



HAL
open science

Study on the mechanism of HIV-1 maturation inhibitors on viral Gag precursor

Xiaowei Chen

► **To cite this version:**

Xiaowei Chen. Study on the mechanism of HIV-1 maturation inhibitors on viral Gag precursor. Structural Biology [q-bio.BM]. Université Paris Cité, 2020. English. NNT: 2020UNIP5005. tel-03867126

HAL Id: tel-03867126

<https://theses.hal.science/tel-03867126>

Submitted on 23 Nov 2022

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.

Université de Paris

École doctorale : Médicament- toxicologie-Chimie-Imagerie (MTCI)

*Laboratoire : Cibles Thérapeutiques et Conception de Médicaments (CiTCoM),
CNRS UMR8038, Faculté de pharmacie, université de Paris*

Study on the mechanism of HIV-1 maturation inhibitors on viral Gag precursor

Xiaowei Chen

Thèse de doctorat de Biologie Structurale

Dirigée par Serge Bouaziz

Présentée et soutenue publiquement le 25 Septembre 2020

Devant un jury composé de :

Dr Nathalie Chazal	MCU, Université de Montpellier	Rapportrice
Dr Christina Sizun	DR, Université Paris-Saclay	Rapportrice
Dr Benjamin Bardiaux	CR, Institut Pasteur	Examineur
Dr Jean-Christophe Paillart	DR, Université de Strasbourg	Examineur
Dr Serge Bouaziz	DR, Université de Paris	Directeur de thèse



Acknowledgements

This work was carried out in the *Cibles Thérapeutiques et Conception de Médicaments (CiTCoM)* at the *Faculté de Pharmacie* in Université de Paris. Firstly, I would like to thank the members of the jury for reading and judging my work. Thank you to Dr. Nathalie Chazal and Dr. Christina Sizun for agreeing to be reviewers of my thesis. Thanks also to Dr. Benjamin Bardiaux and Dr. Jean-Christophe Paillart for agreeing to review my work.

I would like to thank Prof. Nicolas Leulliot, the director of the CiTCoM, for accepting me in your laboratory and allowing me to have this precious experience. I would also like to thank the people I have met during the four years since 2016.

Firstly, thank you to my advisor Serge Bouaziz, head of the research team, Etudes structurales et fonctionnelles de nouvelles cibles thérapeutiques par RMN. Thank you very much for your encouragement, support and advice throughout the years. Thank you for teaching me a lot of things, including Nuclear Magnetic Resonance (NMR), which is completely a new area for me when I arrived at the lab. Thank you for helping me to solve tons of problems during my PhD especially first year, the French language learning, different daily life problems and my project. Thank you for your patience in many situations. For example, the software installation and running for protein structure calculation, the paper writing and submission, and English and French transformation in group meeting. Thank you for being open to all my ideas and giving me suggestions to help me advance my project.

I would also like to thank the members of my team for supporting and helping me in the four years. Thanks to Pascale Coric for giving me many information about this project and helping me with ARIA running and data analysis. Thanks to Sylvie for helping me in the expression and purification of protein. Thank you for your useful advices in the NMR experiments and paper writing. Thanks to Valéry Larue for the handling of the cryo-600 spectrometer and the help in structural calculation. Thanks to Édith Gincel for your help in the analysis of the calculated structures. Thanks to Philippe Laugâa for your docking results, which gives me more information about the EP39 binding. Thanks to Hesna Kara for showing me the manipulation of protein expression and purification. Thank you for your kindness and optimistic attitude that gives me a pleasant beginning of the PhD. Thanks to Xiao Wang for your help and accompany during the three years.

I also want to thank all the people in this lab that have helped and supported me. Thanks to Lila Delbos and Luc Ponchon for your help in my attempt of protein crystallization. Thanks to Nathalie Ulryck for helping me construct plasmid in my first year. Thanks to Magali Blaud, Cathy Bonamour Du Tartre, Cyril Garnier, Michèle Salem, Gille Phan, Jean-Francois Gaucher for your help during my thesis.

Besides, I also want to thank Nathalie Chazal, IRIM UMR9004 CNRS, for providing me the plasmids of his tagged CA_{CTD}SP1NC. Thanks to Prof. Micheal F. Summers, Department of Chemistry and Biochemistry UMBC, for kindly providing the plasmids for NC and wild type/mutated CA_{CTD}SP1NC. Thanks to Serge Turcaud, LCBPT UMR8601 CNRS, for providing EP39. Thanks to Zhao Dang, Department of Surgery, Duke University Medical Center, for providing the compound CB310F3. Thanks to Prof. Johnson Mak (Institute for Glycomics, Griffith University QLD 4111, Australia) for providing PF46396.

Finally, I would like to thank my families and friends for their love and support, especially my husband Xiao Wang. Thanks for your accompany and support during the past and upcoming years.

Contents

Acknowledgements	3
Abbreviations	9
List of Figures and Tables	11
Introduction	19
I The human immunodeficiency virus (HIV)	20
1. Overview of HIV-1 life cycle	20
1.1 Viral Attachment and Fusion	24
1.2 Reverse Transcription	24
1.3 Nuclear Entry	25
1.4 Integration	25
1.5 Transcription	25
1.6 Translation and Assembly	26
1.7 Budding	27
1.8 Maturation	27
2. HIV-1 particle maturation.....	28
II HIV-1 Gag precursor	31
1. Matrix	32
2. Capsid protein.....	36
2.1 CA monomer structure	36
2.2 Role of CA in the assembly of the immature HIV-1 Gag lattice.....	37
2.3 Role of CA in the formation of the mature capsid.....	38
3. Nucleocapsid.....	40
3.1 Structure of NC.....	40
3.2 Functions of NC.....	41
4. p6	44
4.1 structure of p6.....	44
4.2 Functions of p6.....	44
5. Spacer peptides (SP1 and SP2).....	47
5.1 Structure of the two spacer peptides	47
5.2 Functions of the two spacer peptides	47
6. Structure of CA _{CTD} -SP1-NC domain of Gag.....	48
III HIV-1 inhibitors	50
1. Reverse Transcriptase Inhibitors	52
1.1 Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs):.....	52

1.2 Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs)	52
2. Entry (attachment and fusion) Inhibitors	52
3. Protease Inhibitors (PIs)	53
4. Integrase Inhibitors (INIs)	53
5. Maturation Inhibitors (MIs).....	53
IV HIV-1 maturation inhibitors	54
1. Bevirimat, the first-generation maturation inhibitor.....	54
2. Second-generation maturation inhibitors.....	59
Objectives of the Thesis	61
Materials and methods	65
I Expression and purification of HIV-1 CA _{CTD} -SP1-NC, CA _{CTD} ^{W184A, M185A} -SP1-NC and NC	66
1. Overexpression of HIV-1 CA _{CTD} -SP1-NC, mutated CA _{CTD} ^{W184A, M185A} -SP1-NC and NC.....	66
1.1 Plasmid transformation	66
1.2 Cell culture	67
2. Purifications of HIV-1 CA _{CTD} -SP1-NC, mutated CA _{CTD} ^{W184A, M185A} -SP1-NC and NC.....	68
2.1 Lysis of bacteria	68
2.2 Elimination of nucleic acid	68
2.3 Ion exchange chromatography	68
2.4 Size exclusion chromatography	69
II Maturation inhibitors	71
1. EP39.....	71
2. PF-46396 and CB310F3	71
III Nuclear magnetic resonance (NMR)	72
1. NMR Introduction	72
2. Assignment of protein by NMR	73
2.1 General assignment strategy.....	73
2.2 Assignment of HIV-1 Gag CA _{CTD} ^{W184A, M185A} -SP1-NC.....	80
3. Chemical Shift Perturbation Analysis	80
4. Protein Structure Calculation.....	82
5. Protein dynamics	82
Results	85
I Expression and purification of HIV-1 NC, CA _{CTD} -SP1-NC and CA _{CTD} ^{W184A, M185A} -SP1-NC	86

1. Expression and purification of wild type HIV-1 (NL4-3) CA _{CTD} -SP1-NC and mutated CA _{CTD} ^{W184A, M185A} -SP1-NC.....	86
1.1 Expression of proteins CA _{CTD} -SP1-NC and CA _{CTD} ^{W184A, M185A} -SP1-NC.....	86
1.2 Purification of protein CA _{CTD} -SP1-NC and CA _{CTD} ^{W184A, M185A} -SP1-NC.	87
2. Expression and purification of HIV-1 (NL4-3) NC.	90
2.1 Expression of protein HIV-1 NC.	90
2.2 Purification of protein HIV-1 NC.	90
II Nuclear Magnetic Resonance (NMR) Analysis of Protein HIV-1 CA_{CTD}-SP1-NC, CA_{CTD}^{W184A, M185A}-SP1-NC and NC.....	94
1. Folding of protein HIV-1 CA _{CTD} -SP1-NC, CA _{CTD} ^{W184A, M185A} -SP1-NC and NC	94
2. The two mutations, W184A and M185A, abolish CA dimerization	95
3. NC does not seem to interact with CA _{CTD} ^{W184A, M185A} or SP1	96
III Backbone assignment of protein CA_{CTD}^{W184A, M185A}-SP1-NC	98
1. Expression and purification of ¹⁵ N labeled CA _{CTD} ^{W184A, M185A} -SP1-NC	98
2. Backbone assignment of the protein ¹⁵ N CA _{CTD} ^{W184A, M185A} -SP1-NC.....	99
IV Interaction between EP39 and protein CA_{CTD}^{W184A, M185A}-SP1-NC	104
1. EP39 binds mainly to the SP1 domain on protein CA _{CTD} ^{W184A, M185A} -SP1-NC	104
2. Assignment of the protein CA _{CTD} ^{W184A, M185A} -SP1-NC in the presence of EP39....	106
3. CSP and intensity ratio	108
4. Attempt to evaluate the K _d of the interaction.....	109
V Structure of CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39	111
VI Relaxation studies of CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39.	116
VII Interaction between PF46396/CB310F3 and CA_{CTD}^{W184A, M185A}-SP1-NC.....	120
1. PF46396.....	120
2. CB310F3.....	124
Conclusions and discussion	127
I Effect of EP39 on the structure and dynamics of CA_{CTD}^{W184A, M185A}-SP1-NC.....	128
1. Structure of protein CA _{CTD} ^{W184A, M185A} -SP1-NC in the presence of EP39	128
2. EP39 disturbs the natural helix-coil equilibrium on each side of the SP1 domain..	129
3. The most dynamic part in SP1, QVT motif, is the first target of EP39.....	129
4. MHR, especially CA-K158, could be an important target for EP39 binding.....	130
5. EP39 stabilizes the N-terminal end of NC and causes the formation of a helical structure within the SP1-NC junction.....	131
6. The dynamics of SP1 domain impedes the precise positioning of EP39 on protein CA _{CTD} ^{W184A, M185A} -SP1-NC.	131

II Interaction between CA _{CTD} ^{W184A, M185A} -SP1-NC and PF46396.....	133
1. PF46396 interacts with CA _{CTD} ^{W184A, M185A} -SP1-NC mainly at CA-SP1 junction and N-terminal of NC.....	133
2. PF46396 is less potent than EP39.....	133
3. MHR is an important domain for PF46396 binding.....	134
4. The N-terminal end of NC is stabilized in the presence of PF46396	135
III CB310F3 does not have significant effect on protein CA _{CTD} ^{W184A, M185A} -SP1-NC in NMR experiments.	136
IV Perspectives.....	136
References	139
Appendices	156
I Assigned chemical shifts	157
II Titration of ¹⁵ N-CA _{CTD} ^{W184A, M185A} -SP1-NC by EP39.....	175
III ¹⁵ N relaxation data for CA _{CTD} ^{W184A, M185A} -SP1-NC	179
IV Published scientific articles.....	182

Abbreviations

aa:	Amino Acid
ARIA:	Ambiguous Restraints for Iterative Assignment
ASP:	Antisense Protein
AZT:	Zidovudine
BA:	Betulinic Acid
BME:	2-Mercaptoethanol
BVM:	Bevirimat
CA:	Capsid
CD:	Circular Dichroism
CNS:	Crystallography & NMR System
CO:	Carbonyl group
Cryo-EM:	Cryogenic Electron Microscopy
Cryo-ET:	Electron cryotomography
CTD:	C-terminal domain
CV:	Column Volume
CypA:	Cyclophilin A
C _α :	Alpha Carbon
DMSO:	Dimethyl Sulfoxide
DPC:	Dodecylphosphocholine
DTG:	Dolutegravir
DTT:	Dithiothreitol
ELISA:	Enzyme-Linked Immunosorbent Assay
env:	envelop
ESCRT-I:	Endosomal Sorting Complex Required For Transport I
EVG:	Elvitegravir
FDA:	Food and Drug Administration
gp:	Glycoprotein
gRNA:	Genomic RNA
HBR:	Highly Basic Region
HetNOE:	Heteronuclear NOEs
HIV:	Human Immunodeficiency Virus
HMQC:	Heteronuclear Multiple Quantum Coherence
HSQC:	Heteronuclear Single Quantum Coherence
IN:	Integrase
INIs:	Integrase Inhibitors
IPTG:	Isopropyl β-d-1-thiogalactopyranoside
LB:	Luria Broth
LTR:	Long Terminal Repeats
MA:	Matrix
MAS:	Magic Angle Spinning
MD:	Molecular Dynamics
MHR:	Major Homology Region
MI:	Maturation Inhibitors
mRNA:	Messenger RNA

MVC:	Maraviroc
NC:	Nucleocapsid
Nef:	Negative factor
NFAT:	Nuclear Factor of Activated T-cells
NF- κ B:	Nuclear Factor - Kappa B
NMR:	Nuclear Magnetic Resonance
NNRTIs:	Non-Nucleoside Reverse Transcriptase Inhibitors
NOE:	Nuclear Overhauser Effect
NOESY:	Nuclear Overhauser Spectroscopy
NRTIs:	Nucleoside/Nucleotide Reverse Transcriptase Inhibitors
NTD:	N-Terminal Domain
PBS:	Primer Binding Site
PDB:	Protein Data Bank
PEI:	Polyethyleneimine
PI:	Phosphatidylinositol
PICs:	Pre-Integration Complexes
PIs:	Protease Inhibitors
PM:	Plasma Membrane
PMSF:	Phenylmethylsulfonyluoride
pol:	Polymerase
PR:	Protease
RAL:	Raltegravir
Rev:	Regulator of Expression of Virion Proteins
RNP:	Ribonucleoprotein
RT:	Reverse Transcriptase
RTCs:	Reverse-Transcription Complexes
SE:	Sedimentation Equilibrium
SL:	Stem Loop
SOFAST:	Selective Optimized Flip-Angle Short-Transient
SP:	Spacer Peptide
ssNMR:	solid-state NMR
Tat:	Trans-Activator of Transcription
TFE:	Trifluoroethanol
TM:	Transmembrane
TSG101:	Tumor Susceptibility Gene 101
Vif:	Virion infectivity factor
VLP:	Virion Like Particles
Vpr:	Viral protein R
Vpu:	Viral Protein U
ZF:	Zinc Finger
$\Delta\delta$:	chemical shifts perturbation

List of Figures and Tables

Figure 1. Organization of the HIV-1 genome (top) and Schematic representation of a single HIV-1 particle (bottom) adapted from [1].

Figure 2. Schematic representation of HIV replication cycle adapted from [2].

Figure 3. Overview of HIV-1 attachment and fusion adapted from [3]. To deliver the viral genome into cells, HIV-1 gp120 subunit first attaches to the host cell by binding CD4. This causes conformational changes of gp120, which allows coreceptor binding mediated in part by the variable loop 3 (V3) of gp120. This initiates the membrane fusion process as the fusion peptide of gp41 inserts into the target membrane, followed by six-helix bundle formation and complete membrane fusion.

Figure 4. The structure and functions of Gag adapted from [1] and the protease cleavage during the maturation of HIV particles. High-resolution structures of the major domains of Gag, matrix (MA), capsid (CA), nucleocapsid (NC) and p6, are presented. The main functions of each domain are listed. The Protein Data Bank (PDB) accessions of MA, CA, NC, p6 are 1UPH, 2M8P, 1ESK and 2C55 respectively. The fastest to slowest cleavage rate of Gag by protease is SP1-NC (1) > MA-CA (2) / SP2-p6 (2) > CA-SP1 (3) / NC-SP2 (3).

Figure 5. HIV-1 maturation adapted from [1]. During virus release, the viral protease cleaves the Gag and GagPol polyproteins at several sites which trigger the maturation of virus. These cleavages result in virion morphology changes, including the generation of the conical capsid core. (a) A cryoelectron tomogram (left), an illustration (middle) and a cryoelectron tomography (cryo-ET) reconstruction (right) of the immature HIV-1 virion, with a zoomed-in view of the immature Gag lattice (far-right). In the far-right panel, the CA N-terminal domain (CA_{NTD}) is depicted in blue and the CA C-terminal domain (CA_{CTD}) is in orange. (b) A cryoelectron tomogram (left), an illustration (middle) and a cryo-ET reconstruction (right) of the mature HIV-1 virion, together with a zoomed-in view of the mature CA lattice (far-right; the CA_{NTD} and CA_{CTD} are colored as in part a. (c) The all-atom structure of an HIV-1 capsid core with twelve CA pentamers (green) located at the two sides of the core, in the otherwise hexameric lattice. (d) A cryo-ET reconstruction of a virion produced from cells that treated with the maturation inhibitor bevirimat, which shows disruption of capsid core formation.

Figure 6. Matrix structure and functions. (A), Matrix crystal structures as hexamer and trimer (PDB accession 1HIW); and NMR structure as monomer (PDB accession 1UPH). (B) Membrane-binding model predicted from the structural studies. The myristyl group and residues that contact the phosphatidylinositide (yellow, with red phosphates) are colored green and blue, respectively. The exposed 1'-fatty acids and myristyl groups project from a highly basic surface (Arg and Lys side chains shown in blue) in a manner expected to synergistically promote membrane binding. Adapted from [4]. (C) HIV-1 Gag protein assumes compact conformations in solution, with both ends near each other in 3-dimensional space. Both the N-terminal of MA (cornflower blue) and the C-terminal of NC (pink) are positively charged, but the MA has a much higher affinity for negatively charged phospholipids than for RNA. Adapted from [5].

Figure 7. Structure of CA domain of HIV-1 Gag precursor. (A) The crystal structure of CA in monomeric form. CA_{NTD} in blue, CA_{CTD} in orange. (PDB ID 2M8N [6], dimeric form (code PDB 2M8L [6]. pentameric form, PDB ID 3P05 [7], and hexameric form, PDB ID 3H4E [8]. (B) Model of the mature capsid in conical shape. CA in pentameric (green) and hexameric form constitutes assembly blocks to form the mature capsid.

Figure 8. Arrangement of CA in immature and mature HIV-1 virions, adapted from [9]. (A) Structural arrangement of the CA in immature (left) and mature (right) HIV-1 virions. CA_{NTDS} and CA_{CTDS} of two CA monomers from neighbouring hexamers are colored in green and magenta, respectively. All other CA_{NTDS} and CA_{CTDS} are in blue and in orange, respectively. Some important interfaces involved in the formation of the two structures are shown. Hexagons indicate a hexameric interface in both structures. In the immature lattice (left) homo-dimeric (black ovals) and homo-trimeric (black triangles) interfaces important for connecting neighboring hexamers are formed by helices 1 and 2 of CA_{NTD}. (PDB ID: 4USN) [10]. In the mature virion (right), inter-hexamer interactions are formed by helices 10 and 11 residing at trimeric interfaces (triangles) and by residues from helix 9 at dimeric interfaces (ovals). (PDB ID: 5MCX) [11]. (B) CA_{CTD} and the N-terminal seven residues of SP1 (side view, left; top view, right) in the six-helix bundle. Residues of SP1 (light green), CA_{CTD} (orange), MHR loop (yellow), loop connecting helices 9 and 10 (cyan), and a β -turn (magenta) are key elements for the formation and stabilization of the CA-SP1 six-helix bundle in the immature Gag lattice. (PDB ID: 5I4T) [12]. The scissor indicates the cleavage site between CA and SP1.

Figure 9. HIV-1 nucleic capsid (NC) protein. Adapted from [13]. **A**, the amino acid sequence of NC. The two zinc fingers are colored green. The N- and C- terminal ends of NC are colored gray and purple respectively. The linker between two zinc fingers is colored orange. **B**, solution NMR structure of NC, PDB ID 1AAF. The color code for NC is same as A. **C**, The structure of NC binding with SL2 and SL3.

Figure 10. HIV-1 p6. (A) The amino acid sequence of p6 HIV-1, NL4-3. The motifs interacted with TSG101, Vpr and ALIX are underlined. (B) The NMR structure of p6 in 50% TFE. PDB ID 2C55.

Figure 11. NMR structure of CA(211-231)-SP1(1-14)-NC(1-13) (A) and crystal structure of CA_{CTD}-SP1 (B).

Figure 12. Antiretroviral drugs against different stages of the viral life cycle adapted from [2].

Figure 13. Schematic of full-length polyprotein Gag; wild type CA_{CTD}-SP1-NC, which contains the C-terminal domain of CA, the spacer peptide SP1 and the nucleocapsid protein NC; CA_{CTD}^{W184A, M185A}-SP1-NC, contains two mutations (W184A, M185A) in the C-terminal of CA to prevent CA dimerization. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

Figure 14. 1D NMR spectrum of CA_{CTD}^{W184A, M185A}-SP1-NC. Different proton regions are presented. The molecular weight of this protein is 17 kD. It is complicated to identify the protons in this spectrum.

Figure 15. 2D SOFAST-HMQC spectrum of HIV-1 Gag CA_{CTD}^{W184A, M185A}-SP1-NC. This spectrum shows the ¹H-¹⁵N correlations of the amide groups. 1 amino acid = 1 amide bond = 1 peak on the spectrum. but Trp side chain N ϵ -H ϵ groups and Asn/Gln side chain N δ -H δ 2/N ϵ -H ϵ 2 groups are also visible. Trp side chain N ϵ -H ϵ peaks are usually in the blue square part of the spectrum. Asn/Gln side chain NH₂ group peaks are generally in the purple square part of the spectrum and logically result in two peaks at the same nitrogen frequency with two different hydrogen frequencies.

Figure 16. Extraction of several ¹H-¹³C planes from the 3D HNCA experiment showing the ¹H-¹⁵N correlation of an amino acid (each band corresponds to 1 peak of the HSQC) to a C_{ai}

(most intense peak of the band) or $C_{\alpha i-1}$ (least intense peak of the band). In 3D HN(CO)CA experiment, the $^1\text{H}_i$ - $^{15}\text{N}_i$ correlation of an amino acid only to a $C_{\alpha i-1}$.

Figure 17. Sequential assignment of amino acids from the backbone of a protein using a 3D experiment ^{15}N NOESY-HSQC. ^1H - ^{15}N of an amino acid i has correlation to the amide protons of the previous and next residues, HN_{i-1} and HN_{i+1} . Besides, the ^1H - ^{15}N also correlate to the alpha proton of the previous residue, $\text{H}_{\alpha i-1}$. Adapted from [14].

Figure 18. Example of the sequential assignment of residue 197-201 of the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ -SP1-NC with the CCPNMR program [15].

Figure 19. Side chain assignment using ^{15}N TOCSY-HSQC. Adapted from [16].

Figure 20. Example of the side chain assignment of the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ -SP1-NC with the CCPNMR program.

Figure 21. The dependence of 2D NMR peak shape on exchange rate with gradual addition of the ligand. Adapted from [17]. Fast exchange (Left): peaks move smoothly from free (blue) to bound (red) state. In the limit of very fast exchange, peaks have the same shape throughout. As they move out of this limit, peaks may become broader when in equilibrium between free and bound, and then sharpen up again close to saturation. Slow exchange (Right): the free state (blue) decreases in intensity as the bound state (red) increases.

Figure 22. Timescales involved in protein dynamics and NMR observable. Spin relaxation experiment covers most of the timescales where the effective protein internal (and overall) motions take place. The fastest movement can be detected by nuclear spin relaxation measurements (R_1 , R_2 , NOE) or by analyzing the dynamic contribution of the residual dipolar couplings (RDC). The line-shape analysis and the relaxation dispersion experiment account for motions involving a larger number of atoms and are most relevant in biology. The slower motions can be characterized in NMR using EXSY and ZZ exchange experiments [18].

Figure 23. Expression of protein CA_{CTD} -SP1-NC and $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ -SP1-NC.

Figure 24. Purification of protein CA_{CTD} -SP1-NC and $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ -SP1-NC with anion exchange chromatography.

Figure 25. Purification of protein CA_{CTD}-SP1-NC and CA_{CTD}^{W184A, M185A}-SP1-NC with cation exchange chromatography.

Figure 26. Purification of protein CA_{CTD}-SP1-NC and CA_{CTD}^{W184A, M185A}-SP1-NC by size exclusion chromatography.

Figure 27. Expression of protein HIV-1 NC.

Figure 28. Purification of protein HIV-1 NC with anion exchange chromatography.

Figure 29. Purification of protein HIV-1 NC with cation exchange chromatography.

Figure 30. Purification of protein HIV-1 NC with size exclusion chromatography.

Figure 31. 1D spectra (left) and natural abundance 2D ¹H-¹⁵N SOFAST-HMQC spectra (right) of proteins HIV-1 CA_{CTD}-SP1-NC (A, B), CA_{CTD}^{W184A, M185A}-SP1-NC (C, D) and NC (E, F).

Figure 32. Natural abundance 2D ¹H-¹⁵N SOFAST-HMQC spectra of CA_{CTD}-SP1-NC (magenta) and CA_{CTD}^{W184A, M185A}-SP1-NC (black).

Figure 33. 2D ¹H-¹⁵N SOFAST-HMQC spectra of CA_{CTD}^{W184A, M185A}-SP1-NC (black) and NC (green).

Figure 34. Expression and purification of ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC.

Figure 35. ¹H-¹⁵N SOFAST-HMQC spectrum of ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC. (A) The full ¹H-¹⁵N SOFAST-HMQC spectrum. (B) Zoom in the black square in A.

Figure 36. Secondary structure prediction of CA_{CTD}^{W184A, M185A}-SP1-NC. (A) the secondary chemical shift of H_α, chemical shift index (CSI), DANGLE predicted secondary structure and the secondary structure of protein CA_{CTD}-SP1 (PDB code 5I4T) is shown. (B) the chart table shows a graphical representation of NOEs observed in the 3D NOESY-HSQC experiment.

Figure 37. Solution NMR study of EP39 binding to CA_{CTD}^{W184A, M185A}-SP1-NC. (A) Comparison of ¹H-¹⁵N SOFAST-HMQC spectra recorded on ¹⁵N labelled CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39 from 1 to 6 equivalents. (B) Resonances corresponding to residue CA-G²²⁰, CA-G²²⁵ and NC-G²⁴⁹ appear after adding 1 equivalent

EP39. Resonances of CA-H²²⁶ appear after adding 4 equivalents of EP39. After 4 eq EP39, there is no more new resonances appears. Residues CA-T²¹⁰, CA-G²²² and CA-L²³¹ show a chemical shift perturbation when EP39 binds to the protein.

Figure 38. Solution NMR of EP39 binding to CA_{CTD}^{W184A, M185A}-SP1-NC. (A) Superposition of ¹H-¹⁵N SOFAST-HMQC spectra recorded on ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC in the absence (red) and in the presence of 6 equivalents of EP39 (blue) at 303K. Resonances are labeled with amino acids numbers. (B) Zoom in the black square in A.

Figure 39. Chemical shift perturbations (CSP) and resonance intensity ratios (I/I₀) of the amide groups (¹H/¹⁵N combined chemical shift) in the presence and in the absence of EP39. The chemical shift perturbations and resonance intensity ratios of the residues whose ¹H-¹⁵N resonance are seen only in the presence of EP39 were set at 0.1 and 8 respectively. The red line indicates the resonance intensity ratio of 1. Proline residues are labeled as P and unassigned residues are labeled as *. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

Figure 40. Variations of chemical shift perturbations (CSP) of ten residues (CA-V²²¹, CA-L²³¹, SP1-A²³², SP1-E²³³, SP1-M²³⁵, SP1-S²³⁶, SP1-Q²³⁷, SP1-V²³⁸, SP1-T²³⁹, NC-T²⁵⁷) isolated in the ¹H-¹⁵N SOFAST-HMQC spectrum as a function of EP39 whose concentration increase from one to six equivalents of protein. Kds were computed as described in Materials and Methods.

Figure 41. NMR structure of CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39. The secondary structure of the protein is indicated at the bottom. H: helix; ZF: zinc finger. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively. (A) the structure which has the lowest energy is represented. From N-terminal to C-terminal, this structure contains six helices: H1 (residue CA¹⁶¹⁻¹⁷²), H2 (residue CA¹⁷⁹⁻¹⁹²), H3 (residue CA¹⁹⁶⁻²⁰⁸), H4 (residue CA²¹²⁻²¹⁸), H5 (residue CA^{227-SP1²³⁹}), H6 (residue SP1^{241-NC²⁵¹}) and two zinc fingers: ZF1 (residue NC²⁶⁰⁻²⁷³) and ZF2 (residue NC²⁸¹⁻²⁹⁴). (B) CA_{CTD}^{W184A, M185A}-SP1-NC was divided into three domains. The ten best structures are superimposed based on each domain. Left: superimposition on the first domain, CA¹⁵¹⁻²²⁰, shows a RMSD of 1.06 Å calculated on the backbone atoms. Middle: Superimposition on the second domain, CA^{221-NC²⁵¹}, performed on H5 shows a RMSD of 0.74 Å. Right: superimpositions on the third domain, NC²⁵²⁻²⁹⁴, performed on ZF1 and ZF2 respectively show RMSD of 0.92 Å and 1.19 Å

respectively. (C) Superimposition on the region CA¹⁴⁸-SP1²³⁸ between the NMR structure (Figure 40A) and chain G (green) in CA_{CTD}-SP1 hexamer (PDB code: 5I4T). H5 in our structure and in chain G of the crystal structure has different orientations as indicated by black arrows. PDB code for CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39 is 6RWG.

Figure 42. Dynamics characterization of CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of 6 eq EP39 from ¹⁵N NMR relaxation data recorded at 600 MHz. T1 (A, B); T2 (C, D); T1/T2 (E, F) and ¹H-¹⁵N HetNOE (G, H). Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively. The boxes in light blue indicate the four α -helices in CA_{CTD} (H1, H2, H3 and H4). The two α -helices in the second domain (CA-SP1 and SP1-NC junctions) are indicated by the light purple boxed (H5 and H6). The two zinc fingers in NC are shown in pink boxes (ZF1 and ZF2). The missing residues are proline or unassigned.

Figure 43. Effect of EP39/PF46396 on protein CA_{CTD}^{W184A, M185A}-SP1-NC detected by ¹H - ¹⁵N SOFAST-HMQC NMR experiment. The chemical structure of EP39 and PF46396 is shown in A and B. (A) Superposition of ¹H - ¹⁵N SOFAST-HMQC NMR spectrum recorded on protein CA_{CTD}^{W184A, M185A}-SP1-NC in the absence (red) and presence of EP39 (blue). (B) Superposition of ¹H - ¹⁵N SOFAST-HMQC NMR spectrum recorded on protein CA_{CTD}^{W184A, M185A}-SP1-NC in the absence (red) and presence of PF46396 (green). (C) Superposition of ¹H - ¹⁵N SOFAST-HMQC NMR spectrum recorded on protein CA_{CTD}^{W184A, M185A}-SP1-NC in the absence (red) and presence of EP39 (blue) and PF46396 (green). (D) Zoom in the brown circle in figure 43C.

Figure 44. Chemical shift perturbation of the amide groups (¹H/¹⁵N combined chemical shift) in the presence and in the absence of EP39 and PF46396. The chemical shift perturbations of the residues whose ¹H-¹⁵N resonance are seen only in the presence of inhibitor were set at 0.1. Proline residues are labeled as P and unassigned residues are labeled as *. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

Figure 45. Effect of CB310F3 on protein CA_{CTD}^{W184A, M185A}-SP1-NC detected by ¹H-¹⁵N SOFAST-HMQC NMR experiment. (A) Superposition of ¹H - ¹⁵N SOFAST-HMQC NMR spectra recorded on protein CA_{CTD}^{W184A, M185A}-SP1-NC in the absence (red) and presence of CB310F3 (green). The chemical structure of CB310F3 is shown. (B) Zoom in the brown circle in figure 45A. The missing residue is labeled with *.

Figure 46. Residue CA-K¹⁵⁸ (green) and CA-I²⁰¹ (red) mapped on NMR structure of CA_{CTD}^{W184A, M185A}-SP1-NC (PDB code 6RWG) (A) and crystal structure of CA_{CTD}-SP1 hexamer (PDB code 5I4T) (B) respectively with top view (up) and side view (down). The residues in CA_{CTD} and SP1 are colored in blue and purple respectively.

Table 1. HIV-1 proteins and their functions are listed below and adapted from [19].

Table 2. Conservation of the QVT motif in HIV-1 subtype B, C, A, D, 01_AE and 02_AG sequences retrieved from the Los Alamos database. Adapted from [20].

Table 3. Chemical structure of BA, BVM and maturation inhibitor candidates listed in order of publication time of related articles.

Table 4. Composition of M9 medium for labeled protein expression

Table 5. NMR experiments used for the assignment of the main chain (top of the table).

Table 6. Statistics for the top 10 NMR structures of monomeric mutant HIV-1 CA_{CTD}^{W184A, M185A}-SP1-NC

Introduction

I The human immunodeficiency virus (HIV)

The human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), is the result of multiple viral cross-species transmissions from nonhuman primates to humans. Simian immunodeficiency virus (SIV) from central chimpanzees and western lowland gorillas from Cameroon has crossed the species barrier on at least four occasions leading to HIV-1 groups M and N and HIV-1 groups O and P in humans. HIV-2 viruses result from at least nine independent transmissions of SIVs infecting sooty mangabeys from West Africa, leading to HIV-2 groups A-I. HIV-1 group M is responsible for the global pandemic.

The first case of AIDS was identified in 1980 in the United States among male homosexuals. This disease causes lung and oral infections, leukemias and death of the patient by completely weakening his immune system. Thus, the virus was named as “human immunodeficiency virus”.

Since the beginning of the epidemic, 76 million people have been infected with HIV virus and about 33 million people have died of HIV/AIDS. At the end of 2019, 38 million people in the world are living with HIV. 67% of these people had access to antiretroviral therapy. 7.1 million of them did not know that they have HIV. which has caused about 33 million of deaths. (<https://www.who.int/data/gho/data/themes/hiv-aids>).

1. Overview of HIV-1 life cycle

The human immunodeficiency virus type I (HIV-1) is one of the most devastating viral pandemics in history and still represents a major public health threat worldwide. HIV-1 belongs to the *retroviridae* family and is a member of the genus of *lentiviruses*. With electron microscopy, the virus particles were shown to be spherical and with a diameter of 100-120 μm [21].

The genetic information of HIV-1 is carried by two identical positive-sense single stranded RNA. HIV-1 genomic RNA (gRNA) is a molecule of about 9200 nucleotides that contains 9 reading frames coding for 15 proteins [22], but recently a 10th reading frame which codes antisense protein (ASP) was discovered. This genome is made up of three major genes: *gag* (group-specific antigen) which codes for polyprotein Gag precursor (Pr55^{Gag}), *pol* (polymerase)

coding for polyprotein GagPol precursor (Pr160^{GagPol}) and *env* (envelop) coding for virus envelop p90 which was glycosylated to glycoprotein gp160 precursor. Gag precursor is cleaved by HIV-1 protease into matrix (MA), capsid (CA), nucleocapsid (NC), p6 and spacer peptides (SP1 and SP2) in the late stage of HIV-1 replication cycle. GagPol precursor is also processed by HIV-1 protease and releases protease (PR), reverse transcriptase (RT) and integrase (IN). gp160 precursor is cleaved by cellular enzymes to give docking glycoprotein gp120 and transmembrane glycoprotein gp41 [19] (Figure 1). This genome also codes for accessory proteins, like Nef, Vif, Vpr and Vpu and for regulatory proteins such as Tat and Rev [23]. The gRNA also has two long terminal repeats (LTR) sequences, involved in the RNA polymerase II transcription (Figure 1) [24]. The antisense transcripts also give rise to the production of proteins whose functions are still unknown. The existence of antisense protein (ASP) expressed from a polyadenylated antisense transcript is displayed in Figure 1.

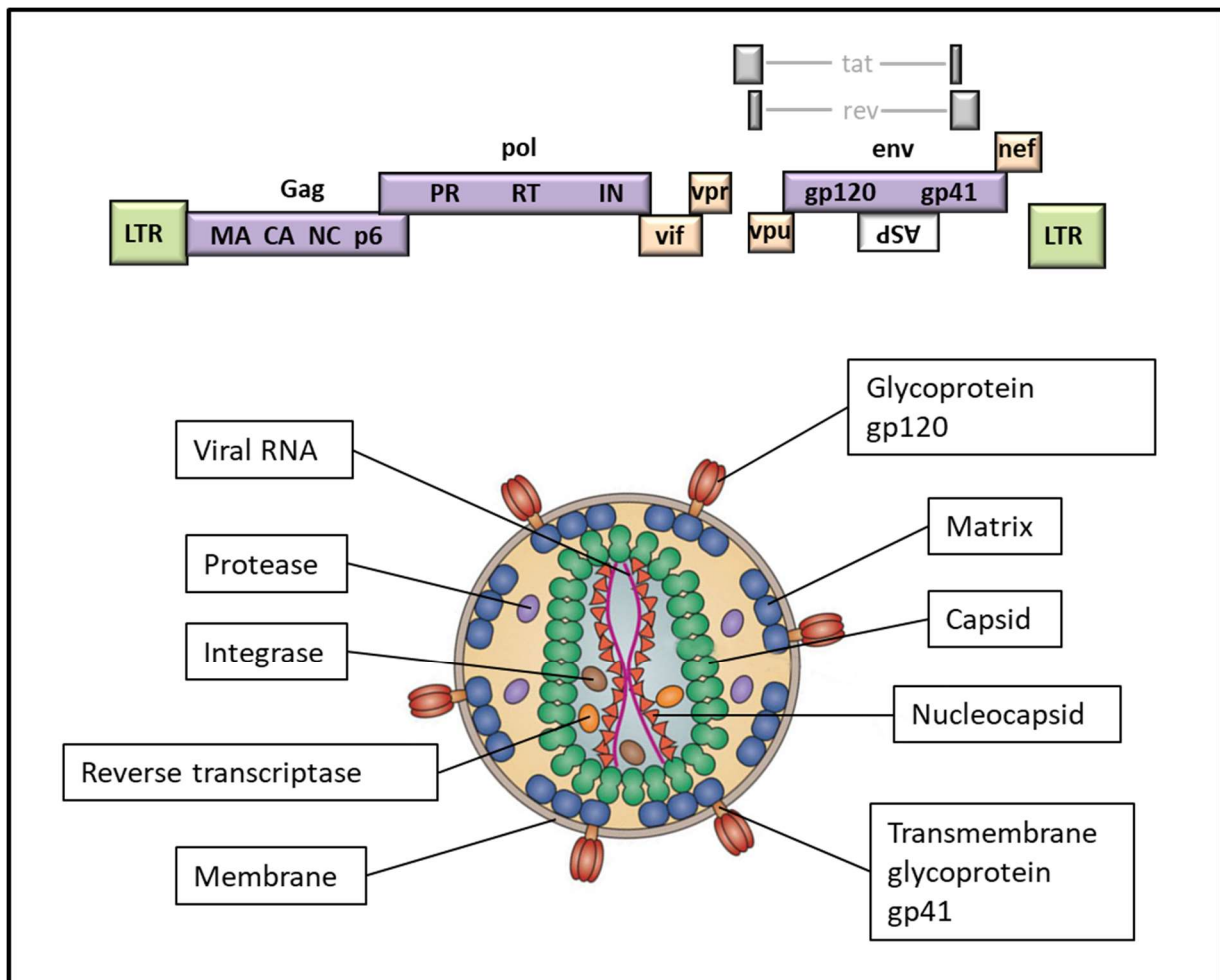


Figure 1. Organization of the HIV-1 genome (top) and Schematic representation of a single HIV-1 particle (bottom) adapted from [1].

The replication cycle of HIV-1 can be divided into two phases: the early and late phases of replication. In the early phase, HIV-1 virus firstly attaches at the surface of the host cell. Then, the envelope of virus fuses with the host cell membrane which induces the virus core enters the cell cytoplasm. Viral RNA is reverse transcribed to double-stranded DNA by reverse transcriptase. When the reverse transcription is finished, the cell nuclear membrane disassembles in dividing cells allowing the proviral DNA transported through the nuclear pore and inserted into host cell genome by integrase. In some long-living cells, the integrated viral genome may remain silent for many years which makes the eradication of viral very difficult. In the late phase of HIV-1 replication, proviral DNA is transcribed into messenger RNA (mRNA) and this mRNA will be translated into viral proteins. Two copies of viral genomic RNA and newly synthesized viral proteins are then recruited at the cell membrane and released from cell forming non-infectious HIV-1 particles. After the cleavage of the polyprotein Gag by HIV-1 protease into several proteins, these proteins reassemble and finally form the mature HIV-1 particles (Figure. 2).

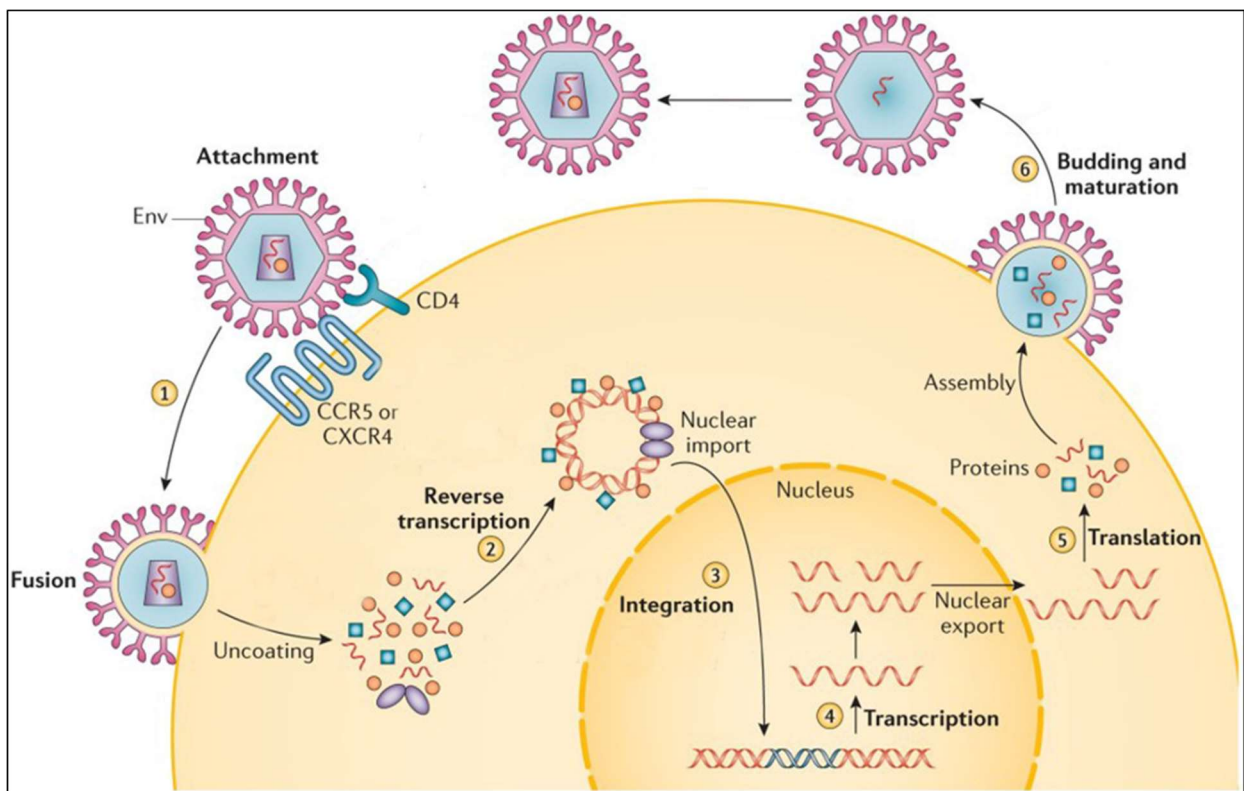


Figure 2. Schematic representation of HIV-1 replication cycle adapted from [2].

Table 1: HIV-1 proteins and their functions are listed below and adapted from [19].

Virus Protein	# Copies / Virion	Interactions with Other Viral Factors	Virus Protein Function
Matrix, MA (p17 ^{Gag})	~5000	Transmembrane glycoprotein (TM)	Plasma membrane targeting of Gag for virion assembly; Env incorporation; post-entry events
Capsid, CA (p24 ^{Gag})	~5000		Virion core structure and assembly
Nucleocapsid, NC (p7 ^{Gag})	~5000	Viral RNA genome (gRNA)	Virion packaging of genome RNA; RNA chaperone; virion assembly
p6 ^{Gag}	~5000	Vpr	Promotes virion budding
Protease, PR	~250	Gag, Pol	Proteolytic processing of Gag and Gag-Pol polyproteins
Reverse Transcriptase, RT	~250	gRNA, IN	cDNA synthesis; RNaseH domain degrades RNA
Integrase, IN	~250	Viral cDNA, RT	Covalent insertion of virus cDNA into cellular DNA
Surface Glycoprotein, SU (gp120 ^{Env})	4 to 35 trimers	TM	Binds cell-surface receptors; mediates virus attachment and entry
Transmembrane Glycoprotein, TM (gp41 ^{Env})	4 to 35 trimers	SU, MA	Contains fusion peptide; mediates membrane fusion and virus entry
Virion Infectivity Factor, Vif	1 to 150		Suppresses APOBEC3G/APOBEC3F, host factors that inhibit infection
Viral Protein R, Vpr	~700	p6	Moderate enhancer of post-entry infectivity; G2/M cell-cycle arrest
trans-Activator of Transcription, Tat	No	Viral RNA via trans-acting response (TAR) element	Potent activator of viral transcription elongation
Regulator of Expression of Virion Proteins, Rev	No	Intron-containing viral RNAs via Rev response element (RRE)	Induces nuclear export of intron-containing viral RNAs
Viral Protein U, Vpu	No		CD4/MHC downregulation; induces virion release from host cell surface
Negative Factor, Nef	Yes, cleaved by PR		CD4/MHC downregulation; T-cell activation; moderate enhancer of viral infectivity; blocks apoptosis; pathogenicity determinant

1.1 Viral Attachment and Fusion

HIV-1 enters the host cell by first binding of gp120 protein to the CD4 receptor which is expressed by CD4⁺ T cell, microglia, dendritic cell and macrophages [25]. Thus, this receptor is the main target for HIV-1. After this step, the viral gp120 subunit undergoes a conformational change which leads to the expose of an epitope. This epitope will allow the gp120 binding to coreceptor. The most essential coreceptors are CCR5 and CXCR4 [26]. There are five variable regions (V1-V5) and four constant regions (C1-C4) in gp120. The amino acid sequence of gp120, especially the sequence of V3 and V1-V2, determines which coreceptor HIV-1 uses [27–29]. After the binding of virus to CD4 and coreceptor, gp120 adopts a conformational change which enables gp41 to reorient parallel to viral and cellular membranes and finally leads to the fusion of two membranes [30] (Figure 3).

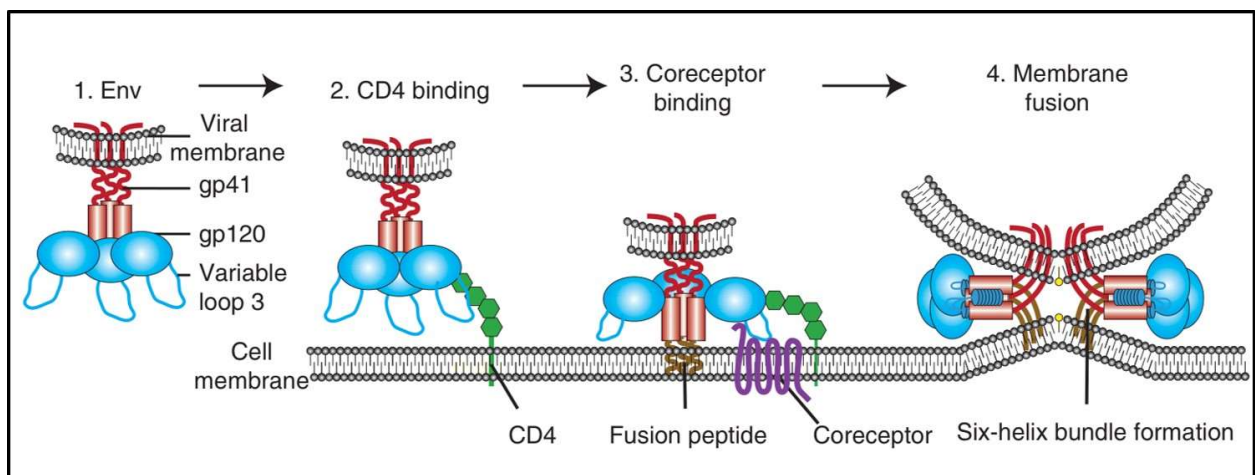


Figure 3. Overview of HIV-1 attachment and fusion adapted from [3]. To deliver the viral genome into cells, HIV-1 gp120 subunit first attaches to the host cell by binding CD4. This causes conformational changes of gp120, which allows coreceptor binding mediated in part by the variable loop 3 (V3) of gp120. This initiates the membrane fusion process when the fusion peptide of gp41 inserts into the target membrane, followed by six-helix bundle formation and complete membrane fusion.

