « Les protéines de nucléocapside du VIH-1 : Structures, dynamiques, propriétés de fixation et de déstabilisation des acides nucléiques »

Mots clés : Nucléocapside, VIH-1, RMN, Protéines chaperonnes, Dynamique, Interactions moléculaires

Résumé : L'un des obstacles majeurs à l'éradication du VIH-1, est sa capacité à faire évoluer rapidement son matériel génétique. Ceci lui permet de contourner le système immunitaire de l'hôte ainsi que la pharmacologie antirétrovirale, du fait de l'émergence de formes mutantes et résistantes des enzymes ciblées. Parmi les facteurs à l'origine de ses recombinaisons génétiques, se trouvent les processus de transfert de brins, facilités par la protéine de nucléocapside (NC). Notre objectif principal est de mieux comprendre les propriétés de protéine chaperon de la NC, durant les transferts de brin en général et le premier transfert en particulier. Au cours de ce dernier, la NC déstabilise les structures secondaires des acides nucléiques impliqués (TAR ARN et cTAR ADN) afin de les hybrider pour former un hétéroduplex. D'autre part, nous souhaitons déterminer si la NC est capable de reconnaître la polarité des chaînes d'acides nucléiques selon leur nature (ADN ou ARN) et si cela module les propriétés déstabilisatrices de la protéine.

Pour réaliser cette étude, nous avons étudié différents complexes entre la NC et différents types d'acides nucléiques (ADN et ARN), en utilisant principalement la résonance magnétique nucléaire couplée à d'autres méthodes biophysiques et biochimiques, dans le but d'obtenir des informations à l'échelle atomique.

Dans un premier temps, l'étude de la déstabilisation de la structure secondaire de TAR ARN par NC à faible ratio (1 NC: 59 nt) montre que les "premières" paires de bases déstabilisées sont localisées dans la tige "apicale" suggérant un site de nucléation : la G26. Par ailleurs, l'étude du mécanisme d'hybridation du couple TAR-cTAR en présence de NC à faibles concentrations, montre que l'effet de la NC sur les vitesses d'hybridation est toujours perceptible. Cependant, la vitesse de conversion vers le duplexe étendu est fortement dépendante d'un rapport de NC/nt important.

En outre, on a démontré que la position des boucles internes de TAR-cTAR, influait sur les vitesses d'association. Ces données sont en accord avec une fixation et une déstabilisation ordonnée de la NC sur les acides nucléiques et influent sur les voies d'hybridation empruntées « zipper ou kissing ».

Nos études sur l'affinité de liaison de la NC avec différentes séquences d'acides nucléiques nous ont permis de trouver un motif de haute affinité (GGC). Également, la comparaison de l'affinité de liaison entre la séquence SL3 ARN et ADN (contenant un motif GXG sur la boucle apicale) montre des différences importantes avec une plus forte affinité de liaison pour l'ARN. Cela démontre que la présence de guanines non appariées n'est pas suffisante pour garantir une haute affinité. Il faut qu'en plus ces guanines soient bien accessibles (exemptes d'interactions d'appariement et d'empilement). Si ces conditions sont respectées, les substrats contenant deux guanines avec la séquence GXG ont une très forte affinité pour la NC.

Nos travaux sur la dynamique interne de la NC ont permis de montrer que certains mouvements étaient corrélés avec les modalités de fixation sur l'ARN et permettaient de mieux comprendre le fonctionnement coordonné des deux doigts de zinc. Les mesures des paramètres dynamiques ont été effectuées sur la protéine mature NCp7 et les formes immatures NCp9 et NCp15. Ces deux dernières formes se distinguent de la NCp7 par la présence de domaines additionnels: p1 (NCp9) et p1+p6 (NCp15). Les domaines p1 et p6 apparaissent capables de moduler de façon importante la dynamique du domaine NC sur différentes échelles de temps. Ces modulations proviendraient principalement de contacts transitoires entre les domaines p1 et p6, qui sont largement désordonnés. En retour les propriétés d'interaction avec les acides nucléiques des trois protéines diffèrent largement.