1.2 Reverse Transcription

After fusion, the virus core containing HIV-1 genome is released into cytoplasm of the target cell. This viral core undergoes a partial and progressive disassembly called uncoating which accompanies the transient between reverse-transcription complexes (RTCs) and pre-integration

complexes (PICs) [31]. In the RTCs, HIV-1 reverse transcriptase (RT) transcribes HIV-1 RNA to a double stranded DNA containing a flap [31]. HIV-1 PICs contain PR, RT, IN and Vpr, the presence of structural proteins is controversial [32,33].

1.3 Nuclear Entry

The PICs of most retroviruses can not enter intact nuclei until the breakdown of the nuclear membrane occurring in mitosis [34,35]. While, lentiviruses such as HIV-1 can productively infect non-dividing cells [36]. Four viral components are identified to contribute to the nuclear entry of HIV-1 DNA. IN, MA, Vpr and HIV-1 DNA forming PIC are suspected to have a significant role in the complex process [37] either by direct or indirect ways although the exact function of these constituents are not well defined.

1.4 Integration

After the HIV-1 DNA enters nuclear membrane, HIV-1 must insert its genome to the host genome. With the help of virus-encoded IN, HIV-1 DNA will be inserted into the cellular DNA and the cell will be infected for the remainder of its life. The viral DNA is replicated with the host DNA. Thus, the infection spreading will be achieved either by multiplication of cells that already containing viral DNA or by infecting new cells [38].

1.5 Transcription

In the infected cell, the integrated HIV-1 provirus serves as a template for the transcription of both viral messengers and genomic RNA by host polymerase II (Pol II). The transcription of proviral is triggered by the viral promoter at the 5' LTR. The viral gene expression is strictly dependent on cellular transcription factors, such as Nuclear Factor Kappa-light-chain-enhancer of activated B cells (NF- κ B) and Nuclear Factor of Activated T-cells (NFAT) [39]. The presence of these cellular factors in the nucleus of the infected cell often depends on the state of cellular activation. This is the reason why the viral replication rate is closely related to the activation state of the infected cell. The transcriptional output is low at this initially stage, because the elongation of viral transcripts is inefficient and viral transactivator protein Tat is required for the effective viral gene expression. Tat allows the efficient synthesis of full-length HIV-1 transcripts, and more than 25 different mRNAs in three size classes are generated by alternative splicing: (a), unspliced RNA (9 kb) serving as a genomic RNA or to be translated

to Gag or Gag-Pol precursors; (b), singly spliced RNA (4 kb) which encoding Vpr, Vpu, Vif and Env; (c), fully spliced RNA (2 kb) expressing Rev, Nef and Tat. The spliced and un-spliced mRNAs are transported to the cytoplasm mediated by viral Rev [40].

1.6 Translation and Assembly

In the early phase of translation, fully spliced mRNAs are exported to the cytoplasm and translated to Rev, Tat and Nef. Tat and Rev go to the nucleus. Tat increases the translation rate [41]. In the late phase, Rev transports the transcribed single spliced and the genome RNA from the nucleus to the cytoplasm for translation [42]. Once freed from the perinuclear space, the singly spliced and multiply spliced viral RNAs are committed to translation. Un-spliced RNA serves as both template for Gag / GagPol synthesis and substrate for packaging through the interaction with the nucleocapsid in Gag precursor [42]. The initiation of HIV-1 translation occurs mainly through the cap-dependent scanning mechanism. The 7-methyl-guanosine cap at the 5' end of mRNA is recognized by eukaryotic translation initiation factor eIF4E cap-binding protein and forms a complex with eIF4G and eIF4A. The 40S ribosomal subunit associated with eIF3, eIF2, GTP and Met-tRNA forms the 43S ribosome complex, binds the capped mRNA and scans it until an initiator AUG codon is found [43]. Another mechanism, internal ribosome entry site (IRES) dependent, is also used in HIV-1 translation. IRES recruits the 40S ribosome unit directly. In HIV-1, IRES has been identified in the 5'UTR, upstream of the HIV-1 Gag AUG [43]. The synthesis of Pol protein in HIV-1 is dependent on a ribosomal -1 frameshifting to produce the GagPol fusion protein. This mechanism uses a heptanucleotide slippery sequence upstream of a structured frameshift stimulatory signal forming an RNA pseudoknot. The heptanucleotide interacts with the ribosome, which unwinds the lower part of the structured stimulatory signal. This stable upper part of the structure becomes an effective frameshift stimulatory signal triggering an incomplete translocation of two nucleotides instead of three [44].

Viral assembly is a highly ordered and complex process. The Gag and GagPol multimerized via interaction between Gag proteins. The highly basic region of Gag interacts directly with the negatively charged inositol headgroup of phosphatidylinositol (PI) 4,5-bisphosphate [PI(4,5)P₂] [4,45]. The binding of MA to PI(4,5)P₂ may trigger the exposure of myristate which facilitates its insertion into the inner leaflet of plasma membrane [4]

1.7 Budding

The release of progeny virions from infected cells is called budding. The late p6 domain of Gag contains a highly conserved PTAP motif near its N-terminus [46]. This motif is reported to interact with a cellular protein tumor susceptibility gene 101 (TSG101), which is a part of endosomal sorting complex required for transport I (ESCRT-I) [47]. This ESCRT machinery has an important role in a variety of membrane budding and scission processes in cell.

1.8 Maturation

When Gag is expressed, it is competent to drive the assembly of non-infectious virion like particles (VLP). During or shortly after the budding of HIV-1 particles from the cell surface, the maturation of HIV-1 particles is triggered by the cleavage of Gag or GagPol by HIV-1 protease, which is expressed and brought into virions as part of GagPol. The HIV-1 protease functions as a dimer whose active site locates at the cleft of the dimer interface [48]. The full-length transframe region protease precursor appears to be monomeric yet undergoes maturation via intramolecular cleavage of a transient precursor dimer, concomitant with the appearance of mature-like catalytic activity [49]. After the cleavage, the proteins in Gag reassembly and form the infectious virions. The detail information for maturation will be explained later.

2. HIV-1 particle maturation

For HIV-1, Gag is competent to drive the assembly and release of noninfectious, immature VLPs [50,51]. The infectivity of HIV-1 particle requires the proteolytic activity of HIV-1 protease, which is expressed and brought into virions as part of GagPol. Accompany with virus release, HIV-1 protease cleaves Gag and GagPol polyproteins and triggers the maturation of HIV-1.

HIV-1 protease is an aspartyl protease that functions as a dimer. The active site of the protease locates in a cleft at the dimer interface [52]. HIV-1 protease is brought into virions as part of GagPol precursor and undergoes an autocleavage, concomitant with the appearance of mature-like catalytic activity [49].

The efficiency of HIV-1 protease cleavage of Gag varies considerably, which giving rise to a highly ordered processing cascade (Figure 4). The Gag protein is made up of six domains: matrix (MA), capsid (CA), a spacer peptide 1 (SP1), nucleocapsid (NC), a spacer peptide 2 (SP2), and p6. During this stepwise reaction, the SP1-NC junction is cleaved with highest efficiency. Then, the SP2-p6 and MA-CA junctions are cleaved, which release the free MA and p6 together with p25 (CA-SP1) and NC-SP2 intermediates. The separation of spacer peptides is the final step in HIV-1 maturation. After the cleavage at CA-SP1 and NC-SP1 junction, p25 intermediate is turned into mature p24 capsid protein and NC is released [53][54][55]. A structural and dynamic study by Nuclear magnetic resonance (NMR) on the Gag protein (without the SP2 and p6 domains) has shown that the MA, CA and NC domains are folded within Gag and retain their structures when isolated [56] (Figure 4).

The protease mediated process is accompanied by a major change in virion morphology (Figure 5). In the HIV-1 immature particle, Gag polyprotein is packed in a radial manner (Figure 5a). Following the liberation of individual Gag domains, CA protein reassembles and form the conical capsid core (Figure 5b). The organization of the mature capsid core is well defined. The capsid core assembles with fullerene-like geometry. CA forms predominantly hexameric rings to generate a lattice that is closed at both ends by the inclusion of pentamers. Seven CA pentamers are located at the wide end and five at the narrow end (Figure 5c) [57]. Many studies have contributed to the understanding of the core structure. A combination of cryo-EM and molecular dynamics simulation generated an atomic model for the HIV-1 capsid core [7]. The inter- and intro-hexamer contacts have been defined. Although the immature Gag and mature

CA assemble into hexameric lattices, the intermolecular interfaces in the two structures differ considerably (Figure 5a, b) [10].

In the next round of infection, HIV-1 capsid core provides a protection for viral RNA as it undergoes reverse transcription. The details of the capsid core disassemble, which is known as uncoating, after infection is still being investigated. An emerging view is that at least a remnant of hexagonal CA lattice stay intact as the reverse transcription complex traffics to the cell nucleus. This lattice serves as a signal for the recognition for cell restriction factors and is also recognized by nuclear pore components that play an active part in the nuclear import of viral DNA.

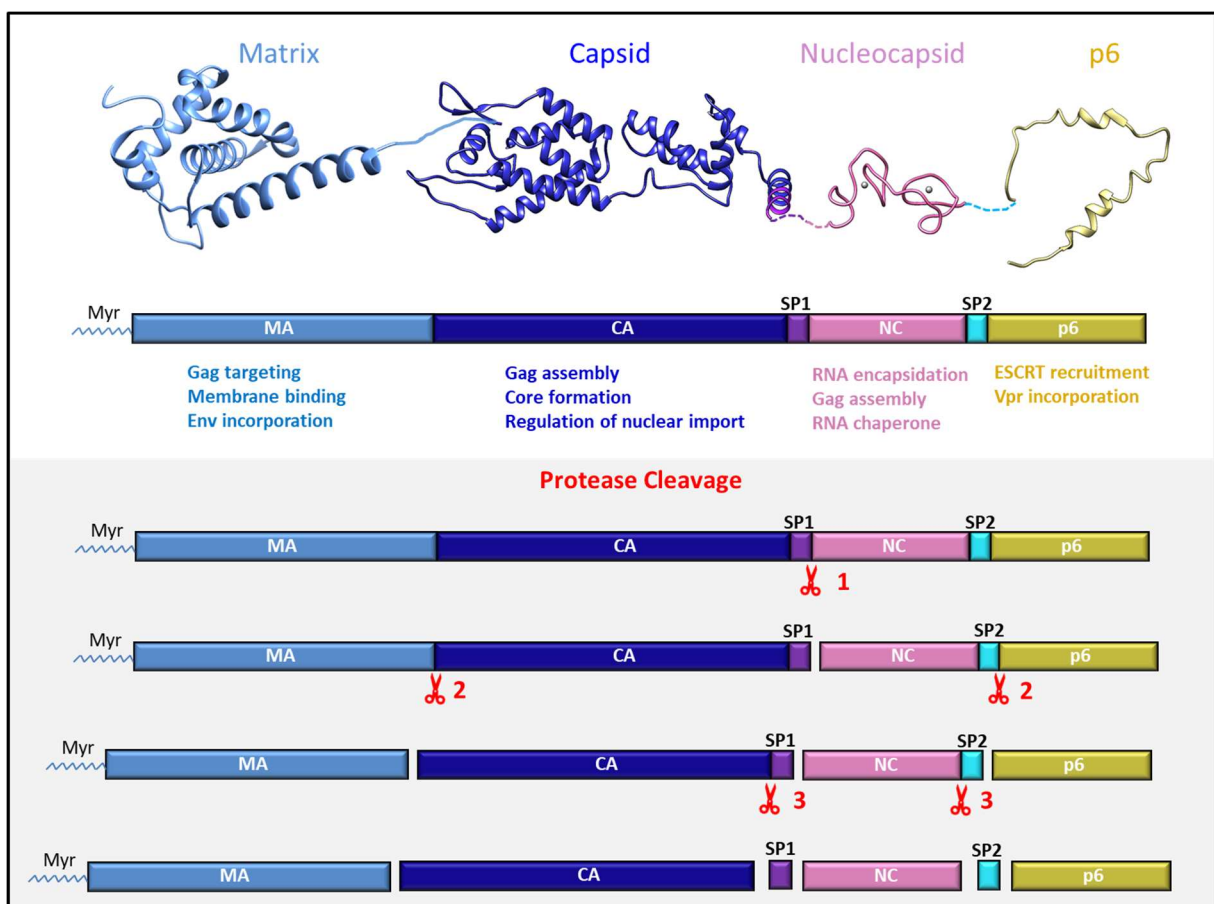


Figure 4. The structure and functions of Gag adapted from [1] and the protease cleavage during the maturation of HIV-1 particles. High-resolution structures of the major domains of Gag, matrix (MA), capsid (CA), nucleocapsid (NC) and p6, are presented. The main functions of each domain are listed. The Protein Data Bank (PDB) accessions of MA, CA, NC, p6 are 1UPH, 2M8P, 1ESK and 2C55 respectively. The fastest to slowest cleavage rate of Gag by protease is SP1-NC (1) > MA-CA (2) / SP2-p6 (2) > CA-SP1 (3) / NC-SP2 (3).

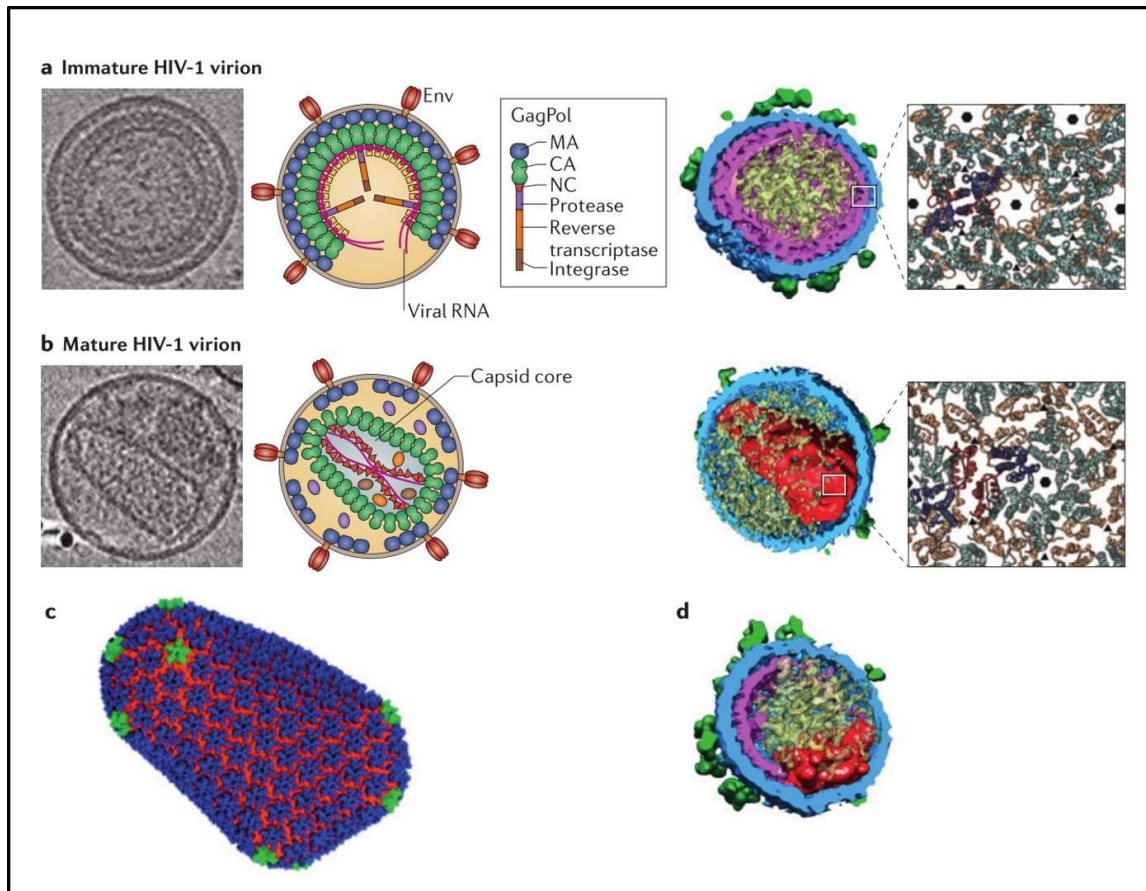


Figure 5. HIV-1 maturation adapted from [1]. During virus release, the viral protease cleaves the Gag and GagPol polyproteins at several sites which trigger the maturation of virus. These cleavages result in virion morphology changes, including the generation of the conical capsid core. (a) A cryoelectron tomogram (left), an illustration (middle) and a cryoelectron tomography (cryo-ET) reconstruction (right) of the immature HIV-1 virion, with a zoomed-in view of the immature Gag lattice (far-right). In the far-right panel, the CA N-terminal domain (CA_{NTD}) is depicted in blue and the CA C-terminal domain (CA_{CTD}) is in orange. (b) A cryoelectron tomogram (left), an illustration (middle) and a cryo-ET reconstruction (right) of the mature HIV-1 virion, together with a zoomed-in view of the mature CA lattice (far-right; the CA_{NTD} and CA_{CTD} are colored as in part a. (c) The all-atom structure of an HIV-1 capsid core with twelve CA pentamers (green) located at the two sides of the core, in the otherwise hexameric lattice. (d) A cryo-ET reconstruction of a virion produced from cells that treated with the maturation inhibitor bevirimat, which shows disruption of capsid core formation.

II HIV-1 Gag precursor

In retroviruses, the 55 kDa Gag precursor is the only protein required for the formation and structural organization of the viral particle. The Gag protein is composed of four functionally distinct domains (MA, CA, NC, p6) as well as two spacer peptides (SP1 and SP2). Each domain of Gag is involved in the assembly and release process of HIV-1 particles. The MA domain of Gag directs Pr55Gag to the plasma membrane of the host cell where it anchors Gag to the inner leaflet of the lipid bilayer [58]. The CA domain of Gag, together with SP1 and NC, homooligomerizes to drive the multimerization of Gag and promote particle assembly. The NC domain of Gag binds to the viral genomic RNA, leading to the packaging of two copies of viral genomic RNC into each virus particle [59][60]. The p6 domain of Gag interact with cellular ESCRT machinery to promote the fission and release of virions from the cell surface [61].

In the immature virus particle, about 5000 Gag molecules, together with GagPol at lower levels relative to Gag (~5%) due to an infrequent ribosomal frameshift event, are tightly packed to form a roughly spherical protein shell [62][63]. High resolution cryoelectron microscopy analyses have demonstrated that rod-shaped Gag molecules are packed within immature particles side-by-side in a radial fashion, with the N-terminal MA domain associated with the membrane and the C-terminal portion of Gag oriented toward the center of the particle [63].

The synthesis of HIV-1 GagPol precursor results from a -1 frameshifting event occurs at a frequency of 5 to 10 % during translation of unspliced Gag or GagPol mRNA [64]. Thus, the intracellular Gag/GagPol ratio of around 20:1 is found during the replication of retroviruses. The multimerization of Gag gives rise to viral particles, while expression of the GagPol fusion protein ensures viral enzymes are incorporated into viral particles during viral assembly. During and after release of virions from cells, PR is released from GagPol by autocleavage [49]. Gag precursor is cleaved by PR into mature proteins: MA, CA, NA, p6 and two spacer peptides, SP1 and SP2. GagPol is cleaved to yield MA, CA, SP1, NC, as well as transframe protein, PR, RT and IN. Gag alone is sufficient for assembly and release of VLP [1]. Incorporation of GagPol or its mature products is required for infectivity, as they mediate the synthesis and integration of viral cDNA in infected cells [65]. The morphological maturation of the virion core also requires the cleavage of precursor proteins by PR [1].

1. Matrix

The MA domain of HIV-1 Gag precursor is a protein of 132 amino acids (aa) and consists of five major α -helices capped by three-stranded mixed β -sheet [66]. This protein is mainly responsible for the recruitment of Gag and Gag-Pol precursors to host cell plasma membrane (PM) [24].

The N-terminal domain of MA is positively charged and undergoes a post-translational modification, myristoylation. This myristoylation involves recognition of a specific N-terminal, six amino acid signal by a cytoplasmic enzyme, resulting in co-translational addition of myristate, a 14-carbon saturated fatty acid, to the nascent polypeptide N-terminus [67]. This modification is required for targeting Gag to PM.

The attachment of MA to the membrane is carried out by (Figure 6B): (1) the hydrophobic interaction, which is mediated by the insertion of the myristoyl group of MA into the lipid bilayer of PM; (2) the electrostatic interaction between a basic region conserved in the N-terminal of the MA, called Highly Basic Region (HBR) (Figure 6A, right), and the PI(4,5)P₂ [58,68–72]. *In vivo*, these phospholipids are necessary for the specific targeting of Gag proteins to the membrane. The myristoyl group has two conformations, a sequestered state and an exposed state [71,73]. The sequestered state is described by the insertion of myristic acid within a globular domain formed by the helices of the MA. The exposed state is favored by the interaction of Gag with PI(4,5)P₂. Indeed, one of the aliphatic chains of PI(4,5)P₂ is sequestered in the same hydrophobic pocket as that of the myristoyl group, this promoting the change in conformation of myristic acid and causing its exposure for its attachment to the membrane [70]. Studies in solution or by X-ray crystallography show that the MA can associate in trimer [66,68] (Figure 6A, middle). *In vitro* studies show a hexamer association of MA in the presence of phospholipids, Figure 6A left [74]. Indeed, the oligomerization of Gag favors the exposure of myristoyl group, which in turn promotes the interaction of MA with PI(4,5)P₂ and therefore attachment of Gag to the membrane. This conformational change is supposed to prevent non-specific interactions of Gag to cell membranes [68]. Indeed, the formation of correct sized VLPs are formed *in vitro* only in the presence of a compound resembling the polar heads of membrane phospholipids [75].

MA has also been described as able to interact with nucleic acids. This has been shown *in vitro* [76] (Figure 6C). In the presence of nucleic acids and lipids, MA prefers the lipids of the

membrane and the NC fixes the nucleic acids [5,77,78]. The MA-RNA interaction is a weak secondary interaction [56] and would allow a specific recognition of membranes made up of PI(4,5)P₂ [69,79]. This interaction may not be specific for an RNA sequence or structure [80,81]. In fact, it has been shown by CLIP-seq studies *in vivo* that the HBR domain could regulate the binding of Gag to cell membranes by binding to a set of tRNAs in the cytosol, mainly by electrostatic interactions [82]. Almost 90% of the MA-RNA complexes observed are made up of tRNA [82]. The latter are very abundant within a cell. Retroviruses use cellular tRNA as a primer to initiate reverse transcription. tRNA^{Lys3} is the natural primer for HIV-1 and related lentiviruses [83]. In the case of HIV-1, MA selectively and predominantly fixes tRNA^{Lys}, tRNA^{Glu}, tRNA^{Gly} and tRNA^{Val} [82]. Deletion and substitution of lysines from the HBR region inhibits the interaction between Gag and tRNA. Thus, the HIV-1 Gag protein would use cellular tRNAs to regulate membrane binding, possibly to prevent unproductive intracellular assembly [82].

Furthermore, it has been shown *in vitro* that the affinity of MA for the tRNA^{Lys3} is greater than that for liposomes enriched in PI(4,5)P₂ [84]. NMR studies indicate that binding of the tRNA^{Lys3} blocks the association of MA with liposomes, including those enriched in PI(4,5)P₂ [84]. However, the affinity of MA for the tRNA^{Lys3} is reduced when the myristoyl group is under exposure state. Since Gag-Gag interactions are known to favor the exposure of the myristoyl group, this leads to an assembly model of the virus in which the targeting of the membrane by Gag and its binding to genome are two coupled steps[84].

Another function of MA is to interact with the Env glycoprotein for their incorporation into the membrane of the future viral particle [1,85]. MA in the trimeric form interacts with the cytoplasmic domain of transmembrane glycoproteins (TM-gp41) in order to guide them towards the membrane [86,87]. Several mechanisms regarding the incorporation of Env proteins have been described [1,86,88]. However, the interaction between Gag and TM-gp41 is not yet well defined.

In addition to the roles of MA domain of Gag precursor in the late stage of virus life cycle, mature MA protein might also be involved in the nuclear import of the pre-integration complex (PIC). It has been reported that HIV-1 MA is present in the viral core at low levels [89]. The presence of MA in the viral pre-integration complex functions to transport the viral DNA into the nucleus [32,89]. The association of MA with core structures suggests a role for MA following membrane fusion. The mutations that impair virus replication without affecting

assembly and release or Env incorporation have been described. These mutations caused a defect in the amount of viral DNA synthesized post-infection [90,91], which suggests a defect early in the virus life cycle.

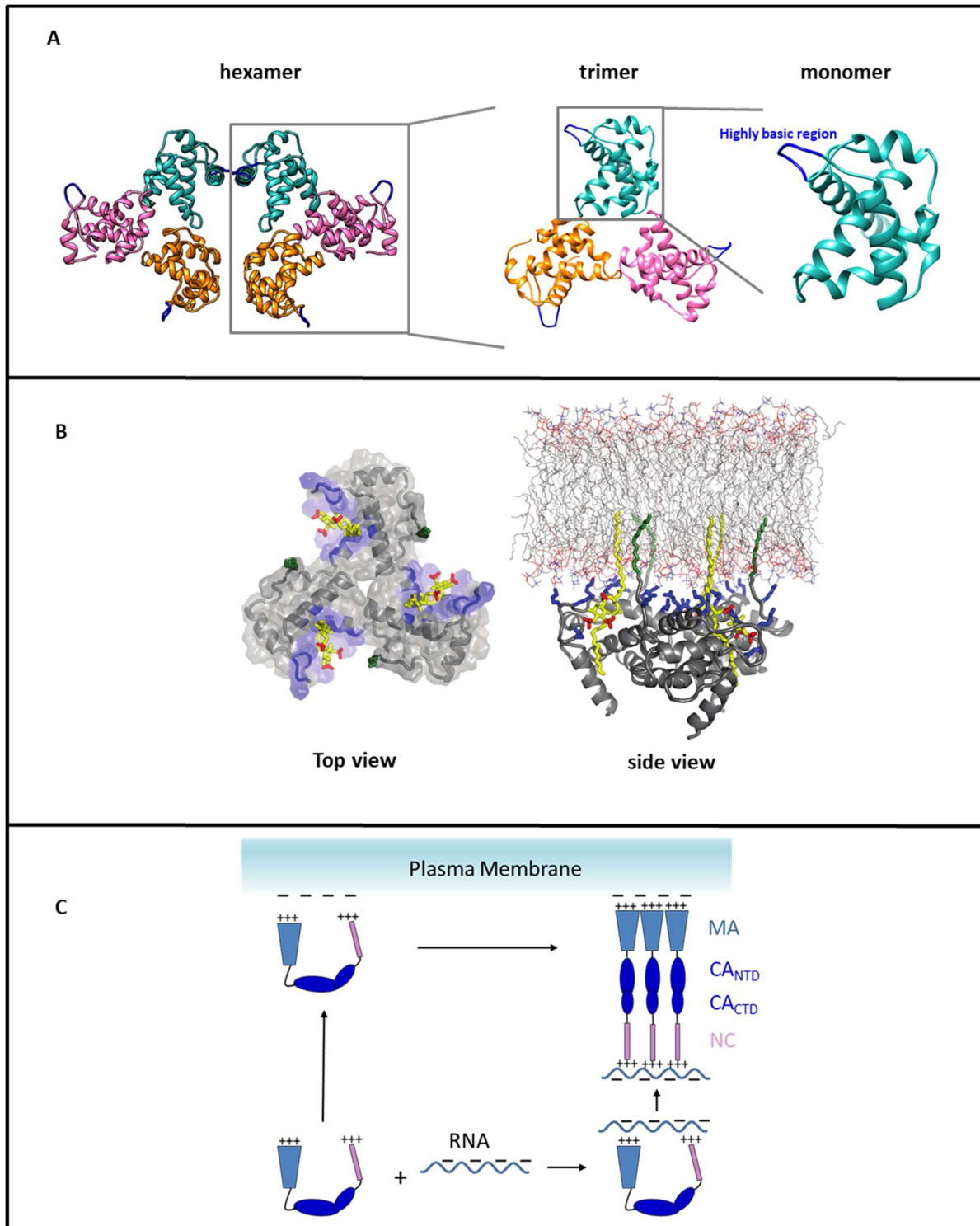


Figure 6. Matrix structure and functions. (A), Matrix crystal structures as hexamer and trimer (PDB accession 1HIW); and NMR structure as monomer (PDB accession 1UPH). (B)

Membrane-binding model predicted from the structural studies. The myristyl group and residues that contact the phosphatidylinositide (yellow, with red phosphates) are colored green and blue, respectively. The exposed 1'-fatty acids and myristyl groups project from a highly basic surface (Arg and Lys side chains shown in blue) in a manner expected to synergistically promote membrane binding. Adapted from [4]. (C) HIV-1 Gag protein assumes compact conformations in solution, with both ends near each other in 3-dimensional space. Both the N-terminal of MA (cornflower blue) and the C-terminal of NC (pink) are positively charged, but the MA has a much higher affinity for negatively charged phospholipids than for RNA. Adapted from [5].

2. Capsid protein

The CA domain of HIV-1 Gag precursor is a 231 aa protein involved in the oligomerization of Gag proteins and in the assembly of immature particles [70,71,78]. The assembly and maturation of virions is highly dependent on the CA and anything that can interfere with CA-CA interactions has a severe effect on viral replication.

2.1 CA monomer structure

CA is divided into two structured and independent sub-domains [8]: (a) the N-terminal domain called CA_{NTD} (residue 1-145) (blue in Figure 7A); (b) the C-terminal domain called CA_{CTD} (residue 150-221) (orange in Figure 7A). The CA_{NTD} domain contains seven α -helices and an extended cyclophilin A (CypA)-binding loop [92,93]. In the immature Gag lattice, the N-terminus of CA_{NTD} is connected with MA and is unstructured. Upon the cleavage at MA-CA junction by protease, the N-terminus of CA_{NTD} is released and folds into a β -hairpin which is stabilized by the salt bridge between the N-terminus (Pro1) and the highly conserved aspartate residue (Asp51) [94]. The flexible "linker" (residue 146-149) that separates the CA_{NTD} and the CA_{CTD} is important for the correct assembly of the mature capsid which takes place during the maturation stage of the virus [95]. The CA_{CTD} contains a short 3_{10} helix, the major homology region (MHR) and four α -helices [96]. The MHR is an 18- to 19-amino-acid motif in CA_{CTD} that exhibits a most conserved sequence. This motif is formed by an extended strand, a β -turn and helix 8 (Figure 7a). The MHR is critical for assembly of HIV-1 and other retroviruses. The MHR deletion, like conserved MHR residue substitution, leads to a dramatic reduction in particle production in human and nonhuman primate cells expressing HIV-1 proviruses [97]. The conservation of MHR residues is important for the stable CA_{CTD} folding [98]. The formation of CA_{CTD} also adopts a significant structural rearrangement during maturation. Before maturation, the C-terminus of CA_{CTD} is connected with SP1. The CA-SP1 junction in the immature virions forms a six-helix bundle. This bundle is a critical structural element in Gag hexamer [99]. With the cleavage at CA-SP1 junction by protease, the C-terminus of CA_{CTD} is liberated and becomes disordered [96]. The structure of CA subunits is highly conserved. However, the arrangement of CA_{NTD} and CA_{CTD} varies significantly in immature lattice, mature lattice and mature core [100,101].

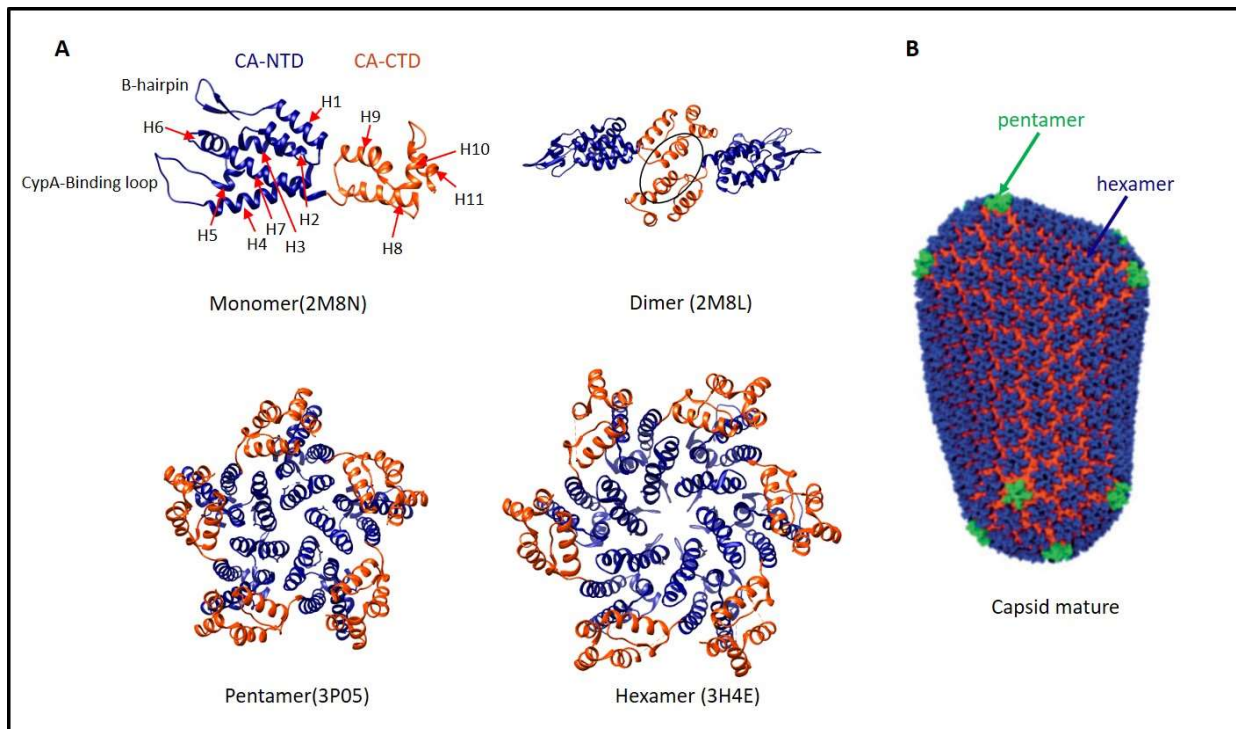


Figure 7. Structure of CA domain of HIV-1 Gag precursor. (A) The crystal structure of CA in monomeric form. CA_{NTD} in blue, CA_{CTD} in orange. (PDB ID 2M8N [6], dimeric form (code PDB 2M8L [6]. pentameric form, PDB ID 3P05 [7], and hexameric form, PDB ID 3H4E [8]. (B) Model of the mature capsid in conical shape. CA in pentameric (green) and hexameric form constitutes assembly blocks to form the mature capsid.

2.2 Role of CA in the assembly of the immature HIV-1 Gag lattice

For the immature viral CA lattice, CA_{NTD} forms a network of interactions between different CA_{NTD}. This network forms CA hexamer and neighbouring hexamers as shown in Figure 8A, left. The inter-hexamer contacts are mediated by the first two helices in CA_{NTD}, which form homo-dimeric and homo-trimeric interfaces (Figure 8A) [10]. Besides, the CA_{CTD} also plays an important role in the assembly of immature virions. MHR and the CA_{CTD} dimeric interface (Figure 7A, circled in black) are essential for generating Gag hexamer. In the immature virions, CA-SP1 junction forms a six-helix bundle as shown in Figure 8B [12,102]. Several residues in MHR (Figure 8B yellow), the loop between helix 9 and 10 (Figure 8B cornflower blue), GVGGP β -turn motif (Figure 8B magenta) and the CA-SP1 junction helix are required to maintain and stabilize the hexamer structure [12].

The CA-SP1 junction is also essential for the maturation of HIV-1. After the PR cleavages at MA-CA and SP1-NC junctions, the CA-SP1 six-helix bundle is destabilized and exposed to the PR (Figure 6B) [10,12]. The dynamics of CA-SP1 junction is a key determinant in the process of maturation [103].

2.3 Role of CA in the formation of the mature capsid

After the protease mediated cleavage of Gag, the CA is assembled into a cone-shaped capsid. The mechanism for the mature capsid formation is still incompletely understood. The structure of mature capsid with 250 CA hexamers and 12 pentamers is well determined in Cryogenic Electron Microscopy (cryo-EM), X-ray crystallography and Electron cryotomography (cryo-ET) [7,8,11,57,104,105] (Figure 7; Figure 8A right). The CA hexamer in mature lattice is linked together by several contacts: several residues at the N-terminus of CA_{CTD} and helix 9 form a dimeric interface; helices 10 and helix 11 are essential structural elements at the trimeric interface (Figure 8A right). The CA pentamers located at the sites of high curvature [11] suggest a relationship between the angle of curvature and the positions of 12 pentamers. The movements between CA_{NTD} and CA_{CTD} and between the dimeric and trimeric interfaces in mature capsid provide twists and tilts necessary for variable curvature of the conical capsid structure [11].

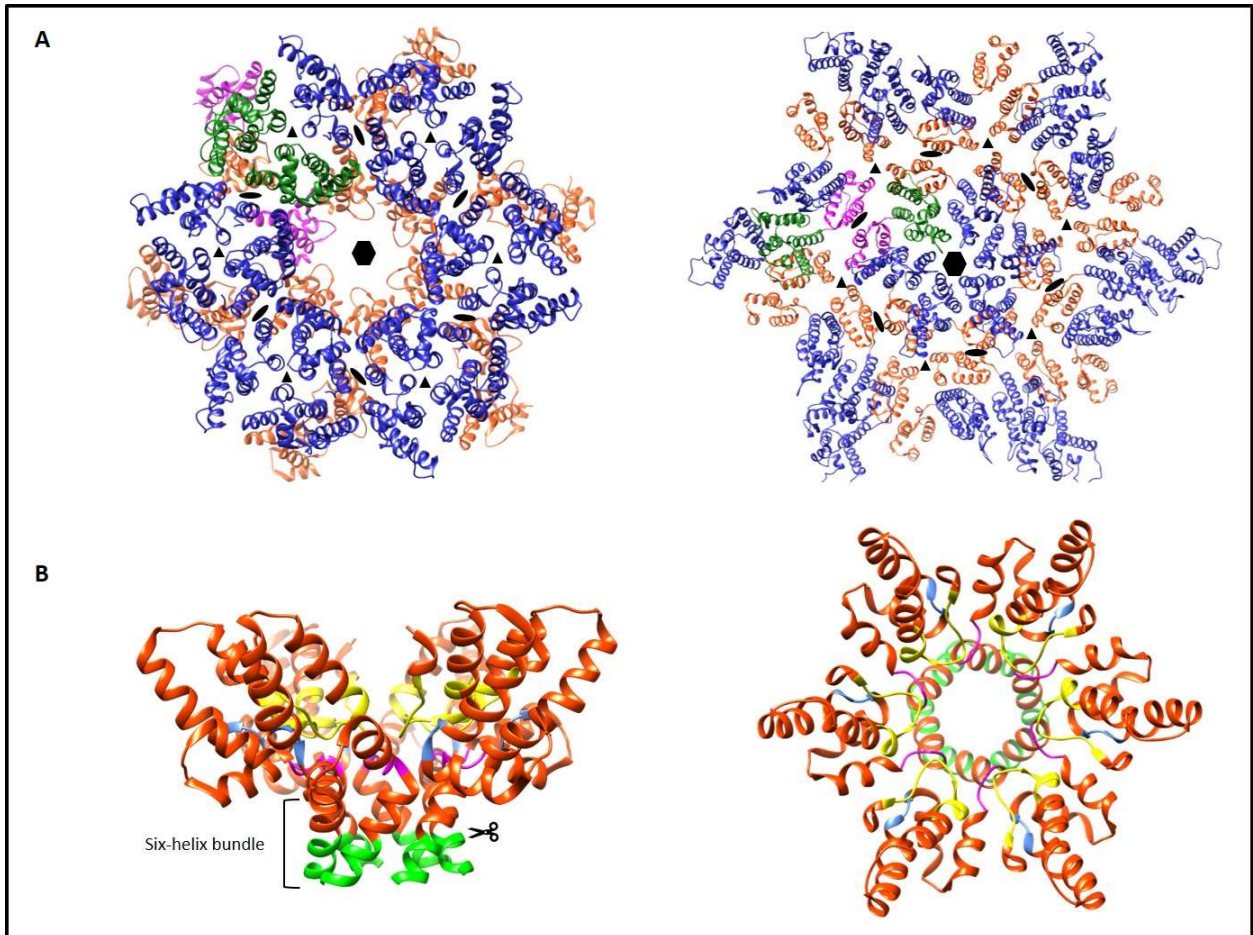


Figure 8. Arrangement of CA in immature and mature HIV-1 virions, adapted from [9]. (A) Structural arrangement of the CA in immature (left) and mature (right) HIV-1 virions. CA_{NTDS} and CA_{CTDS} of two CA monomers from neighbouring hexamers are colored in green and magenta, respectively. All other CA_{NTDS} and CA_{CTDS} are in blue and in orange, respectively. Some important interfaces involved in the formation of the two structures are shown. Hexagons indicate a hexameric interface in both structures. In the immature lattice (left) homo-dimeric (black ovals) and homo-trimeric (black triangles) interfaces important for connecting neighboring hexamers are formed by helices 1 and 2 of CA_{NTD} . (PDB ID: 4USN) [10]. In the mature virion (right), inter-hexamer interactions are formed by helices 10 and 11 residing at trimeric interfaces (triangles) and by residues from helix 9 at dimeric interfaces (ovals). (PDB ID: 5MCX) [11]. (B) CA_{CTD} and the N-terminal seven residues of SP1 (side view, left; top view, right) in the six-helix bundle. Residues of SP1 (light green), CA_{CTD} (orange), MHR loop (yellow), loop connecting helices 9 and 10 (cyan), and a β -turn (magenta) are key elements for the formation and stabilization of the CA-SP1 six-helix bundle in the immature Gag lattice. (PDB ID: 5I4T) [12]. The scissor indicates the cleavage site between CA and SP1.

3. Nucleocapsid

The nucleocapsid (NC) is a small basic protein with 55 aa. NC consists of two domains structured in zinc fingers (ZF) (Figure 9A). The motif of binding to the zinc atom is highly conserved [106,107]: CX₂CX₄HX₄C, X represents the variable amino acids other than cysteine (C) or histidine (H). This CCHC motif can chelate Zn²⁺ with a high affinity [108–110].

3.1 Structure of NC

The structure of the mature form NC was determined by solution NMR (Figure 9B) [111–113]. The structures show that the overall folding of each ZF is very similar [114]. It has been shown that the two HIV-1 ZFs are not functionally equivalent [115] and that their spatial arrangement, in particular the proximity between certain residues of the two ZFs, seems to be an important factor for the biological activity of the protein [116]. In fact, mutants were constructed in which the native sequence of each ZF was maintained, but their positions in the NCp7 protein have been changed. The mutant in which the first ZF (ZF1) has been duplicated (ZF1 / ZF1) is capable of packaging up to 70% of the RNAs and allows replication necessary for acquiring infectivity of virions. While no infectious virus was detected for the mutants for which the second ZF (ZF2) was duplicated (ZF2 / ZF2) or for the case where the position of the two ZFs has been reversed (ZF2 / ZF1) [115].

The NMR experiments demonstrate the existence of weak NOEs (Nuclear Overhauser Effect: experiment allowing to highlight the close protons in space) between protons of two residues belonging to each ZF. This suggests that the two domains are transiently close to each other [109,111,112,117,118]. For example, low intensity NOEs are observed between the side chains of two aromatic residues, F16 belonging to ZF1 and W37 belonging to ZF2 [111].

On the other hand, the rotation correlation times differ between the two ZFs confirming that the interaction between the two ZFs is transient and that they move independently [111]. The dynamics of ZFs differ from each other [119]. The ZF in C-terminal is more flexible than that on the N-terminal side [111,114] and the N- and C-terminal ends also appear very flexible [111]. Within the sequence of the ‘linker’ -RAPRKKG- (orange in Figure 9) separating the two ZFs, the residue P31 plays a critical role in bringing the two zinc fingers together [116,118]. The two ZFs within the NC undergo nanosecond movements (confirming the dynamic nature of the NC) around the ‘linker’ [119]. On the other hand, flexibility is only observed for the G35

residue, the other residues of the ‘linker’ appear rather rigid, and some of them interact with the residues of ZF1 [119]. This observation would explain the nucleic acid binding properties of the two ZFs characterized by asymmetry [119] where each of the ZFs performs a specific function. The ZF2 would be responsible for fixing an unpaired guanine with an RNA, while the ZF1 would fix a second unpaired guanine, if one is available, or would destabilize the structure of the nucleic acid by hydrophobic interactions [119].

The formation and arrangement of the ZFs give the protein a particular folding, which allows the appearance of a hydrophobic plateau on the surface of the ZFs involving the residues, V13, F16, I24, A25, W37, G45 and M46 [113]. Among these residues, F16 in ZF1 and W37 in ZF2 play a crucial role since they are largely involved in the interaction of the protein with nucleic acids [113]. In addition, the basic N-terminal residues of NC are all important for the recognition properties and the chaperone activity of NC on nucleic acids [120].

3.2 Functions of NC

- RNA encapsidation

During virus assembly, HIV-1 encapsidates two copies of viral RNA genome per virion. This encapsidation involves contacts between the HIV-1 Gag precursor and viral RNA during assembly. The process of encapsidation strongly favors viral RNA over cellular RNA and full-length, unspliced viral RNA over spliced viral RNA. Mutations in the highly conserved Cys and His residues of HIV-1 NC zinc-fingers caused major defects in the specific encapsidation of full-length viral RNA and *in vitro* RNA binding activity [121]. Mutations in basic residues flanking the zinc fingers also reduced RNA binding *in vitro* [122].

The primary domain in viral RNA that is responsible for specific RNA packaging, known as the Psi site, locates between the 5’ long terminal repeat and the beginning of the *gag* open reading frame [121]. In HIV-1, this domain folded into four stem-loops (SL1-SL4) with SL3 being particularly well conserved among virus isolates [123]. Although HIV-1 NC binds single-stranded RNA relatively non-specifically, it displays a strong preference for RNAs containing the Psi site. The binding affinity of NC on SL2 and SL3 is identical and on the order of nanomolar. This affinity is lower for SL1 and SL4 [124]. The structures of NC complexed with SL2 (Figure 9C, left) and SL3 (Figure 9C, right) were resolved by NMR [125,126]. The mode of recognition of these two RNAs by NCp7 is very similar.

HIV-1 NC has been reported to facilitate viral RNA dimerization [127] and mediate the conversion of the dimeric RNA genome to a more stable form, a process known as “RNA maturation” [128].

- Virus assembly

Many HIV-1 NC mutations caused defects in virus release suggesting that the NC domain is critical for the assembly of the virion core [129]. The N-terminal basic domain of NC, rather than the zinc fingers, appeared to be particularly important in this regard [130]. Furthermore, NC domain is involved in Gag-Gag multimerization through its association with the genomic RNA [131,132]. The RNA is thought to play a role in scaffolding particle formation via either the gRNA or cellular RNAs by creating a platform onto which the NC can find support for Gag multimerization [131–133].

- Early post-entry steps and reverse transcription

It is established that the NC domain of Gag precursor is important for the RNA encapsidation and virus assembly. However, the mutations in NC can block virus infectivity without affecting particle assembly and release or the levels of full-length genomic RNA in virions. This indicates that the mature NC protein plays a role in the early steps of virus life cycle. The primer for HIV-1 reverse transcription is tRNA^{Lys} that is specifically incorporated into virions during assembly and binds to a sequence known as the primer binding site (PBS) located near the N-terminus of the genome. NC stimulates the binding of tRNA^{Lys} to the PBS, specific initiation of reverse transcription from the bound tRNA^{Lys} and efficient strand transfer during reverse transcription [134–136].

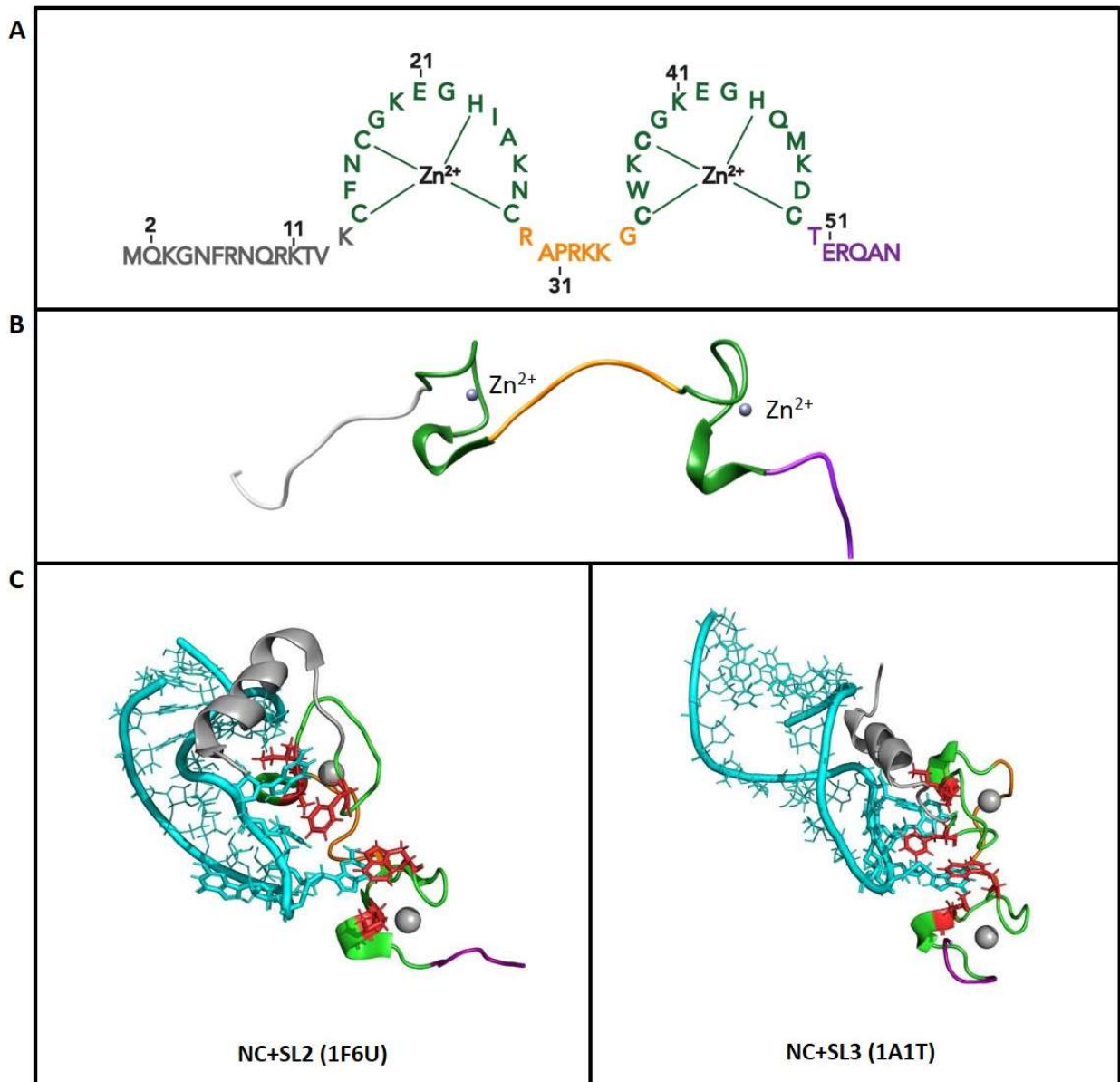


Figure 9. HIV-1 nucleic capsid (NC) protein. Adapted from [13]. **A**, the amino acid sequence of NC. The two zinc fingers are colored green. The N- and C- terminal ends of NC are colored gray and purple respectively. The linker between two zinc fingers is colored orange. **B**, solution NMR structure of NC, PDB ID 1AAF. The color code for NC is same as A. **C**, The structure of NC binding with SL2 and SL3.

4. p6

4.1 structure of p6

p6 is the final domain of Gag with 52 aa, Figure 10A. The solution NMR structure of p6 was determined in 50% 2,2,2-Trifluoroethanol (TFE) with two short helices [137] as shown in Figure 10B. The use of TFE made it possible to mimic the hydrophobic environment of the cytoplasmic membrane where p6 could be found at the time of assembly of the virions [137]. However, the use of TFE can induce (or enhance) the formation of helical structure [138,139]. In this hydrophobic environment, p6 adopts a helix-flexible helix structure (Figure 10B). The two helices appear in N- and C-terminal of the p6 domain and extend from residue 14 to 18 (Helix 1) and residue 33 to 44 (Helix 2) (Figure 10B). The two helices are connected by a flexible region allowing free orientation of the protein. In the presence of Dodecylphosphocholine (DPC) micelles, the p6 domain also has the two helices located at residues 14-18 and 33-44 [140].

In aqueous solution, although the protein p6 is soluble, it does not have any rigid conformation [137,141]. However, the p6 domain exhibits two weak helical domains connected by a flexible region as Figure 10B [140]. This weak helical structure is enforced in hydrophobic conditions [140].

The two helical regions relate to the interaction regions of p6 with the viral protein Vpr (Helix 1 and Helix 2), cell protein TSG 101 (helix 1) and the ALIX (helix 2) [137,140] (Figure 10A).

4.2 Functions of p6

The budding of membrane virions is facilitated by the cellular machinery associated with ESCRT protein complexes hijacked by HIV-1. This system consists of more than 20 proteins organized into several families known as ESCRT-0, ESCRT-I, ESCRT-II and ESCRT-III [1,142–145]. All these complexes and other accessory proteins will allow the remodeling of cellular PM favoring viral budding. HIV-1 exploits only two complexes of this ESCRT machinery. The p6 domain is essential for interaction with the TSG101 proteins ('Tumor Susceptibility Gene 101') belonging to complex I, and with the cellular protein ALIX ('Apoptosis-Linked gene 2 Interacting protein X') of complex III [144–147].

The conserved PTAP (residue 7-10) motif of p6 (Figure 10A) allows recruitment and binding at the level of the structured N-terminal domain called UEV (Inactive Ubiquitin Lase E2 Variant) of TSG101 [144,146]. The recruitment of TSG101 takes place at the same time as the Gag proteins migrate to the PM for assembly [148,149]. The first structural study of this interaction was carried out by NMR showing that the peptide -PTAP- binds in a hydrophobic region of TSG101-UEV [150]. The crystal structure of TSG101-UEV in the presence of the peptide -PEPTAPPEE- has been resolved and differs from the NMR structure [61]. The peptide binds in the same region of TSG101-UEV, but the positioning of the residues differs. The second proline and the threonine of the motif -PTAP- of p6 carry out hydrogen bonds with the main chain of UEV which are not observed in the NMR structure [61]. The mutation of the residues in the motif -PTAP- results in the decrease or inhibition of the binding of the peptide to TSG101-UEV [61]. This explains the importance and the conservation of the motif -PTAP- for the recruitment of TSG101.

The ALIX protein is a central node for the endosomal-lysosomal trafficking and the budding of HIV-1. This protein contains three domains: Bro1, V and PRD [151,152]. ALIX domain V is fixed by hydrophobic and electrostatic interactions to the conserved motif -LYPX_nL- of p6 of HIV-1 (X_n corresponds to any sequence) (Figure 10A) [153]. The crystallographic structure of the Bro1 and V domains of ALIX was resolved in the presence of the peptide -LYPLTSLRSL- corresponding to the interaction motif of p6 (Figure 10A) [154]. The Bro1 domain would interact with Gag NC via either the gRNA [153,155–157] or the membrane [158].

Apart from its role in viral particles budding, the p6 domain also interacts with another partner, Vpr, allowing its incorporation into the virions during the assembly and budding.

The Vpr protein is necessary for the effective infection of non-dividing cells, including macrophages [159,160]. Many cellular processes are also influenced by the presence of Vpr [161,162]. In particular, Vpr blocks the cell cycle in G2 / M phase by activating the endonuclease complex SLX4, regulator of innate immunity [163]. This protein is also involved in improving the fidelity of reverse transcriptase [161]. It also plays a role in the nuclear import of the viral genome [161].

Vpr can be incorporated into virions and allows them maximum infectivity. The N-terminal region of Vpr which is structured in α -helix interacts with the C-terminal region of p6 with the conserved motif -LRSLF- (Figure 10A) [164]. The oligomerization of Vpr seems to be crucial

for its recognition by Gag and its accumulation at the level of the PM [165]. Indeed, the identified functions of the p6 domain are likely to occur near the PM of the cells. The 15-FRFG-18 motif (Figure 10A), resembling the FXFG repeat sequences, commonly found in the proteins of nuclear pore complexes, corresponds to another region allowing the recruitment of Vpr [166]. Vpr can recognize the motif -FxFG- to be packaged in virions and also to associate with nuclear pores for the translocation of the preintegration complex into the nucleus [166].

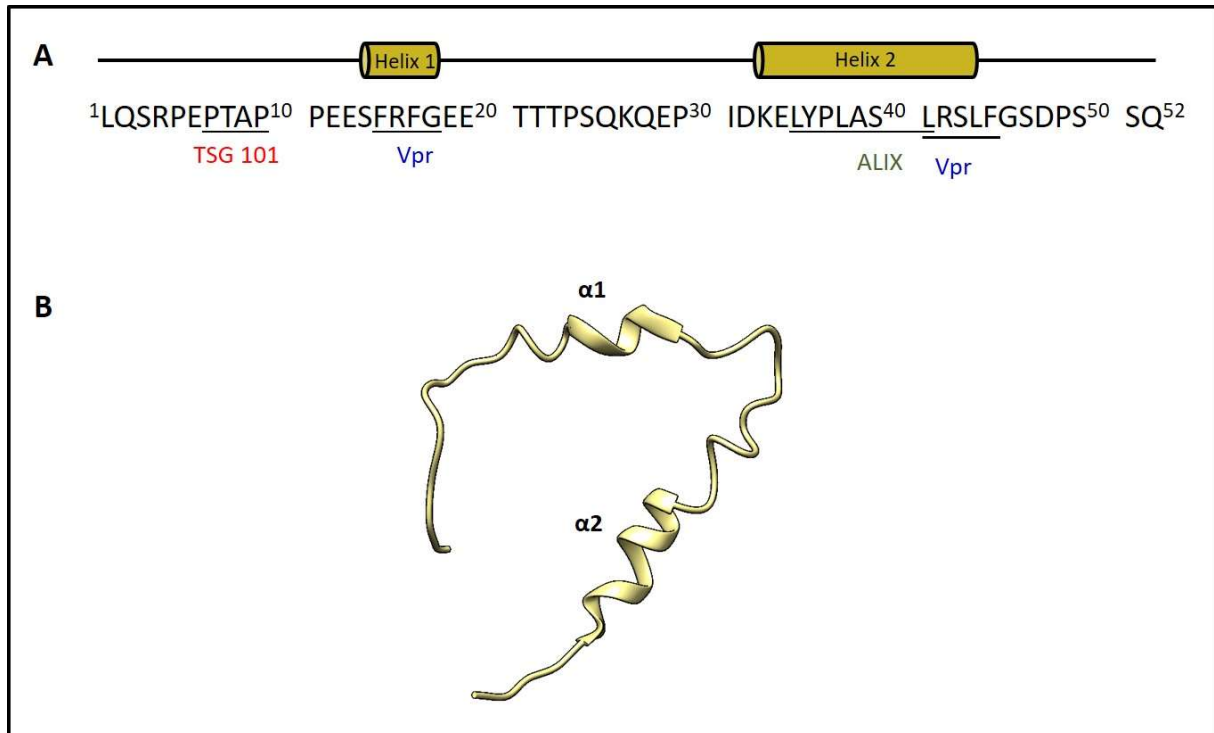


Figure 10. HIV-1 p6. (A) The amino acid sequence of p6 HIV-1, NL4-3. The motifs interacted with TSG101, Vpr and ALIX are underlined. (B) The NMR structure of p6 in 50% TFE. PDB ID 2C55.

5. Spacer peptides (SP1 and SP2)

5.1 Structure of the two spacer peptides

In HIV-1 Gag precursor, CA and NC are separated by SP1, a peptide containing 14 aa. NC and p6 are separated by SP2, a 16 aa peptide. SP1 is very flexible in Gag precursor. It has been shown to exist in a dynamic helix–coil equilibrium [167,168]. When the polarity of the environment is reduced by the addition of DMSO/TFE/ethanol, SP1 domain will form helical structure. For example, in 30% TFE, SP1 formed a long helix encompassing the seven last residues of CA_{CTD}, SP1 and the two residues at the N-terminus of NC [169]. SP2 is disordered in solution [170].

5.2 Functions of the two spacer peptides

SP1 domain is essential for viral assembly and infectivity. The mutations in the first seven residues of SP1 perturb the viral assembly and led to the generation of tubular structures with unprocessed Gag at the PM [171,172]. SP1 also acts as a switch for activation of Gag multimerization. The oligomerization of Gag at the early stage of viral assembly induces a conformational change of CA-SP1 junction from coil state to a six α -helix bundle that favors Gag auto-assembly via SP1-SP1 interactions [68,171–174]. The role of SP1 in Gag multimerization also relates to its membrane binding [175]. Recent structural and functional studies have shown that SP1 helix bundle is stabilized by inositol hexakisphosphate (IP₆). Upon protease cleavage, IP₆ interaction also promotes capsid maturation [176]. Thus, CA-SP1 cleavage site acts as a regulatory switch of HIV-1 particles maturation. Thr-to-Ile substitution at SP1 residue 8 (T8I) results in impaired CA-SP1 processing and stabilizes the immature-like CA-SP1 lattice [177]. In addition, SP1 favors the recognition of gRNA Psi site by the NC domain [178]. The minimal Gag context, SP1 favors specific packaging of gRNA but not of spliced forms of viral RNA [179].

The function of SP2 domain is less known. Mutagenesis studies suggest a possible implication of its proline residues in the processing of Gag polyprotein, the control of gRNA dimer stability and the packaging of Gag-Pol into nascent virions. These mutations also abolish the infectivity of multiple HIV-1 strains in peripheral blood monolayer cells [180]. While the proper processing between NC and p6 appears crucial for viral infectivity and maturation, the SP2 itself seems dispensable, and its deletion has only minor effect on the viral infectivity [181].

6. Structure of C_{ACTD}-SP1-NC domain of Gag

The SP1 domain is a short peptide (14 aa) located between CA and NC. This spacer peptide is essential for the proper assembly and infectivity of HIV-1 virions [173,182,183]. The deletion of SP1 severely reduces the ordered assembly and infectivity of HIV-1 virions.

In the crystal structure of C_{ACTD}-SP1 domain in 1999, C_{ACTD} is well folded with a short 3₁₀ helix followed by an extend strand and four α -helices. However, there is no electron density at C_{ACTD}-SP1 junction and SP1 domain, indicating these residues are disordered or highly dynamic [184]. With NMR relaxation and sedimentation equilibrium (SE), the research group of Michael F. Summers studied the structural property of C_{ACTD}-SP1-NC [167]. The two domains in protein C_{ACTD}-SP1-NC, C_{ACTD} and NC, are folded. The residues of SP1, the adjoining thirteen C-terminal residues of C_{ACTD} and thirteen N-terminal residues of NC are flexible, but have a slight propensity to form α -helix conformation [167]. The NMR structure of a short peptide encompassing the SP1 with the upstream 21 residues in C_{ACTD} and the downstream 13 residues in NC was determined in 30% TFE by Nelly Morellet *et al.* in our group [169]. The main structure is a well-defined α -helix including SP1, the seven last residues of C_{ACTD} and the two first residues of NC [169] (Figure 11A).

With molecular dynamics (MD) simulations and circular dichroism (CD), SP1 is shown to be unstructured in aqueous solution [172]. While, when the polarity of the environment is reduced by addition of dimethyl sulfoxide (DMSO), TFE, or ethanol, the SP1 domain undergoes a concerted change to an α -helix. In 80% DMSO, there is a high helical content between residue SP1-A¹ to NC-M¹ (essentially, the SP1 domain) with a relatively low helical propensity at SP1-N9. Besides, the coil to helix transition is also observed in aqueous solution at high peptide concentration [172].

With Magic Angle Spinning (MAS) NMR spectroscopy, the assembly of CA and CA-SP1 was examined [185]. In assembled CA-SP1, SP1 peptide is in a random coil conformation and dynamically disordered [185]. Bharat *et al.* assembled HIV-1 Gag-derived protein to form tubular arrays and solved the structure of these arrays with cryo-EM and tomography. The arrangement of CA-SP1 junction is approximately six-fold symmetric. However, the resolution of cryo-EM is not high enough to model this region [186]. In 2016, the crystal structure of C_{ACTD}-SP1 was solved [12]. In this structure, a goblet-shaped hexamer conformation is formed with C_{ACTD} comprising the cup and the six-helix bundle of CA-SP1 junction comprising the

stem. The C_{ACTD} adopts canonical fold: a short 3₁₀ helix, a strand element named MHR and then four α -helices. Between the C_{ACTD} and CA-SP1 junction, there is a type II β -turn. The CA-SP1 junction forms an α -helix (spanning the final nine residues of C_{ACTD} and the first seven residues of SP1) (Figure 11B) [12].

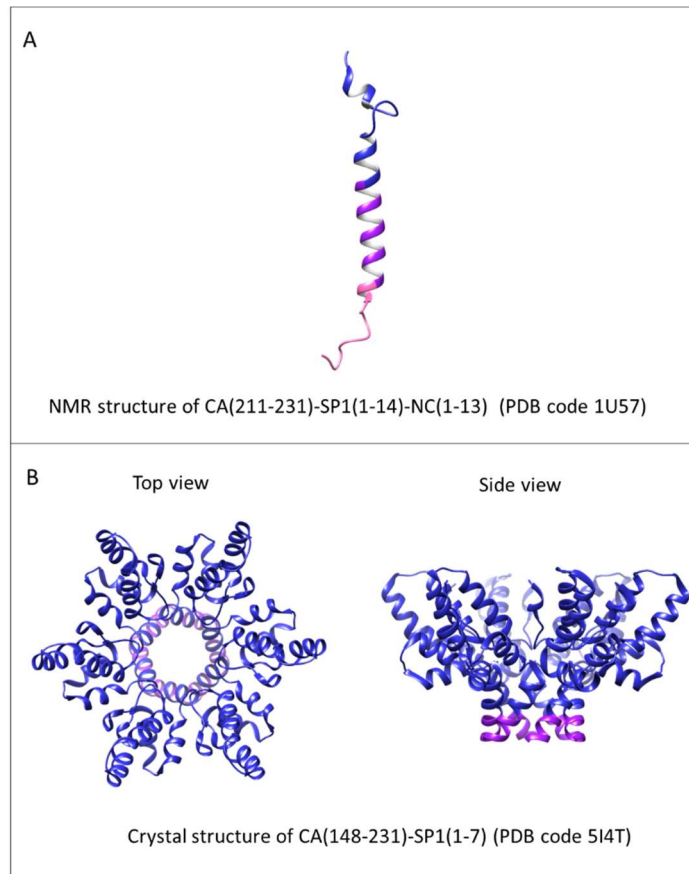


Figure 11. NMR structure of CA(211-231)-SP1(1-14)-NC(1-13) (A) and crystal structure of C_{ACTD}-SP1 (B).

III HIV-1 inhibitors

Since the identification of the first acquired immunodeficiency syndrome (AIDS) patient in 1981, many antiviral drugs have been developed targeting different stages of viral life cycle (Figure 12). The first generation of anti-HIV-1 drugs, such as Zidovudine (AZT) approved by U.S. Food and Drug Administration (FDA) in 1987, are all HIV-1 reverse transcriptase inhibitors (RTIs) [187]. Subsequently, with the efforts by scientists, several additional targets for antiviral therapy were identified, such as HIV-1 protease and integrase [188][189]. In the mid-1990s, the development of HIV-1 protease inhibitors and the introduction of combination antiretroviral therapy (cART) revolutionized the treatment of HIV-1 infection. This cART strongly suppressed viral replication and reduced plasma HIV-1 viral load resulting in a significant reconstitution of the patients immune system [190][191]. The life expectancy of HIV-1 infected patients improved significantly after 1996 [192]. Thus, the successful development of cART has changed HIV-1 infection and AIDS from an inevitable fatal disease into a manageable chronic infection. However, some obstacles were encountered when conducting cART in the early days, for example drug related toxicities and the emergence of drug resistance.

To overcome these problems, some new generation RTIs (*eg.* tenofovir) and protease inhibitors (PIs) (*eg.* darunavir), which are more effective against the existing drug resistant HIV-1 strains, have been developed [193][194]. At the same time, some new drugs focusing on other targets have been developed, such as integrase inhibitors (INIs) (*eg.* dolutegravir) and entry inhibitors. Now, the HIV-1 patients can be treated with a new recommended combination antiretroviral therapy: two “backbone” RTIs and one “key-drug” INI or PI. For the people in the risk of getting HIV, pre-exposure prophylaxis (PrEP) is an efficient way to reduce the risk of getting HIV from sex by about 99% when taken daily. Among people who inject drugs, PrEP reduces the risk by at least 74% when taken daily.

The Joint United Nations Programme on HIV/AIDS (UNAIDS) has committed to the goal of ending the AIDS pandemic by 2030. To achieve this, UNAIDS has developed an ambitious treatment goal named “90-90-90”, to help end the AIDS epidemic: by 2020, 90% of people living with HIV know their HIV status, 90% of all HIV-infected people tested receive long-term antiretroviral therapy and 90% of people on antiretroviral therapy have a permanently suppressed viral load [195]. Now, there are 38 million people living with HIV. 81% of the

people know they are HIV positive. Two out of three people living with HIV are on antiretroviral therapy. Only 59% of HIV positive people have undetectable levels of the virus [195]. Hence, a continuous effort is still needed to establish a way to provide effective antiretroviral drugs around the world.

At present, there are seven classes of HIV inhibitors that have been approved by FDA (Figure 12): nucleoside and nucleotide reverse transcriptase inhibitors (eg. Zidovudine AZT); non-nucleoside reverse transcriptase inhibitors (eg. Nevirapine NVP); protease inhibitors (eg. saquinavir SQV); integrase inhibitors (eg. Raltegravir RAL); fusion inhibitor (Enfuvirtide T-20); CCR5 inhibitor (Maraviroc MVC); attachment inhibitor (ibalizumab-uyyk IBA). Besides these approved drugs, multiple drugs are under investigation.

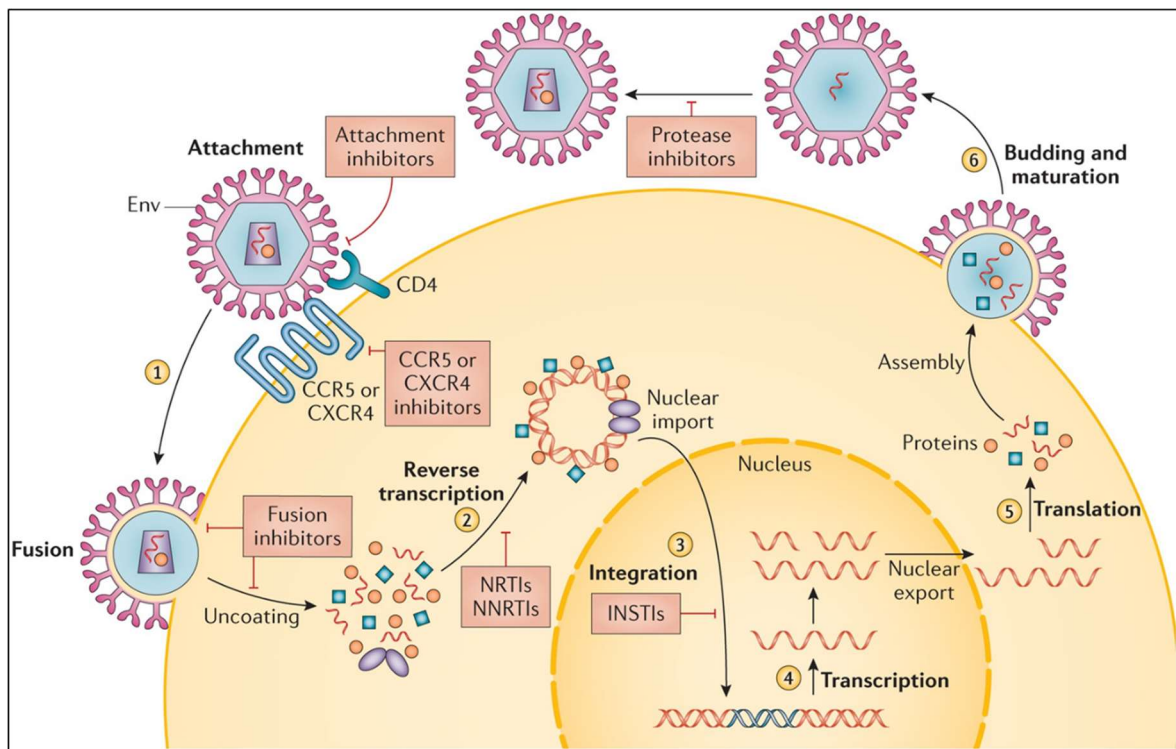


Figure 12. Antiretroviral drugs against different stages of the viral life cycle adapted from [2].

1. Reverse Transcriptase Inhibitors

1.1 Nucleoside/Nucleotide Reverse Transcriptase Inhibitors (NRTIs):

This group of drugs are nucleoside or nucleotide analogs, which are incorporated into the synthesized DNA by RT. NRTIs are pro-drugs that are transformed into tri-phosphorylated compounds by cellular enzymes after entering the cell. The tri-phosphorylated compounds inhibit the transcription of viral RNA into complementary DNA by interference with nucleic acids. Due to the absence of the 3' - hydroxyl group in their structure, unlike normal nucleic acids, they inhibit chain elongation during the formation of viral DNA from RNA. Resistance to these molecules occurs in the form of accumulation of mutations in the reverse transcriptase which prevent their incorporation into the chain of newly synthesized viral DNA. There are six FDA approved NRTIs: zidovudine (AZT, ZDT), lamivudine (3TC), abacavir (ABC), tenofovir disoproxil fumarate (TDF), emtricitabine (FTC) and tenofovir alafenamide (TAF).

1.2 Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

NNRTIs block the reverse transcriptase directly by fixing themselves in a hydrophobic pocket close to the catalytic site of reverse transcriptase. Once the inhibitors are bound to the RT, it impairs the RT's flexibility and results in the inability to synthesize DNA. NNRTIs must be used in combination with other anti-HIV agents because one single mutation in the hydrophobic pocket is enough to create a conformational change in RT resulting in the non-binding of the inhibitor to reverse transcriptase. There are five FDA approved NNRTIs: nevirapine (NVP), efavirenz (EFV), etravirine (ETR), rilpivirine (RPV) and doravirine (DOR).

2. Entry (attachment and fusion) Inhibitors

Entry Inhibitors disrupt the virus replication cycle at the first stage: entry of the viral core into the cytoplasm. Of all the known HIV inhibitors, entry inhibitors are the only inhibitor that target at a host protein. Enfuvirtide (T-20) is the first and only approved fusion inhibitor [196]. Maraviroc (MVC) is an attachment inhibitor. It targets the CCR5 coreceptor and interferes the virus life cycle at the beginning [197]. Ibalizumab-uiyk (IBA) is an inhibitor of HIV-1

attachment, which blocks the binding of HIV-1 envelope molecule, gp120 to CD4 on the cell surface.

3. Protease Inhibitors (PIs)

HIV-1 PR is a virus encoded aspartyl protease. Two pol precursor polyproteins interact and form a transient dimer which trigger the autocleavage and release the mature PR [49]. The HIV-1 protease prefers tyrosine-proline and phenylalanine-proline containing substrates [198]. It has two 99 residues monomers arranged in a homodimer. The active site is the cleft of the dimer interface [52]. PIs are competitive inhibitors of the catalyzed cleavage. They do not block the infection of viral particles but stop the maturation of HIV-1 particles and leads to the production of non-infectious viral particles. There are seven FDA approved PIs: saquinavir (SQV), ritonavir (RTV), lopinavir (LPV), atazanavir (ATV), fosmaprenavir (FOS-APV), tipranavir (TPV) and darunavir (DRV).

4. Integrase Inhibitors (INIs)

Integrase inhibitors prevent the integrase (IN) from integrating the viral DNA into the host genome. IN functions as a tetramer and its N-terminal domain contains a HHCC zinc finger motif that is essential for the multimerization, optimal activity and protein stability. The DDE motif in the core domain of IN forms the catalytic triad. The C-terminal domain of IN binds to DNA in a non-specific manner [199]. Currently, FDA approved INIs include raltegravir (RAL), dolutegravir (DTG) and elvitegravir (EVG). They are all strand transfer inhibitors. RAL interferes with the integration process by binding to the DDE motif in the catalytic domain [200]. There are four FDA approved INIs: raltegravir (RAL), elvitegravir (EVG), dolutegravir (DTG) and bictegravir (BIC).

5. Maturation Inhibitors (MIs)

Until now, there is no maturation inhibitor approved by FDA. But we still give a brief introduction of the status of MIs investigation. More details about MIs will be described later. Unlike PIs targeting the HIV-1 protease, MIs bind to the substrate of protease [201]. The most advanced MI, *bevrimat* (BVM), inhibits the cleavage between CA and SP1, which will impact the virus maturation. However, some natural and BVM induced polymorphisms in Gag inactivate the antiviral activity of BVM [202].

IV HIV-1 maturation inhibitors

1. Bevirimat, the first-generation maturation inhibitor

Betulinic Acid (BA) (Table 3) isolated from the leaves of *Syzygium claviflorum* was firstly reported to be inhibitors of HIV replication in H9 lymphocyte cells in 1994 [203]. Then, several BA derivatives were synthesized to enhance their anti-HIV activities [204]. During these BA derivatives, 3-*O*-(3',3'-dimethylsuccinyl) betulinic acid which was named as YK-FH312/PA-457/DSB/Bevirimat (BVM) (Table 3) in later research has the highest anti-HIV activity. But its mechanism remained unknown.

In 2001, Japanese researchers found that BVM did not show inhibitory effect against HIV-1 replication. However, no virions were detected in the culture of HIV-1-infected cells which were treated with BVM. Besides, with HIV capsid protein (p24) enzyme-linked immunosorbent assay (ELISA), the virus expression in the supernatant of cell culture was inhibited by BVM. While, the viral protein production was not completely inhibited by BVM with Western blot analysis. Thus, they hypothesized that BVM affects the virion assembly and/or budding in HIV replication cycle which is a novel mechanism of an anti-HIV compound [205]. In 2003, BVM was confirmed for the first time disrupting the late step in Gag processing involving conversion of the capsid precursor (p25) to mature capsid protein (p24). Consistent with this result, the resistance mutations caused by BVM locate at the cleavage of p25 to p24 [206]. This novel family of drugs are termed as maturation inhibitors (MI).

Phase I/II, placebo-controlled, monotherapy studies show that BVM is a dose-dependent inhibitor against virus replication [207]. While, approximately 30% of the treatment-naïve patient population harbored an HIV-1 subtype B isolates with at least one mutation at CA-SP1 junction (CA-H226Y, CA-L231M, SP1-Q6H, SP1-V7A/M/del and SP1-T8del). These mutations associate with a reduced susceptibility to bevirimat [208]. In HIV-1 subtype C isolates, the native polymorphism in residue 6 to 8 of Gag SP1 induces resistance to BVM, especially the residue 7 of SP1 which is the primary determinant residue associated drug resistance to BVM [209]. The conservation of the SP1-Q⁶V⁷T⁸ motif in different subtypes of HIV-1 is shown in table 2.

Table 2. Conservation of the QVT motif in HIV-1 subtype B, C, A, D, 01_AE and 02_AG sequences retrieved from the Los Alamos database. Adapted from [20].

Subtype	Prevalence ^a of:			Frequency (%) of QVT motif ^b
	Q	V	T	
C (<i>n</i> = 553)	Q _{96.0} H _{2.0}	V _{11.6} A _{65.8}	T _{3.1} N _{81.4}	0.5
B (<i>n</i> = 265)	Q _{97.4} H _{1.1}	V _{74.3} A _{15.5}	T _{86.8} N _{8.3}	66.0
A (<i>n</i> = 134)	Q _{96.3} H _{3.0}	V _{64.2} A _{27.6}	T _{0.7} Q _{91.0}	0.0
D (<i>n</i> = 76)	Q _{97.4} H/I _{1.3}	V _{13.2} A _{85.5}	T _{73.7} S _{21.1}	9.2
01_AE (<i>n</i> = 72)	Q _{62.5} H _{23.6}	V _{33.3} A _{63.9}	T _{0.0} Q _{98.6}	0.0
02_AG (<i>n</i> = 59)	Q _{100.0}	V _{69.5} A _{27.1}	T _{0.0} Q _{94.9}	0.0

^a The prevalence of Q, V, and T as a percentage is given as the subscript. Also, the most prevalent amino acid (or secondmost prevalent amino acid when Q, V, or T was the most prevalent) is presented. The most prevalent motif per subtype is indicated in boldface type.

^b Frequency of the complete QVT motif.

Point mutations corresponding to commonly observed polymorphisms at SP1 residues 6-8 in HIV-1 clinical isolates confer varying degrees of susceptibility to BVM in the NL4-3 background. SP1-Q6A, -Q6H and -T8A, do not interfere with BVM susceptibility. While, SP1-V7M and T8Δ mutations confer intermediate BVM resistance. The SP1-V7A mutation confers high-level BVM resistance [210].

With cryo-electron tomography technology, the structure of HIV-1 virus particles was examined. The virus particles treated with BVM contain a remnant of immature Gag which has been processed by protease, except CA-SP1. The mutations at CA-SP1 junction (CA-L231I and SP1-M4I) have similar effect on the virus particles by blocking the cleavage between CA and SP1 [211].

The *in vitro* resistances to BVM, CA-H226Y, -L231F, and -L231M, SP1-A1V -A3T, and -A3V, were identified by the team of Eric O. Freed. In these mutations, SP1-A3V and -A3T severely impair virus replication and inhibit virion core condensation [212].

In these years, many research groups try to decipher the mechanism of BVM in blocking the cleavage of the CA-SP1 junction by protease.

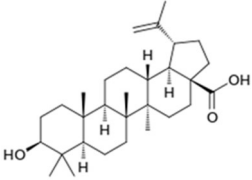
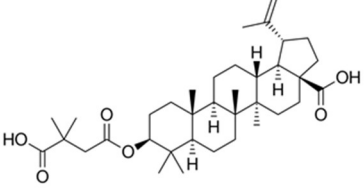
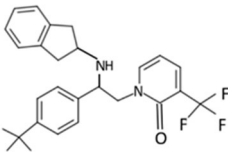
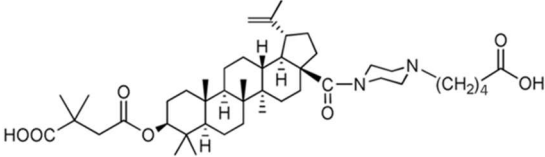
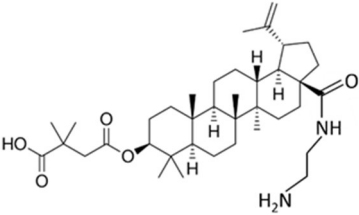
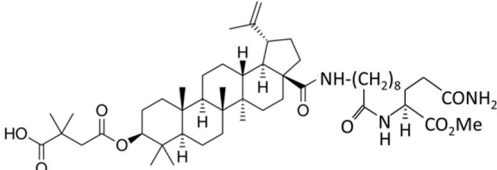
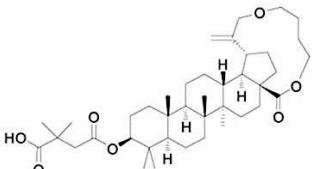
In the Cryo-EM structure of CA(A92E)-SP1, there is no detectable electron density for SP1 which means SP1 is conformationally or dynamically disordered. With molecular dynamics (MD) simulation, Space peptide 1 (SP1) was shown to exist in a dynamic helix-coil equilibrium.

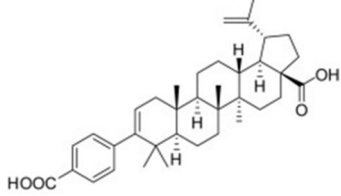
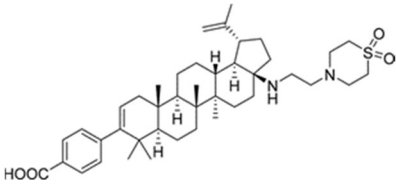
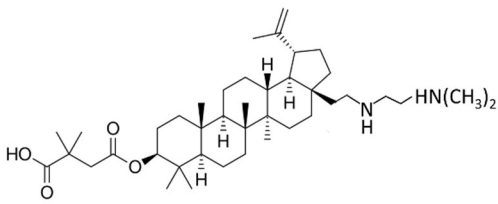
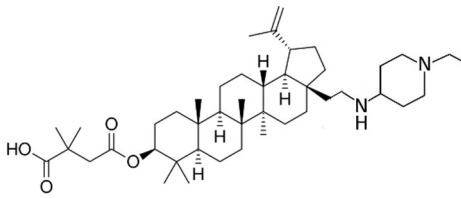
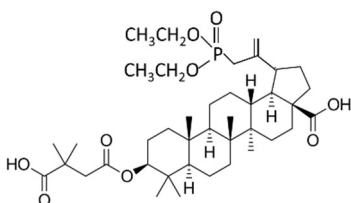
The SP1-T8I mutation modulates the helix-coil equilibrium and induces a helical structure in SP1. Besides, BVM binds to CA_{CTD}-SP1 junction by following the chemical shift perturbation of CA_{CTD}-SP1 with solution NMR [168].

The structure of the complex CA_{CTD}-SP1 and BVM determined by electron diffraction measurements reveals that BVM binds to the center of CA_{CTD}-SP1 hexamer by electric and hydrophobic interaction. CA-K227 interacts electrostatically with the carboxyl group of bound BVM [213].

In the solid-state NMR (ssNMR) measurements on the virus-like particles (VLPs), ¹⁵N and ¹³C chemical shifts of CA-SP1 junction (CA-C218 to SP1-A3) do not have obvious perturbation after BVM binding. Only the ¹⁵N chemical shift of residue SP1-A1 changes significantly. The slight decrease in the ¹⁵N spin relaxation rates reveals the weak rigidification of the six-helix bundle in CA-SP1 junction caused by BVM binding [214].

Table 3. Chemical structure of BA, BVM and maturation inhibitor candidates listed in order of publication time of related articles.

Number	Name	Molecular Structure	Reference	Year
1	Betulinic acid (BA)		T. Fujioka, <i>et al.</i> , [203]	1994
2	Bevirimat (BVM)		Y. Kashiwada, <i>et al.</i> , [204]	1996
3	PF-46396		W. S. Blair, <i>et al.</i> , [215]	2009
4	Compound 41		K. Qian, <i>et al.</i> , [216]	2012
5	EP39		P. Coric, <i>et al.</i> , [217]	2013
6	CB310F3		Z. Dang, <i>et al.</i> , [218]	2013
7	20b		J. Tang <i>et al.</i> , [219]	2014

8	9a		Z. Liu, <i>et al.</i> , [220]	2016
9	GSK-3532795		B. Nowicka-Sans, <i>et al.</i> , [221]	2016
10	7m		E. Urano <i>et al.</i> , [222]	2016
11	7r		E. Urano <i>et al.</i> , [222]	2016
12	14a		E. Chrobak, <i>et al.</i> , [223]	2019

2. Second-generation maturation inhibitors

With the appearance of resistance mutations, many research groups have tried to optimize BVM or screen molecules to find other maturation inhibitor candidates.

PF46396, a pyridone-based compound {1-[2-(4-tert-butylphenyl)-2-(2,3-dihydro-1H-inden-2-ylamino) ethyl] -3-(trifluoromethyl) pyridine -2(1H)-one} (Table 3), was first discovered to be a hit from high-throughput full replication screen that incorporate all the HIV-1 targets required for viral replication in T-cell lines [224]. This molecule was then identified to target the virus maturation. It inhibits the cleavage process between CA and SP1, which block the maturation of HIV-1 particles same like BVM [215]. For HIV-1 clade B, PF-46396 is less efficient than BVM. It induces mutations at CA-SP1 junction, CA-I201 and CA MHR [225][226]. These results suggest that PF-46396 and BVM bind to the same region of HIV-1 Gag. In HIV-1 clade C, PF-46396 can also interfere the cleavage between CA - SP1 and block the virus maturation [227].

Qian *et al.* have synthesized 47 new betulinic acid derivatives. Among them, compound 41 (Table 3) with C-28 piperazine substitution exhibited a much higher anti-HIV activity compared to BVM [216].

EP39 (Table 3) is a BVM derivative synthesized and evaluated by our team and our collaborators. The C-28 group of BVM has been modified by grafting hydrophilic substituents to increase its water solubility. EP39, having a higher hydro-solubility compared to BVM, shows a 2.5 fold increase in activity, a better antiviral profile and a higher selectivity index [217]. The EP39 resistant mutations were determined: CA-A194T, CA-T200N, CA-V230I, CA-V230A and SP1-A1V. By *in-silico* studied, EP39 and BVM appear to bind in the same pocket but exhibit different predicted positions in CA_{CTD}-SP1 crystal hexamer [228].

Dang *et al.* synthesized 28 new derivatives of BVM and tested their anti-virus activities against BVM-resistant HIV-1 variant. The compound 6 (CB310F3) (Table 3) with a C-28 methyl nonanoyl-glutamate side chain showed a much improved activity against HIV-1 strains carrying BVM-resistant polymorphism [218]. Besides, CB310F3 is 20 folds more potent than BVM against the replication of NL4-3/SP1-V7A, which is the most prevalent clinical BVM resistant polymorphism in Gag-SP1 [218].

Tang *et al.* established efficient synthetic routes to macrocyclized BVM derivatives. The derivative, 20b (Table 3), was evaluated biologically to have equal antiviral activities compared to BVM. However, this new derivative is not efficient to overcome the polymorphisms at SP1 [219].

In 2006, Liu *et al.* designed and synthesized a series of modified C-3 BA [220]. The compound 4-benzoic acid 9a (Table 3) has comparable antiviral potency and pharmacokinetic profile to BVM, which demonstrate the 4-benzoic acid motif introduced in C-3 position of BA acts as a suitable replacement for the dimethyl succinic acid side chain in BVM. While, this compound does not have significant improvement in the potency against SP1-V7A polymorphism and solubility compared to BVM [220].

To improve the potency against SP1-V7A, Regueiro-Ren *et al.* synthesized the derivatization of the C-28 position of 4-benzoic acid 9a [229]. BMS-955176 (Table 3) is a potent HIV-1 inhibitor, which exhibit broad spectrum antiviral activities that encompass SP1-V7 Δ and SP1-V7A containing polymorphisms [229]. All these data suggest that BMS-955176 is a second-generation MI with greatly improved preclinical profile compared to BVM. However, the development of this compound was ended in Phase 2b for the gastro intestinal intolerability and drug resistance [230].

Urano *et al.* synthesized alkyl amine derivatives modified at C-28 position of BVM. They found a set of molecules that are more potent than BVM against HIV-1 clade B (NL4-3) and has higher antiviral activity against NL4-3 V7A polymorphism [222]. These molecules also displayed a potent antiviral activity against HIV-1 clade C which presents more polymorphism in SP1 compared to HIV-1 clade B [231]. With *de novo* selection experiments, the mutations that confer resistance to these novel compounds (7m and 7r Table 3) were identified at the first residue of SP1 (A1V) and CA residue 157 (P157A) in major homology region (MHR) with HIV-1 clade B. For HIV-1 clade C, more mutations were identified: CA-V230M, SP1-A1V, SP1-S5N and SP1-G10R [232].

In 2019, Chrobak synthesized and evaluated the phosphate and phosphonate derivatives of BVM. The 30-diethylphosphonate analog of BVM, compound 14a (Table 3), has similar antiviral effect and more selective compared to BVM [223].

Objectives of the Thesis

With the drug resistance occurring, it is urgent to find a novel HIV inhibitor. The HIV maturation inhibitor is a new direction for drug development. The clinical assays of two maturation inhibitor candidates, BVM and BMS-955176, both terminated at phase IIb for drug resistance and side effects. To propose new active molecules, despite the polymorphism of the SP1 domain, we have to decipher the mechanism of action of these maturation inhibitors. Thus, our principal objective was to study the interaction between the BVM derivative, EP39, designed in our group, and the CA_{CTD}-SP1-NC domain of HIV-1 Gag.

EP39 retains the C-28 group and presents a substitution on the C-3 group of BVM, which increases the solubility of the compound. BVM and EP39 have been demonstrated to interfere the cleavage process between CA and SP1. EP39 induces resistant mutations at CA-SP1 junction and the C-terminal of CA. Thus, we studied the HIV-1 Gag domain encompassing the C-terminal of CA, the full SP1 and the full NC to study its structure in the absence and presence of EP39. With the structural and dynamics differences of CA_{CTD}-SP1-NC, we can tell the effect of EP39 on HIV-1 Gag domain. Two mutations, CA-W184A, -M185A, were introduced to avoid the dimerization of CA and, at the same time, keep the structure of CA.

Besides EP39, we also studied the effect of PF-46396 and CB310F3 on HIV-1 Gag domain CA_{CTD}-SP1-NC. PF-46396 has a completely different structure compared to BVM. However, it can also inhibit the cleavage at CA-SP1 junction. So, we want to study the interaction between PF-46396 and CA_{CTD}-SP1-NC to get a more complete model of the mechanism of HIV-1 maturation inhibitor.

The first part of the manuscript is the “Materials and Methods”, which described all the expression and purification protocols for the various proteins studied in my thesis. In this part, I will also present the principle of the NMR techniques.

The second part concerns the results, which showed the different effects of EP39, PF46396 and CB310F3 on the structure and dynamics of protein CA_{CTD}-SP1-NC, especially the SP1 domain. There is already a crystal structure of CA_{CTD}-SP1 hexamer but the structure of SP1 domain is not complete and there is no available dynamic information about the protein. Thus, in this research, we will calculate the structure of CA_{CTD}-SP1-NC domain of Gag and check the effect of the three inhibitors on this Gag domain, structurally and dynamically. BVM is not used in this research because the poor water solubility makes it difficult to study it by NMR.

The third part is the “Discussion and conclusion”. In this part, I will compare and analyze the mechanism of different HIV-1 maturation inhibitors.

Materials and methods

I Expression and purification of HIV-1 CA_{CTD}-SP1-NC, CA_{CTD}^{W184A, M185A}-SP1-NC and NC.

The wild type HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC constructions were kindly provided by Prof. Michael F. Summers. The amino acids sequence of the three domains are shown in Figure 13.

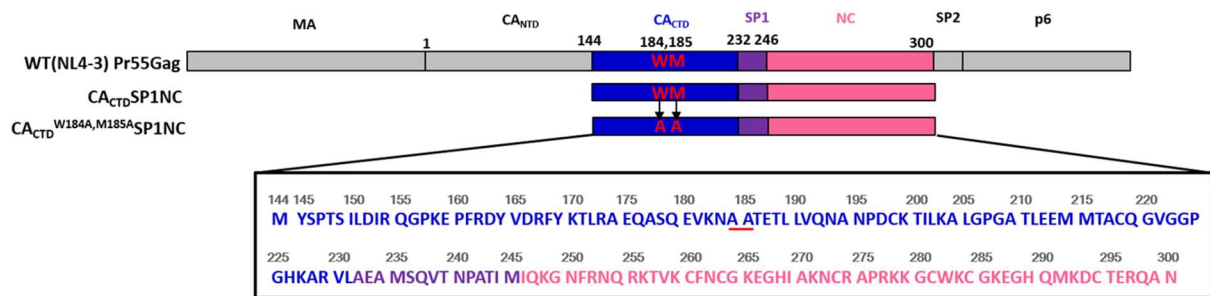


Figure 13. Schematic of full-length polyprotein Gag; wild type CA_{CTD}-SP1-NC, which contains the C-terminal domain of CA, the spacer peptide SP1 and the nucleocapsid protein NC; CA_{CTD}^{W184A, M185A}-SP1-NC, contains two mutations (W184A, M185A) in the C-terminal of CA to prevent CA dimerization. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

1. Overexpression of HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC

1.1 Plasmid transformation

HIV-1 (NL4-3) CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC are cloned to pLATE11 (Thermo Scientific). HIV-1 (NL4-3) NC is cloned to pET-3a (Novagen). Their expression is driven by T7 promoter which can be induced by Isopropyl β-d-1-thiogalactopyranoside (IPTG). These three plasmids are transformed chemically into competent bacterial BL21(DE3)pLysE (Invitrogen). Add 500 μl Luria Broth (LB) to the bacterial and grow in 37 °C shaking incubator for 45 min. Then plate the bacterial onto a LB agarose plate containing 34 μg/ml chloramphenicol (this resistance gene is contained in the plasmid pLysE) and 50 μg/ml ampicillin (this resistance gene is contained in vector pLATE11 and pET-3a). Incubate plates at 37 °C overnight.

1.2 Cell culture

One colony of the bacterial with plasmid transformed was taken and added to 10 ml LB with 34 µg/ml chloramphenicol and 50 µg/ml ampicillin. After growing in 37 °C shaking incubator overnight, the cell culture was moved into one litre LB with 34 µg/ml chloramphenicol and 50 µg/ml ampicillin and grew in 37 °C shaking incubator until the optical density (OD) at 600 nm reached 0.5~0.6. IPTG was added into the culture to reach the final concentration of 1mM. The expression of HIV-1 (NL4-3) CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC was induced at 30 °C for 4 hours. For NC, the protein expression was induced at 30 °C for 4 hours. After induction, the bacterial is collected by centrifuge at 5000 rpm for 30 min.

For ¹⁵N or ¹⁵N/¹³C labeled protein expression, the bacterial cultured overnight was transferred to minimum medium M9 (Table 4) enriched with 0.1 mM ZnCl₂ (to form the zinc finger in NC), ¹⁵N (with ¹⁵N-NH₄Cl) or ¹⁵N/¹³C (with ¹⁵N-NH₄Cl and ¹³C-glucose) and cultured in 37 °C shaking incubator from the initial OD₆₀₀ 0.1 to 0.5~0.6. IPTG was added to reach the final concentration of 1mM. Then, the protein expression was induced as the previous paragraph.