Ceci nous amène à reconsidérer le rôle des différentes parties de la polyprotéine Gag, sur les propriétés du domaine NC au cours de l'étape de reconnaissance et de la sélection du génome viral ARN, pour l'encapsider dans les particules néosynthétisées. Ce travail peut également être utile dans la conception d'une nouvelle classe d'inhibiteurs dirigés contre la NC, sachant que celle-ci représente une cible thérapeutique attrayante, étant très conservée et très importante pour l'infectiosité virale.

Title : « The HIV-1 nucleocapsid proteins : Structures, dynamics, binding and destabilization properties of nucleic acids »

Keywords : Nucleocapsid, HIV-1, NMR, Chaperone protein, Dynamic, Molecular interactions

Abstract : A major limitation in eradicating HIV-1 is in part due to the rapid evolution of its genome. Therefore, the virus escapes the host's immune response and the antiretroviral pharmacology due to the emergence of mutant and resistant forms of the targeted enzymes. Strand transfer processes facilitated by nucleocapsid protein (NC) have been reported as a genetic recombination causative factor.

Our goal is to understand the chaperonni properties of NC protein during the strand transfers in general, and during the first strand transfer in particular; where NC destabilizes the involved RNA and DNA secondary structures to anneal them and form a hybrid RNA/DNA duplex.

Furthermore, this study would enable us to assess whether NC have the property to recognize or not the polarity of nucleic acids chains depending on their nature (DNA or RNA) and if this modulates its destabilization properties.

In order to answer this question, we studied different complexes between NC and nucleic acid sequences (RNA and DNA) using mainly NMR spectroscopy and biophysical or biochemical methods to obtain information at the atomic scale.

First, we studied the destabilization mechanism of TAR RNA secondary structure by NC at low ratio (1 NC: 59 nt). We further demonstrated that the "first" destabilized base pairs are on the "apical" stem suggesting a site of nucleation: the G26. Next, we investigated the mechanism of hybridization of the TAR-cTAR in the presence of NC at low concentrations. Our results indicated that the effect of the NC on the hybridization rates is always perceptible. However, the rate of conversion to the extended duplex is strongly dependent on a high NC/nt ratio.

Finally, we demonstrated that the position of the internal loops of TAR-cTAR affected the association rates. These data are in agreement with an orderly destabilization and fixation of the NC on the nucleic acids chains and influence the hybridization paths borrowed "zipper or kissing".

Our studies on the binding affinity of NC with different nucleic acid sequences enabled us to find a high affinity motif (GGC). Also, the comparison of the binding affinity between SL3 RNA and DNA sequences (containing a GXG motif on the apical loop) shows significant differences with a higher binding affinity for the RNA. This demonstrates that the presence of unpaired guanines is not sufficient to ensure a high affinity. In addition, these guanines must be easily accessible (free from matching and stacking interactions). If these conditions are respected, the substrates containing two guanines with the GXG sequence have a very high affinity for the NC.

Our work on the internal dynamic of the protein demonstrating that some motions are correlated to its binding to RNA and DNA has the potential to improve our understanding of coordinated action of the two zinc finger domains. Here, we compared the properties of the different matured forms of NC protein during the viral life cycle and we concluded that p1 and p6 domains present within the immature forms (NCp9 and NCp15) tunes in an important manner the dynamic properties of the NC domain at different time scales and thus adjusted differently the interaction properties to nucleic acids compared to the mature form (NCp7).

Together these finding suggest a reconsideration of the role of the different parts of the Gag polyprotein on the properties of the NC during the recognition and selection steps of the viral genomic RNA, in order to package it in the newly formed viral particles.

This work can also be useful in the design of a new class of inhibitors against NC protein, which constitutes an attractive therapeutic target due to its conservation and importance in viral infectivity.