Table 4. Composition of M9 medium for labeled protein expression

The composition of medium M9 5×
- 86 g Na ₂ HPO ₄ ·12H ₂ O
- 15 g KH ₂ PO ₄
- 2.5 g NaCl

The composition of medium M9 for one litre culture
- 200 ml M9 5×
- 790 ml H ₂ O
- 2 ml MgSO ₄ 1M
- 0.1 ml CaCl ₂ 1M
- 10 ml vitamin MEM®
- 2 g glucose (¹³ C-D-glucose Eurisotop® for ¹³ C labeled protein)
- 1 g ammonium (¹⁵ N-NH ₄ Cl Eurisotop® for ¹⁵ N labeled protein)

2. Purifications of HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC.

2.1 Lysis of bacteria

HIV-1 CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC:

Bacteria are resuspended in 20 ml lysis buffer: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mM ZnCl₂, 5 mM 2-Mercaptoethanol (BME). The cells were sonicated 5s on, 15s off, totally 30 min. Cellular debris is eliminated by centrifuge at 20,000 rpm for 20 min.

HIV-1 NC:

Bacteria are resuspended in 20 ml lysis buffer: 50 mM Tris-HCl (pH 8.0), 10% (w/v) glycerol, 100 mM NaCl, 0.1 mM ZnCl₂, 5 mM dithiothreitol (DTT), 2 mM EDTA. Add 17 µl of 100 mM PMSF (phenylmethylsulfonyluoride) and 2.1 µl of 1% (w/v) sodium deoxycholate. The cells were sonicated 5s on, 15s off, totally 30 min. Cellular debris is eliminated by centrifuge at 20,000 rpm for 20 min.

2.2 Elimination of nucleic acid

For HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC, we use the same protocol to eliminate nucleic acid.

Add polyethyleneimine (PEI) to the lysis solution to a final concentration of 0.4% (w/v) and stirred for 15 min before centrifuge at 23,000 rpm for 30 min at 4 °C to eliminate nucleic acid. The supernatant was collected and filtered through 0.2 mm pore size.

2.3 Ion exchange chromatography

HIV-1 CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC:

Buffer A: 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1 mM ZnCl₂, 10 mM BME.

Buffer B: 50 mM Tris-HCl pH 7.5, 1 M NaCl, 0.1 mM ZnCl₂, 10 mM BME.

The ion exchange chromatography is carried out with two columns: Mono Q anion exchange chromatography column (GE Healthcare®) and RESOURCE-S cation exchange chromatography column (GE Healthcare®).

These two columns are both pre-equilibrated with buffer A. The sample recovered after nucleic acid elimination is injected to Mono Q. Flow through is collected and be injected to RESOURCE-S column. Wash RESOURCE-S column with 10 column volumes (CV) of buffer A, then make a gradient of buffer B from 10%-100% in 10 CV at 1 ml/min. The protein is eluted at around 40% buffer B.

HIV-1 NC:

Buffer A: 50 mM Tris-HCl pH 8.0, 10% glycerol, 100 mM NaCl, 0.1 mM ZnCl₂, 10 mM BME.

Buffer B: 50 mM Tris-HCl pH 8.0, 10% glycerol, 1 M NaCl, 0.1 mM ZnCl₂, 10 mM BME.

The protocol of ion exchange chromatography for HIV-1 NC is same as the protocol for HIV-1 CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC.

2.4 Size exclusion chromatography

HIV-1 CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC:

Buffer C: 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1 mM ZnCl₂, 5 mM BME.

Buffer D: 25 mM Acetate-Na PH6.5, 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM BME.

This size exclusion chromatography is carried out with HiLoad 16/600 Superdex 75 pg (GE Healthcare®). This column is pre-equilibrated with buffer C. The protein eluted in ion exchange chromatography is concentrated to 5 ml and injected into Superdex 75 column. Elution is finished with buffer C at 1 ml/min. After elution, the protein buffer is exchanged to buffer D with Amicon® Ultra Centrifugal Filter and protein is concentrated to 300 µM.

HIV-1 NC:

Buffer C: 50 mM Tris-HCl pH 7.0, 10% glycerol, 100 mM NaCl, 0.1 mM ZnCl₂, 10 mM BME.

Buffer D: 25 mM Acetate-Na PH6.5, 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM BME.

The protocol of size exclusion chromatography for HIV-1 NC is the same as the protocol for HIV-1 CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC. After elution, protein buffer is exchanged to buffer D with Amicon® Ultra Centrifugal Filter and protein is concentrated to 300 μM.

II Maturation inhibitors

1. EP39

EP39 was chemically synthesized by our collaborator, Serge Turcaud (LCBPT, CNRS, UMR 8601).

The protocol for EP39 synthesis [217]: The tert-Butyldimethylsilyltrifluoromethanesulfonate (t-BuMe₂-SiOTf, 150 mL, 0.65 mmol) was added to a stirred solution dropwisely under argon of N-tert-Boc derivatives, 4-({28-[1,1-dimethylethyl-(2-aminoethyl)carbamate]-28-oxolup-20,29-en-3 β -yl}oxy)-2,2-dimethyl-4-oxobutanoic acid (0.19 mg, 0.26 mmol), and 2,6-lutidine (91 mL, 0.78 mmol) in dry CH₂Cl₂ (1 mL). The mixture was stirred for 30 min, quenched with saturated aqueous ammonium chloride solution (2 mL). The mixture was diluted with CH₂Cl₂ (5 mL), the organic layer was washed with water, brine and dried over MgSO₄. The solvent was concentrated under vacuum and the residue was used in the next step without purification. Tetra-butylammonium fluoride (260 mL, 1 M solution in THF, 0.26 mmol) was added to a stirred solution of the N-(tert-butyldimethylsilyloxy) derivatives in dry THF (0.5 mL) at room temperature. The reaction mixture was stirred for 1 h, quenched with saturated aqueous ammonium chloride solution (2 mL). The resulting precipitate was collected, washed with water and dried over vacuum to yield the corresponding compound EP39.

The solid state EP39 was then solubilized in DMSO-*d*₆ in final concentration 60 mM.

2. PF-46396 and CB310F3

PF-46396 is provided by Prof. Johnson Mak (Institute for Glycomics, Griffith University QLD 4111, Australia). We solubilized PF-46396 with DMSO-*d*₆ at the final concentration 60 mM.

CB310F3 is offered by Zhao Dang (Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA). Same as EP39 and PF46396, CB310F3 was also solubilized in DMSO-*d*₆ to final concentration 60 mM.

III Nuclear magnetic resonance (NMR)

In this part, the NMR experiments used to study the structure and dynamics of HIV-1 Gag CA_{CTD}-SP1-NC, CA_{CTD}^{W184A, M185A}-SP1-NC and NC will be explained. I will describe the NMR experiments as well as the method of data analysis for each of the experiences.

1. NMR Introduction

NMR is a technique allowing the analysis at the atomic scale of a wide range of molecules (chemical or biological). Here, we will focus mainly on biological macromolecules, such as proteins. In order to be able to study proteins by NMR, certain nuclei must be labeled specifically. Indeed, the carbon 12 (¹²C) and nitrogen 14 (¹⁴N) nuclei constituting the proteins are naturally abundant isotopes but cannot be observed in NMR. The production of the labeled macromolecules in a recombinant manner takes place in a medium enriched in nitrogen 15 (¹⁵N) and/or carbon 13 (¹³C), which are isotopes observable by NMR.

This spectroscopic method is well suited to characterize the structure and dynamics of proteins in solution. These studies are essential to understand the role of these macromolecules in a cellular process. The structural flexibility of proteins reflects an important mechanistic aspect for their functions including catalysis, regulation, binding, and cellular structure. This dynamics extends over a large time scale (from picosecond to second) [233]. NMR spectroscopy provides a different set of experiments for studying protein dynamics at different time scales. Thereafter, I will introduce the experiences that were used in my thesis.

Before starting the structural and / or dynamics studies of protein by NMR, the assignment of protein resonances (or chemical shifts) is necessary. This assignment lies in correlating each peak of the spectrum with the amino acid atom of the protein sequence.

2. Assignment of protein by NMR

2.1 General assignment strategy

Proteins are made up of hundreds of hydrogens (^1H), carbons (^{13}C) and nitrogen atoms (^{15}N). The one-dimensional spectrum (1D) of the $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ is presented in Figure 14 and shows the chemical shift of the different types of ^1H present within the protein. The 1D spectrum can give a basic idea of the protein's folding. Generally, the protein is well folded if the peaks in 1D NMR spectrum are sharp and narrow and spread over a large range of chemical shifts. Whereas, if the protein is unfolded or partially folded, the peaks in 1D spectrum will be broader and not widely dispersed. The 1D spectrum with good dispersion, eg. Peaks in a negative chemical shift (methyl groups) or peaks with chemical shift above 10 ppm, is a good sign of well folded structure.

Protons have specific chemical shift regions. Overall, between 0 and 1.5 ppm are methyl groups protons; aliphatic protons range from 1 to 4 ppm; the protons H_α between 3.5 and 5 ppm; aromatic protons between 6 and 8 ppm and finally the region between 6.5 and 10 ppm corresponds to the amide protons (1 H_N : proton bound to the nitrogen of the amide group) Figure 14.

Protein assignment is based on heteronuclear and multidimensional experiments on a doubly labeled $^{15}\text{N} / ^{13}\text{C}$ sample [234]. Indeed, a set of 3-dimensional (3D) experiments is recorded (Table 5) and is based on the transfer of magnetization by scalar couplings between carbon ^{13}C or nitrogen ^{15}N to their respective proton of protein backbone. Each atom of an amino acid has a different chemical shift, which can facilitate its assignment. In addition, the chemical shift of the different nuclei of the backbone depends on the secondary structures present in the protein.

After full assignment and structure calculation of $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ in the presence of EP39, the data (chemical shifts, NMR restraints) and coordinates of the atoms of the protein are deposited at the Protein Data Bank (PDB).

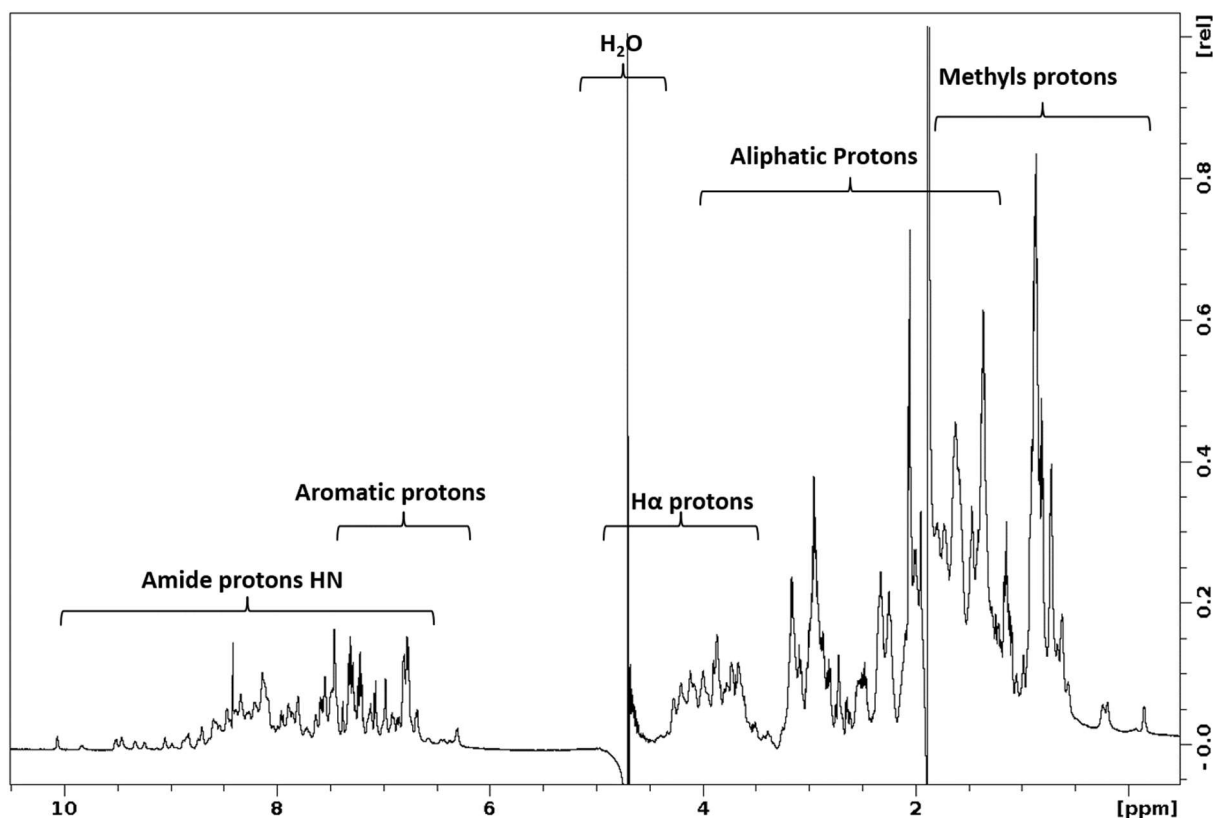


Figure 14. 1D NMR spectrum of $CA_{CTD}^{W184A, M185A}$ -SP1-NC. Different proton regions are presented. The molecular weight of this protein is 17 kD. It is complicated to identify the protons in this spectrum.

All amino acids have a ^1H - ^{15}N group except for prolines. The 2D ^{15}N -HSQC experiment (Heteronuclear Single Quantum Coherence) is a standard experiment for two dimensions (^1H and ^{15}N) showing the proton-nitrogen correlations [235,236]. The peaks observed on this spectrum therefore correspond to the amide groups of the protein backbone and those of side chains of the amino acids, tryptophan, asparagine and glutamine. The Arginine side chain $\text{N}\epsilon$ - $\text{H}\epsilon$ peaks are in principle also visible, but because the $\text{N}\epsilon$ chemical shift is outside the region usually recorded, the peaks are folded and appear as negative peaks. If working at low pH, the Arginine $\text{N}\eta$ - $\text{H}\eta$ and Lysine $\text{N}\zeta$ - $\text{H}\zeta$ groups can also be visible and folded. The prolines are therefore not visible on this type of experiment.

The 2D ^{15}N -HSQC spectrum is usually the first heteronuclear experiment performed on proteins. From it we can assess whether other experiments are likely to work and for instance, whether it is worth carbon labelling the protein before spending the time and money on it. If

the protein is reasonably large, we might be able to judge whether deuteration might be necessary.

The new band-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence (SOFAST-HMQC) records protein ^1H - ^{15}N correlation spectra in a few minutes of acquisition time [237,238] (Figure 15). This experiment yields significantly increased sensitivity with a much shorter time compared to standard techniques, sensitivity-enhanced ^{15}N -HSQC, fast-HMQC [239], or FHSQC [240].

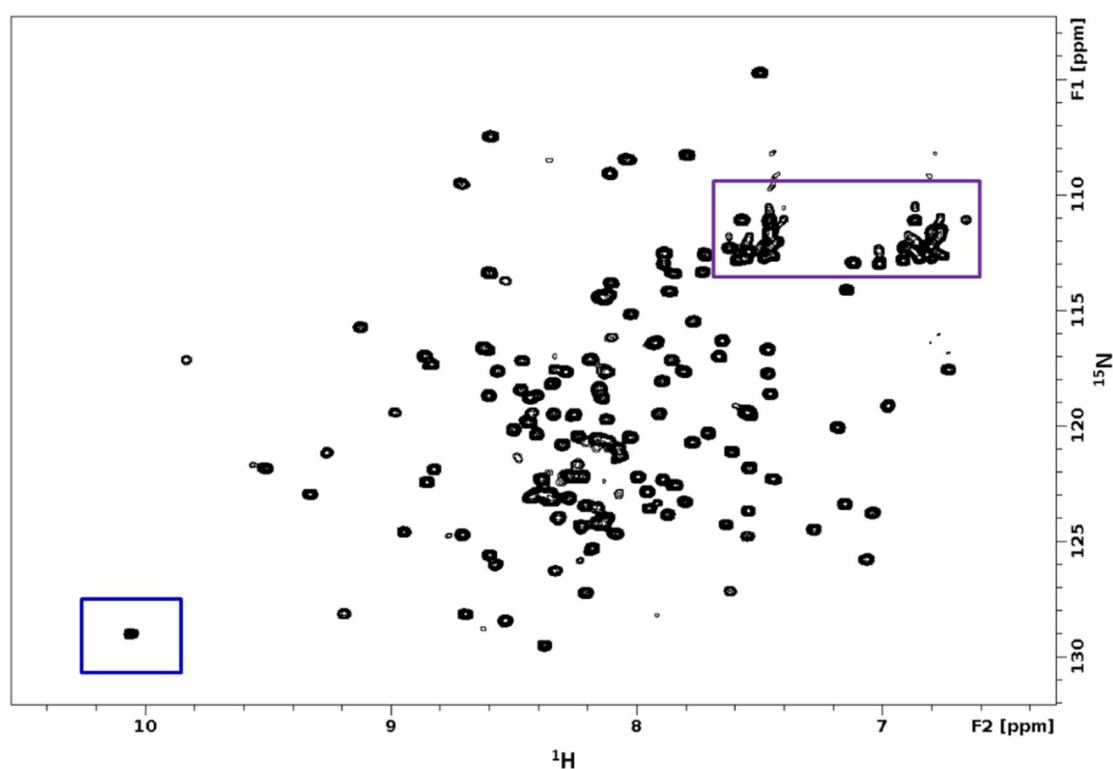


Figure 15. 2D SOFAST-HMQC spectrum of HIV-1 Gag CA_{CTD}^{W184A, M185A}-SP1-NC. This spectrum shows the ^1H - ^{15}N correlations of the amide groups. 1 amino acid = 1 amide bond = 1 peak on the spectrum. but Trp side chain $\text{N}\epsilon$ - $\text{H}\epsilon$ groups and Asn/Gln side chain $\text{N}\delta$ - $\text{H}\delta$ 2/ $\text{N}\epsilon$ - $\text{H}\epsilon$ 2 groups are also visible. Trp side chain $\text{N}\epsilon$ - $\text{H}\epsilon$ peaks are usually in the blue square part of the spectrum. Asn/Gln side chain NH_2 group peaks are generally in the purple square part of the spectrum and logically result in two peaks at the same nitrogen frequency with two different hydrogen frequencies.

The assignment of ^{15}N -HSQC spectrum resonances is essential to initiate dynamic studies and obtain information for each amino acid in a protein. This is done by recording 3D resonance experiments (^1H , ^{15}N and ^{13}C) (Table 5). A ^1H - ^{15}N group of an amino acid i is correlated with a carbon atom, $^{13}\text{C}_\alpha$ (alpha carbon) or ^{13}CO (carbonyl group) of amino acid i and / or the

preceding amino acid $i-1$. For example, the 3D HNCA experiment correlates $^1\text{H}_\text{N}$ to the chemical shift of ^{15}N from residue i with that of $^{13}\text{C}_\alpha$ from residues $i-1$ and i [241–243]. In contrast, the HN(CO)CA experiment correlates $^1\text{H}_\text{N}$ to the chemical displacement of ^{15}N from residue i with that of $^{13}\text{C}_\alpha$ from residue $i-1$ only [242,244] (example of correlation illustrated in Figure 16). By recording the HNCA and HN(CO)CA experiments, we can distinguish the intra and inter-residue resonances on the spectrum and start the sequential assignment of the protein backbone.

As a rule, the main chain of the protein is assigned (C_α , CO and HN) then the side chains (H_β - C_β , H_γ - C_γ , H_δ - C_δ , H_ϵ - C_ϵ , H_ζ - C_ζ and H_η - C_η). Table 5 presents a list of NMR experiments which can be used for the sequential assignment. Some amino acids can be used as starting points in attribution, for example: alanine, glycine, serine and threonine. They are easily identifiable thanks to their specific chemical displacements of $^{13}\text{C}_\alpha$ or $^{13}\text{C}_\beta$.

Table 5. NMR experiments used for the assignment of the main chain (top of the table).

NMR experiments	Observed atoms
2D ^{15}N -HSQC	$^1\text{H}/^{15}\text{N}$
3D HNCA	$^{13}\text{C}_{\alpha i-1} / ^{13}\text{C}_{\alpha i} / ^{15}\text{N} / ^1\text{H}_\text{N}$
3D HN(CO)CA	$^{13}\text{C}_{\alpha i-1} / ^{15}\text{N} / ^1\text{H}_\text{N}$
3D HNCACB	$^{13}\text{C}_{\alpha i} / ^{13}\text{C}_{\beta i} / ^{13}\text{C}_{\alpha i-1} / ^{13}\text{C}_{\beta i-1} / ^{15}\text{N} / ^1\text{H}_\text{N}$
3D CBCA(CO)NH	$^{13}\text{C}_{\alpha i-1} / ^{13}\text{C}_{\beta i-1} / ^{15}\text{N} / ^1\text{H}_\text{N}$
3D HNCO	$^{13}\text{CO}_{i-1} / ^{15}\text{N} / ^1\text{H}_\text{N}$
3D HN(CA)CO	$^{13}\text{CO}_{i-1} / ^{13}\text{CO}_i / ^{15}\text{N} / ^1\text{H}_\text{N}$

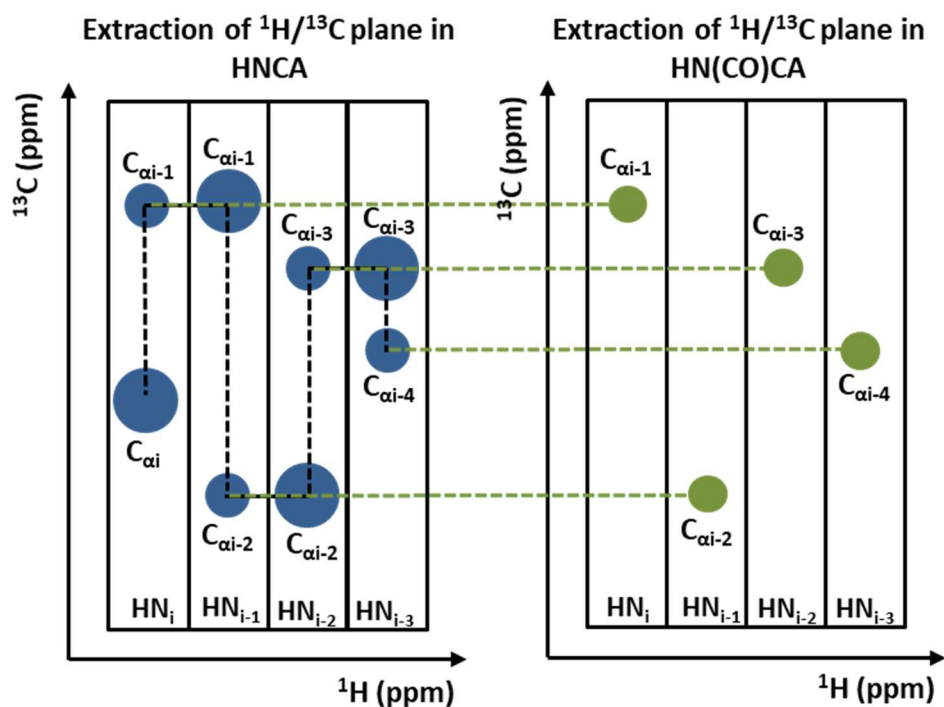


Figure 16. Extraction of several ^1H - ^{13}C planes from the 3D HNCA experiment showing the ^1H - ^{15}N correlation of an amino acid (each band corresponds to 1 peak of the HSQC) to a $\text{C}_{\alpha i}$ (most intense peak of the band) or $\text{C}_{\alpha i-1}$ (least intense peak of the band). In 3D HN(CO)CA experiment, the $^1\text{H}_i$ - $^{15}\text{N}_i$ correlation of an amino acid only to a $\text{C}_{\alpha i-1}$.

The NOESY experiment (Nuclear Overhauser Spectroscopy) is an experiment allowing the assignment of resonances through dipolar coupling and provides essential information for structural determination [234]. The dipole-dipole coupling is the interaction across the space between two spins, for example two ^1H . The strength of this interaction depends on the distance between the two atoms. The intensity of the NOE correlation peak between two ^1H is approximately proportional to $1/d^6$, with d the distance between these two atoms.

The assignment is not done sequentially but it is based on the spatial connectivities between pairs of protons separated by less than 5-6 Å. The NOESY experiment makes it possible to correlate the ^1H linked to a labeled heteroatom, ^{15}N (3D ^{15}N NOESY-HSQC) or ^{13}C (3D ^{13}C NOESY-HSQC), to all nearby protons through space

This spectrum is essential to obtain the distance restraints for structure calculations. It can also be used to help assignment, and for small to medium-sized proteins, assignment using ^{15}N NOESY-HSQC and ^{15}N TOCSY-HSQC only is possible if the spectra are well resolved.

Indeed, using the ^{15}N NOESY-HSQC experiment, we can place the amino acids one behind the other because the amide proton of an amino acid has NOE peaks to the amide proton $\text{HN}_{i\pm 1}$. In helical sections, HN_i shows correlation not only with $\text{HN}_{i\pm 1}$, but also with $\text{HN}_{i\pm 2}$, sometimes $\text{HN}_{i\pm 3}$. The β -sheet structures contain short HN-HN distances between the strands. That means there is a strong cross-strand NOE observed between HNs that are not necessarily close to each other in the sequence. The side-chain NOEs in the ^{15}N NOESY-HSQC is also useful during the assignment process, as $\text{HN}_i\text{-H}\alpha_{i-1}$ are generally very strong, especially in helical and β -sheet sections (Figure 17, 18) [14].

^{15}N TOCSY-HSQC should show links from the backbone HN group to all side-chain hydrogens of the same residue. Using this spectrum, the amino acid type can be identified. (Figure 19, 29) [16]

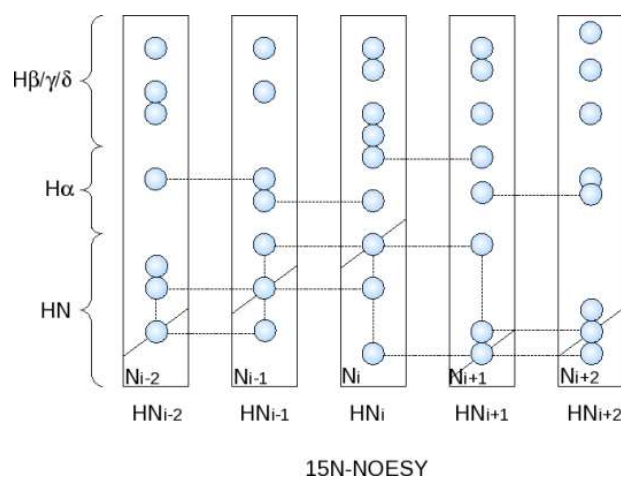


Figure 17. Sequential assignment of amino acids from the backbone of a protein using a 3D experiment ^{15}N NOESY-HSQC. $^1\text{H-}^{15}\text{N}$ of an amino acid i has correlation to the amide protons of the previous and next residues, HN_{i-1} and HN_{i+1} . Besides, the $^1\text{H-}^{15}\text{N}$ also correlate to the alpha proton of the previous residue, $\text{H}\alpha_{i-1}$. (source www.protein-nmr.org)

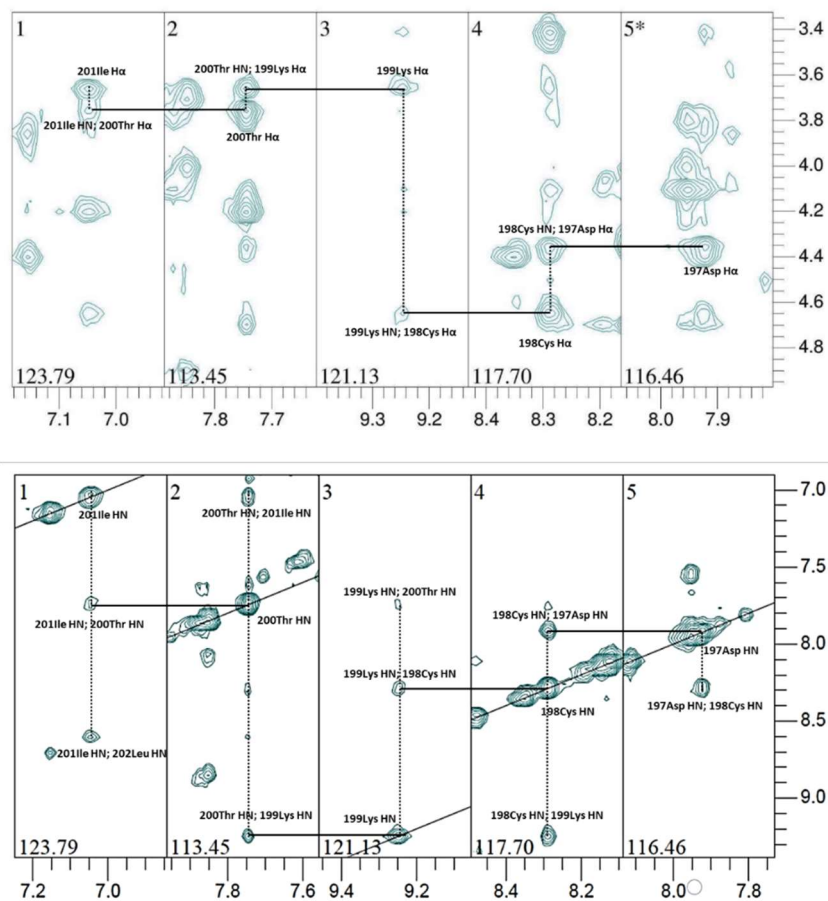


Figure 18. Example of the sequential assignment of residue 197-201 of the protein CA_{CTD}^{W184A} . $M185A$ -SP1-NC with the CCPNMR program [15].

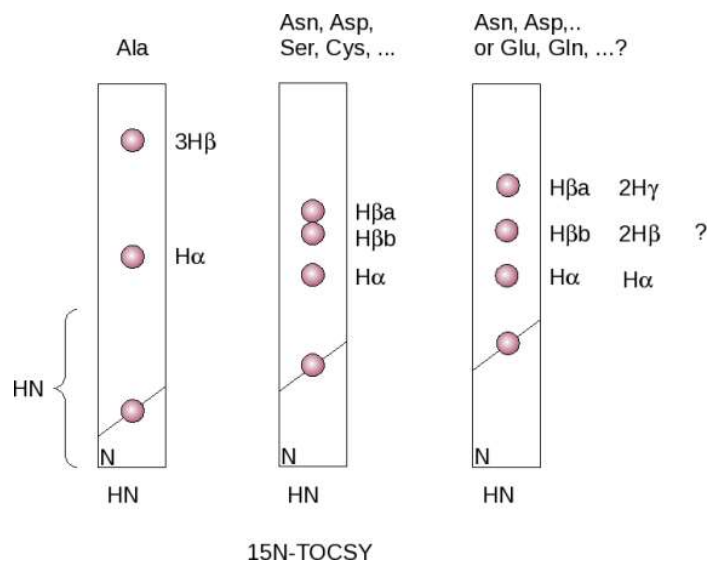


Figure 19. Side chain assignment using ^{15}N TOCSY-HSQC. (source www.protein-nmr.org).

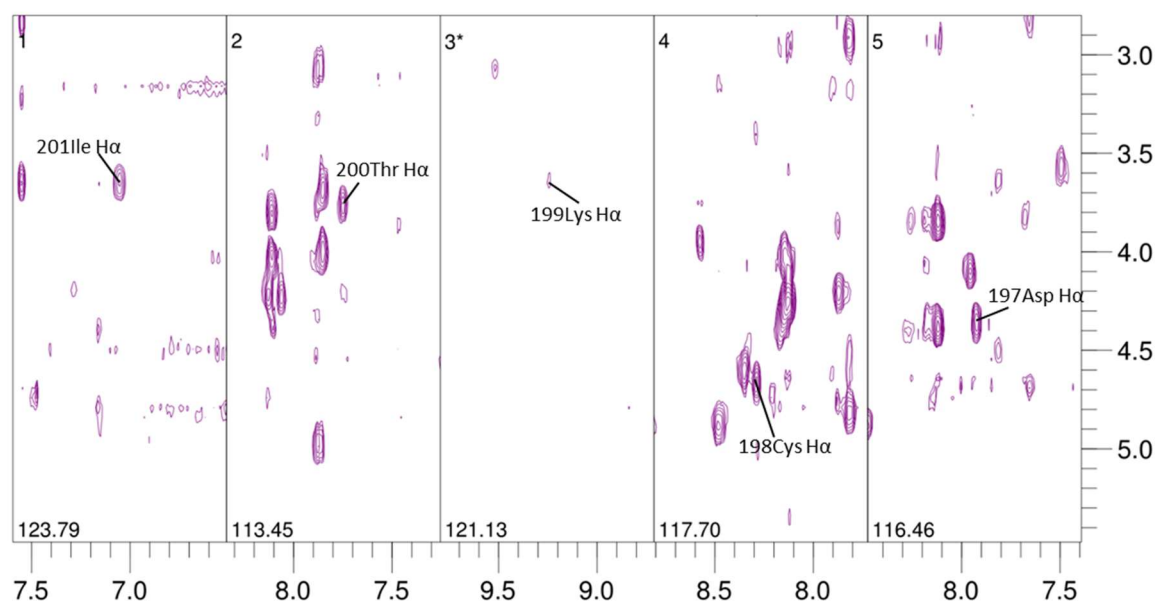


Figure 20. Example of the side chain assignment of the protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC with the CCPNMR program.

2.2 Assignment of HIV-1 Gag $CA_{CTD}^{W184A, M185A}$ -SP1-NC

The 2D SOFAST-HMQC [238] and 3D ^{15}N NOESY-HSQC [243][245], 3D ^{15}N TOCSY-HSQC [245] and 3D HNHA [246] were recorded on a Bruker spectrometer operating at 600 MHz, equipped with a cryoprobe, at 303 K on the protein HIV-1 Gag $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence and presence of the maturation inhibitor EP39. The experiments were processed with software Topspin 3.5pl7. The chemical shift assignment of the protein was carried out with CcpNmr 2.4.2 [15]. The chemical shift and NMR restraints were deposited to PDB under code 6RWG. We also recorded the 2D SOFAST-HMQC spectra with protein HIV-1 Gag NC to identify the structural differences between the isolated NC and the NC in $CA_{CTD}^{W184A, M185A}$ -SP1-NC.

The restraints obtained in 3D NOESY-HSQC were used for structure calculation in next section.

3. Chemical Shift Perturbation Analysis

Chemical shift is very sensitive to the electronic environment of a residue. The disruption of a chemical shift can be caused either by a change in conformation of the protein or by an interaction with a partner (protein, nucleic acid, ligand...). The identification of an interaction

surface on a protein can be easily identified by measuring the variation in chemical shifts [17]. Typically, an HSQC spectrum ^1H - ^{15}N of the ^{15}N labeled protein is recorded. By adding the interacting partner, the chemical shift of some resonances varies. This indicates that these resonances located on the interaction surface, or close to this region.

The measurement of the chemical shifts perturbation ($\Delta\delta$) between two states of a protein is carried out by combining the variation of chemical shift of the proton ^1HN ($\Delta\delta_{\text{HN}}$) and that of nitrogen ^{15}N ($\Delta\delta_{\text{N}}$) from the amide group using the following equation [247]:

$$\Delta\delta = [(\Delta\delta_{\text{N}}/6.5)^2 + (\Delta\delta_{\text{HN}})^2]^{1/2}$$

The NMR fingerprint of a protein and its interaction partner can also provide information on exchange kinetics. The speed of this exchange between two states (free state and bound state) of the protein influences the type of signal obtained by NMR. If we consider the HSQC spectrum of a protein with ligand titrated in, when the exchange rate is fast, the signals will move smoothly from the free state to the bound state with gradual addition of the ligand. On the other hand, if the exchange rate is slow, we could observe two peaks for each interacting residue, corresponding to the bound and free state (Figure 21). Between the fast exchange and the slow exchange, a third exchange could be observed. In the event of an intermediate exchange, a coalescence of the resonance peaks can be observed. At this rate the two separate peaks in a slow exchange coalesce and merge into a single broad peak. This peak depending on the case may disappear completely.

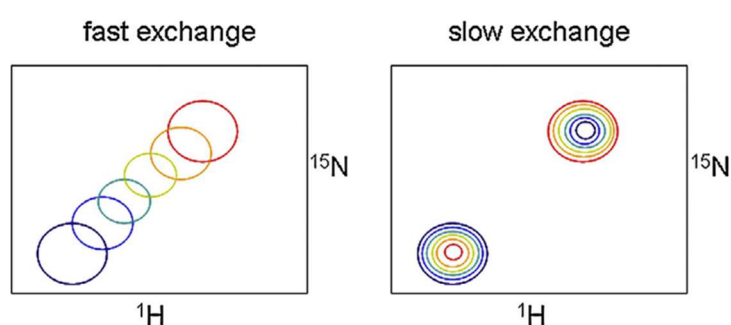


Figure 21. The dependence of 2D NMR peak shape on exchange rate with gradual addition of the ligand. Adapted from [17]. Fast exchange (Left): peaks move smoothly from free (blue) to bound (red) state. In the limit of very fast exchange, peaks have the same shape throughout. As

they move out of this limit, peaks may become broader when in equilibrium between free and bound, and then sharpen up again close to saturation. Slow exchange (Right): the free state (blue) decreases in intensity as the bound state (red) increases.

4. Protein Structure Calculation

Ambiguous Restraints for Iterative Assignment (ARIA) is a software for automated NOE assignment and NMR structure calculation [248,249]. ARIA does not perform the structure generation by itself: it drives the structure generation with the software Crystallography & NMR System (CNS) [250] by analyzing the conformers obtained in the previous step to update the restraints and obtain a set of improved conformers. CNS performs the NMR structure calculation using NOEs, J-coupling, chemical shift, dihedral angles and dipolar coupling data [250,251]. The last stage is the refinement of the calculated structures dynamics, which is performed in explicit solvent (water or DMSO) [249].

5. Protein dynamics

NMR is a very powerful method for characterizing protein dynamics. Indeed, it offers a panel of experiments to characterize the different movements of proteins on a large time scale, going from the picosecond (ps) to the second (s) (Figure 22) [233]. Slow movements are located on a microsecond (μ s) to millisecond (ms) timescale and can be characterized by NMR with CPMG dispersion relaxation experiments [252–254] or by CEST type experiments [255,256]. Slow movements make it possible to determine the conformational changes of a macromolecule, and to highlight non-visible minor states (sparsely populated and short lifetime). On the other hand, there are the fast movements which are located on a picosecond (ps) to nanosecond (ns) time scale, which are characterized by spin relaxation experiments. These experiments make it possible to measure the relaxation constants T1 and T2 as well as heteronuclear NOEs (HetNOE). These studies provide information regarding the movement of bonds within a macromolecule.

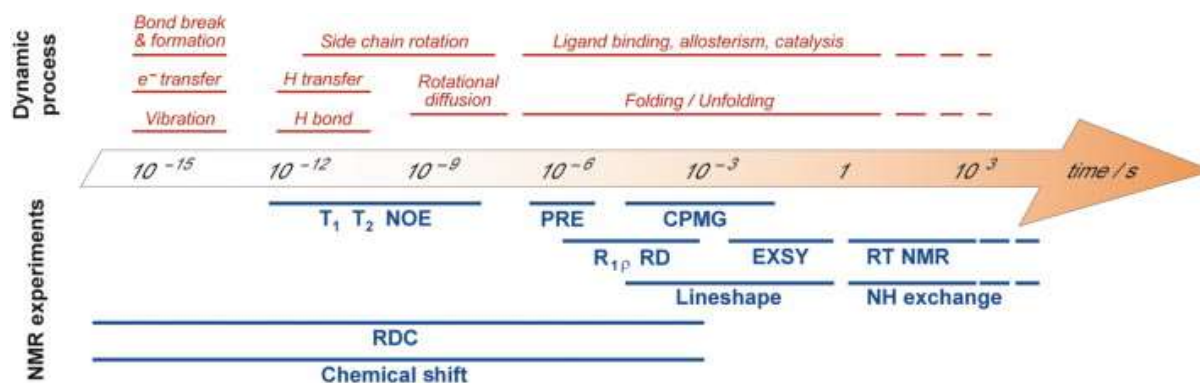


Figure 22. Timescales involved in protein dynamics and NMR observable. Spin relaxation experiment covers most of the timescales where the effective protein internal (and overall) motions take place. The fastest movement can be detected by nuclear spin relaxation measurements (R_1 , R_2 , NOE) or by analyzing the dynamic contribution of the residual dipolar couplings (RDC). The line-shape analysis and the relaxation dispersion experiment account for motions involving a larger number of atoms and are most relevant in biology. The slower motions can be characterized in NMR using EXSY and ZZ exchange experiments [18].

The ^{15}N backbone relaxation experiments were performed on 300 μM CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of 6 equivalents of EP39. T₁, T₂ and HetNOE ^{15}N experiments used standard Bruker pulse sequences[257]. The T₁ experiments were recorded with 5, 10, 20, 30, 40 (repeated), 60, 80, 100 (repeated), 200, 300, 400, 600, 800 and 1200 ms for recovery delay. T₂ experiments were recorded with 1, 2, 4, 8, 10 (repeated), 12, 14, 16, 18, 20 (repeated), 22, 24, 28 and 32 ms for recovery delay. The ^{15}N - $\{^1\text{H}\}$ NOE (HetNOE) values were taken as the ratio between the intensities recorded with and without saturation of the amide protons. The experiments were processed with Dynamics Center 2.3 in Topspin 3.5 pl 7. The fit function for T₁ is $f(t) = I_0 * [e^{-t/T_1}]$. The fit function for T₂ is $f(t) = I_0 * [e^{-t/T_2}]$. NOE: intensity ratio ($I_{\text{sat}}/I_{\text{un-sat}}$).

Results

I Expression and purification of HIV-1 NC, CA_{CTD}-SP1-NC and CA_{CTD}^{W184A, M185A}-SP1-NC

To determine the structure of CA_{CTD}-SP1-NC and study the effect of HIV-1 maturation inhibitors on HIV-1 Gag, we expressed and purified wild type HIV-1 (NL4-3) CA_{CTD}-SP1-NC and mutated HIV-1 (NL4-3) CA_{CTD}^{W184A, M185A}-SP1-NC. We checked the global folding of the two proteins by NMR.

We expressed and purified HIV-1 (NL4-3) NC to study the structure of HIV-1 NC in the absence and presence of CA_{CTD}-SP1 and also the effect of NC on the structure and dynamics of the SP1 domain which has never been fully observed by X-rays or Cryo-EM due to its great dynamics.

1. Expression and purification of wild type HIV-1 (NL4-3) CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC.

The two proteins, wild type CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC, only have two residues difference (CA-W184A, -M185A). Thus, they have almost same protein characteristics. The molecular weights for two proteins are both 17 Kd. The theoretical isoelectric point (pI) are both 9.44. So, they have same protocols for expression and purification.

1.1 Expression of proteins CA_{CTD}-SP1-NC and CA_{CTD}^{W184A, M185A}-SP1-NC.

The two proteins, wild type CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC, were both expressed in *Escherichia coli* (*E. coli*) BL21 (DE3) pLysE and induced by 1 mM IPTG at 30 °C for 3 hours (Figure 23).

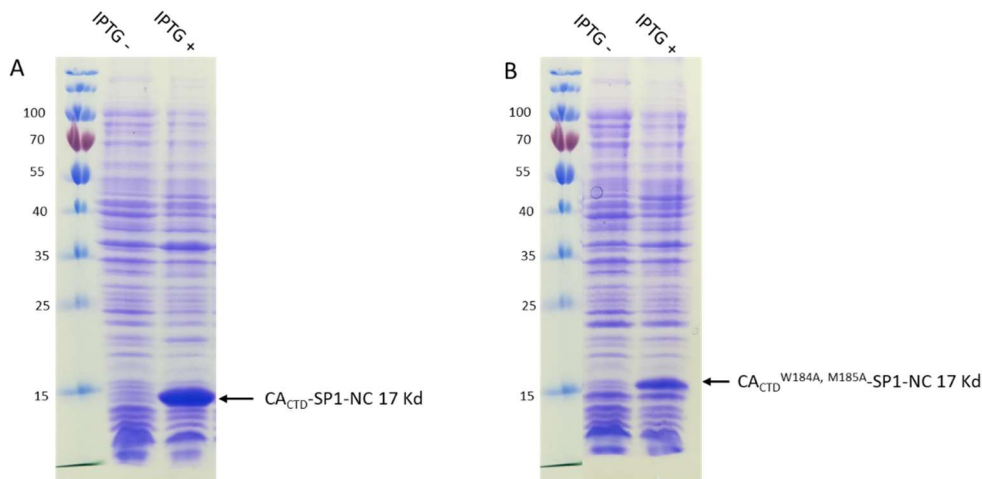


Figure 23. Expression of protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC.

1.2 Purification of protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC.

After the protein was well expressed in bacteria, we released it from bacteria by sonication with buffer pH 7.5. In this step, there is also a lot of nucleic acid released and binding with highly basic NC. Thus, we used 0.4 % polyethyleneimine (PEI) to eliminate nucleic acid before the purification.

- **ion exchange chromatography**

We use ion exchange chromatography to separate the protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC from the other constituents. Ion exchange chromatography is a process that separate ions and polar molecules based on their affinity to the ion exchanger. There are two types of ion exchange chromatography, anion exchange and cation exchange chromatography.

In the purification, we used anion exchange chromatography firstly to eliminate the negatively charged protein. When the sample was loaded to the column, the negatively charged protein bind on the column and eluted with the gradiently increased sodium. The positively charged protein cannot bind on the column and go through the column directly.

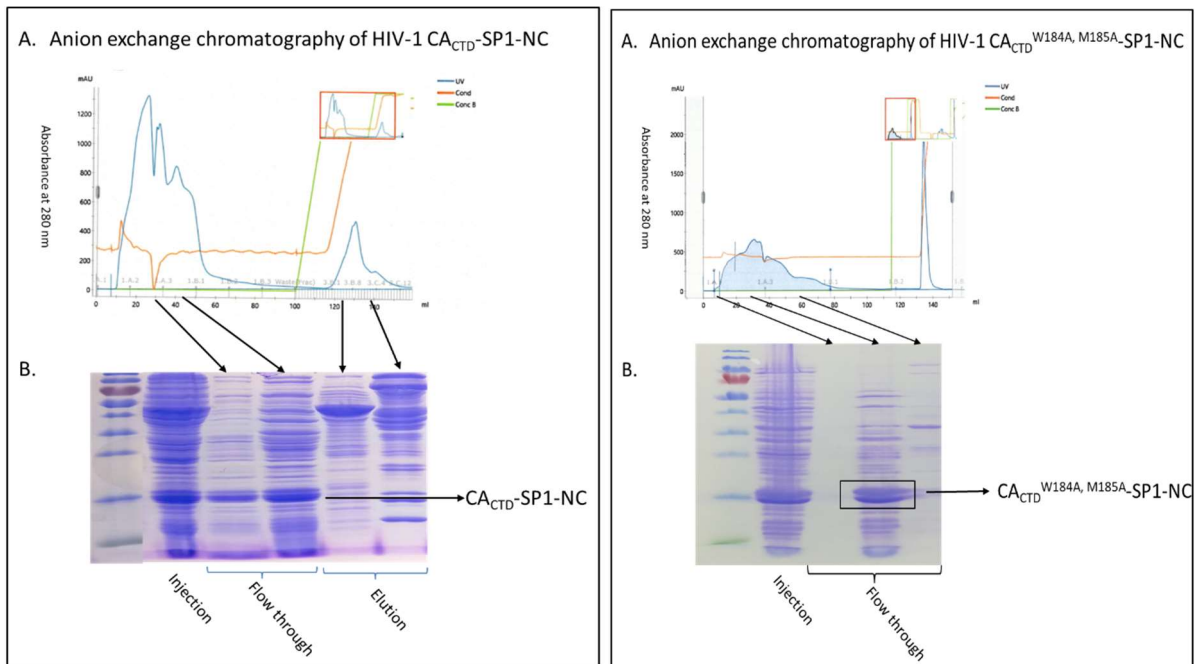


Figure 24. Purification of protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC with anion exchange chromatography.

As shown in Figure 24, the positively charged protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC are in the flow through.

Then, we collected the flow through from the anion exchange chromatography and used cation exchange chromatography to separate different positively charged protein.

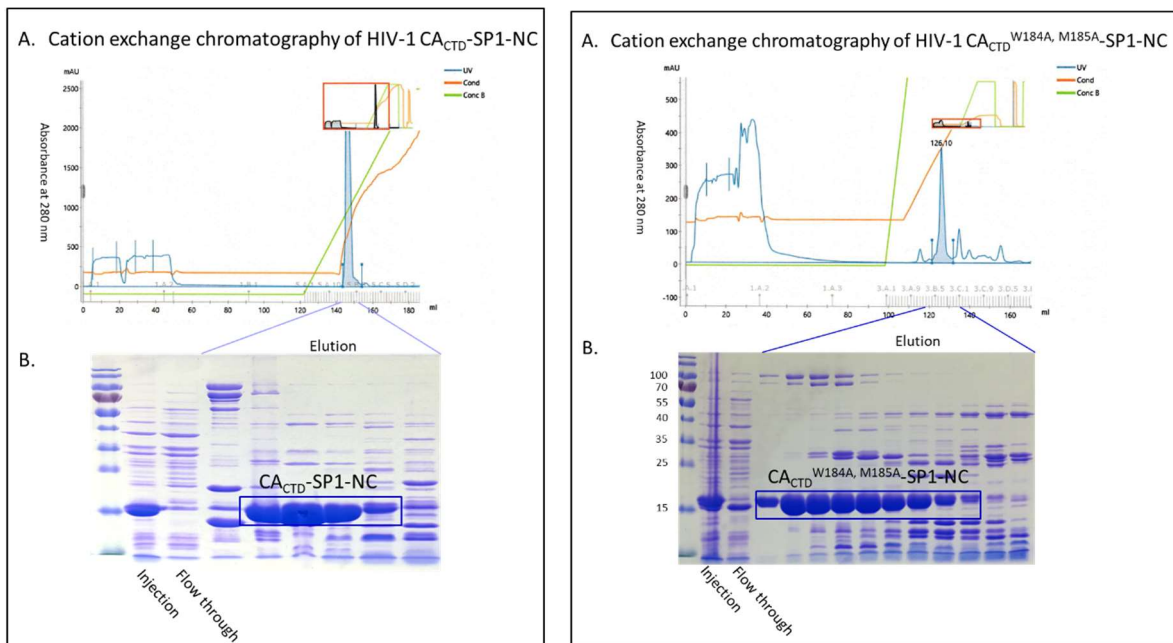


Figure 25. Purification of protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC with cation exchange chromatography.

In figure 25, the protein, CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC, are eluted along with some bacterial proteins, which requires further purification by size exclusion chromatography.

- **Size exclusion chromatography**

Size exclusion chromatography is a technique of separating the molecules by their size. The sample is injected into a separation column which contains a molecular sieve made up of a network of Sephadex or Sepharose type polysaccharide beads. Large molecules are less retained by this network and therefore are eluted first. Small molecules remain within the pore for a longer period of time and are eluted later.

The final step in the purification of CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC is size exclusion chromatography with Superdex75g (16/60) ® column pre-equilibrated with Tris buffer at pH 7.5 at a flow rate of 1.0 mL / min.

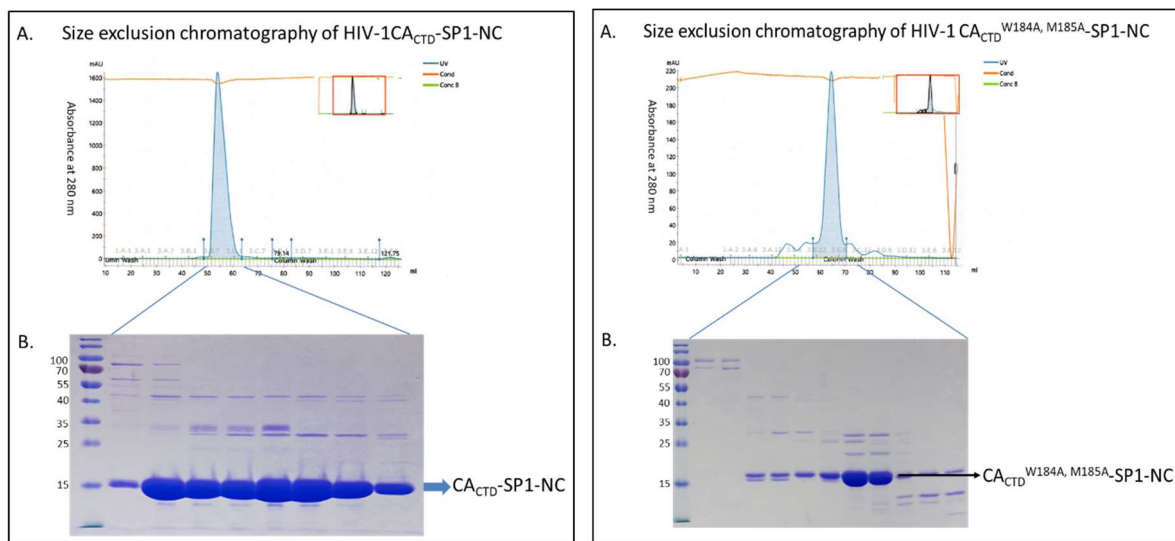


Figure 26. Purification of protein CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC by size exclusion chromatography.

Then, we collected the protein, CA_{CTD} -SP1-NC and $CA_{CTD}^{W184A, M185A}$ -SP1-NC, in the elution of size exclusion chromatography and change the buffer to acetate buffer at pH 6.5 and concentrated the protein to 500 μ M.

2. Expression and purification of HIV-1 (NL4-3) NC.

The protocol for the expression and purification of HIV-1 NC is similar with the protocol for proteins C_{ACTD}-SP1-NC and C_{ACTD}^{W184A, M185A}-SP1-NC.

2.1 Expression of protein HIV-1 NC.

We used same bacteria BL21(DE3)pLysE for the expression of HIV-1 NC. In the induction of protein expression, we set the temperature as 30 °C.

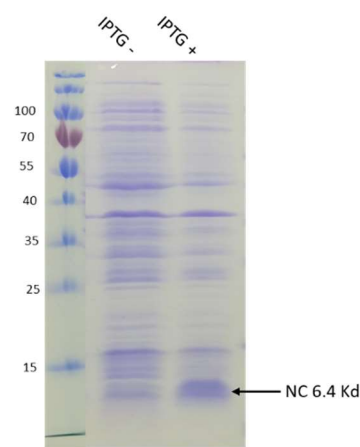


Figure 27. Expression of protein HIV-1 NC.

2.2 Purification of protein HIV-1 NC.

- **Ion exchange chromatography**

The protein HIV-1 NC is also a positively charged protein. With theoretical pI = 9.93. We used the same protocol to purify HIV-1 NC.

First step is anion exchange chromatography. The protein does not bind to the column and eluted in the flow through directly. Negatively charged proteins will bind on the column.

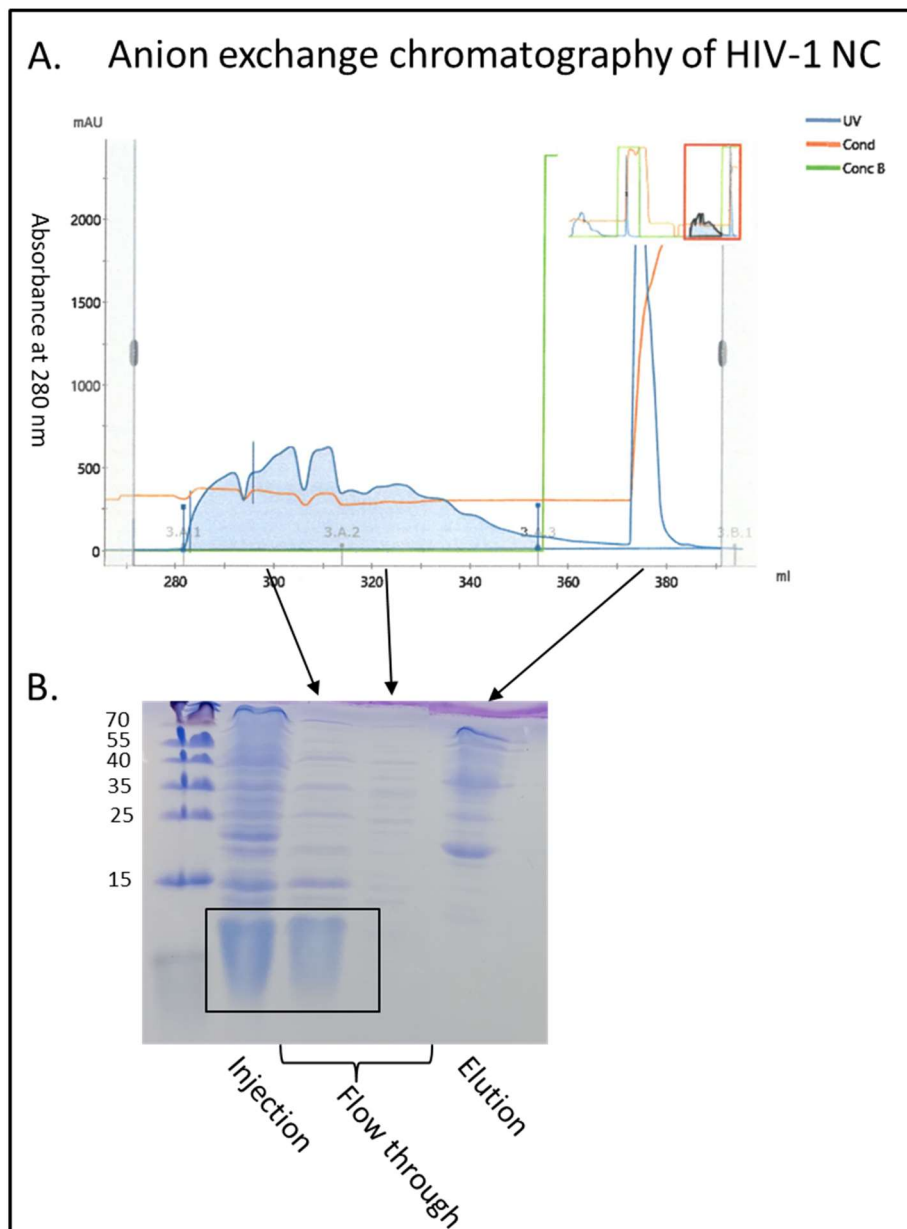


Figure 28. Purification of protein HIV-1 NC with anion exchange chromatography.

We collected the flow through from anion exchange chromatography and loaded it to the cation exchange column. With the gradient increase of sodium concentration, different positively charged proteins are eluted from the column.

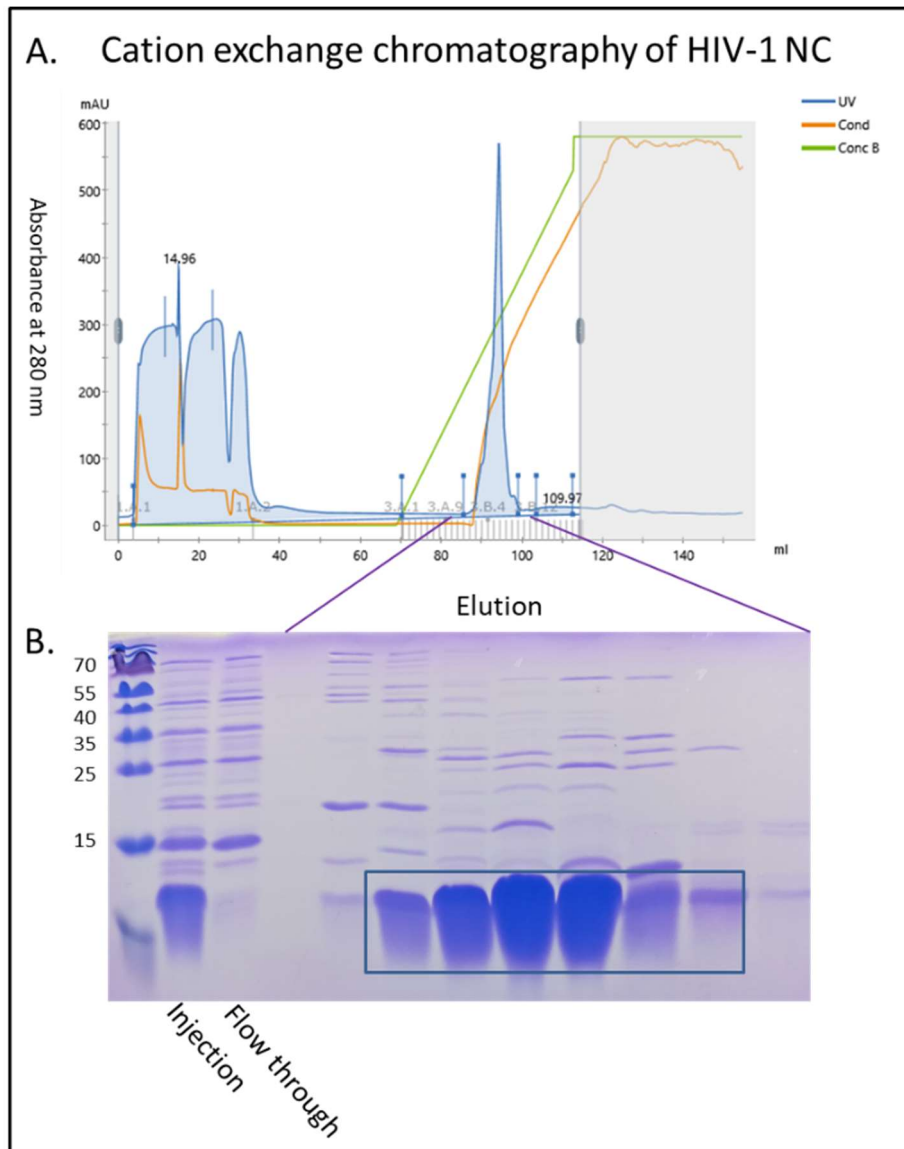


Figure 29. Purification of protein HIV-1 NC with cation exchange chromatography.

In figure 29, there are still some other proteins eluted along with protein HIV-1 NC. So, we need size exclusion chromatography to eliminate other proteins.

- **size exclusion chromatography**

We collected the fractions containing HIV-1 NC and loaded it on the size exclusion column, Superdex75g (16/60) ®. Proteins with different size were eluted at different time.

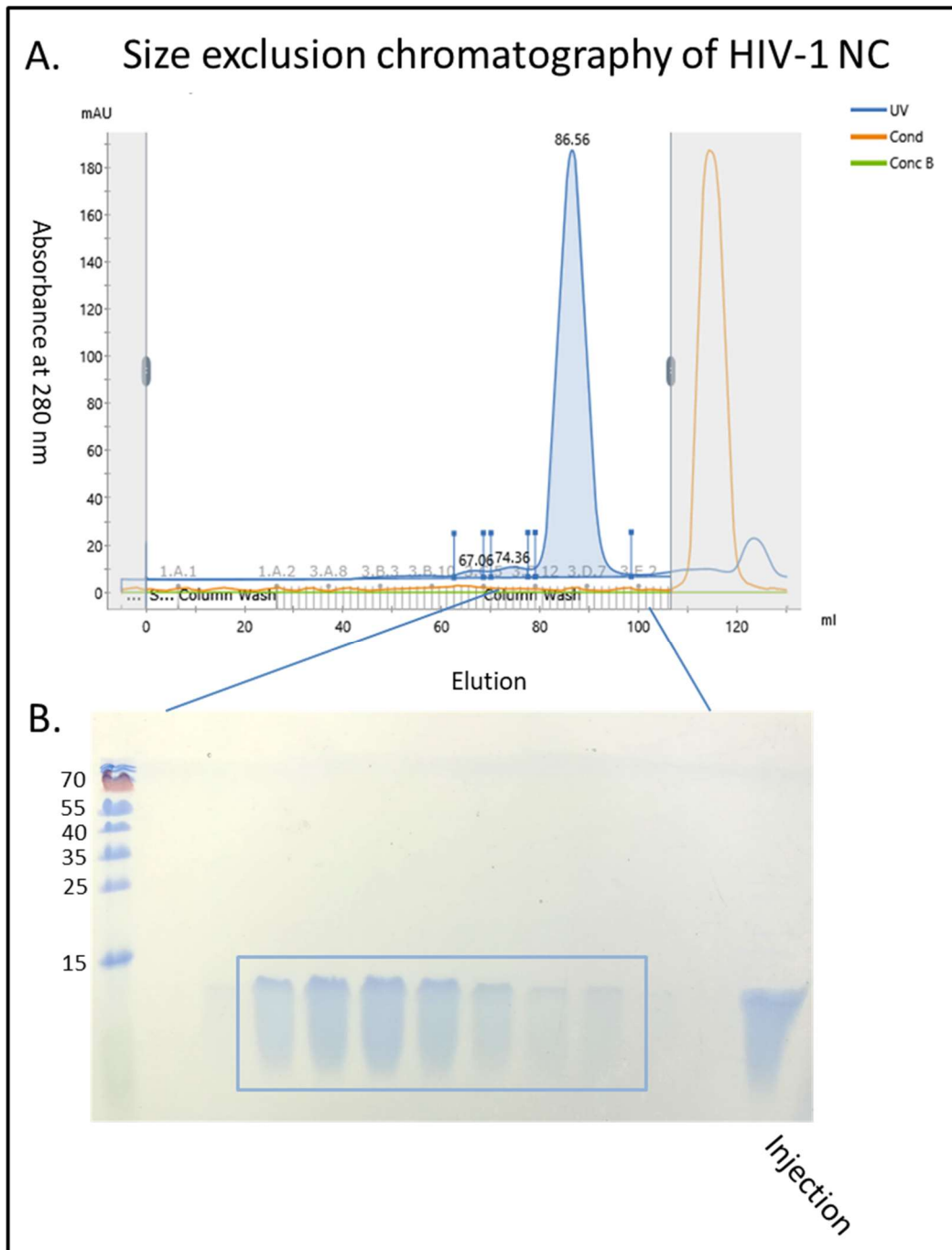


Figure 30. *Purification of protein HIV-1 NC with size exclusion chromatography.*

After this size exclusion chromatography, the protein HIV-1 NC is pure. We collected the protein and changed the buffer to acetate buffer at pH 6.5. Finally, we concentrated the protein to 500 μ M.

II Nuclear Magnetic Resonance (NMR) Analysis of Protein HIV-1 CA_{CTD}-SP1-NC, CA_{CTD}^{W184A}, M185A-SP1-NC and NC.

To verify the folding of protein HIV-1 CA_{CTD}-SP1-NC, CA_{CTD}^{W184A}, M185A-SP1-NC and NC, we added 10 % of D₂O to the protein to reach the final concentration 500 μM, final volume 500 μL. 1D NMR spectra and natural abundance 2D ¹H-¹⁵N SOFAST-HMQC were recorded using an AVANCE 600 MHz spectrometer equipped with a cryoprobe. For the 2D ¹H-¹⁵N SOFAST-HMQC experiment, it takes longer acquisition time for the unlabeled protein compared to the ¹⁵N labeled protein, but it is a good method to check the folding of the protein when the concentration of protein is high enough.

1. Folding of protein HIV-1 CA_{CTD}-SP1-NC, CA_{CTD}^{W184A}, M185A-SP1-NC and NC

The spectra (Figure 31) showed that the three proteins are well-structured. Their resonances are all spread over a large range of chemical shifts (from 0 to 10.5 ppm). In the 1D spectrum, the peaks between 0 and 1.5 ppm correspond to the protons carried by the methyl groups of the protein. the region between 6.5 and 10 ppm corresponds to the chemical shifts of the amide protons.

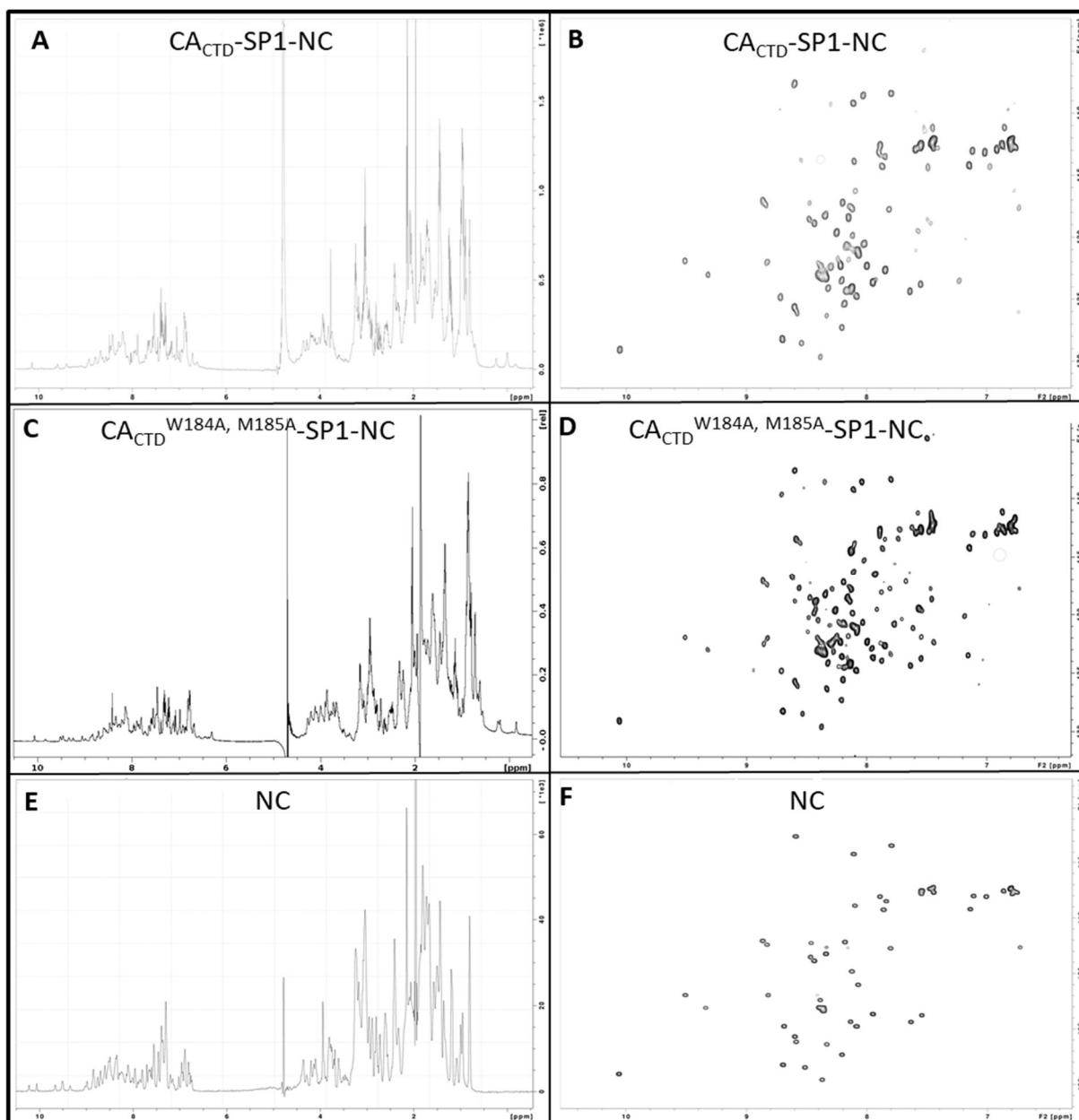


Figure 31. 1D spectra (left) and natural abundance 2D ^1H - ^{15}N SOFAST-HMQC spectra (right) of proteins HIV-1 $\text{CA}_{\text{CTD}}\text{-SP1-NC}$ (A, B), $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ (C, D) and NC (E, F).

2. The two mutations, W184A and M185A, abolish CA dimerization

To check the difference between proteins $\text{CA}_{\text{CTD}}\text{-SP1-NC}$ and $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$, we superimposed the spectra of $\text{CA}_{\text{CTD}}\text{-SP1-NC}$ on $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$. Spectra were recorded with samples at the same concentration and the NMR experimental parameters were identical.

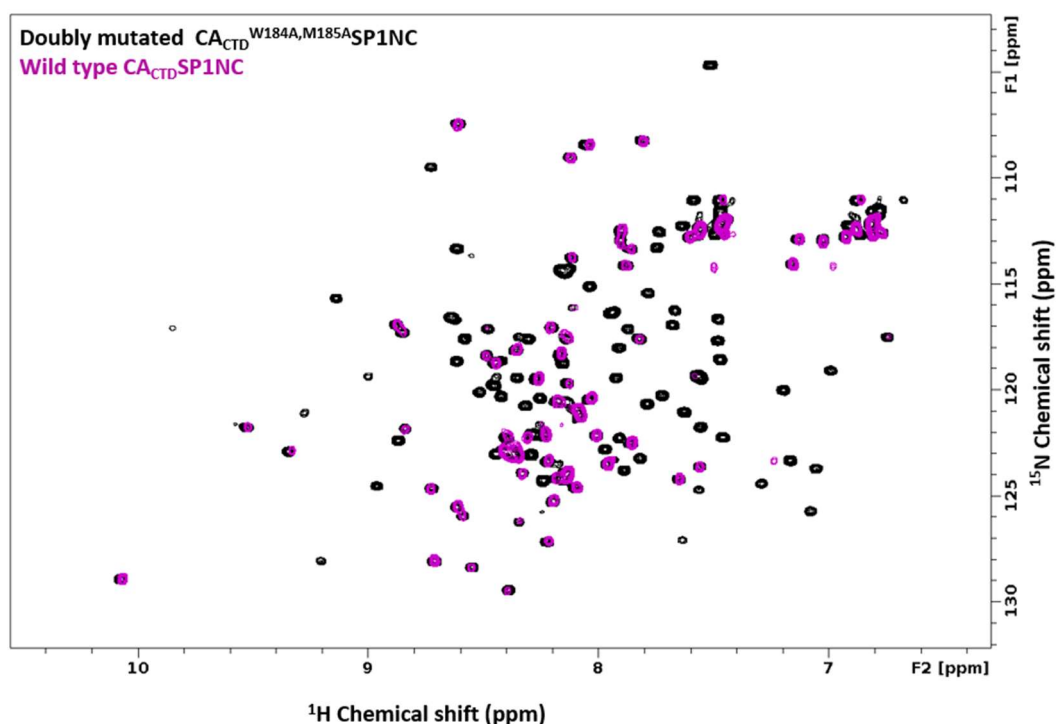


Figure 32. Natural abundance 2D ^1H - ^{15}N SOFAST-HMQC spectra of $\text{CA}_{\text{CTD}}\text{-SPI-NC}$ (magenta) and $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SPI-NC}$ (black).

Comparison between natural abundance ^1H - ^{15}N SOFAST-HMQC spectra obtained for the wild type $\text{CA}_{\text{CTD}}\text{-SPI-NC}$ and doubly mutated $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SPI-NC}$ (Figure 32) shows that their resonances superimposed well, especially the resonance of NC (by reference to the resonances labeled with residue number in Figure 35). However, many resonances of the CA_{CTD} domain (residue 144-231) are missing in wild type $\text{CA}_{\text{CTD}}\text{-SPI-NC}$ compared with the mutated $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SPI-NC}$ because of the line broadening due to a monomer-dimer equilibrium [167].

Thus, in the subsequent experiments, we used the mutated protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SPI-NC}$ to obtain the full assignment of the protein resonances, to determine its structure and to study its interaction with the maturation inhibitors.

3. NC does not seem to interact with $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ or SP1

To verify the structure of HIV-1 NC alone and in the protein $\text{CA}_{\text{CTD}}\text{-SPI-NC}$, we recorded the spectrum of HIV-1 NC and compared it with the spectrum of $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SPI-NC}$.

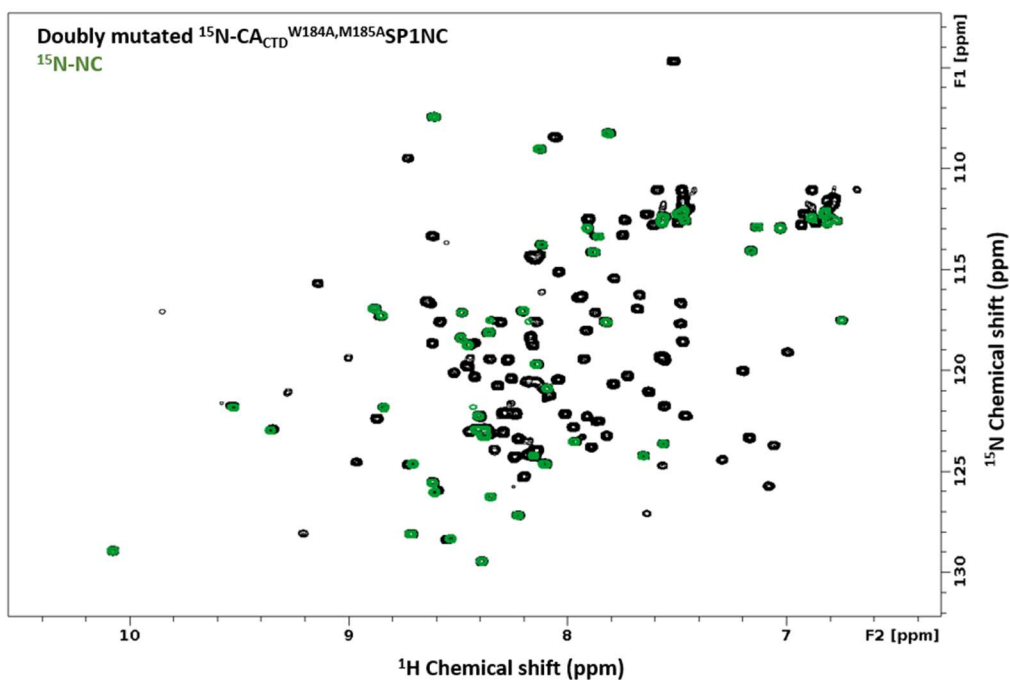


Figure 33. 2D ^1H - ^{15}N SOFAST-HMQC spectra of $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ (black) and NC (green).

Comparison of the natural abundance ^1H - ^{15}N SOFAST-HMQC spectra of HIV-1 NC and $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ (Figure 33) shows that the resonances of NC of both proteins superimpose well, which indicates that the conformation of NC is preserved in NC alone and in $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$. We can conclude that in $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$, NC does not seem to interact with the rest of the protein, $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}$ or SP1.

III Backbone assignment of protein CA_{CTD}^{W184A, M185A}-SP1-NC

In this part, we expressed and purified the ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC. We performed the backbone assignment of protein CA_{CTD}^{W184A, M185A}-SP1-NC with NMR experiments 2D SOFAST-HMQC, 3D NOESY-HSQC and 3D TOCSY-HSQC.

1. Expression and purification of ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC

In order to obtain good quality 3D spectra, thereby facilitating the assignment, the labeled protein must be produced. To do this, ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC has been expressed in *E. coli* BL21(DE3)pLysE. The cell cultures were carried out at 37 °C in a minimum medium containing ammonium ¹⁵N as the only source of nitrogen. When the OD 600 reached 0.5~0.6, the temperature was lowered to 30 °C and the induction of protein expression was achieved by adding 1.0 mM IPTG. In order to control this induction, samples were taken before the addition of IPTG and 3 hours later and then analyzed on SDS-PAGE gel.

The protocol for protein purification is the same as for the unlabeled protein, so we do not repeat the process again. The result of the purification is indicated on the SDS-PAGE gel (Figure 34).

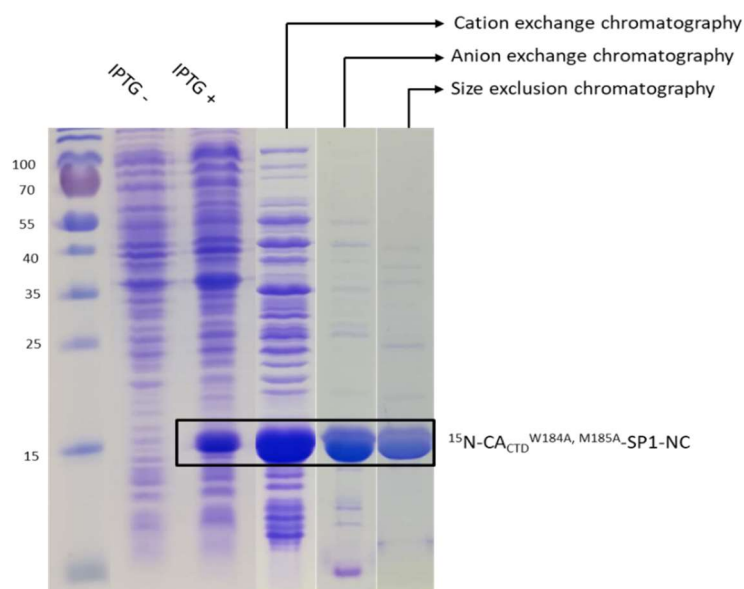


Figure 34. Expression and purification of ^{15}N labeled $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$.

After purification, the buffer for the pure protein ^{15}N $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ was exchanged to acetate buffer (25 mM Acetate Sodium, pH 6.5, 25 mM NaCl, 0.1 mM BME, 0.1 mM ZnCl_2). Protein was concentrated to 300 μM .

2. Backbone assignment of the protein ^{15}N $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$

The sample for NMR experiments contains the ^{15}N labeled $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ in 90 % of acetate buffer and 10 % D_2O at the concentration of 300 μM in a 3 mm NMR tube. All the experiments were performed in our group on an Avance III Bruker $\text{\textcircled{R}}$ spectrometer operating at 600.13 MHz and equipped with a cryoprobe. The spectrometer is equipped with cryogenic probe. All the experiments were performed at 303 K. The acquisition and processing of the experiments were obtained by Topspin $\text{\textcircled{R}}$ 3.5.7. The analysis of the spectra and the assignment were performed with CCPNMR $\text{\textcircled{R}}$ 2.4.2. We first recorded the 2D SOFAST-HMQC spectra to verify that the protein is folded. Then, we recorded 3D NOESY-HSQC, 3D TOCSY-HSQC and 3D HNHA experiments for backbone assignment.

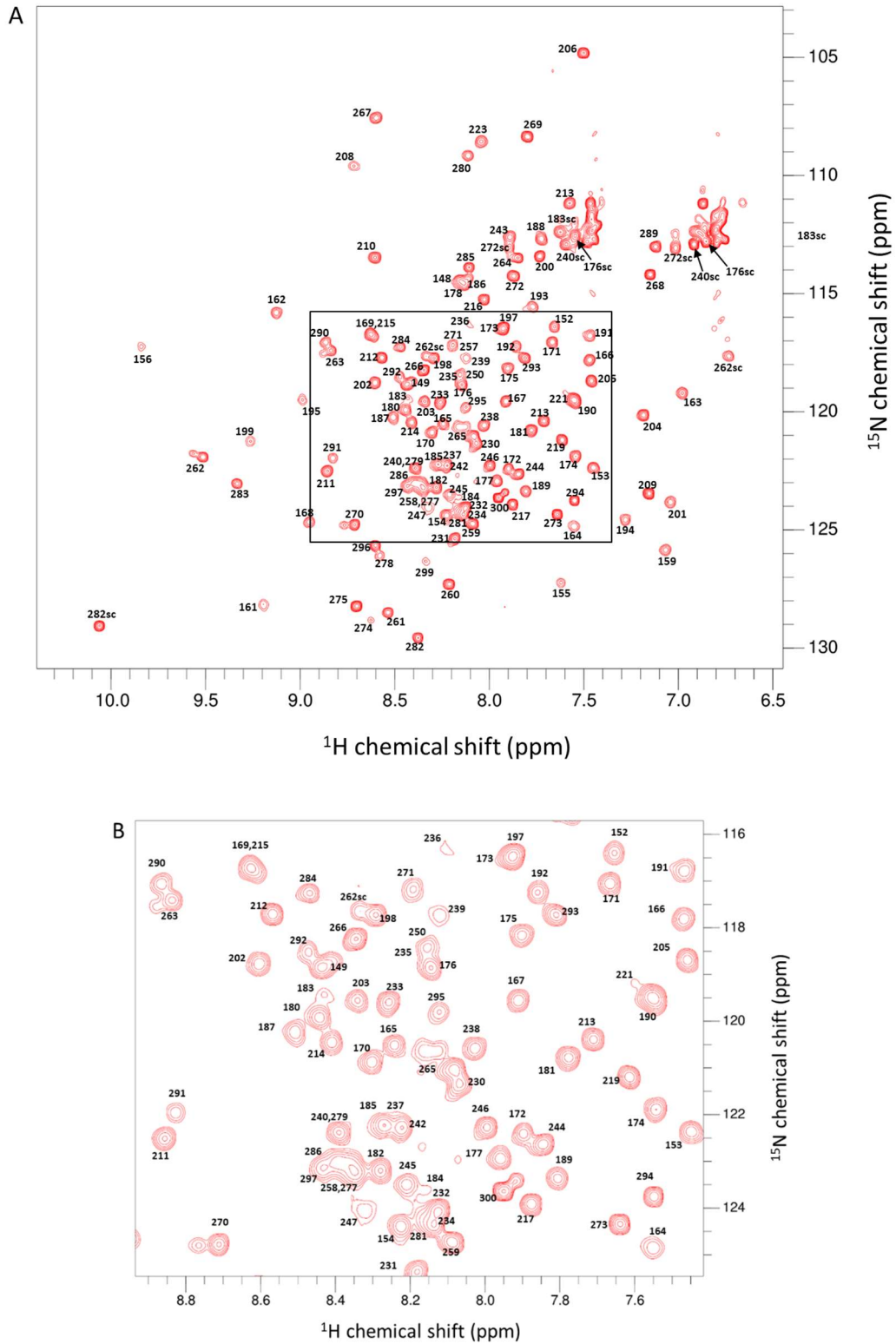


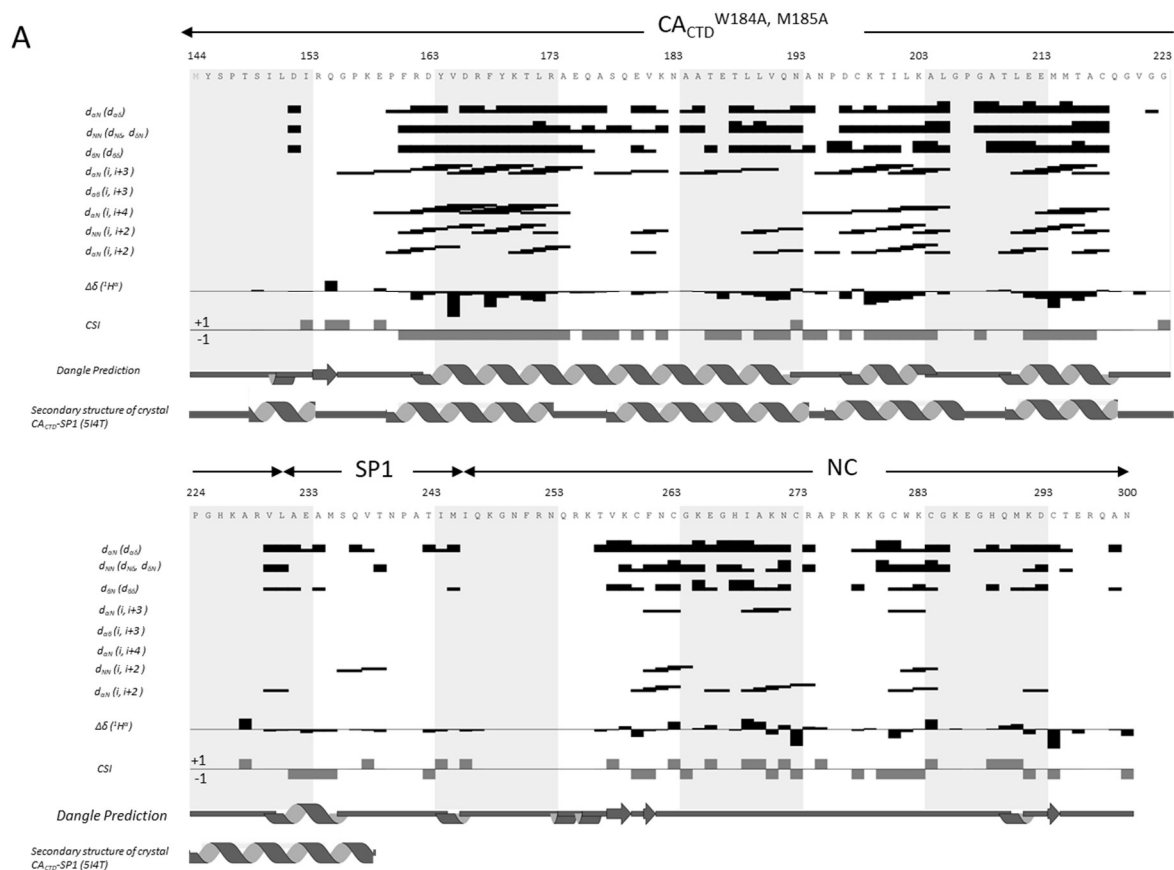
Figure 35. ^1H - ^{15}N SOFAST-HMQC spectrum of ^{15}N labeled $\text{CACTD}^{W184A, M185A}$ -SPI-NC. (A) The full ^1H - ^{15}N SOFAST-HMQC spectrum. (B) Zoom in the black square in A.

The ^1H - ^{15}N SOFAST-HMQC spectrum of ^{15}N labeled $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ is shown in Figure 35a. A zoom of the region in black square is shown in Figure 35b. In this spectrum, we assigned 126 residues, 80 % of all the residues. There are 18 residues we did not find the resonances and eight prolines do not have signal in the spectrum. The missing residues (except Proline) are divided into three groups: six residues at the N-terminal end of the protein; 14 residues at CA-SP1 and SP1-NC junctions; three residues at the C-terminal of NC.

The secondary chemical shifts for H_α protons [258] were calculated as the difference between measured chemical shifts and random-coil chemical shifts for each amino acid of protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ [259] and were corrected for next-neighbor effects [260]. These secondary chemical shifts are combined to give the chemical shift index (CSI) values that indicates the secondary structure type: -1 indicates alpha-helix and +1 beta-strand.

The secondary structure was predicted by DANGLE [261] with NMR derived information: through-space (NOEs) connectivity; secondary chemical shift and combined chemical shift index (CSI).

As predicted by DANGLE in figure 36, there is a long helix covering residues $\text{CA}^{162-193}$, two short helices spanning $\text{CA}^{197-203}$ and $\text{CA}^{210-218}$, a short helix at CA-SP1 junction ($\text{CA}^{230}\text{-SP1}^4$). In this figure, we added the secondary structure of the protein $\text{CA}_{\text{CTD}}\text{-SP1}$ (PDB code 5I4T). The secondary structure of the CA_{CTD} domain in $\text{CA}_{\text{CTD}}\text{-SP1}$ is very similar to our structure predicted by DANGLE. The helix in CA-SP1 junction is longer than that in our protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$. From residue 219 to 256 ($\text{CA}^{219}\text{-NC}^{11}$), there is much less information compared to the other domain. The secondary chemical shift of H_α in this domain is close to zero that indicates the disorder of this domain. The two zinc fingers in NC (ZF1: C260-C263-H268-C273; ZF2: C281-C284-H289-C294) is structured with $\Delta\delta$ ($^1\text{H}^\alpha$) varying from -0.954 to 0.494 ppm.



B

$d_{\alpha N} i, i+1$	residue that show an alpha(i) to amide(i+1) through-space/NOE connection
$d_{NN} i, i+1$	residue that show an amide(i) to amide(i+1) through-space/NOE connection
$d_{\beta N} i, i+1$	residue that show a beta(i) to amide(i+1) through-space/NOE connection
$d_{\alpha N} i, i+3$	residue that show an alpha(i) to amide(i+3) through-space/NOE connection
$d_{\alpha\beta} i, i+3$	residue that show an alpha(i) to beta(i+3) through-space/NOE connection
$d_{\alpha N} i, i+4$	residue that show an alpha(i) to amide(i+4) through-space/NOE connection
$d_{NN} i, i+2$	residue that show an amide(i) to amide(i+2) through-space/NOE connection
$d_{\alpha N} i, i+2$	residue that show an alpha(i) to amide(i+2) through-space/NOE connection
$\Delta\delta 1H\alpha$	Hydrogen alpha secondary chemical shift (difference to random coil value)
CSI	indicating the secondary structure type: -1 indicates alpha-helix and +1 beta-strand.

Figure 36. Secondary structure prediction of CA_{CTD}^{W184A, M185A}-SPI-NC. (A) the secondary chemical shift of H_α, chemical shift index (CSI), DANGLE predicted secondary structure and

the secondary structure of protein C_{ACTD}-SP1 (PDB code 5I4T) is shown. (B) the chart table shows a graphical representation of NOEs observed in the 3D NOESY-HSQC experiment.

IV Interaction between EP39 and protein

CA_{CTD}^{W184A, M185A}-SP1-NC

In this part, we made the titration of protein ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC by EP39 up to the limit of solubilization of the ligand. When the molecule is saturated, we did the assignment again and analyzed the effect of molecule EP39 on the folding and dynamics of protein CA_{CTD}^{W184A, M185A}-SP1-NC.

1. EP39 binds mainly to the SP1 domain on protein

CA_{CTD}^{W184A, M185A}-SP1-NC

A 300 μM ¹⁵N labeled CA_{CTD}^{W184A, M185A}-SP1-NC was prepared in 200 μL acetate buffer (25 mM sodium acetate, pH 6.5, 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM 2-mercaptoethanol, 10% D₂O) in a 3 mm NMR tube. A ¹H-¹⁵N SOFAST-HMQC reference spectrum was recorded.

The addition of 1 μL DMSO-*d*₆ 0.3% solution containing EP39 (corresponding to 1 equivalent of protein) in CA_{CTD}^{W184A, M185A}-SP1-NC was carried out and allowed to record a ¹H-¹⁵N SOFAST-HMQC spectrum without significant chemical shift perturbation (CSP) compared to the spectrum of the protein alone. This procedure was repeated 5 times, allowing 6 equivalents of EP39 to be added to the protein. Due to the low water solubility of EP39, a visible precipitate was observed in the NMR tube after adding 6 equivalents of EP39. The titration was interrupted and the NMR data were collected and analyzed.

As shown in Figure 36, the addition of EP39 led to the appearance of some resonances and the chemical shift perturbation of several residues. For example, the three residues in Figure 37B (1), which are missing in the spectrum of the protein alone, appeared when one equivalent of EP39 was added into the protein. With the addition of more EP39, the intensities of the three amino acids increased. For the residue in Figure 37B (2) (indicated by the arrow), there is no signal until 4 equivalents of EP39 were added to the protein. With the addition of EP39, the intensity of this residue increased slowly. At the end of the titration, the intensity of this residue is still very weak indicating the binding affinity varies for different residues. Besides the appearance of new resonances, there are also some residues whose chemical shift is perturbed

by the addition of EP39 as shown in figure 37B (3,4) indicated by the arrow.

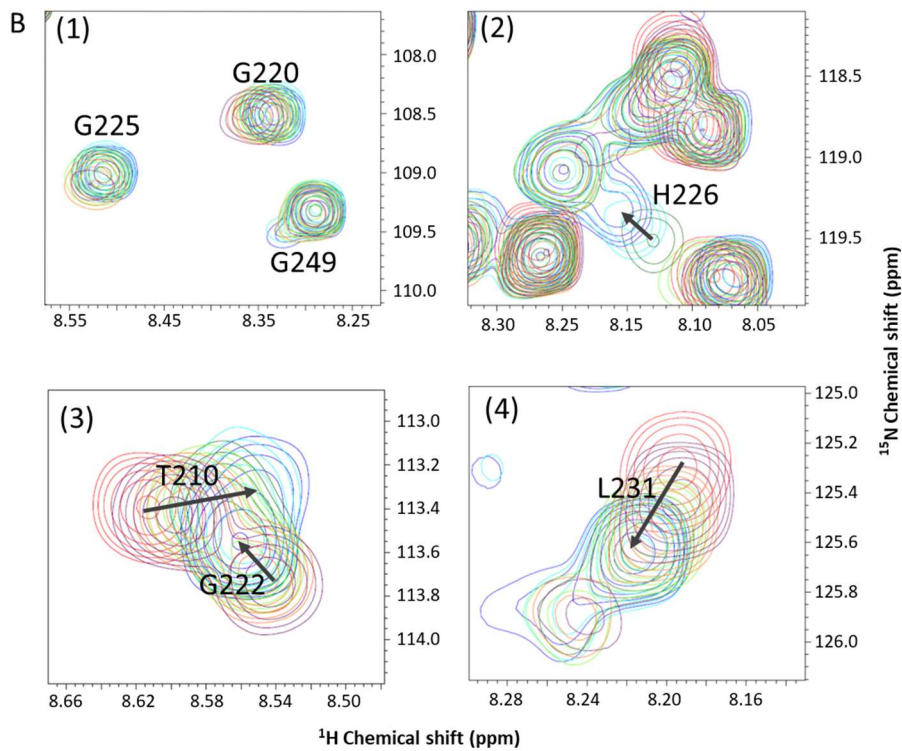
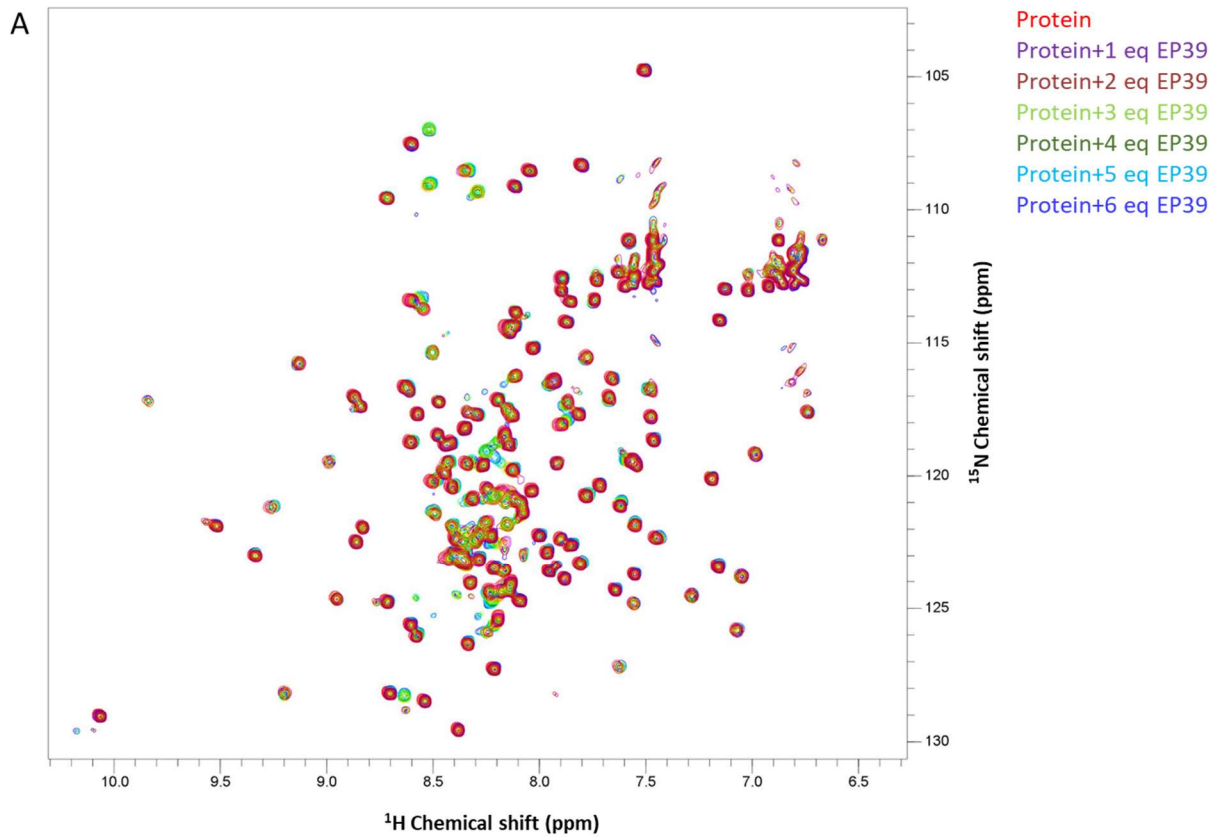


Figure 37. Solution NMR study of EP39 binding to $CA_{CTD}^{W184A, M185A}$ -SP1-NC. (A) Comparison of 1H - ^{15}N SOFAST-HMQC spectra recorded on ^{15}N labelled $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence and presence of EP39 from 1 to 6 equivalents. (B) Resonances corresponding to residue CA-G²²⁰, CA-G²²⁵ and NC-G²⁴⁹ appear after adding 1 equivalent EP39. Resonances of CA-H²²⁶ appear after adding 4 equivalents of EP39. After 4 eq EP39, there is no more new resonances appears. Residues CA-T²¹⁰, CA-G²²² and CA-L²³¹ show a chemical shift perturbation when EP39 binds to the protein.

2. Assignment of the protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the presence of EP39.

The newly resonances that have appeared are still unknown. To get the full NMR information of the protein, we did a new assignment with protein ^{15}N $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the presence of six equivalents of EP39. The NMR experiments recorded for the assignment is the same as for the protein ^{15}N $CA_{CTD}^{W184A, M185A}$ -SP1-NC alone.

The superposition of 1H - ^{15}N SOFAST-HMQC spectra recorded on ^{15}N labeled $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence and in the presence of 6 equivalents of EP39 is shown in Figure 37 with the residue number labeled. In the absence of EP39, the amide resonances of residues CA²¹⁸ to NC²⁵⁷ (encompassing the fourteen C-terminal residues of CA, SP1 and the twelve N-terminal residues of NC) exhibit weak dispersion with narrow linewidths, indicating that this region is highly flexible, in accordance with the structural properties of CA_{CTD} -SP1-NC [167]. Many resonances in this region are missing due to residue dynamics and fast proton exchange with the solvent. Upon EP39 addition, new resonances corresponding to residues CA^{220, 225, 226, 227, 228, 229} and NC^{248, 249, 250, 251, 252, 253, 256} appear as shown in Figure 38, which means the regions encompassing these residues are stabilized by EP39.

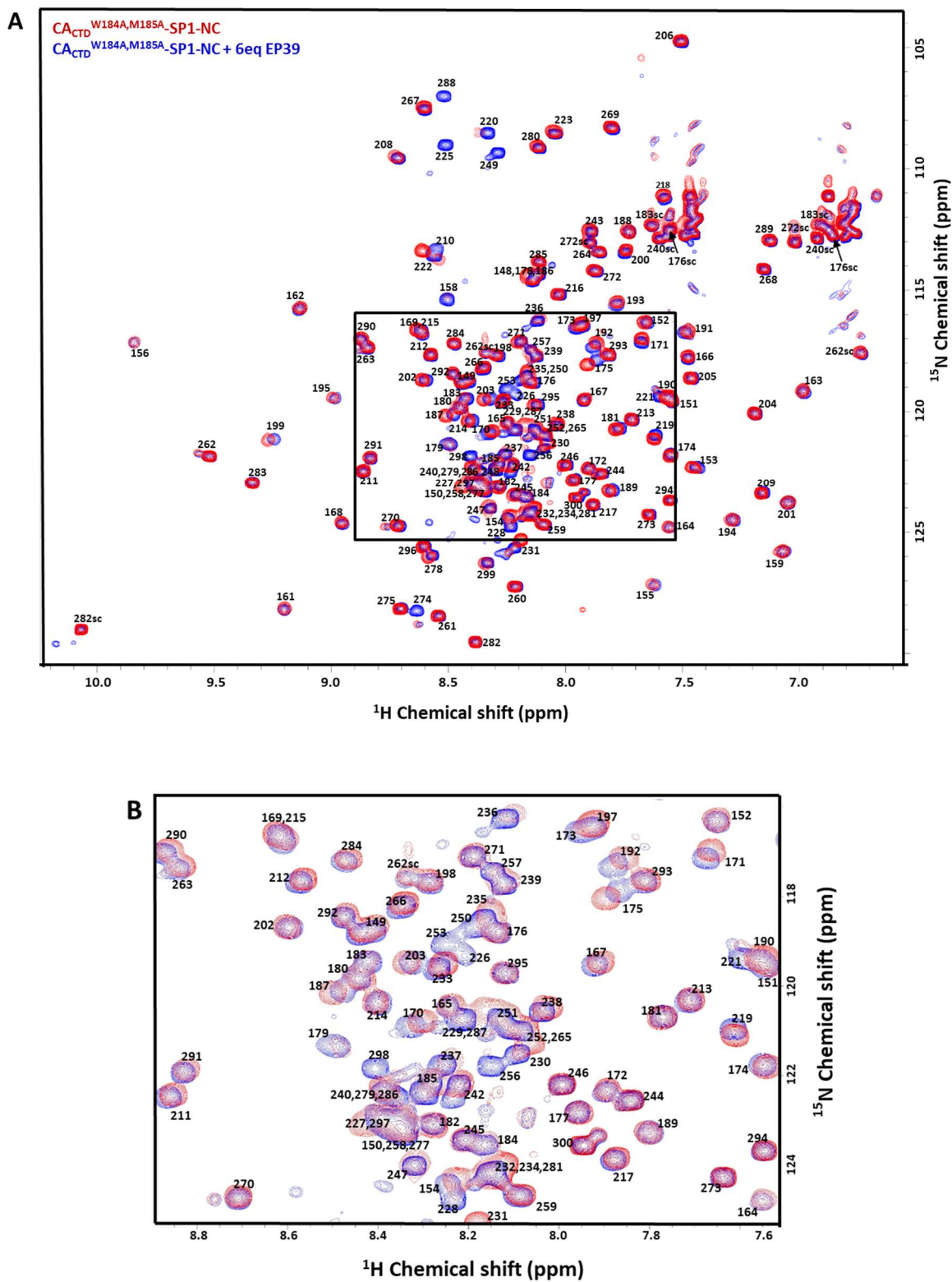


Figure 38. Solution NMR of EP39 binding to $CA_{CTD}^{W184A, M185A-SPI-NC}$. (A) Superposition of 1H - ^{15}N SOFAST-HMQC spectra recorded on ^{15}N labeled $CA_{CTD}^{W184A, M185A-SPI-NC}$ in the

absence (red) and in the presence of 6 equivalents of EP39 (blue) at 303K. Resonances are labeled with amino acids numbers. (B) Zoom in the black square in A.

3. CSP and intensity ratio

To better analyze the effect of EP39 on protein CA_{CTD}^{W184A, M185A}-SP1-NC, we summarised the chemical shift perturbation (CSP) and peak intensity ratios (I/I₀) with the protein CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39 (I: peak intensity in the presence of EP39, I₀: peak intensity in the absence of EP39) (Figure 39). The CSP of the amide moiety were normalized according to the equation: $\Delta\delta = [(\Delta\delta_N/6.5)^2 + (\Delta\delta_{H_N})^2]^{1/2}$ where $\Delta\delta_N$ and $\Delta\delta_{H_N}$ are the chemical shift changes in the nitrogen and proton dimensions respectively [247].

In figure 38, the domain (CA²¹⁸-NC²⁵⁷) undergoes a significant CSP and I/I₀ suggesting that EP39 modifies the dynamics and/or the conformation of the corresponding domain, consistent with the previously described effect of BVM on CA_{CTD}-SP1 [168]. We also note that CA-T²¹⁰ undergoes a relatively high CSP of 0.06 ppm. It could be caused by an allosteric structural change following the MI blocking of the transient interaction between SP1 and the first α -helix of CA_{CTD} [168]. The resonance intensity of residue CA-K¹⁵⁸ (in major homology region MHR) and CA-G²²² is six and five times stronger after EP39 addition consistent with our previously docking results of EP39 on CA^{CTD}-SP1 showing that the carbonyl in position C-28 of EP39 has a close contact with CA-K¹⁵⁸ [228].

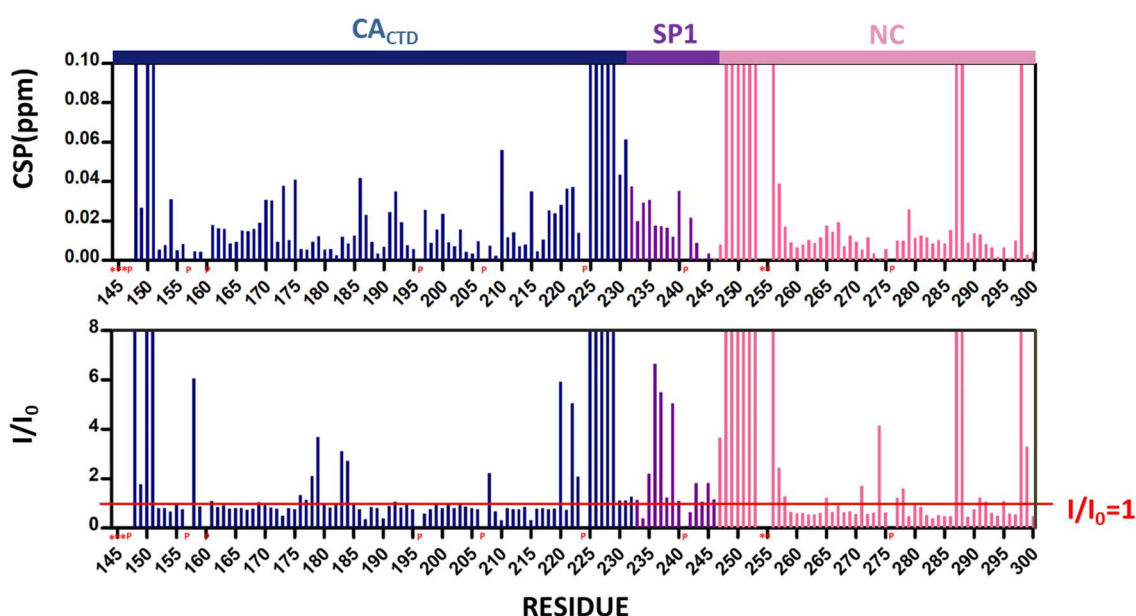


Figure 39. Chemical shift perturbations (CSP) and resonance intensity ratios (I/I_0) of the amide groups ($^1H/^{15}N$ combined chemical shift) in the presence and in the absence of EP39. The chemical shift perturbations and resonance intensity ratios of the residues whose $^1H-^{15}N$ resonance are seen only in the presence of EP39 were set at 0.1 and 8 respectively. The red line indicates the resonance intensity ratio of 1. Proline residues are labeled as P and unassigned residues are labeled as *. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

4. Attempt to evaluate the Kd of the interaction

We attempted to evaluate the dissociation constants (Kd) of EP39 on protein CA_{CTD}^{W184A, M185A}-SP1-NC by following the CSP of ten residues (CA-V²²¹, CA-L²³¹, SP1-A²³², SP1-E²³³, SP1-M²³⁵, SP1-S²³⁶, SP1-Q²³⁷, SP1-V²³⁸, SP1-T²³⁹, NC-T²⁵⁷) whose resonances are isolated in the spectrum. Kds were calculated using one-site binding model. Non-linear fitting was carried out with Graphpad Prism software with the equation: $Y = B_{max} * X / (Kd + X) + NS * X$. X is the ligand concentration, Y is the chemical shift perturbation, B_{max} is the maximum specific binding in the same units as Y and Kd is the equilibrium binding constant, in the same units as X. NS is the slope of non-specific binding in Y units divided by X units. These Kds could be approximate because of the slight precipitation of EP39 at high concentration during the experiment, but they give us ideas about the binding affinity of EP39 to different residues of SP1.

Of these ten residues (Figure 40), SP1-T²³⁹ has the highest binding affinity (52.4 μ M); SP1-S²³⁶, SP1-Q²³⁷ and SP1-V²³⁸ have relatively high binding affinities (179.6 μ M, 154.5 μ M and 204.4 μ M respectively); SP1-E²³³, SP1-M²³⁵ and NC-T²⁵⁷ have medium binding affinities (237.0 μ M, 235.4 μ M, 308.4 μ M respectively); CA-V²²¹, CA-L²³¹ and SP1-A²³² have weak binding affinities (669.6 μ M, 678.4 μ M and 691.9 μ M respectively). The higher binding affinity at SP1-Q²³⁷V²³⁸T²³⁹ is consistent with previous study which shows that SP1-Q²³⁷V²³⁸T²³⁹ motif polymorphism confers intrinsic resistance to BVM [209].

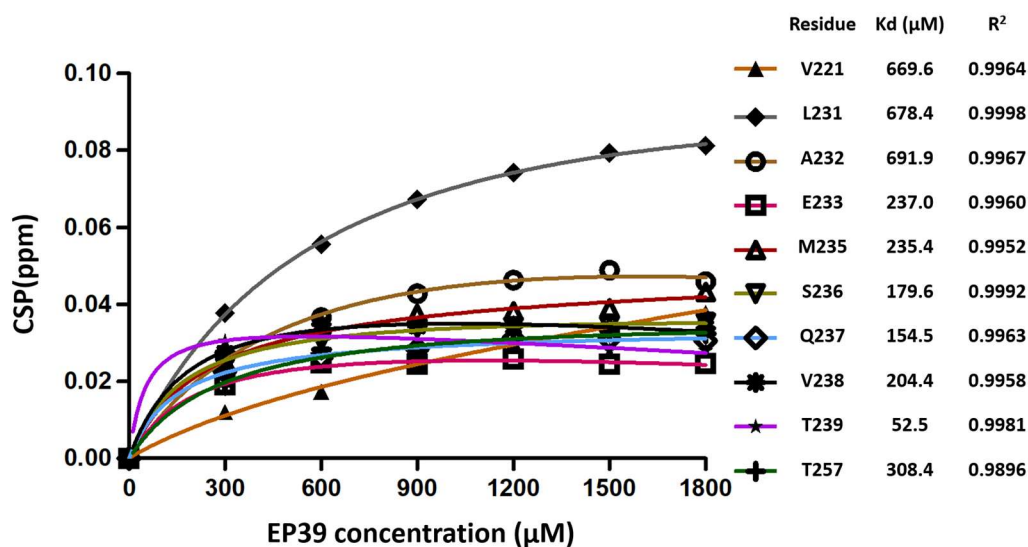


Figure 40. Variations of chemical shift perturbations (CSP) of ten residues (CA-V²²¹, CA-L²³¹, SPI-A²³², SPI-E²³³, SPI-M²³⁵, SPI-S²³⁶, SPI-Q²³⁷, SPI-V²³⁸, SPI-T²³⁹, NC-T²⁵⁷) isolated in the ¹H-¹⁵N SOFAST-HMQC spectrum as a function of EP39 whose concentration increase from one to six equivalents of protein. Kds were computed as described in Materials and Methods.

V Structure of CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39

In this part, we calculated the structure of protein CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39 with the NMR information we got in Part 4.

The NMR structures were calculated with ARIA 2.3.2 software [262] coupled to CNS 1.2.1 [250] based on the inter-proton distance restraints from NOESY spectrum and the dihedral angle restraints from TALOS+ program [263]. The experimental NMR restraints are summarized in Table 6. Two zinc coordination were added (C²⁶⁰-C²⁶³-H²⁶⁸-C²⁷³ and C²⁸¹-C²⁸⁴-H²⁸⁹-C²⁹⁴) to form the zinc finger for structure calculation of NC. ARIA was run with explicit water refinement using standard protocols. The ten lowest total energy conformers were selected out of the 500 structures calculated in the final run. The average RMSD was calculated on the backbone atoms between the best structure which has the lowest energy, and each of the remaining nine structures on the first, second and third domain respectively.

The coordinates of the ten best structures, the chemical shifts, the constraints and the list of peaks have been deposited in the Protein Data Bank under accession code 6RWG.

CA_{CTD}^{W184A, M185A}-SP1-NC contains six alpha helices: H1 (CA¹⁶¹⁻¹⁷²), H2 (CA¹⁷⁹⁻¹⁹²), H3 (CA¹⁹⁶⁻²⁰⁸), H4 (CA²¹²⁻²¹⁸), H5 (CA²²⁷-SP1²³⁹) and H6 (SP1²⁴¹-NC²⁵¹); two zinc fingers: ZF1 (NC²⁶⁰⁻²⁷³) and ZF2 (NC²⁸¹⁻²⁹⁴) connected by seven amino acids (Figure 41A).

To better analyze the structure of CA_{CTD}^{W184A, M185A}-SP1-NC, we structurally divided the protein into three domains: the first one, CA¹⁵¹⁻²²⁰, encompassing the four α -helices in CA; the second one, CA²²¹-NC²⁵¹, constituted by CA-SP1 (H5) and SP1-NC (H6) helical junctions and finally the third one, NC²⁵²⁻²⁹⁴, formed by the two zinc fingers connected by seven amino acids (Figure 41B). The N and C ends of the protein were excluded because of their high flexibility. Then, each domain of the 10 structures was superimposed to evaluate their convergence. The first domain CA¹⁵¹⁻²²⁰ displays a good convergence with a backbone atoms RMSD of 1.06 Å (left Figure 41B). Six amino acids, CA²²¹⁻²²⁶, starting the second domain are very flexible and display different orientations (Figure 41B middle). A flexible elbow between H5 and H6, corresponding to the QVT motif, allows these two helices to adopt different relative orientations. (Figure 41B middle). The backbone atoms RMSDs for H5 and H6 are 0.74 Å and

1.34 Å respectively. The last domain contains two zinc fingers, ZF1 and ZF2, connected by seven residues linker (NC²⁷⁴⁻²⁸⁰). The flexibility of this linker allows the two zinc fingers to adopt variable orientations with respect to each other. This relative motion can modulate the nucleic acid binding property [119]. RMSDs for ZF1 and ZF2 are 0.92 Å and 1.19 Å respectively on the backbone atoms for the 10 structures (Figure 41B right).

The structure which has the lowest energy was superimposed on the hexameric crystal structure of CA_{CTD}SP1 (PDB code: 5I4T). The RMSD calculated on CA¹⁴⁸-SP1²³⁹ backbone atoms between our structure and the chain G of the hexamer is 1.15 Å (Figure 41C). This result demonstrates that the structure of CA_{CTD} in monomeric form is consistent with the structure in hexameric form and the two mutations (W184A and M185A) do not modify the structure of CA_{CTD}. We also observed two different orientations for H5 (CA²²¹⁻²²⁶) relative to H4 (CA²¹²⁻²¹⁸) between our structure and the hexameric one as shown in Figure 40C, demonstrating again the flexibility of the residues between H4 and H5. H6 which is missing in the crystal and cryo-EM structures is well folded in our NMR structure (Figure 41A) and adopts different orientations compared to H5.

Table 6. Statistics for the top 10 NMR structures of monomeric mutant HIV-1 CA_{CTD}^{W184A}, M185A-SP1-NC

Restrains	
Total number of NMR restraints	2121
Total number of NOE distance restraints	1898
Long range ($ i-j > 5$)	90
Medium range ($2 \leq i-j \leq 4$)	548
Sequential range $ i-j =1$	718
Intra-residue NOEs	542
Total number of dihedral restraints	223
Restraint statistics	
Distance violations per structure	
0.2-0.3 Å	2.90
0.3-0.5 Å	0.20
>0.5 Å	0.00
Root mean square (r.m.s.) on distance violations per constraint (Å)	0.02
Maximum distance violation (Å)	0.42
Dihedral angle violations per structure	
5-10°	0.80
>10°	0.00
r.m.s. on dihedral violations per constraint (°)	0.73
Maximum dihedral angle violation (°)	8.78
Ramachandran analysis of residues	
Most favored regions (%)	87
Allowed regions (%)	10
Disallowed regions (%)	3

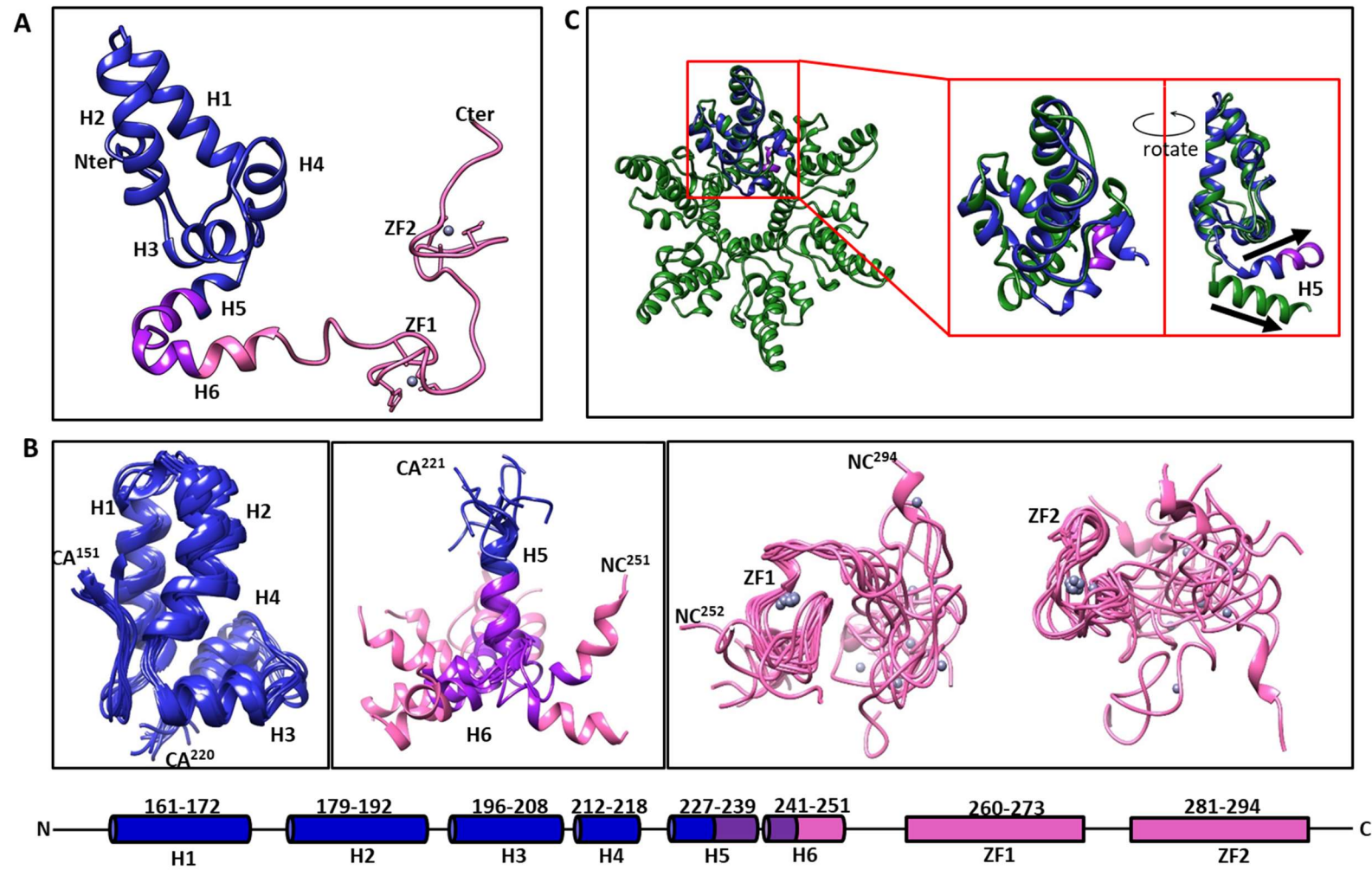


Figure 41. NMR structure of $CA_{CTD}^{W184A, M185A}$ -SPI-NC in the presence of EP39. The secondary structure of the protein is indicated at the bottom. H: helix; ZF: zinc finger. Residues of $CA_{CTD}^{W184A, M185A}$, SPI and NC are colored in blue, purple and pink respectively. (A) the structure which has the lowest energy is represented. From N-terminal to C-terminal, this structure contains six helices: H1 (residue $CA^{161-172}$), H2 (residue $CA^{179-192}$), H3 (residue $CA^{196-208}$), H4 (residue $CA^{212-218}$), H5 (residue CA^{227} -SPI²³⁹), H6 (residue SPI²⁴¹-NC²⁵¹) and two zinc fingers: ZF1 (residue NC²⁶⁰⁻²⁷³) and ZF2 (residue NC²⁸¹⁻²⁹⁴). (B) $CA_{CTD}^{W184A, M185A}$ -SPI-NC was divided into three domains. The ten best structures are superimposed based on each domain. Left: superimposition on the first domain, $CA^{151-220}$, shows a RMSD of 1.06 Å calculated on the backbone atoms. Middle: Superimposition on the second domain, CA^{221} -NC²⁵¹, performed on H5 shows a RMSD of 0.74 Å. Right: superimpositions on the third domain, NC²⁵²⁻²⁹⁴, performed on ZF1 and ZF2 respectively show RMSD of 0.92 Å and 1.19 Å respectively. (C) Superimposition on the region CA^{148} -SPI²³⁸ between the NMR structure (Figure 40A) and chain G (green) in CA_{CTD} -SPI hexamer (PDB code: 5I4T). H5 in our structure and in chain G of the crystal structure has different orientations as indicated by black arrows. PDB code for $CA_{CTD}^{W184A, M185A}$ -SPI-NC in the presence of EP39 is 6RWG.

VI Relaxation studies of CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39.

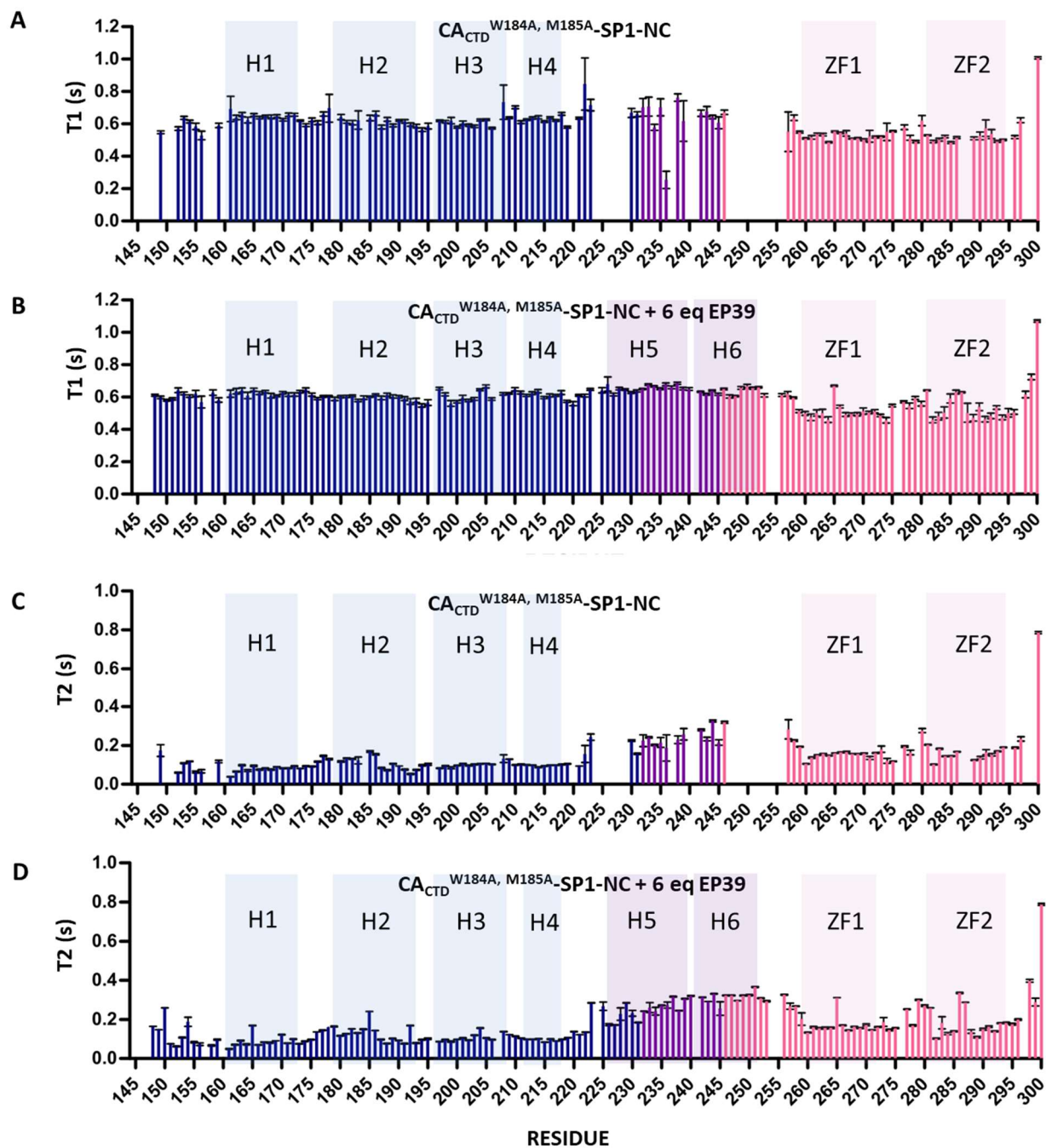
The ¹⁵N backbone relaxation experiments were performed on 300 μM CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of 6 equivalents of EP39. T1, T2 and HetNOE ¹⁵N experiments used standard Bruker pulse sequences [264]. The T1 experiments were recorded with 5, 10, 20, 30, 40 (repeated), 60, 80, 100 (repeated), 200, 300, 400, 600, 800 and 1200 ms for recovery delay. T2 experiments were recorded with 1, 2, 4, 8, 10 (repeated), 12, 14, 16, 18, 20 (repeated), 22, 24, 28 and 32 ms for recovery delay. The 15N-{1H} NOE (HetNOE) values were taken as the ratio between the intensities recorded with and without saturation of the amide protons. The experiments were processed with Dynamic Center 2.3 in Topspin 3.5 pl 7. The fit function for T1 calculation is $f(t) = I_0 * [e^{-t/T1}]$, fit function for T2 calculation is $f(t) = I_0 * [e^{-t/T2}]$, NOE: intensity ratio (I_{sat}/I_{un-sat}).

¹⁵N T1, T2 relaxation times, T1/T2 ratio and heteronuclear NOE (HetNOE) values are presented in Figure 42. T1/T2 ratio provides a good estimation of the rate at which each N-H vector reorients with global tumbling. HetNOE values are indicative of the magnitude of local sub-nanosecond motions with high values corresponding to restricted motions and low values to high-amplitude motions.

As shown in figure 42, in the absence of EP39, the N- and C-terminal ends of CA_{CTD}^{W184A, M185A}-SP1-NC are highly flexible. The first domain containing four α-helices is the most stable part in CA_{CTD}^{W184A, M185A}-SP1-NC. The second domain is the most flexible part and many amino acids ¹H-¹⁵N correlation peaks are even undetectable. The third domain containing the two stable zinc fingers connected by seven amino acids is more flexible than the first domain but is more ordered than SP1.

In the presence of EP39, changes are observed in the second domain. New correlation peaks corresponding to the residues (CA²²⁵, CA²²⁶, CA²²⁷, CA²²⁸, CA²²⁹, NC²⁴⁸, NC²⁴⁹, NC²⁵⁰, NC²⁵¹) of the second domain appear in ¹H-¹⁵N SOFAST-HMQC spectrum. Besides, the average values of T1/T2 and HetNOE for H5 (Figure 42F, H) are slightly higher than that of H6 suggesting H5 is less dynamic than H6. The residues SP1-Q²³⁷, SP1-V²³⁸ and SP1-T²³⁹, forming the QVT motif located at the elbow between H5 and H6, have lowest HetNOE values compared to other residues in the second domain, especially SP1-T²³⁹ with lowest HetNOE value. After adding

EP39, HetNOE value of SP1-T²³⁹ increases from -0.324 to -0.164 which means this residue is less dynamic in the presence of EP39. Conversely, the first and third domains do not seem to be affected by the addition of EP39.



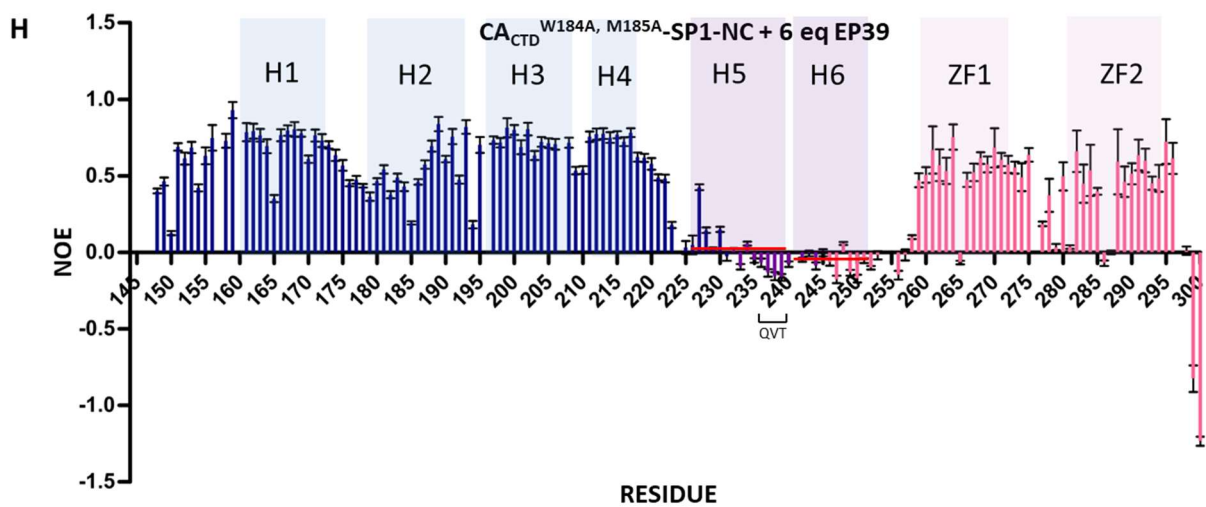
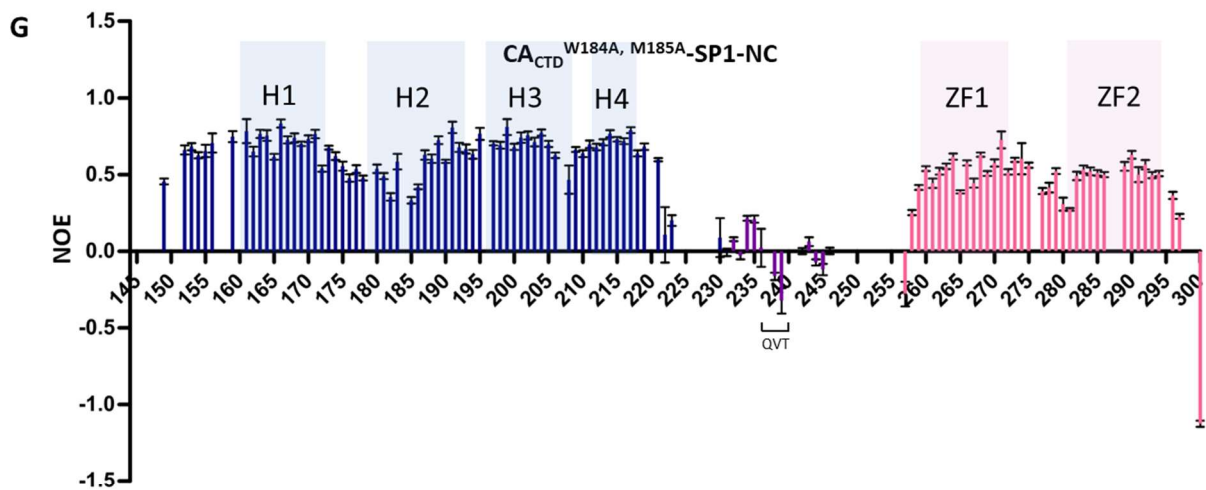
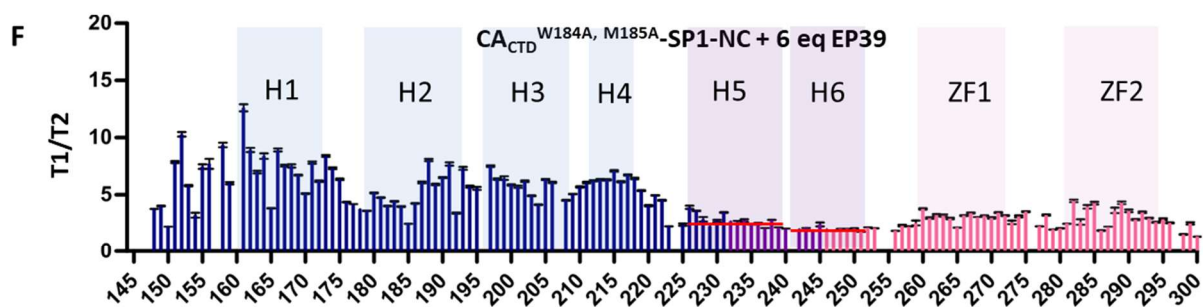
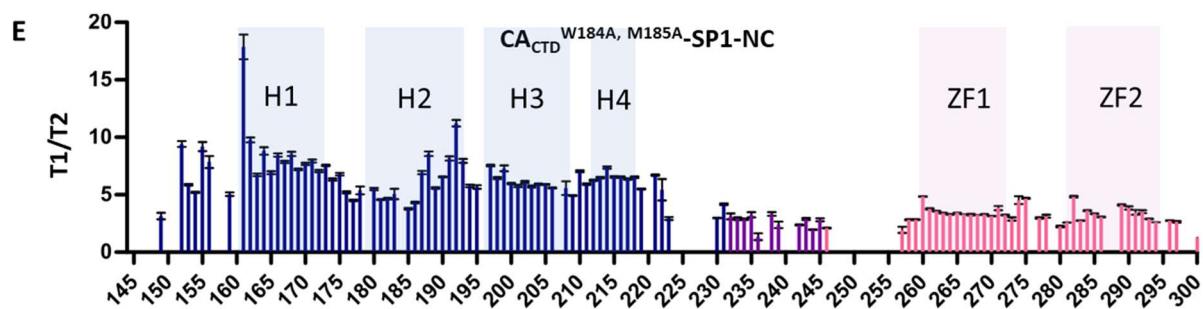


Figure 42. Dynamics characterization of $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence and presence of 6 eq EP39 from ^{15}N NMR relaxation data recorded at 600 MHz. T1 (A, B); T2 (C, D); T1/T2 (E, F) and 1H - ^{15}N HetNOE (G, H). Residues of $CA_{CTD}^{W184A, M185A}$, SP1 and NC are colored in blue, purple and pink respectively. The boxes in light blue indicate the four α -helices in CA_{CTD} (H1, H2, H3 and H4). The two α -helices in the second domain (CA-SP1 and SP1-NC junctions) are indicated by the light purple boxed (H5 and H6). The two zinc fingers in NC are shown in pink boxes (ZF1 and ZF2). The missing residues are proline or unassigned.

VII Interaction between PF46396/CB310F3 and CA_{CTD}^{W184A, M185A}-SP1-NC.

The first-in-class maturation inhibitor, BVM, performed well in the clinical trials. However, the naturally occurring polymorphisms within SP1-QVT motif are associated with the lack of response in other patients [202,208,265]. The lack of response of patients to BVM in a significant percentage leads to the discontinuation of efforts to develop BVM for clinical use. Thus, many research groups take efforts to modify the BVM or screen molecules to find other maturation inhibitor candidates.

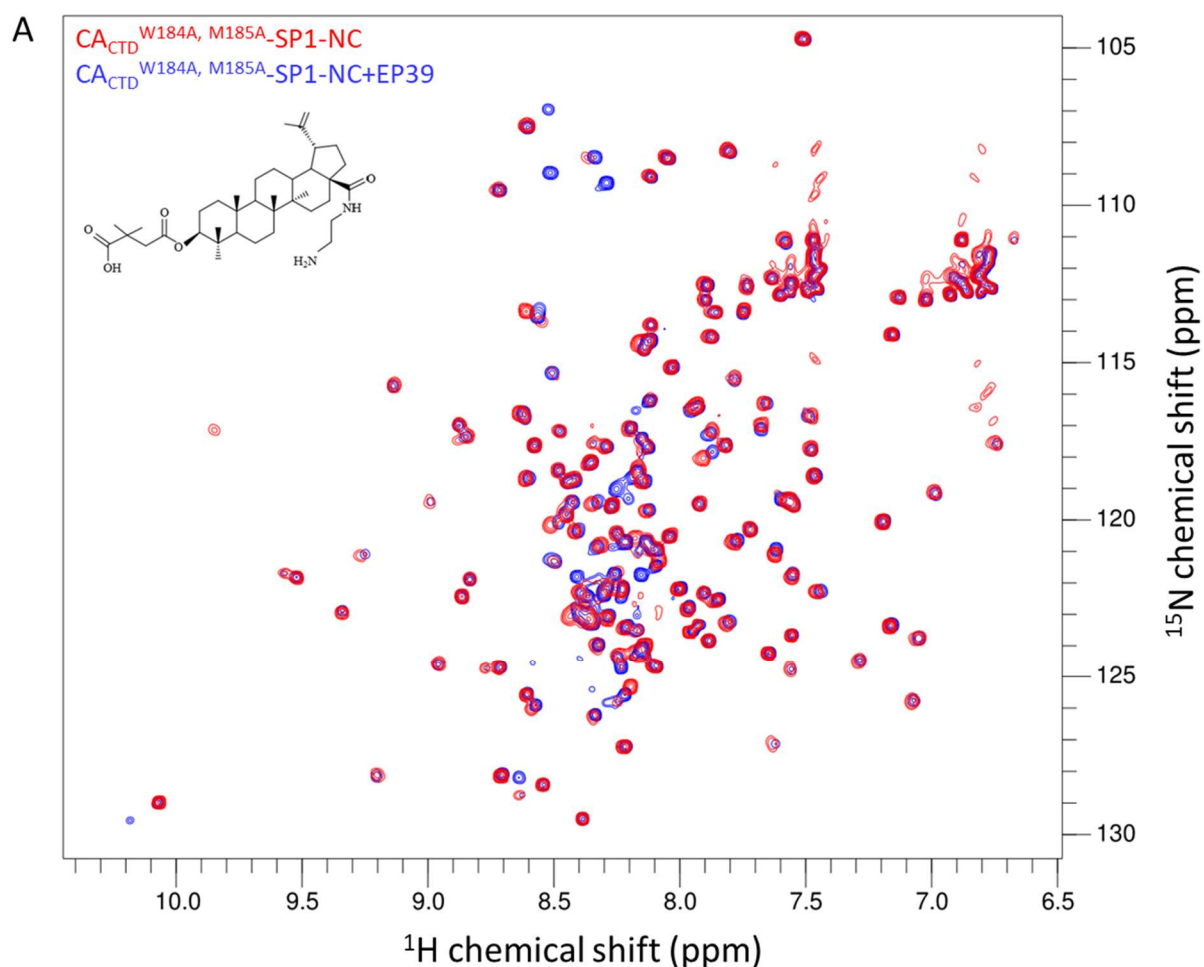
1. PF46396

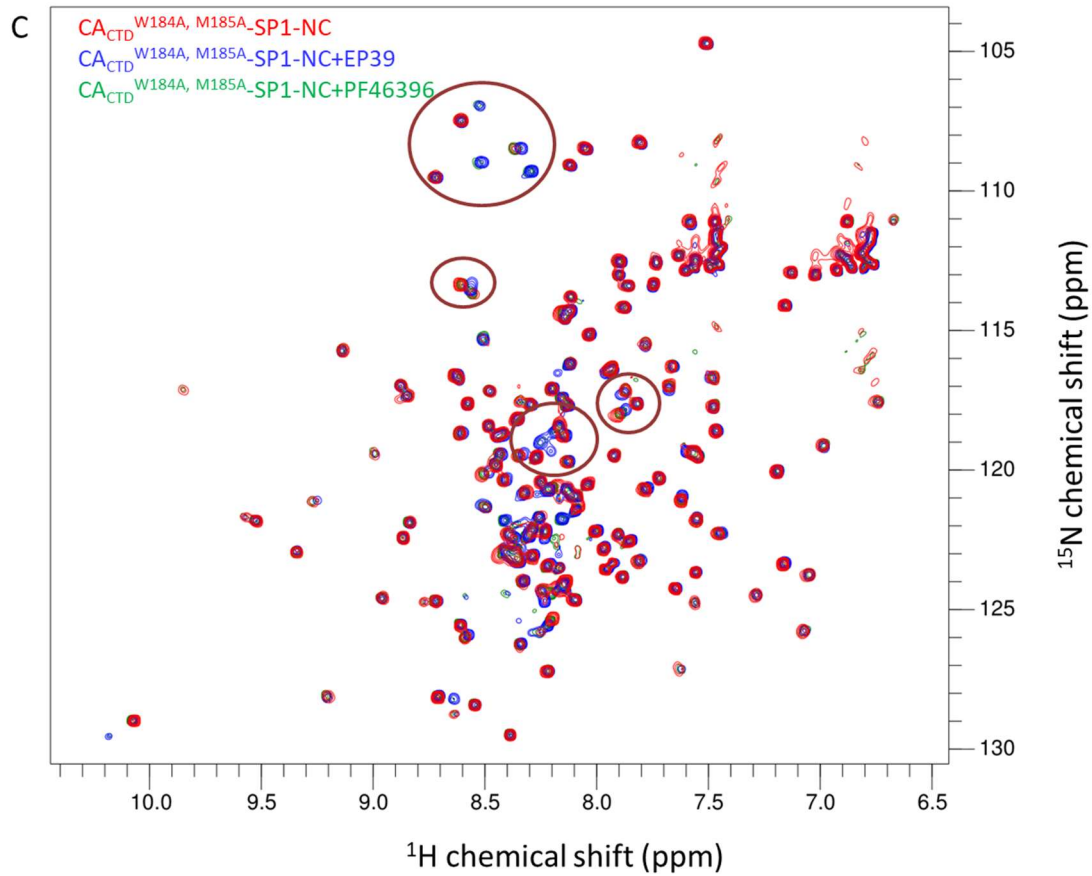
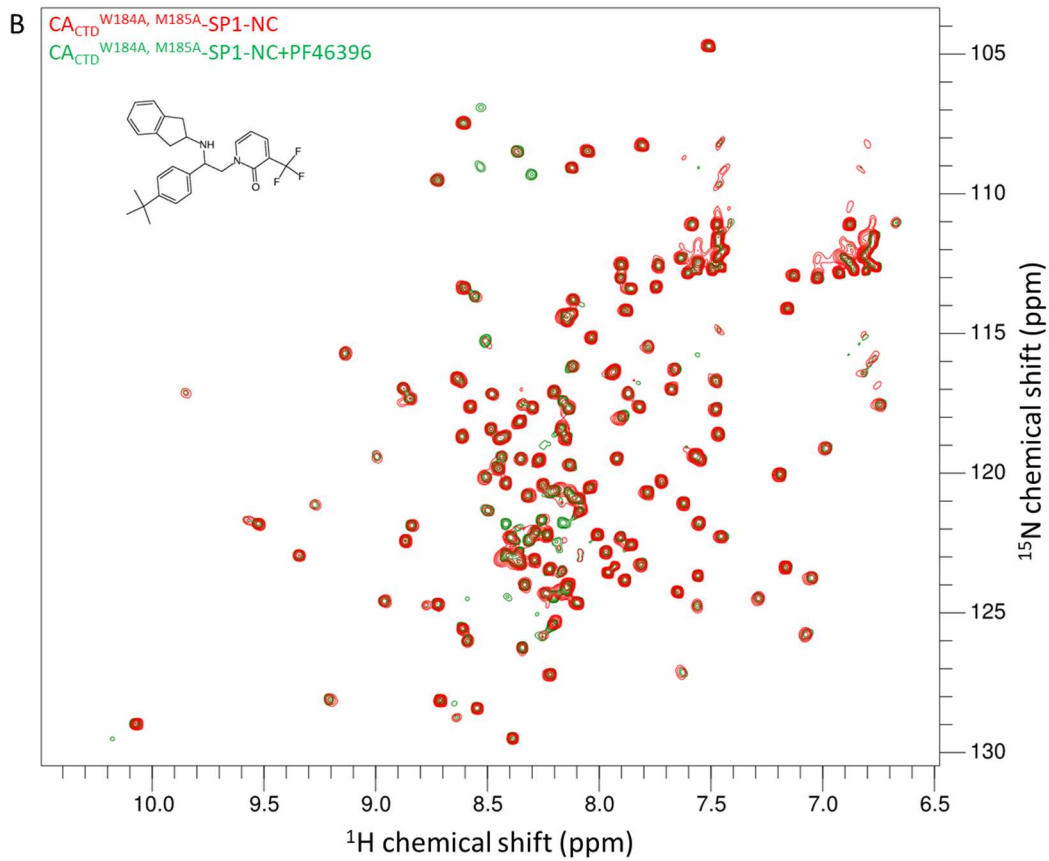
PF46396, a pyridone-based compound, blocks CA-SP1 processing and inhibits the maturation of HIV-1 particles [215]. The SP1 polymorphisms have minimal effect on susceptibility of PF-46396. PF-46396 resistance maps to three domains of HIV-1 Gag: CA-SP1 cleavage site where BVM resistance maps, residue CA-I201 and CA MHR [225].

PF46396 has same effect on the cleavage process of Gag even it has a completely different structure compared to BVM. To better understand the mechanism of different MIs, we checked the effect of PF46396 on protein ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC by NMR experiments at 303 K. PF46396 was solubilized in DMSO-*d*₆ at final concentration 60 mM. Six equivalents of PF46396 were added to ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC.

As shown in Figure 43B, new resonances appeared in ¹H - ¹⁵N SOFAST-HMQC spectrum after the addition of PF46396. This effect is similar with the effect caused by EP39 (Figure 43A). To better analyze and compare the effects of PF46396 and EP39, we superimposed the spectra recorded by protein ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39 and PF46396 (Figure 43C). The two spectra, ¹H - ¹⁵N SOFAST-HMQC recorded on CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39 and PF46396, are very close too each other suggesting that PF46396 and EP39 have similar effect on protein CA_{CTD}^{W184A, M185A}-SP1-NC. Then, we zoomed in on several residues. The two spectra both have new resonances (CA-G²²⁰, CA-G²²⁵ and NC-G²⁴⁹). While, the chemical shift of the new resonances in the two spectra are slightly different (Figure 43D). This indicates the binding of PF46396 decrease the dynamics of these residues like the effect of EP39, but in different ways. The resonance of residue CA-H²²⁶

appeared in the spectrum of CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39, but not appeared in the presence of PF46396 (Figure 43D). Combined with the fact that this resonance appeared only in the presence of 4 eq of EP39, we conclude that EP39 binds to the protein CA_{CTD}^{W184A, M185A}-SP1-NC in a more efficient way than PF46396. The chemical shift perturbations for residues CA-E¹⁷⁵, CA-Q¹⁹², CA-T²¹⁰, CA-G²²² and NC-K²⁹³ are different in the presence of PF46396 and EP39 (Figure 43D). All these results indicate that PF46396 and EP39 have similar effect on protein CA_{CTD}^{W184A, M185A}-SP1-NC, but EP39 is more efficient than PF46396.





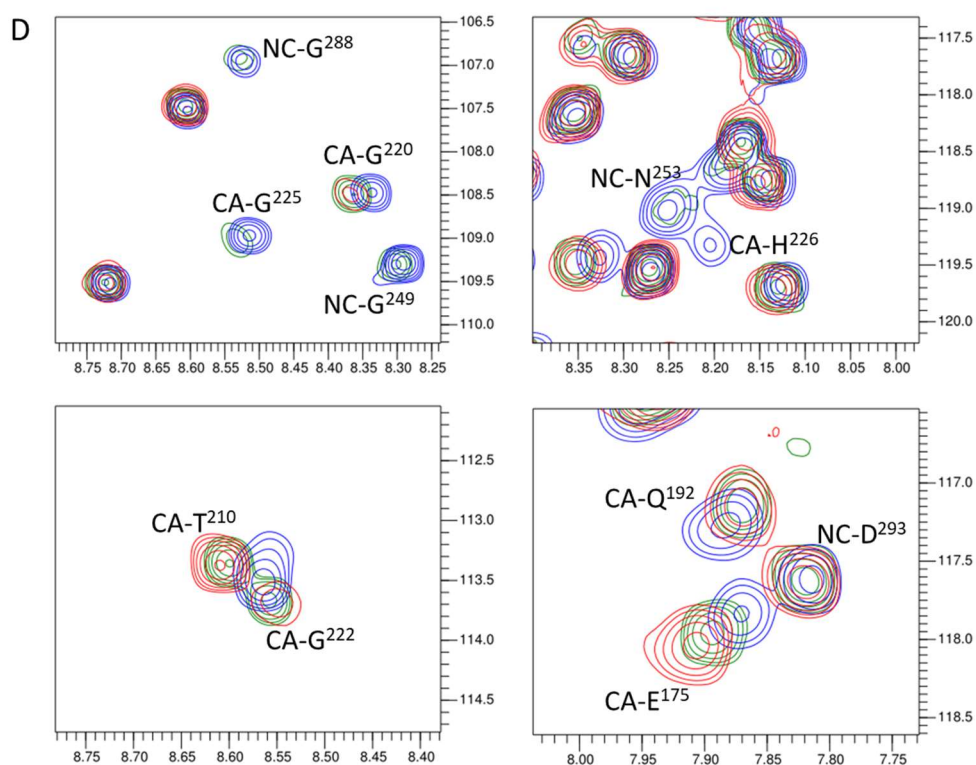


Figure 43. Effect of EP39/PF46396 on protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC detected by $^1H - ^{15}N$ SOFAST-HMQC NMR experiment. The chemical structure of EP39 and PF46396 is shown in A and B. (A) Superposition of $^1H - ^{15}N$ SOFAST-HMQC NMR spectrum recorded on protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence (red) and presence of EP39 (blue). (B) Superposition of $^1H - ^{15}N$ SOFAST-HMQC NMR spectrum recorded on protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence (red) and presence of PF46396 (green). (C) Superposition of $^1H - ^{15}N$ SOFAST-HMQC NMR spectrum recorded on protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC in the absence (red) and presence of EP39 (blue) and PF46396 (green). (D) Zoom in the brown circle in figure 43C.

We collected and compared the CSP caused by EP39 and PF46396 (Figure 44). The addition of the two molecules both lead to the appearances of several resonances, especially at CA-SP1 junction and N-terminal end of NC. The appearance of these resonances indicates the decrease of their dynamics. Thus, like EP39, the binding of PF46396 to protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC decreases the dynamics of CA-SP1 and SP1-NC junctions. The intensities of newly appeared resonances in the presence of PF46396, for example CA-G²²⁵, CA-G²⁴⁹ and NC-N²⁵³ (Figure 43D), are much weaker than that in the presence of EP39 indicating PF46396 decreases the dynamics of CA-SP1 and SP1-NC junctions in a much weaker way than EP39. In addition to the newly appeared resonances, there are also significant CSPs at SP1 domain, which suggest

that PF46396 interacts with CA_{CTD}^{W184A, M185A}-SP1-NC mainly at SP1 domain (including CA-SP1 junctions, SP1 and SP1-NC junction).

Apart from the SP1 domain, the other part of the protein CA_{CTD}^{W184A, M185A}-SP1-NC adopted less CSP in the presence of PF46396 compared to EP39. Combined with the weak dynamics decrease of CA-SP1 and SP1-NC junctions in the presence of PF46396, we conclude that PF46396 interacts with the protein CA_{CTD}^{W184A, M185A}-SP1-NC and decreases the dynamics of SP1 domain in a much weaker way than EP39.

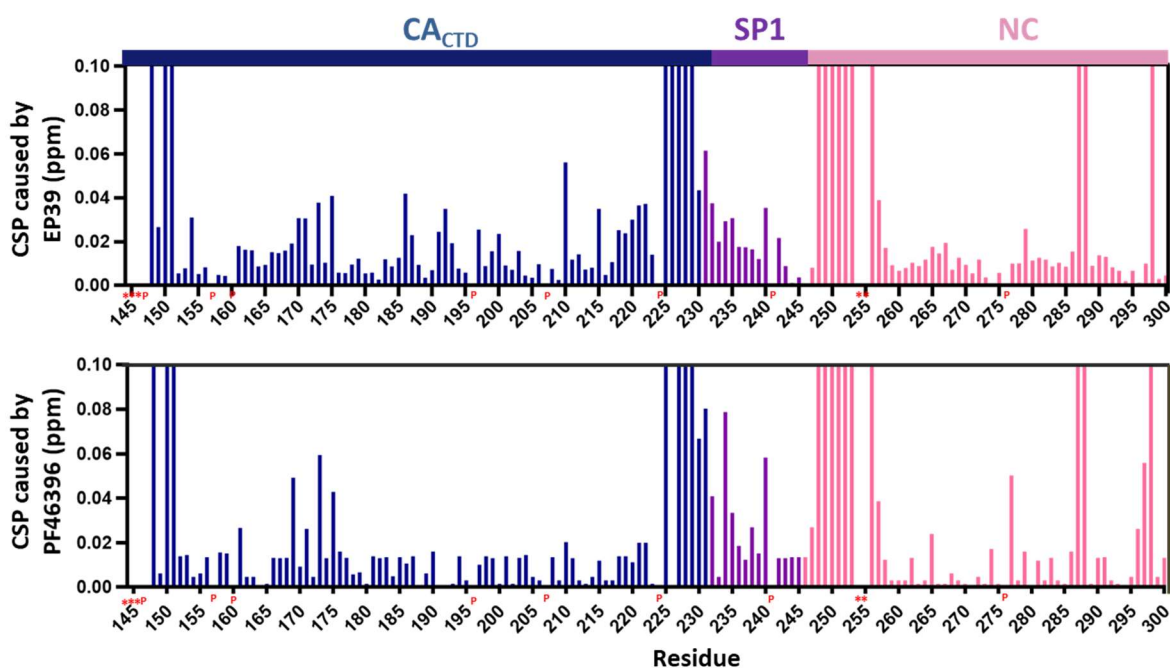


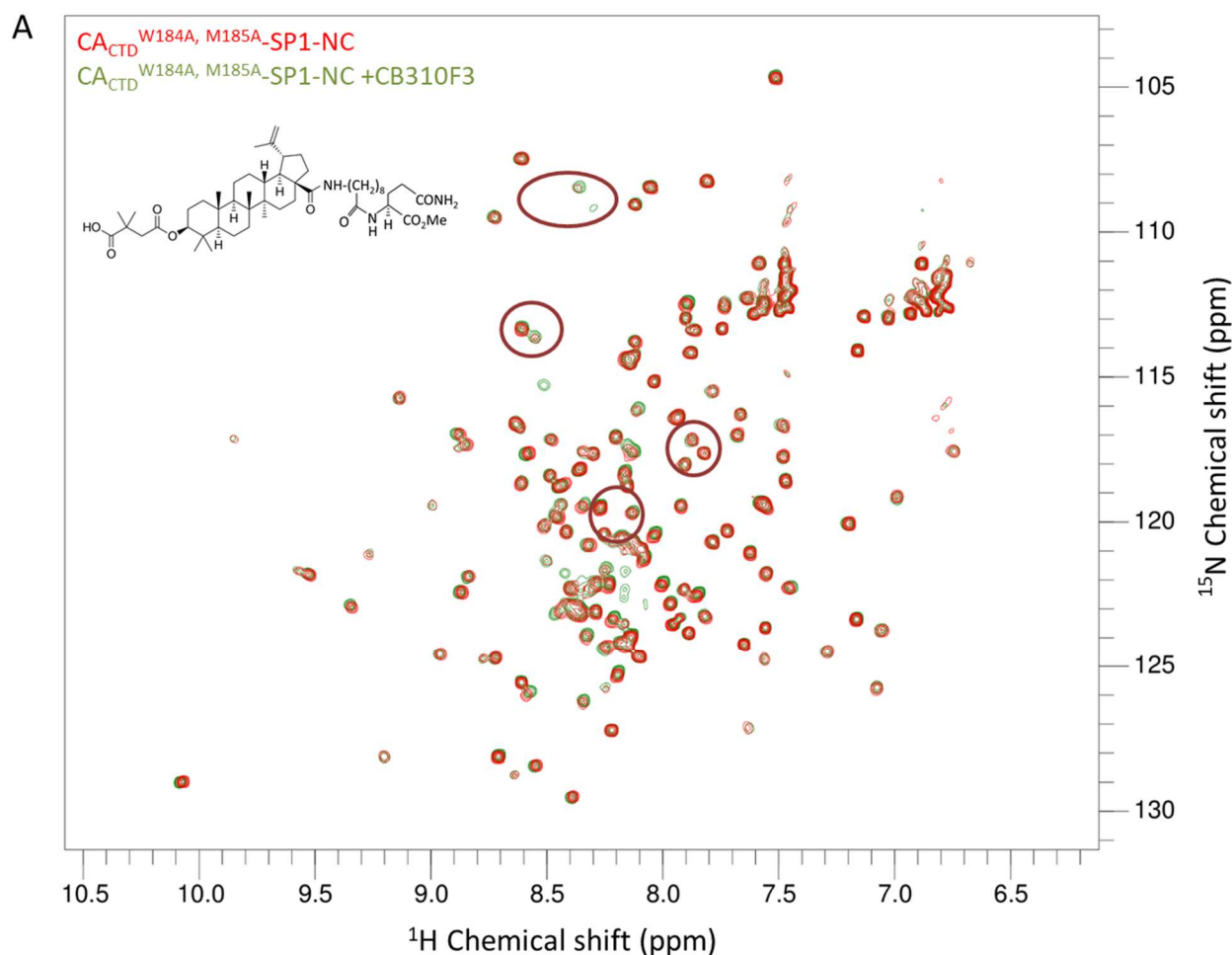
Figure 44. Chemical shift perturbation of the amide groups ($^1\text{H}/^{15}\text{N}$ combined chemical shift) in the presence and in the absence of EP39 and PF46396. The chemical shift perturbations of the residues whose $^1\text{H}-^{15}\text{N}$ resonance are seen only in the presence of inhibitor were set at 0.1. Proline residues are labeled as P and unassigned residues are labeled as *. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively.

2. CB310F3

CB310F3, derivative of BVM which has a long hydrophobic chain in C-28 position, shows an improved activity against HIV-1 carrying BVM-resistant polymorphisms. It has been proved that CB310F3 could inhibit the cleavage process between CA and SP1. Here, we checked the

effect of CB310F3 on protein CA_{CTD}^{W184A, M185A}-SP1-NC with the ¹H-¹⁵N SOFAST-HMQC spectrum.

The same protocol as EP39 and PF46396 was used. We solubilized CB310F3 in DMSO-*d*6 and added this molecule to protein ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC. Then, we recorded the NMR data of ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of CB310F3. After the addition of six equivalents CB310F3, there is no significant changes in the ¹H-¹⁵N SOFAST-HMQC spectrum of ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC (Figure 45A). We zoomed in the brown circle region and analysis specific residues. In the presence of CB310F3, CA-G²²⁵, CA-H²²⁶ and NC-N²⁵³ are still missing in the spectrum (Figure 45B). The resonance for NC-G²⁴⁹ appeared with much weaker signal than that in the presence of EP39 or PF46396 (Figure 45B). Chemical shift for residues CA-E¹⁷⁵, CA-Q¹⁹², CA-T²¹⁰, CA-G²²² and NC-K²⁹³ do not have significant changes in the presence of CB310F3 compared with EP39 and PF46396 (Figure 45B). The weak effect of CB310F3 on protein CA_{CTD}^{W184A, M185A}-SP1-NC indicate the weak binding of the molecule to the protein.



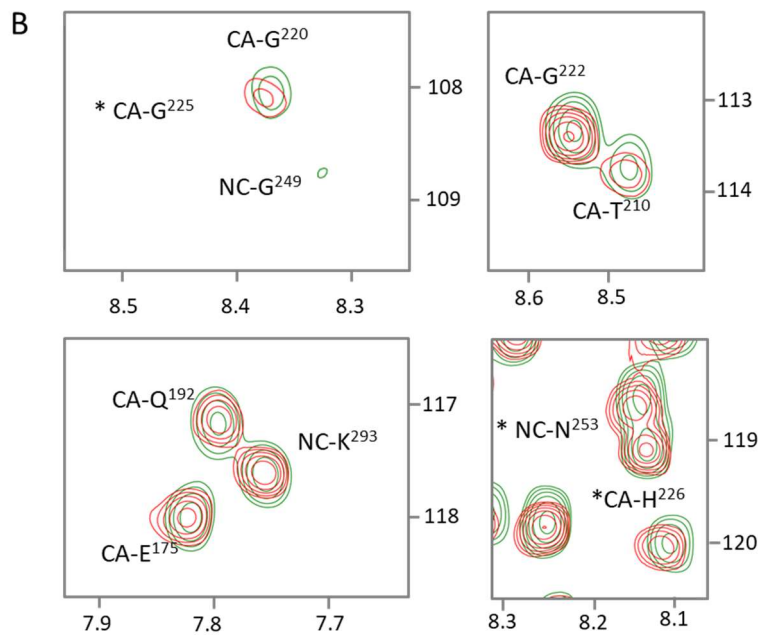


Figure 45. Effect of CB310F3 on protein $CA_{CTD}^{W184A, M185A}$ -SPI-NC detected by 1H - ^{15}N SOFAST-HMQC NMR experiment. (A) Superposition of 1H - ^{15}N SOFAST-HMQC NMR spectra recorded on protein $CA_{CTD}^{W184A, M185A}$ -SPI-NC in the absence (red) and presence of CB310F3 (green). The chemical structure of CB310F3 is shown. (B) Zoom in the brown circle in figure 45A. The missing residue is labeled with *.

Conclusions and discussion

Maturation of HIV-1 particle is an essential step for viral infectivity. During the maturation, the HIV-1 Gag precursor is cleaved by HIV-1 protease into several proteins [53]. The final cleavage at CA-SP1 junction is the limiting step for HIV-1 maturation. Bevirimat (BVM), the first-in-class maturation inhibitor, can block the cleavage between CA and SP1 [206]. However, the mechanism of BVM is still not well-known. To investigate this mechanism, we resolved the structure of protein HIV-1 CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39 with NMR derived information. With these results, we analyzed the structural and dynamic properties of CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of EP39. Besides EP39, we also studied the effects of CB310F3 (derivatives of BVM) and PF46396 (a pyridone-based compound) on protein CA_{CTD}^{W184A, M185A}-SP1-NC with NMR experiments.

I Effect of EP39 on the structure and dynamics of CA_{CTD}^{W184A, M185A}-SP1-NC

1. Structure of protein CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39

The structure of CA-SP1 junction has been reported in different studies. In 2004, the residues in SP1, the final thirteen residues of CA_{CTD} and the thirteen N-terminal residues of NC were demonstrated to be flexible and have a low propensity of forming an α -helix [167]. Similar results were found in the cryo-EM structure of CA-SP1: no detectable electron density was present in SP1 which means SP1 is probably highly dynamics [186][168]. In 2016, the first crystal structure of CA_{CTD}-SP1 hexamer was published with an α -helix at CA_{CTD}-SP1 junction encompassing the nine C-terminal residues of CA_{CTD} and seven N-terminal residues of SP1 [12]. In 30% Trifluoroethanol (TFE) solution, the short peptide CA²¹¹⁻²³¹-SP1²³²⁻²⁴⁵-NC²⁴⁶⁻²⁵⁸ can form a long α -helix extending from CA²²⁶ to NC²⁴⁸ [169]. However, TFE is known to stabilize the *alpha*-helical structure in proteins and the long helix in SP1 could be structurally stabilized by TFE.

In this work, we solved the first monomeric solution structure of CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of the maturation inhibitor EP39 by NMR in physiological conditions. In this structure, from N-terminus to C-terminus, four α -helices are formed at the CA_{CTD} domain, as in the hexameric crystal structure of CA_{CTD}-SP1. The two junctions CA-SP1 and SP1-NC form two α -helices (H5 and H6 respectively) separated by an elbow constituted by the QVT motif.

This is the first time that the whole SP1 is demonstrated to be in helical structure at physiological condition. The NC folds into two zinc fingers separated by a flexible linker.

2. EP39 disturbs the natural helix-coil equilibrium on each side of the SP1 domain.

BVM is known to stabilize the immature Gag lattice [211]. In the microED structure of CA_{CTD}-SP1 complexed with BVM, the inhibitor binds to the hydrophobic center of CA_{CTD}-SP1 hexamer [213]. In our study, EP39 interacts with CA_{CTD}^{W184A, M185A}-SP1-NC, mainly with the two alpha helices of the second domain (H5 and H6). The second domain CA²²¹-NC²⁵¹ has been reported to be highly dynamic and presents an equilibrium between coil and alpha helix [167][185]. Our NMR data shows that many resonances corresponding to the second domain CA²²¹-NC²⁵¹ cannot be detected in ¹H-¹⁵N SOFAST-HMQC spectrum in the absence of EP39 (Figure 35), because of the high dynamics of this domain. However, in the presence of EP39, this domain seems to be stabilized and showed both a CSP of existing peaks and the appearance of new peaks due to decrease in the rate of exchange with the solvent and the stabilization of this domain (Figure 35) that makes this domain to adopt a helical structure in the presence of EP39. Thus, we have shown that the stabilization of this domain by EP39 leads to the formation of two alpha helices, H5 and H6, with a short flexible elbow between them containing the very important QVT motif.

3. The most dynamic part in SP1, QVT motif, is the first target of EP39.

Polymorphism of the SP1 domain, especially the QVT motif (SP1-Q²³⁷V²³⁸T²³⁹) forming an elbow between helices H5 and H6, has been associated with reduced BVM activity *in vivo* [209][266] suggesting this motif is essential for BVM binding. The mutation SP1-T239I which mimics the effect of BVM binding could attenuate dynamics and increase helicity of residues CA²²⁶-SP1²³⁷ [168]. In this work, residues SP1-Q²³⁷V²³⁸T²³⁹ are the most dynamic amino acids in SP1 and have high binding affinities to EP39 compared to other residues (Figure 40, 42), especially residue SP1-T²³⁹ whose dynamics is decreased after adding EP39. These results suggest that EP39 binds to the protein firstly at the SP1-QVT motif and that on each side of the SP1-QVT motif, H5 and H6 are stabilized to helical structure. These results could explain why

the polymorphism of the QVT motif leads to a defect in binding of EP39 / BVM and therefore to a high dynamics of the CA-SP1 junction allowing optimal activity of the protease.

In the crystal and microED structures of CA_{CTD}-SP1, the helix formed at the CA-SP1 junction starts at CA²²⁵ and stop at SP1²³⁸ [213]. As we have shown by our relaxation experiments, H6 is more dynamic than H5 (Figure 42F, H). The residue SP1-T²³⁹ at the end of H5 is the most dynamic residue in the second domain which could be the reason why the CA-SP1 junction helix stops at SP1-V²³⁸ and why H6 is not detectable in the crystal and microED structures of CA_{CTD}-SP1.

In the latest published paper about the effect of BVM on the structure and dynamics of CA-SP1 junction [214], the assigned signals of SP1 ends at SP1-A²³² (A282 in their numbering scheme) because of the high dynamics of SP1. Their data show that BVM involves in the subtle changes in the structure and dynamics of CA-SP1 junction. However, the lost of SP1 information makes it impossible to clarify the essential role of SP1-QVT motif in the maturation inhibitor binding.

4. MHR, especially CA-K158, could be an important target for EP39 binding.

MHR plays a critical role in HIV-1 assembly, maturation and infectivity [267]. There are frequent, short lived contacts between SP1 and MHR [168]. In the docking result between EP39 / BVM and CA_{CTD}-SP1, one common hydrogen bond involving the carbonyl in position C-28 of EP39/BVM and the ϵ -amino of CA-K¹⁵⁸ is observed [228]. The microED structure of CA_{CTD}-SP1 complexed with BVM indicates that the negatively charged dimethylsuccinyl carboxyl group on BVM is located at the center of the positive charges generated by the double ring of lysines (K¹⁵⁸ and K²²⁷) in the hexameric structure [213]. These results can easily be extrapolated to demonstrate the essential role of CA-K¹⁵⁸ in EP39 binding since EP39 and BVM have the same dimethylsuccinyl carboxyl group. In our NMR experiments, the intensity of the resonance of residue CA-K¹⁵⁸ in MHR is much stronger after EP39 addition (Figure 39), which means this residue could be an essential target for EP39 binding.

5. EP39 stabilizes the N-terminal end of NC and causes the formation of a helical structure within the SP1-NC junction

The SP1-NC junction (residue SP1²⁴¹-NC²⁵¹) in the C-terminal of the second domain also adopted an α -helix in the presence of EP39. This finding is surprising and can be explained firstly by the presence of NC and secondly by the effect of EP39 binding. The N-terminal of NC (residues 246-257) is known to form a 3_{10} helix by electrostatic contacts with SL2 / SL3 (Figure 9C) [268,269]. In 30% TFE solution, the peptide CA²¹¹⁻²³¹SP1²³²⁻²⁴⁵NC²⁴⁶⁻²⁵⁸ adopts a helical structure spanning domain CA²²⁶-NC²⁴⁸ [169]. In the present work, the α -helix in the second domain starts from CA²²¹ and stops at NC²⁵¹ with a kink at SP1²⁴⁰ which could be caused by the presence of a proline at this position. This suggests that the N-terminal domain of NC has the propensity to form an alpha helix even in the absence of nucleic acids. Similar with the conformational stabilization of CA-SP1 junction, the helix formed at the SP1-NC junction results from the decrease in the dynamics of the QVT motif caused by EP39 binding. The microED structure of CA_{CTD}-SP1 complexed with BVM contains a helical structure in the N-terminal of SP1 and the C-terminal part of SP1 is invisible [213]. Here, we observe that the C-terminal domain of SP1 and the N-terminal domain of NC adopt an alpha helical structure (H6). This demonstrates that, in the presence of NC, EP39 causes the formation of a helical structure within the SP1-NC junction. Various NMR studies, including ours, have shown that SP1 is poorly structured in aqueous solution but adopts a helical structure if the polarity of the solvent is reduced. This property of the SP1 domain has also been demonstrated in an aqueous medium at high concentrations of the SP1 peptide [172]. These results suggest that SP1 could act as a switch during the assembly of Gag. It seems that we observe the same phenomenon in the presence of EP39 which would stabilize the association of Gag and its stabilization thus disturbing the binding of the protease and then the maturation.

6. The dynamics of SP1 domain impedes the precise positioning of EP39 on protein CA_{CTD}^{W184A, M185A}-SP1-NC.

In this work, we have observed the effect of EP39 on the structure and dynamics of the protein CA_{CTD}^{W184A, M185A}-SP1-NC, but unfortunately it was not possible to observe any intermolecular

correlations between the protein and EP39 on the NOESY experiments. Several reasons make it difficult to precisely position EP39 at the junction CA-SP1-NC, starting firstly with the low water solubility of the EP39 molecule. After adding six equivalents of EP39 to the protein, a slight precipitation of EP39 appeared in the NMR tube, which makes the amount of EP39 interacting with the protein uncertain. Thus, it was impossible to use large amounts of EP39 which would have allowed us to shift the equilibrium of the interaction to the bound form of EP39 to the protein. Secondly, the dynamic property of the second domain prevents the positioning of EP39 on the protein although we observe the effect of EP39 on its dynamics. After interaction with EP39, the second domain is less flexible, but remains very dynamic compared to other domains of the protein, as we have shown by the ^{15}N relaxation studies. This means the binding between EP39 and its target is also very dynamic. The intensity of the NMR signals that we observe for this domain is relatively weak compared to the other amino acids. The NMR resonances corresponding to EP39 are also invisible due to its low solubility or to its intermediate exchange rate between the free and bound form with the protein. Therefore, it is difficult to identify specific restraints between EP39 and the protein. In the 3D NOESY-HSQC experiment, the signal intensity is weak when the distance between two protons is greater than 5 Å and this will also make it difficult to obtain the correlations between the protein and EP39.

To conclude, the binding of EP39 to the second domain appears to particularly disturb the dynamic balance of the helix-coil of CA-SP1 and SP1-NC junctions. Indeed, by binding to the QVT motif of the SP1 domain, the dynamics of this motif is reduced and allows the alpha-helix H5 and H6 to form at the CA-SP1 and SP1-NC junctions respectively. We assert that the decrease in dynamics and the change in conformation of these domains caused by EP39 makes it difficult for the protease to access and cleave the CA-SP1 junction. This therefore results in an inhibition of the maturation of HIV-1 particles by the protease.

II Interaction between CA_{CTD}^{W184A, M185A}-SP1-NC and PF46396.

1. PF46396 interacts with CA_{CTD}^{W184A, M185A}-SP1-NC mainly at CA-SP1 junction and N-terminal of NC.

PF46396 is a structurally different molecule compared to EP39. It was demonstrated disrupting the cleavage between CA and SP1 [215]. In our study, PF46396 interact with CA_{CTD}^{W184A, M185A}-SP1-NC mainly at SP1 domain. The resonances corresponding to CA-SP1 junction and N-terminal end of NC were missing in ¹H-¹⁵N SOFAST-HMQC spectrum in the absence of inhibitor. In the presence of PF46396, this domain exhibits CSP of existing peaks and the appearances of new peaks, which is similar to the protein in the presence of EP39 (Figure 43, 44). Thus, we conclude that PF46396 has similar effect as EP39 on protein CA_{CTD}^{W184A, M185A}-SP1-NC. Its binding stabilizes the CA-SP1 junction and the N-terminal end of NC that will lead to the formation of the helical structure in SP1 domain. The formation of the helical structure in CA-SP1 domain will interfere with the access of HIV-1 protease and block the maturation of HIV-1 particles.

2. PF46396 is less potent than EP39

In our result, PF46396 caused similar effect on protein CA_{CTD}^{W184A, M185A}-SP1-NC as EP39. However, when we analyze some residues more carefully, there are still some differences. In the figure 43D, the resonance intensities of residue CA-G²²⁵, NC-G²⁴⁹ and NC-N²⁵³ in the presence of PF46396 are weaker than that in the presence of EP39. The resonance of residue CA-H²²⁶ was detected in the presence of EP39 but can not be detected in the presence of PF46396. Besides the difference at intensities of the residues appearing in the presence of inhibitors, there are also differences in the level of CSPs of existing peaks in the presence of EP39 and PF46396. In figure 44, the CSPs of residue CA²¹⁸ to CA²²² in the absence and presence of EP39 range from 0.025 to 0.037 ppm. In the presence of PF46396, CSPs for residue CA²¹⁸ to CA²²² range from 0.011 to 0.019 ppm. Residues CA^{149, 186, 192, 210, 215} have more significant CSPs in the presence of EP39 than that in the presence of PF46396. All these differences indicate that the interaction between protein CA_{CTD}^{W184A, M185A}-SP1-NC and

PF46396 is less strong than that between protein CA_{CTD}^{W184A, M185A}-SP1-NC and with EP39, which is consistent with the result that the EC₅₀ of PF46396 is higher than that of BVM [215].

3. MHR is an important domain for PF46396 binding

Similar with EP39 binding, PF46396 binding also caused a significant intensity increase at residue CA-K¹⁵⁸ in MHR as shown in figure 43D indicating that this residue could also be an essential target for PF46396 binding consistent with the resistance map to PF46396 in 2012 by Kayoko Waki *et al* [225]. Besides MHR, Kayoko Waki *et al.* also mapped another resistance mutation, CA-I²⁰¹, which doesn't have significant CSP or intensity change in our NMR experiments. To better analyse the binding of PF46396, we mapped residue CA-K¹⁵⁸ and CA-I²⁰¹ to our NMR structure of CA_{CTD}^{W184A, M185A}-SP1-NC (PDB code 6RWG) and the crystal structure of CA_{CTD}-SP1 hexamer (PDB code 5I4T) respectively (Figure 46). In our NMR structure, residue CA-K¹⁵⁸ and CA-I²⁰¹ are far away from each other. The distance between the amino group on the side chain of CA-K158 and the γ 2-methyl group of CA-I201 is 20.7 Å (Figure 46) that is much longer than the length of PF46396. Combined the result that CA-SP1 junction is an important binding region of PF46396, we conclude that PF46396 binds at the CA-K158 side of protein CA_{CTD}^{W184A, M185A}-SP1-NC in figure 46A. In the hexamer structure of CA_{CTD}-SP1, CA-K¹⁵⁸ located in the inner side of the hexamer (Figure 46B). CA-I²⁰¹ located at the outside of the hexamer (Figure 46B). One PF46396 molecule can not interact with CA-K¹⁵⁸ and CA-I²⁰¹ at the same time. Thus, we conclude that PF46396 interacts with protein CA_{CTD}-SP1 hexamer in the inner side of the “cup” which is a hydrophobic center.

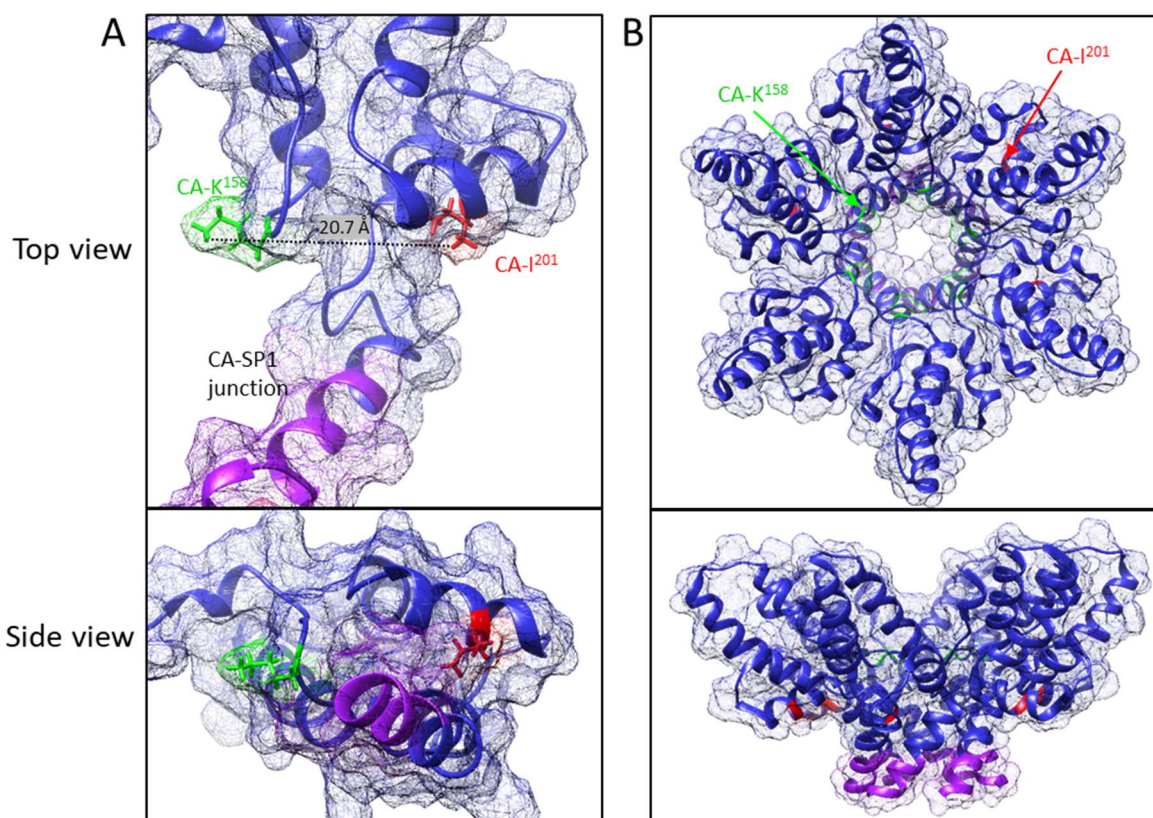


Figure 46. Residue CA-K¹⁵⁸ (green) and CA-I²⁰¹ (red) mapped on NMR structure of CA_{CTD}^{W184A, M185A}-SP1-NC (PDB code 6RWG) (A) and crystal structure of CA_{CTD}-SP1 hexamer (PDB code 5I4T) (B) respectively with top view (up) and side view (down). The residues in CA_{CTD} and SP1 are colored in blue and purple respectively.

4. The N-terminal end of NC is stabilized in the presence of PF46396

The N-terminal end of NC is highly dynamic in the absence of inhibitor and it has the propensity to form helical structure. When we added PF46396, the residues, NC²⁴⁸⁻²⁵⁶, are stabilized with the appearance of resonances in ¹H-¹⁵N SOFAST-HMQC (Figure 44). Similar with the effect of EP39, stabilization in SP1 domain caused by PF46396 binding may also lead to the dynamic decrease of the N-terminal end of NC, which will help the formation of helical structure at the N-terminal of NC.

To conclude, PF46396 has similar effect as EP39 on the structure and dynamics of protein CA_{CTD}^{W184A, M185A}-SP1-NC. PF46396 has two interacting regions on CA_{CTD}^{W184A, M185A}-SP1-NC: SP1 domain that encompassing the CA-SP1 junction and SP1-NC junction, and MHR in

CA. With PF46396 binding, the CA-SP1 and SP1-NC junctions are stabilized and form helical structure. The helical structure in CA-SP1 junction will impede the access of protease, which finally inhibits the maturation of HIV-1 particles. For the precise binding mode of PF46396 on CA_{CTD}^{W184A, M185A}-SP1-NC, we did not get more information. So, in our later work, we will try to discover the precise binding site of PF46396 on CA_{CTD}^{W184A, M185A}-SP1-NC.

III CB310F3 does not have significant effect on protein CA_{CTD}^{W184A, M185A}-SP1-NC in NMR experiments.

CB310F3 is a derivative of BVM with a modified group at the C-28 position. It exhibits a much improved activity against HIV-1 strains that carrying BVM polymorphisms [218]. To test its effect on CA_{CTD}^{W184A, M185A}-SP1-NC in NMR experiments, we added CB310F3 to CA_{CTD}^{W184A, M185A}-SP1-NC and recorded ¹H-¹⁵N SOFAST-HMQC experiments. As shown in figure 44, there is no significant difference between the ¹H-¹⁵N SOFAST-HMQC spectra recorded in the absence and presence of CB310F3. Thus, we conclude that CB310F3 does not bind to the protein CA_{CTD}^{W184A, M185A}-SP1-NC. There is one reason that can explain the deferent binding results. The modified C-28 group in CB310F3 is much bulkier than that in BVM or EP39. It is not allowed in structural space for CB310F3 to access to the SP1 domain of CA_{CTD}^{W184A, M185A}-SP1-NC.

IV Perspectives

During my research, I have successfully obtained structural and dynamic information regarding the mutated protein CA_{CTD}^{W184A, M185A}-SP1-NC, in the absence and presence of the EP39, a maturation inhibitor targeting the CA-SP1 junction. The 3D structure of the monomer CA_{CTD}^{W184A, M185A}-SP1-NC had never been solved before and the structures comprising only the CA and SP1 domains did not allow to characterize the structure of the full SP1 due to the important dynamics of this domain. During my thesis, I solved the structure of CA_{CTD}^{W184A, M185A}-SP1-NC consisting of the mutated C-terminal domain of the capsid, the spacer peptide SP1 and the nucleocapsid protein NC, in the presence of EP39. I have also identified the essential regions of protein CA_{CTD}^{W184A, M185A}-SP1-NC that is important for binding with EP39. I showed that the QVT motif within the SP1 domain and the CA-K158 residue in the CA MHR domain were directly targeted by the maturation inhibitor EP39.

During this project, I encountered two major difficulties. The first is the assignment of the resonances of the ^1H proton and ^{15}N nitrogen nuclei of the SP1 domain, especially in the absence of a structure stabilizing inhibitor. Although the protein spectrum of 157 amino acids is of excellent quality with particularly spread resonances both in ^{15}N and ^1H , some resonances are absent from the spectrum due to the dynamics of the protein, particularly the residues located in SP1. The addition of EP39 allowed the stabilization of this domain and the observation of new signals, which helped to obtain structural data that is necessary for the determination of the protein's 3D structure. The second difficulty encountered was the low solubility and the low affinity of the inhibitor EP39 to protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$. When titrating $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ by EP39, I was limited in adding EP39 since obvious precipitation appeared in the NMR tube. The low affinity of EP39 makes it impossible to obtain the intramolecular constraints which is necessary for the reconstruction of the complex between EP39 and the protein.

Besides EP39, I also studied the interactions between two other maturation inhibitors, PF46396 and CB310F3, and the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$. I was surprised that although PF46396 had a very different structure to EP39 and lower affinity than EP39 for the protein, its effect on the SP1 domain was comparable to EP39. They both stabilized the SP1 domain with the appearance of new resonances on the spectrum. CB310F3 is a derivative of BVM. It has the same structure as EP39, except for the C-28 group. However, the effect of CB310F3 on the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ is much weaker than that of EP39 or PF46396. Of the three inhibitors (EP39, PF46396 and CB310F3), EP39 is the most effective in stabilizing the SP1 domain of the protein.

Although I succeeded in observing the effect of two inhibitors (EP39 and PF46396) on the dynamics and structure of the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ by NMR, I did not obtain their specific binding mode on protein due to their low affinities. We still have several other molecules derived from bevirimat that are already synthesized and will be tested soon in the laboratory. To identify the mode of interaction between the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ and the maturation inhibitors, we need a new inhibitor that has a much higher affinity for the protein $\text{CA}_{\text{CTD}}^{\text{W184A, M185A}}\text{-SP1-NC}$ for new structural studies.

To this end, docking studies have already been undertaken in the laboratory in order to position EP39 precisely at the level of CA-SP1 junction while satisfying the experimental data obtained by NMR and making it possible to explain the mutations induced by EP39. From these

theoretical complexes, we will try to make functional modifications to the EP39 molecule in order to enhance its interaction with the protein. This can be achieved by grafting functions, which allows the formation of new interactions (hydrophobic, hydrogen, etc.) with protein, to the molecule. The new molecules can be synthesized by our collaborator Serge Turcaud who has extensive experience in the synthesis of these molecules and has synthesized several molecules for our team based on the bevirimat backbone and modified C-28 position.

On the other hand, I expressed the protein CA_{CTD}^{W184A, M185A}-SP1-NC enriched in 15N / 13C. This protein will make it possible to obtain data characterizing more precisely the structure of the protein and to have access to distance information involving other protons than the amide protons due to the 15N labeling only. This information could be useful for the recognition of a complex.

In parallel to this study, Pascale Coric in the laboratory showed that there was an interaction between CB310F3 and CA(211-231)-SP1-NC(1-13), a short peptide encompassing the SP1 with the upstream 21 residues in CA_{CTD} and the downstream 13 residues in NC, in the presence of 30% TFE. Spatial proximities have been identified which should allow the resolution of a first structure of the complex between CA(211-231)-SP1-NC(1-13) and CB310F3 by NMR. This structure could make it possible to identify the residues of the protein which bind to the inhibitors, to establish the specific binding mode of the maturation inhibitors on the protein CA_{CTD}^{W184A, M185A}-SP1-NC and to propose a mechanism of action of these inhibitors. Since the bottleneck in the development of new inhibitors of HIV-1 maturation is the polymorphisms at the SP1 domain, these structural studies may allow to obtain new active molecules despite the native polymorphisms of the spacer peptide SP1.

References

- [1] E.O. Freed, HIV-1 assembly, release and maturation, *Nat. Rev. Microbiol.* 13 (2015) 484–496. <https://doi.org/10.1038/nrmicro3490>.
- [2] S.G. Deeks, J. Overbaugh, A. Phillips, S. Buchbinder, HIV infection, *Nat. Rev. Dis. Primer.* 1 (2015) 1–22. <https://doi.org/10.1038/nrdp.2015.35>.
- [3] C.B. Wilen, J.C. Tilton, R.W. Doms, HIV: Cell Binding and Entry, *Cold Spring Harb. Perspect. Med.* 2 (2012). <https://doi.org/10.1101/cshperspect.a006866>.
- [4] J.S. Saad, J. Miller, J. Tai, A. Kim, R.H. Ghanam, M.F. Summers, Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly, *Proc. Natl. Acad. Sci. U. S. A.* 103 (2006) 11364–11369. <https://doi.org/10.1073/pnas.0602818103>.
- [5] A. Rein, S.A.K. Datta, C.P. Jones, K. Musier-Forsyth, Diverse interactions of retroviral Gag proteins with RNAs, *Trends Biochem. Sci.* 36 (2011) 373–380. <https://doi.org/10.1016/j.tibs.2011.04.001>.
- [6] L. Deshmukh, C.D. Schwieters, A. Grishaev, R. Ghirlando, J.L. Baber, G.M. Clore, Structure and Dynamics of Full Length HIV-1 Capsid Protein in Solution, *J. Am. Chem. Soc.* 135 (2013) 16133–16147. <https://doi.org/10.1021/ja406246z>.
- [7] O. Pornillos, B.K. Ganser-Pornillos, M. Yeager, Atomic-level modelling of the HIV capsid, *Nature.* 469 (2011) 424–427. <https://doi.org/10.1038/nature09640>.
- [8] O. Pornillos, B.K. Ganser-Pornillos, B.N. Kelly, Y. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, M. Yeager, X-ray Structures of the Hexameric Building Block of the HIV Capsid, *Cell.* 137 (2009) 1282–1292. <https://doi.org/10.1016/j.cell.2009.04.063>.
- [9] M. Novikova, Y. Zhang, E.O. Freed, K. Peng, Multiple Roles of HIV-1 Capsid during the Virus Replication Cycle, *Virol. Sin.* 34 (2019) 119–134. <https://doi.org/10.1007/s12250-019-00095-3>.
- [10] F.K.M. Schur, W.J.H. Hagen, M. Rumlová, T. Ruml, B. Müller, H.-G. Kräusslich, J.A.G. Briggs, Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution, *Nature.* 517 (2015) 505–508. <https://doi.org/10.1038/nature13838>.
- [11] S. Mattei, B. Glass, W.J.H. Hagen, H.-G. Kräusslich, J.A.G. Briggs, The structure and flexibility of conical HIV-1 capsids determined within intact virions, *Science.* 354 (2016) 1434–1437. <https://doi.org/10.1126/science.aah4972>.
- [12] J.M. Wagner, K.K. Zadrozny, J. Chrustowicz, M.D. Purdy, M. Yeager, B.K. Ganser-Pornillos, O. Pornillos, Crystal structure of an HIV assembly and maturation switch, *ELife.* 5 (2016) e17063. <https://doi.org/10.7554/eLife.17063>.
- [13] A.M. ouahmed, Étude du domaine C-terminal de la protéine Gag du VIH-1 : interactions multiples et rôle au cours de l'assemblage et du bourgeonnement du virus., 2018. <http://www.theses.fr/s176651> (accessed April 28, 2020).
- [14] Double Resonance Backbone Assignment | Protein NMR, (n.d.). <https://www.protein-nmr.org.uk/solution-nmr/assignment-theory/double-resonance-backbone-assignment/> (accessed May 21, 2020).
- [15] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: Development of a software pipeline, *Proteins Struct. Funct. Bioinforma.* 59 (2005) 687–696. <https://doi.org/10.1002/prot.20449>.
- [16] Double Resonance Side-chain Assignment | Protein NMR, (n.d.). <https://www.protein-nmr.org.uk/solution-nmr/assignment-theory/double-resonance-side-chain-assignment/> (accessed May 21, 2020).

- [17] M.P. Williamson, Using chemical shift perturbation to characterise ligand binding, *Prog. Nucl. Magn. Reson. Spectrosc.* 73 (2013) 1–16. <https://doi.org/10.1016/j.pnmrs.2013.02.001>.
- [18] A.G. Palmer, Chemical exchange in biomacromolecules: Past, present, and future, *J. Magn. Reson. San Diego Calif* 1997. 241 (2014) 3–17. <https://doi.org/10.1016/j.jmr.2014.01.008>.
- [19] C.M. Swanson, M.H. Malim, SnapShot: HIV-1 proteins, *Cell*. 133 (2008) 742, 742.e1. <https://doi.org/10.1016/j.cell.2008.05.005>.
- [20] K. Van Baelen, K. Salzwedel, E. Rondelez, V. Van Eygen, S. De Vos, A. Verheyen, K. Steegen, Y. Verlinden, G.P. Allaway, L.J. Stuyver, Susceptibility of Human Immunodeficiency Virus Type 1 to the Maturation Inhibitor Bevirimat Is Modulated by Baseline Polymorphisms in Gag Spacer Peptide 1, *Antimicrob. Agents Chemother.* 53 (2009) 2185–2188. <https://doi.org/10.1128/AAC.01650-08>.
- [21] D.J. Hockley, R.D. Wood, J.P. Jacobs, A.J. Garrett, Electron Microscopy of Human Immunodeficiency Virus, *J. Gen. Virol.* 69 (1988) 2455–2469. <https://doi.org/10.1099/0022-1317-69-10-2455>.
- [22] M.A. Muesing, D.H. Smith, C.D. Cabradilla, C.V. Benton, L.A. Lasky, D.J. Capon, Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus, *Nature*. 313 (1985) 450–458. <https://doi.org/10.1038/313450a0>.
- [23] F. Kirchhoff, Immune Evasion and Counteraction of Restriction Factors by HIV-1 and Other Primate Lentiviruses, *Cell Host Microbe*. 8 (2010) 55–67. <https://doi.org/10.1016/j.chom.2010.06.004>.
- [24] A.D. Frankel, J.A.T. Young, HIV-1: Fifteen Proteins and an RNA, *Annu. Rev. Biochem.* 67 (1998) 1–25. <https://doi.org/10.1146/annurev.biochem.67.1.1>.
- [25] A.G. Dalgleish, P.C.L. Beverley, P.R. Clapham, D.H. Crawford, M.F. Greaves, R.A. Weiss, The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus, *Nature*. 312 (1984) 763–767. <https://doi.org/10.1038/312763a0>.
- [26] E.A. Berger, P.M. Murphy, J.M. Farber, Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease, *Annu. Rev. Immunol.* 17 (1999) 657–700. <https://doi.org/10.1146/annurev.immunol.17.1.657>.
- [27] S.Y. Chan, R.F. Speck, C. Power, S.L. Gaffen, B. Chesebro, M.A. Goldsmith, V3 Recombinants Indicate a Central Role for CCR5 as a Coreceptor in Tissue Infection by Human Immunodeficiency Virus Type 1, *J. Virol.* 73 (1999) 2350–2358.
- [28] Y. Kiselyeva, R. Nedellec, A. Ramos, C. Pastore, L.B. Margolis, D.E. Mosier, Evolution of CXCR4-Using Human Immunodeficiency Virus Type 1 SF162 Is Associated with Two Unique Envelope Mutations, *J. Virol.* 81 (2007) 3657–3661. <https://doi.org/10.1128/JVI.02310-06>.
- [29] C. Pastore, R. Nedellec, A. Ramos, S. Pontow, L. Ratner, D.E. Mosier, Human Immunodeficiency Virus Type 1 Coreceptor Switching: V1/V2 Gain-of-Fitness Mutations Compensate for V3 Loss-of-Fitness Mutations, *J. Virol.* 80 (2006) 750–758. <https://doi.org/10.1128/JVI.80.2.750-758.2006>.
- [30] J.A. Esté, A. Telenti, HIV entry inhibitors, *Lancet Lond. Engl.* 370 (2007) 81–88. [https://doi.org/10.1016/S0140-6736\(07\)61052-6](https://doi.org/10.1016/S0140-6736(07)61052-6).
- [31] N. Arhel, Revisiting HIV-1 uncoating, *Retrovirology*. 7 (2010) 96. <https://doi.org/10.1186/1742-4690-7-96>.
- [32] M.I. Bukrinsky, N. Sharova, T.L. McDonald, T. Pushkarskaya, W.G. Tarpley, M. Stevenson, Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection, *Proc. Natl. Acad. Sci. U. S. A.* 90 (1993) 6125–6129. <https://doi.org/10.1073/pnas.90.13.6125>.
- [33] M.D. Miller, C.M. Farnet, F.D. Bushman, Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition, *J. Virol.* 71 (1997) 5382–5390.
- [34] P.F. Lewis, M. Emerman, Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus., *J. Virol.* 68 (1994) 510–516.

- [35] T. Roe, T.C. Reynolds, G. Yu, P.O. Brown, Integration of murine leukemia virus DNA depends on mitosis, *EMBO J.* 12 (1993) 2099–2108.
- [36] J.B. Weinberg, T.J. Matthews, B.R. Cullen, M.H. Malim, Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes, *J. Exp. Med.* 174 (1991) 1477–1482. <https://doi.org/10.1084/jem.174.6.1477>.
- [37] B.R. Cullen, Journey to the Center of the Cell, *Cell.* 105 (2001) 697–700. [https://doi.org/10.1016/S0092-8674\(01\)00392-0](https://doi.org/10.1016/S0092-8674(01)00392-0).
- [38] R. Craigie, F.D. Bushman, HIV DNA Integration, *Cold Spring Harb. Perspect. Med.* 2 (2012). <https://doi.org/10.1101/cshperspect.a006890>.
- [39] L. Colin, C. Van Lint, Molecular control of HIV-1 postintegration latency: implications for the development of new therapeutic strategies, *Retrovirology.* 6 (2009) 111. <https://doi.org/10.1186/1742-4690-6-111>.
- [40] G. Li, E. De Clercq, HIV Genome-Wide Protein Associations: a Review of 30 Years of Research, *Microbiol. Mol. Biol. Rev. MMBR.* 80 (2016) 679–731. <https://doi.org/10.1128/MMBR.00065-15>.
- [41] A. Gatignol, Transcription of HIV: Tat and Cellular Chromatin, in: *Adv. Pharmacol.*, Academic Press, 2007: pp. 137–159. [https://doi.org/10.1016/S1054-3589\(07\)55004-0](https://doi.org/10.1016/S1054-3589(07)55004-0).
- [42] M. McLaren, Modulating HIV-1 RNA processing and utilization, *Front. Biosci. Volume* (2008) 5693. <https://doi.org/10.2741/3110>.
- [43] S. de Breyne, R. Soto-Rifo, M. López-Lastra, T. Ohlmann, Translation initiation is driven by different mechanisms on the HIV-1 and HIV-2 genomic RNAs, *Virus Res.* 171 (2013) 366–381. <https://doi.org/10.1016/j.virusres.2012.10.006>.
- [44] S. Burugu, A. Daher, E.F. Meurs, A. Gatignol, HIV-1 translation and its regulation by cellular factors PKR and PACT, *Virus Res.* 193 (2014) 65–77. <https://doi.org/10.1016/j.virusres.2014.07.014>.
- [45] N. Shkriabai, S.A.K. Datta, Z. Zhao, S. Hess, A. Rein, M. Kvaratskhelia, Interactions of HIV-1 Gag with assembly cofactors, *Biochemistry.* 45 (2006) 4077–4083. <https://doi.org/10.1021/bi052308e>.
- [46] M. Huang, J.M. Orenstein, M.A. Martin, E.O. Freed, p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease., *J. Virol.* 69 (1995) 6810.
- [47] D.G. Demirov, A. Ono, J.M. Orenstein, E.O. Freed, Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function, *Proc. Natl. Acad. Sci.* 99 (2002) 955–960. <https://doi.org/10.1073/pnas.032511899>.
- [48] A. Wlodawer, J.W. Erickson, Structure-based inhibitors of HIV-1 protease, *Annu. Rev. Biochem.* 62 (1993) 543–585. <https://doi.org/10.1146/annurev.bi.62.070193.002551>.
- [49] C. Tang, J.M. Louis, A. Aniana, J.-Y. Suh, G.M. Clore, Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease, *Nature.* 455 (2008) 693–696. <https://doi.org/10.1038/nature07342>.
- [50] J. a. G. Briggs, J.D. Riches, B. Glass, V. Bartonova, G. Zanetti, H.-G. Kräusslich, Structure and assembly of immature HIV, *Proc. Natl. Acad. Sci.* 106 (2009) 11090–11095. <https://doi.org/10.1073/pnas.0903535106>.
- [51] S. Scarlata, C. Carter, Role of HIV-1 Gag domains in viral assembly, *Biochim. Biophys. Acta BBA - Biomembr.* 1614 (2003) 62–72. [https://doi.org/10.1016/S0005-2736\(03\)00163-9](https://doi.org/10.1016/S0005-2736(03)00163-9).
- [52] A. Wlodawer, J.W. Erickson, Structure-Based Inhibitors of Hiv-1 Protease, *Annu. Rev. Biochem.* 62 (1993) 543–585. <https://doi.org/10.1146/annurev.bi.62.070193.002551>.
- [53] K. Wieggers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, H.-G. Kräusslich, Sequential Steps in Human Immunodeficiency Virus Particle Maturation Revealed by Alterations of Individual Gag Polyprotein Cleavage Sites, *J. Virol.* 72 (1998) 2846–2854.

- [54] L.V. Coren, J.A. Thomas, E. Chertova, R.C. Sowder, T.D. Gagliardi, R.J. Gorelick, D.E. Ott, Mutational Analysis of the C-Terminal Gag Cleavage Sites in Human Immunodeficiency Virus Type 1, *J. Virol.* 81 (2007) 10047–10054. <https://doi.org/10.1128/JVI.02496-06>.
- [55] S.-K. Lee, M. Potempa, R. Swanstrom, The Choreography of HIV-1 Proteolytic Processing and Virion Assembly, *J. Biol. Chem.* 287 (2012) 40867–40874. <https://doi.org/10.1074/jbc.R112.399444>.
- [56] L. Deshmukh, R. Ghirlando, G.M. Clore, 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 (2015) 3374–3379. <https://doi.org/10.1073/pnas.1501985112>.
- [57] S. Li, C.P. Hill, W.I. Sundquist, J.T. Finch, Image reconstructions of helical assemblies of the HIV-1 CA protein, *Nature.* 407 (2000) 409–413. <https://doi.org/10.1038/35030177>.
- [58] M. Barros, F. Heinrich, S.A.K. Datta, A. Rein, I. Karageorgos, H. Nanda, M. Lösche, Membrane Binding of HIV-1 Matrix Protein: Dependence on Bilayer Composition and Protein Lipidation, *J. Virol.* 90 (2016) 4544–4555. <https://doi.org/10.1128/JVI.02820-15>.
- [59] V. D’Souza, M.F. Summers, How retroviruses select their genomes, *Nat. Rev. Microbiol.* 3 (2005) 643–655. <https://doi.org/10.1038/nrmicro1210>.
- [60] A. Rein, RNA Packaging in HIV, *Trends Microbiol.* 27 (2019) 715–723. <https://doi.org/10.1016/j.tim.2019.04.003>.
- [61] Y.J. Im, L. Kuo, X. Ren, P.V. Burgos, X.Z. Zhao, F. Liu, T.R. Burke, J.S. Bonifacino, E.O. Freed, J.H. Hurley, Crystallographic and functional analysis of the ESCRT-I /HIV-1 Gag PTAP interaction, *Struct. Lond. Engl.* 18 (2010) 1536–1547. <https://doi.org/10.1016/j.str.2010.08.010>.
- [62] B. Houck-Loomis, M.A. Durney, C. Salguero, N. Shankar, J.M. Nagle, S.P. Goff, V.M. D’Souza, An equilibrium-dependent retroviral mRNA switch regulates translational recoding, *Nature.* 480 (2011) 561–564. <https://doi.org/10.1038/nature10657>.
- [63] J.A.G. Briggs, M.N. Simon, I. Gross, H.-G. Kräusslich, S.D. Fuller, V.M. Vogt, M.C. Johnson, The stoichiometry of Gag protein in HIV-1, *Nat. Struct. Mol. Biol.* 11 (2004) 672–675. <https://doi.org/10.1038/nsmb785>.
- [64] T. Jacks, M.D. Power, F.R. Masiarz, P.A. Luciw, P.J. Barr, H.E. Varmus, Characterization of ribosomal frameshifting in HIV-1 gag-pol expression, *Nature.* 331 (1988) 280–283. <https://doi.org/10.1038/331280a0>.
- [65] X. Wu, H. Liu, H. Xiao, J.A. Conway, E. Hunter, J.C. Kappes, Functional RT and IN incorporated into HIV-1 particles independently of the Gag/Pol precursor protein., *EMBO J.* 16 (1997) 5113–5122. <https://doi.org/10.1093/emboj/16.16.5113>.
- [66] C.P. Hill, D. Worthylake, D.P. Bancroft, A.M. Christensen, W.I. Sundquist, Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly, *Proc. Natl. Acad. Sci.* 93 (1996) 3099–3104. <https://doi.org/10.1073/pnas.93.7.3099>.
- [67] M.H. Wright, W.P. Heal, D.J. Mann, E.W. Tate, Protein myristoylation in health and disease, *J. Chem. Biol.* 3 (2009) 19–35. <https://doi.org/10.1007/s12154-009-0032-8>.
- [68] N.M. Bell, A.M.L. Lever, HIV Gag polyprotein: processing and early viral particle assembly, *Trends Microbiol.* 21 (2013) 136–144. <https://doi.org/10.1016/j.tim.2012.11.006>.
- [69] V. Chukkapalli, S.J. Oh, A. Ono, 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 (2010) 1600–1605. <https://doi.org/10.1073/pnas.0908661107>.
- [70] B. Ganser-Pornillos, M. Yeager, W.I. Sundquist, The Structural Biology of HIV Assembly, *Curr. Opin. Struct. Biol.* 18 (2008) 203. <https://doi.org/10.1016/j.sbi.2008.02.001>.
- [71] J.R. Lingappa, J.C. Reed, M. Tanaka, K. Chutiraka, B.A. Robinson, How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle, *Virus Res.* 193 (2014) 89–107. <https://doi.org/10.1016/j.virusres.2014.07.001>.
- [72] P.Y. Mercredi, N. Bucca, B. Loeliger, C.R. Gaines, M. Mehta, P. Bhargava, P.R. Tedbury, L. Charlier, N. Floquet, D. Muriaux, C. Favard, C.R. Sanders, E.O. Freed, J. Marchant, M.F.

- Summers, Structural and Molecular Determinants of Membrane Binding by the HIV-1 Matrix Protein, *J. Mol. Biol.* 428 (2016) 1637–1655. <https://doi.org/10.1016/j.jmb.2016.03.005>.
- [73] C. Tang, E. Loeliger, P. Luncsford, I. Kinde, D. Beckett, M.F. Summers, Entropic switch regulates myristate exposure in the HIV-1 matrix protein, *Proc. Natl. Acad. Sci.* 101 (2004) 517–522. <https://doi.org/10.1073/pnas.0305665101>.
- [74] A. Alfadhli, R.L. Barklis, E. Barklis, HIV-1 MATRIX ORGANIZES AS A HEXAMER OF TRIMERS ON MEMBRANES CONTAINING PHOSPHATIDYLINOSITOL-(4,5)-BISPHOSPHATE, *Virology*. 387 (2009) 466. <https://doi.org/10.1016/j.virol.2009.02.048>.
- [75] S. Campbell, R.J. Fisher, E.M. Towler, S. Fox, H.J. Issaq, T. Wolfe, L.R. Phillips, A. Rein, Modulation of HIV-like particle assembly in vitro by inositol phosphates, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 10875–10879. <https://doi.org/10.1073/pnas.191224698>.
- [76] A. Alfadhli, E. Barklis, The roles of lipids and nucleic acids in HIV-1 assembly, *Front. Microbiol.* 5 (2014) 253. <https://doi.org/10.3389/fmicb.2014.00253>.
- [77] S.A.K. Datta, F. Heinrich, S. Raghunandan, S. Krueger, J.E. Curtis, A. Rein, H. Nanda, HIV-1 Gag extension: conformational changes require simultaneous interaction with membrane and nucleic acid, *J. Mol. Biol.* 406 (2011) 205–214. <https://doi.org/10.1016/j.jmb.2010.11.051>.
- [78] N. Kempf, V. Postupalenko, S. Bora, P. Didier, Y. Arntz, H. de Rocquigny, Y. Mély, The HIV-1 nucleocapsid protein recruits negatively charged lipids to ensure its optimal binding to lipid membranes, *J. Virol.* 89 (2015) 1756–1767. <https://doi.org/10.1128/JVI.02931-14>.
- [79] J. Inlora, D.R. Collins, M.E. Trubin, J.Y.J. Chung, A. Ono, Membrane binding and subcellular localization of retroviral Gag proteins are differentially regulated by MA interactions with phosphatidylinositol-(4,5)-bisphosphate and RNA, *MBio.* 5 (2014) e02202. <https://doi.org/10.1128/mBio.02202-14>.
- [80] M. Comas-Garcia, S.A. Datta, L. Baker, R. Varma, P.R. Gudla, A. Rein, Dissection of specific binding of HIV-1 Gag to the “packaging signal” in viral RNA, *ELife.* 6 (2017). <https://doi.org/10.7554/eLife.27055>.
- [81] G.C. Todd, A. Duchon, J. Inlora, E.D. Olson, K. Musier-Forsyth, A. Ono, Inhibition of HIV-1 Gag-membrane interactions by specific RNAs, *RNA N. Y. N.* 23 (2017) 395–405. <https://doi.org/10.1261/rna.058453.116>.
- [82] S.B. Kutluay, T. Zang, D. Blanco-Melo, C. Powell, D. Jannain, M. Errando, P.D. Bieniasz, Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis, *Cell.* 159 (2014) 1096–1109. <https://doi.org/10.1016/j.cell.2014.09.057>.
- [83] Q. Yu, C.D. Morrow, Essential regions of the tRNA primer required for HIV-1 infectivity, *Nucleic Acids Res.* 28 (2000) 4783–4789.
- [84] C.R. Gaines, E. Tkacik, A. Rivera-Oven, P. Somani, A. Achimovich, T. Alabi, A. Zhu, N. Getachew, A.L. Yang, M. McDonough, T. Hawkins, Z. Spadaro, M.F. Summers, HIV-1 Matrix Protein Interactions with tRNA: Implications for Membrane Targeting, *J. Mol. Biol.* 430 (2018) 2113–2127. <https://doi.org/10.1016/j.jmb.2018.04.042>.
- [85] P.R. Tedbury, E.O. Freed, The role of matrix in HIV-1 envelope glycoprotein incorporation, *Trends Microbiol.* 22 (2014) 372–378. <https://doi.org/10.1016/j.tim.2014.04.012>.
- [86] A. Alfadhli, A. Mack, C. Ritchie, I. Cylinder, L. Harper, P.R. Tedbury, E.O. Freed, E. Barklis, Trimer Enhancement Mutation Effects on HIV-1 Matrix Protein Binding Activities, *J. Virol.* 90 (2016) 5657–5664. <https://doi.org/10.1128/JVI.00509-16>.
- [87] P.R. Tedbury, M. Novikova, S.D. Ablan, E.O. Freed, Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) E182-190. <https://doi.org/10.1073/pnas.1516618113>.
- [88] M.A. Checkley, B.G. Luttge, E.O. Freed, HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation, *J. Mol. Biol.* 410 (2011) 582–608. <https://doi.org/10.1016/j.jmb.2011.04.042>.

- [89] P. Gallay, S. Swingler, J. Song, F. Bushman, D. Trono, HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase, *Cell*. 83 (1995) 569–576. [https://doi.org/10.1016/0092-8674\(95\)90097-7](https://doi.org/10.1016/0092-8674(95)90097-7).
- [90] X. Yu, Q.C. Yu, T.H. Lee, M. Essex, The C terminus of human immunodeficiency virus type 1 matrix protein is involved in early steps of the virus life cycle., *J. Virol.* 66 (1992) 5667–5670.
- [91] C.R. Casella, L.J. Raffini, A.T. Panganiban, Pleiotropic Mutations in the HIV-1 Matrix Protein That Affect Diverse Steps in Replication, *Virology*. 228 (1997) 294–306. <https://doi.org/10.1006/viro.1996.8355>.
- [92] T.R. Gamble, F.F. Vajdos, S. Yoo, D.K. Worthylake, M. Houseweart, W.I. Sundquist, C.P. Hill, Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid, *Cell*. 87 (1996) 1285–1294. [https://doi.org/10.1016/s0092-8674\(00\)81823-1](https://doi.org/10.1016/s0092-8674(00)81823-1).
- [93] R.K. Gitti, B.M. Lee, J. Walker, M.F. Summers, S. Yoo, W.I. Sundquist, Structure of the amino-terminal core domain of the HIV-1 capsid protein, *Science*. 273 (1996) 231–235. <https://doi.org/10.1126/science.273.5272.231>.
- [94] U.K. von Schwedler, T.L. Stemmler, V.Y. Klishko, S. Li, K.H. Albertine, D.R. Davis, W.I. Sundquist, Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly, *EMBO J.* 17 (1998) 1555–1568. <https://doi.org/10.1093/emboj/17.6.1555>.
- [95] J. Jiang, S.D. Aplan, S. Derebail, K. Hercík, F. Soheilian, J.A. Thomas, S. Tang, I. Hewlett, K. Nagashima, R.J. Gorelick, E.O. Freed, J.G. Levin, The interdomain linker region of HIV-1 capsid protein is a critical determinant of proper core assembly and stability, *Virology*. 421 (2011) 253–265. <https://doi.org/10.1016/j.virol.2011.09.012>.
- [96] T.R. Gamble, S. Yoo, F.F. Vajdos, U.K. von Schwedler, D.K. Worthylake, H. Wang, J.P. McCutcheon, W.I. Sundquist, C.P. Hill, Structure of the Carboxyl-Terminal Dimerization Domain of the HIV-1 Capsid Protein, *Science*. 278 (1997) 849–853. <https://doi.org/10.1126/science.278.5339.849>.
- [97] M. Tanaka, B.A. Robinson, K. Chutiraka, C.D. Geary, J.C. Reed, J.R. Lingappa, Mutations of Conserved Residues in the Major Homology Region Arrest Assembling HIV-1 Gag as a Membrane-Targeted Intermediate Containing Genomic RNA and Cellular Proteins, *J. Virol.* 90 (2016) 1944–1963. <https://doi.org/10.1128/JVI.02698-15>.
- [98] R. Bocanegra, M.Á. Fuertes, A. Rodríguez-Huete, J.L. Neira, M.G. Mateu, Biophysical Analysis of the MHR Motif in Folding and Domain Swapping of the HIV Capsid Protein C-Terminal Domain, *Biophys. J.* 108 (2015) 338–349. <https://doi.org/10.1016/j.bpj.2014.11.3472>.
- [99] E.R. Wright, J.B. Schooler, H.J. Ding, C. Kieffer, C. Fillmore, W.I. Sundquist, G.J. Jensen, Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells, *EMBO J.* 26 (2007) 2218–2226. <https://doi.org/10.1038/sj.emboj.7601664>.
- [100] S. Mattei, F.K. Schur, J.A. Briggs, Retrovirus maturation-an extraordinary structural transformation, *Curr. Opin. Virol.* 18 (2016) 27–35. <https://doi.org/10.1016/j.coviro.2016.02.008>.
- [101] W. Zhang, S. Cao, J.L. Martin, J.D. Mueller, L.M. Mansky, Morphology and ultrastructure of retrovirus particles, *AIMS Biophys.* 2 (2015) 343–369. <https://doi.org/10.3934/biophy.2015.3.343>.
- [102] F.K.M. Schur, M. Obr, W.J.H. Hagen, W. Wan, A.J. Jakobi, J.M. Kirkpatrick, C. Sachse, H.-G. Kräusslich, J.A.G. Briggs, An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation, *Science*. 353 (2016) 506–508. <https://doi.org/10.1126/science.aaf9620>.
- [103] S. Mattei, A. Tan, B. Glass, B. Müller, H.-G. Kräusslich, J.A.G. Briggs, High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation, *Proc. Natl. Acad. Sci.* 115 (2018) E9401–E9410. <https://doi.org/10.1073/pnas.1811237115>.
- [104] A.T. Gres, K.A. Kirby, V.N. KewalRamani, J.J. Tanner, O. Pornillos, S.G. Sarafianos, STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability, *Science*. 349 (2015) 99–103. <https://doi.org/10.1126/science.aaa5936>.

- [105] G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics, *Nature*. 497 (2013) 643–646. <https://doi.org/10.1038/nature12162>.
- [106] J.M. Berg, Potential metal-binding domains in nucleic acid binding proteins, *Science*. 232 (1986) 485–487. <https://doi.org/10.1126/science.2421409>.
- [107] L.M. Green, J.M. Berg, Retroviral nucleocapsid protein-metal ion interactions: folding and sequence variants, *Proc. Natl. Acad. Sci. U. S. A.* 87 (1990) 6403–6407. <https://doi.org/10.1073/pnas.87.16.6403>.
- [108] E. Bombarda, N. Morellet, H. Cherradi, B. Spiess, S. Bouaziz, E. Grell, B.P. Roques, Y. Mély, 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.* 310 (2001) 659–672. <https://doi.org/10.1006/jmbi.2001.4770>.
- [109] Y. Mély, H. De Rocquigny, N. Morellet, B.P. Roques, D. Gérard, Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy, *Biochemistry*. 35 (1996) 5175–5182. <https://doi.org/10.1021/bi952587d>.
- [110] T.L. South, P.R. Blake, R.C. Sowder, L.O. Arthur, L.E. Henderson, M.F. Summers, The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers, *Biochemistry*. 29 (1990) 7786–7789. <https://doi.org/10.1021/bi00486a002>.
- [111] B.M. Lee, R.N. De Guzman, B.G. Turner, N. Tjandra, M.F. Summers, Dynamical behavior of the HIV-1 nucleocapsid protein, *J. Mol. Biol.* 279 (1998) 633–649. <https://doi.org/10.1006/jmbi.1998.1766>.
- [112] N. Morellet, N. Jullian, H. De Rocquigny, B. Maigret, J.L. Darlix, B.P. Roques, Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR, *EMBO J.* 11 (1992) 3059–3065.
- [113] M.F. Summers, L.E. Henderson, M.R. Chance, J.W. Bess, T.L. South, P.R. Blake, I. Sagi, G. Perez-Alvarado, R.C. Sowder, D.R. Hare, Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1., *Protein Sci. Publ. Protein Soc.* 1 (1992) 563–574.
- [114] T.L. South, P.R. Blake, D.R. Hare, M.F. Summers, C-terminal retroviral-type zinc finger domain from the HIV-1 nucleocapsid protein is structurally similar to the N-terminal zinc finger domain, *Biochemistry*. 30 (1991) 6342–6349. <https://doi.org/10.1021/bi00239a036>.
- [115] R.J. Gorelick, D.J. Chabot, A. Rein, L.E. Henderson, L.O. Arthur, The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent, *J. Virol.* 67 (1993) 4027–4036.
- [116] N. Morellet, H. de Rocquigny, Y. Mély, N. Jullian, H. Déméné, M. Ottmann, D. Gérard, J.L. Darlix, M.C. Fournie-Zaluski, B.P. Roques, Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by 1H NMR, *J. Mol. Biol.* 235 (1994) 287–301. [https://doi.org/10.1016/s0022-2836\(05\)80033-6](https://doi.org/10.1016/s0022-2836(05)80033-6).
- [117] H. Déméné, C.Z. Dong, M. Ottmann, M.C. Rouyez, N. Jullian, N. Morellet, Y. Mely, J.L. Darlix, M.C. Fournié-Zaluski, S. Saragosti, 1H NMR structure and biological studies of the His23-->Cys mutant nucleocapsid protein of HIV-1 indicate that the conformation of the first zinc finger is critical for virus infectivity, *Biochemistry*. 33 (1994) 11707–11716. <https://doi.org/10.1021/bi00205a006>.
- [118] Y. Mély, N. Jullian, N. Morellet, H. De Rocquigny, C.Z. Dong, E. Piémont, B.P. Roques, D. Gérard, Spatial proximity of the HIV-1 nucleocapsid protein zinc fingers investigated by time-resolved fluorescence and fluorescence resonance energy transfer, *Biochemistry*. 33 (1994) 12085–12091. <https://doi.org/10.1021/bi00206a011>.
- [119] L. Zargarian, C. Tisé, P. Barraud, X. Xu, N. Morellet, B. René, Y. Mély, P. Fossé, O. Mauffret, Dynamics of Linker Residues Modulate the Nucleic Acid Binding Properties of the HIV-1

- Nucleocapsid Protein Zinc Fingers, PLOS ONE. 9 (2014) e102150.
<https://doi.org/10.1371/journal.pone.0102150>.
- [120] J.-L. Darlix, J. Godet, R. Ivanyi-Nagy, P. Fossé, O. Mauffret, Y. Mély, Flexible nature and specific functions of the HIV-1 nucleocapsid protein, *J. Mol. Biol.* 410 (2011) 565–581.
<https://doi.org/10.1016/j.jmb.2011.03.037>.
- [121] A. Aldovini, R.A. Young, Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus., *J. Virol.* 64 (1990) 1920–1926.
- [122] E. Schmalzbauer, B. Strack, J. Dannull, S. Guehmann, K. Moelling, Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro., *J. Virol.* 70 (1996) 771–777.
- [123] J. Clever, C. Sasseti, T.G. Parslow, RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1., *J. Virol.* 69 (1995) 2101–2109.
- [124] M.F. Shubsda, A.C. Paoletti, B.S. Hudson, P.N. Borer, Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein, *Biochemistry.* 41 (2002) 5276–5282.
<https://doi.org/10.1021/bi016045+>.
- [125] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition, *J. Mol. Biol.* 301 (2000) 491–511.
<https://doi.org/10.1006/jmbi.2000.3979>.
- [126] R.N.D. Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 Ψ -RNA Recognition Element, *Science.* 279 (1998) 384–388. <https://doi.org/10.1126/science.279.5349.384>.
- [127] J.-L. Darlix, C. Gabus, M.-T. Nugeyre, F. Clavel, F. Barré-Sinoussi, Cis elements and Trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1, *J. Mol. Biol.* 216 (1990) 689–699. [https://doi.org/10.1016/0022-2836\(90\)90392-Y](https://doi.org/10.1016/0022-2836(90)90392-Y).
- [128] Y.X. Feng, T.D. Copeland, L.E. Henderson, R.J. Gorelick, W.J. Bosche, J.G. Levin, A. Rein, HIV-1 nucleocapsid protein induces “maturation” of dimeric retroviral RNA in vitro., *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996) 7577–7581.
- [129] T. Dorfman, J. Luban, S.P. Goff, W.A. Haseltine, H.G. Göttlinger, Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein, *J. Virol.* 67 (1993) 6159–6169. <https://doi.org/10.1128/JVI.67.10.6159-6169.1993>.
- [130] J.B. Jowett, D.J. Hockley, M.V. Nermut, I.M. Jones, Distinct signals in human immunodeficiency virus type 1 Pr55 necessary for RNA binding and particle formation, *J. Gen. Virol.* 73 (Pt 12) (1992) 3079–3086. <https://doi.org/10.1099/0022-1317-73-12-3079>.
- [131] A. Cimarelli, J.L. Darlix, Assembling the human immunodeficiency virus type 1, *Cell. Mol. Life Sci. CMLS.* 59 (2002) 1166–1184. <https://doi.org/10.1007/s00018-002-8495-6>.
- [132] D. Muriaux, J. Mirro, D. Harvin, A. Rein, RNA is a structural element in retrovirus particles, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 5246–5251. <https://doi.org/10.1073/pnas.091000398>.
- [133] N. Jouvenet, S.M. Simon, P.D. Bieniasz, Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 19114–19119. <https://doi.org/10.1073/pnas.0907364106>.
- [134] J. Guo, L.E. Henderson, J. Bess, B. Kane, J.G. Levin, Human immunodeficiency virus type 1 nucleocapsid protein promotes efficient strand transfer and specific viral DNA synthesis by inhibiting TAR-dependent self-priming from minus-strand strong-stop DNA, *J. Virol.* 71 (1997) 5178–5188. <https://doi.org/10.1128/JVI.71.7.5178-5188.1997>.
- [135] H. De Rocquigny, C. Gabus, A. Vincent, M.C. Fournié-Zaluski, B. Roques, J.L. Darlix, Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require

- only peptide domains outside the zinc fingers, *Proc. Natl. Acad. Sci. U. S. A.* 89 (1992) 6472–6476. <https://doi.org/10.1073/pnas.89.14.6472>.
- [136] J.A. Peliska, S. Balasubramanian, D.P. Giedroc, S.J. Benkovic, Recombinant HIV-1 nucleocapsid protein accelerates HIV-1 reverse transcriptase catalyzed DNA strand transfer reactions and modulates RNase H activity, *Biochemistry*. 33 (1994) 13817–13823. <https://doi.org/10.1021/bi00250a036>.
- [137] F. T. W. V, B. K, R. J, H. P, T. U, M. A, K. P, S. U, Solution Structure of the Human Immunodeficiency Virus Type 1 p6 Protein, *J. Biol. Chem.* 280 (2005). <https://doi.org/10.1074/jbc.M507375200>.
- [138] P. Luo, R.L. Baldwin, Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water, *Biochemistry*. 36 (1997) 8413–8421. <https://doi.org/10.1021/bi9707133>.
- [139] J.K. Myers, C.N. Pace, J.M. Scholtz, Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1, *Protein Sci. Publ. Protein Soc.* 7 (1998) 383–388. <https://doi.org/10.1002/pro.5560070219>.
- [140] S.M.Ø. Solbak, T.R. Reksten, F. Hahn, V. Wray, P. Henklein, P. Henklein, Ø. Halskau, U. Schubert, T. Fossen, HIV-1 p6 - a structured to flexible multifunctional membrane-interacting protein, *Biochim. Biophys. Acta.* 1828 (2013) 816–823. <https://doi.org/10.1016/j.bbamem.2012.11.010>.
- [141] D. Stys, I. Blaha, P. Strop, Structural and functional studies in vitro on the p6 protein from the HIV-1 gag open reading frame, *Biochim. Biophys. Acta.* 1182 (1993) 157–161. [https://doi.org/10.1016/0925-4439\(93\)90137-p](https://doi.org/10.1016/0925-4439(93)90137-p).
- [142] W.M. Henne, N.J. Buchkovich, S.D. Emr, The ESCRT pathway, *Dev. Cell.* 21 (2011) 77–91. <https://doi.org/10.1016/j.devcel.2011.05.015>.
- [143] N. Jouvenet, Dynamics of ESCRT proteins, *Cell. Mol. Life Sci. CMLS.* 69 (2012) 4121–4133. <https://doi.org/10.1007/s00018-012-1035-0>.
- [144] J. Votteler, W.I. Sundquist, Virus budding and the ESCRT pathway, *Cell Host Microbe.* 14 (2013) 232–241. <https://doi.org/10.1016/j.chom.2013.08.012>.
- [145] E.R. Weiss, H. Göttlinger, The role of cellular factors in promoting HIV budding, *J. Mol. Biol.* 410 (2011) 525–533. <https://doi.org/10.1016/j.jmb.2011.04.055>.
- [146] J.E. Garrus, U.K. von Schwedler, O.W. Pornillos, S.G. Morham, K.H. Zavitz, H.E. Wang, D.A. Wettstein, K.M. Stray, M. Côté, R.L. Rich, D.G. Myszka, W.I. Sundquist, Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding, *Cell.* 107 (2001) 55–65. [https://doi.org/10.1016/s0092-8674\(01\)00506-2](https://doi.org/10.1016/s0092-8674(01)00506-2).
- [147] J. Martin-Serrano, T. Zang, P.D. Bieniasz, HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress, *Nat. Med.* 7 (2001) 1313–1319. <https://doi.org/10.1038/nm1201-1313>.
- [148] M. Bleck, M.S. Itano, D.S. Johnson, V.K. Thomas, A.J. North, P.D. Bieniasz, S.M. Simon, Temporal and spatial organization of ESCRT protein recruitment during HIV-1 budding, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 12211–12216. <https://doi.org/10.1073/pnas.1321655111>.
- [149] J. Prescher, V. Baumgärtel, S. Ivanchenko, A.A. Torrano, C. Bräuchle, B. Müller, D.C. Lamb, Super-resolution imaging of ESCRT-proteins at HIV-1 assembly sites, *PLoS Pathog.* 11 (2015) e1004677. <https://doi.org/10.1371/journal.ppat.1004677>.
- [150] O. Pornillos, S.L. Alam, R.L. Rich, D.G. Myszka, D.R. Davis, W.I. Sundquist, Structure and functional interactions of the Tsg101 UEV domain, *EMBO J.* 21 (2002) 2397–2406. <https://doi.org/10.1093/emboj/21.10.2397>.
- [151] R.D. Fisher, H.-Y. Chung, Q. Zhai, H. Robinson, W.I. Sundquist, C.P. Hill, Structural and biochemical studies of ALIX/AIP1 and its role in retrovirus budding, *Cell.* 128 (2007) 841–852. <https://doi.org/10.1016/j.cell.2007.01.035>.
- [152] S. Lee, A. Joshi, K. Nagashima, E.O. Freed, J.H. Hurley, Structural basis for viral late-domain binding to Alix, *Nat. Struct. Mol. Biol.* 14 (2007) 194–199. <https://doi.org/10.1038/nsmb1203>.

- [153] H.G. Göttlinger, How HIV-1 hijacks ALIX, *Nat. Struct. Mol. Biol.* 14 (2007) 254–256. <https://doi.org/10.1038/nsmb0407-254>.
- [154] Q. Zhai, R.D. Fisher, H.-Y. Chung, D.G. Myszka, W.I. Sundquist, C.P. Hill, Structural and functional studies of ALIX interactions with YPX(n)L late domains of HIV-1 and EIAV, *Nat. Struct. Mol. Biol.* 15 (2008) 43–49. <https://doi.org/10.1038/nsmb1319>.
- [155] N.F. Bello, V. Dussupt, P. Sette, V. Rudd, K. Nagashima, F. Bibollet-Ruche, C. Chen, R.C. Montelaro, B.H. Hahn, F. Bouamr, Budding of retroviruses utilizing divergent L domains requires nucleocapsid, *J. Virol.* 86 (2012) 4182–4193. <https://doi.org/10.1128/JVI.07105-11>.
- [156] V. Dussupt, P. Sette, N.F. Bello, M.P. Javid, K. Nagashima, F. Bouamr, Basic residues in the nucleocapsid domain of Gag are critical for late events of HIV-1 budding, *J. Virol.* 85 (2011) 2304–2315. <https://doi.org/10.1128/JVI.01562-10>.
- [157] P. Sette, V. Dussupt, F. Bouamr, Identification of the HIV-1 NC binding interface in Alix Bro1 reveals a role for RNA, *J. Virol.* 86 (2012) 11608–11615. <https://doi.org/10.1128/JVI.01260-12>.
- [158] P. Sette, S.K. O’Connor, V.S. Yerramilli, V. Dussupt, K. Nagashima, K. Chutiraka, J. Lingappa, S. Scarlata, F. Bouamr, HIV-1 Nucleocapsid Mimics the Membrane Adaptor Syntenin PDZ to Gain Access to ESCRTs and Promote Virus Budding, *Cell Host Microbe.* 19 (2016) 336–348. <https://doi.org/10.1016/j.chom.2016.02.004>.
- [159] J.W. Balliet, D.L. Kolson, G. Eiger, F.M. Kim, K.A. McGann, A. Srinivasan, R. Collman, Distinct effects in primary macrophages and lymphocytes of the human immunodeficiency virus type 1 accessory genes vpr, vpu, and nef: mutational analysis of a primary HIV-1 isolate, *Virology.* 200 (1994) 623–631. <https://doi.org/10.1006/viro.1994.1225>.
- [160] R.I. Connor, B.K. Chen, S. Choe, N.R. Landau, Vpr is required for efficient replication of human immunodeficiency virus type-1 in mononuclear phagocytes, *Virology.* 206 (1995) 935–944. <https://doi.org/10.1006/viro.1995.1016>.
- [161] E. Le Rouzic, S. Benichou, The Vpr protein from HIV-1: distinct roles along the viral life cycle, *Retrovirology.* 2 (2005) 11. <https://doi.org/10.1186/1742-4690-2-11>.
- [162] R.Y. Zhao, G. Li, M.I. Bukrinsky, Vpr-host interactions during HIV-1 viral life cycle, *J. Neuroimmune Pharmacol. Off. J. Soc. NeuroImmune Pharmacol.* 6 (2011) 216–229. <https://doi.org/10.1007/s11481-011-9261-z>.
- [163] N. Laguette, C. Brégnard, P. Hue, J. Basbous, A. Yatim, M. Larroque, F. Kirchhoff, A. Constantinou, B. Sobhian, M. Benkirane, Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing, *Cell.* 156 (2014) 134–145. <https://doi.org/10.1016/j.cell.2013.12.011>.
- [164] F. Bachand, X.J. Yao, M. Hrimech, N. Rougeau, E.A. Cohen, 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 (1999) 9083–9091. <https://doi.org/10.1074/jbc.274.13.9083>.
- [165] J.V. Fritz, D. Dujardin, J. Godet, P. Didier, J. De Mey, J.-L. Darlix, Y. Mély, H. de Rocquigny, HIV-1 Vpr oligomerization but not that of Gag directs the interaction between Vpr and Gag, *J. Virol.* 84 (2010) 1585–1596. <https://doi.org/10.1128/JVI.01691-09>.
- [166] H. Zhu, H. Jian, L.-J. Zhao, Identification of the 15FRFG domain in HIV-1 Gag p6 essential for Vpr packaging into the virion, *Retrovirology.* 1 (2004) 26. <https://doi.org/10.1186/1742-4690-1-26>.
- [167] J.L. Newman, E.W. Butcher, D.T. Patel, Y. Mikhaylenko, M.F. Summers, Flexibility in the P2 domain of the HIV-1 Gag polyprotein, *Protein Sci. Publ. Protein Soc.* 13 (2004) 2101–2107. <https://doi.org/10.1110/ps.04614804>.
- [168] M. Wang, C.M. Quinn, J.R. Perilla, H. Zhang, R. Shirra Jr., G. Hou, I.-J. Byeon, C.L. Suiter, S. Ablan, E. Urano, T.J. Nitz, C. Aiken, E.O. Freed, P. Zhang, K. Schulten, A.M. Gronenborn, T. Polenova, Quenching protein dynamics interferes with HIV capsid maturation, *Nat. Commun.* 8 (2017). <https://doi.org/10.1038/s41467-017-01856-y>.

- [169] N. Morellet, S. Druillennec, C. Lenoir, S. Bouaziz, B.P. Roques, Helical structure determined by NMR of the HIV-1 (345–392)Gag sequence, surrounding p2: Implications for particle assembly and RNA packaging, *Protein Sci. Publ. Protein Soc.* 14 (2005) 375–386. <https://doi.org/10.1110/ps.041087605>.
- [170] A. Mouhand, A. Belfetmi, M. Catala, V. Larue, L. Zargarian, F. Brachet, R.J. Gorelick, C. Van Heijenoort, G. Mirambeau, P. Barraud, O. Mauffret, C. Tisné, Modulation of the HIV nucleocapsid dynamics finely tunes its RNA-binding properties during virion genesis, *Nucleic Acids Res.* 46 (2018) 9699–9710. <https://doi.org/10.1093/nar/gky612>.
- [171] S.A.K. Datta, P.K. Clark, L. Fan, B. Ma, D.P. Harvin, R.C. Sowder, R. Nussinov, Y.-X. Wang, A. Rein, Dimerization of the SP1 Region of HIV-1 Gag Induces a Helical Conformation and Association into Helical Bundles: Implications for Particle Assembly, *J. Virol.* 90 (2016) 1773–1787. <https://doi.org/10.1128/JVI.02061-15>.
- [172] S.A.K. Datta, L.G. Temeselew, R.M. Crist, F. Soheilian, A. Kamata, J. Mirro, D. Harvin, K. Nagashima, R.E. Cachau, A. Rein, On the Role of the SP1 Domain in HIV-1 Particle Assembly: a Molecular Switch? *J. Virol.* 85 (2011) 4111–4121. <https://doi.org/10.1128/JVI.00006-11>.
- [173] M.A. Accola, S. Höglund, H.G. Göttlinger, A Putative α -Helical Structure Which Overlaps the Capsid-p2 Boundary in the Human Immunodeficiency Virus Type 1 Gag Precursor Is Crucial for Viral Particle Assembly, *J. Virol.* 72 (1998) 2072–2078.
- [174] H.G. Kräusslich, M. Fäcke, A.M. Heuser, J. Konvalinka, H. Zentgraf, The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity., *J. Virol.* 69 (1995) 3407–3419.
- [175] X. Guo, A. Roldan, J. Hu, M.A. Wainberg, C. Liang, Mutation of the SP1 Sequence Impairs both Multimerization and Membrane-Binding Activities of Human Immunodeficiency Virus Type 1 Gag, *J. Virol.* 79 (2005) 1803–1812. <https://doi.org/10.1128/JVI.79.3.1803-1812.2005>.
- [176] R.A. Dick, K.K. Zadrozny, C. Xu, F.K.M. Schur, T.D. Lyddon, C.L. Ricana, J.M. Wagner, J.R. Perilla, B.K. Ganser-Pornillos, M.C. Johnson, O. Pornillos, V.M. Vogt, Inositol phosphates are assembly co-factors for HIV-1, *Nature.* 560 (2018) 509–512. <https://doi.org/10.1038/s41586-018-0396-4>.
- [177] J. Fontana, P.W. Keller, E. Urano, S.D. Ablan, A.C. Steven, E.O. Freed, Identification of an HIV-1 Mutation in Spacer Peptide 1 That Stabilizes the Immature CA-SP1 Lattice, *J. Virol.* 90 (2016) 972–978. <https://doi.org/10.1128/JVI.02204-15>.
- [178] R.S. Russell, A. Roldan, M. Detorio, J. Hu, M.A. Wainberg, C. Liang, Effects of a Single Amino Acid Substitution within the p2 Region of Human Immunodeficiency Virus Type 1 on Packaging of Spliced Viral RNA, *J. Virol.* 77 (2003) 12986–12995. <https://doi.org/10.1128/JVI.77.24.12986-12995.2003>.
- [179] N. Ristic, M.P. Chin, Mutations in matrix and SP1 repair the packaging specificity of a Human Immunodeficiency Virus Type 1 mutant by reducing the association of Gag with spliced viral RNA, *Retrovirology.* 7 (2010) 73. <https://doi.org/10.1186/1742-4690-7-73>.
- [180] M.K. Hill, A. Bellamy-McIntyre, L.J. Vella, S.M. Campbell, J.A. Marshall, G. Tachedjian, J. Mak, Alteration of the proline at position 7 of the HIV-1 spacer peptide p1 suppresses viral infectivity in a strain dependent manner, *Curr. HIV Res.* 5 (2007) 69–78. <https://doi.org/10.2174/157016207779316323>.
- [181] A. de Marco, A.-M. Heuser, B. Glass, H.-G. Kräusslich, B. Müller, J.A.G. Briggs, Role of the SP2 Domain and Its Proteolytic Cleavage in HIV-1 Structural Maturation and Infectivity, *J. Virol.* 86 (2012) 13708–13716. <https://doi.org/10.1128/JVI.01704-12>.
- [182] H.G. Kräusslich, M. Fäcke, A.M. Heuser, J. Konvalinka, H. Zentgraf, The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity, *J. Virol.* 69 (1995) 3407–3419.
- [183] S.C. Pettit, M.D. Moody, R.S. Wehbie, A.H. Kaplan, P.V. Nantermet, C.A. Klein, R. Swanstrom, The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions, *J. Virol.* 68 (1994) 8017–8027.

- [184] D.K. Worthylake, H. Wang, S. Yoo, W.I. Sundquist, C.P. Hill, Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution, *Acta Crystallogr. D Biol. Crystallogr.* 55 (1999) 85–92. <https://doi.org/10.1107/S0907444998007689>.
- [185] Y. Han, G. Hou, C.L. Suiter, J. Ahn, I.-J.L. Byeon, A.S. Lipton, S. Burton, I. Hung, P.L. Gor'kov, Z. Gan, W. Brey, D. Rice, A.M. Gronenborn, T. Polenova, Magic Angle Spinning NMR Reveals Sequence-Dependent Structural Plasticity, Dynamics, and the Spacer Peptide 1 Conformation in HIV-1 Capsid Protein Assemblies, *J. Am. Chem. Soc.* 135 (2013) 17793–17803. <https://doi.org/10.1021/ja406907h>.
- [186] T.A.M. Bharat, L.R.C. Menendez, W.J.H. Hagen, V. Lux, S. Igonet, M. Schorb, F.K.M. Schur, H.-G. Kräusslich, J.A.G. Briggs, Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly, *Proc. Natl. Acad. Sci.* 111 (2014) 8233–8238. <https://doi.org/10.1073/pnas.1401455111>.
- [187] H. Mitsuya, K.J. Weinhold, P.A. Furman, M.H. St Clair, S.N. Lehrman, R.C. Gallo, D. Bolognesi, D.W. Barry, S. Broder, 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro., *Proc. Natl. Acad. Sci. U. S. A.* 82 (1985) 7096–7100.
- [188] M.C. Graves, J.J. Lim, E.P. Heimer, R.A. Kramer, An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity, *Proc. Natl. Acad. Sci.* 85 (1988) 2449–2453. <https://doi.org/10.1073/pnas.85.8.2449>.
- [189] W.G. Farmerie, D.D. Loeb, N.C. Casavant, C.A. Hutchison, M.H. Edgell, R. Swanstrom, Expression and processing of the AIDS virus reverse transcriptase in *Escherichia coli*, *Science*. 236 (1987) 305–308. <https://doi.org/10.1126/science.2436298>.
- [190] B. Autran, G. Carcelain, T.S. Li, C. Blanc, D. Mathez, R. Tubiana, C. Katlama, P. Debré, J. Leibowitch, Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease, *Science*. 277 (1997) 112–116. <https://doi.org/10.1126/science.277.5322.112>.
- [191] M.M. Lederman, E. Connick, A. Landay, D.R. Kuritzkes, J. Spritzler, M. St Clair, B.L. Kotzin, L. Fox, M.H. Chiozzi, J.M. Leonard, F. Rousseau, M. Wade, J.D. Roe, A. Martinez, H. Kessler, Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: results of AIDS Clinical Trials Group Protocol 315, *J. Infect. Dis.* 178 (1998) 70–79. <https://doi.org/10.1086/515591>.
- [192] K. Bhaskaran, O. Hamouda, M. Sannes, F. Boufassa, A.M. Johnson, P.C. Lambert, K. Porter, CASCADE Collaboration, Changes in the risk of death after HIV seroconversion compared with mortality in the general population, *JAMA*. 300 (2008) 51–59. <https://doi.org/10.1001/jama.300.1.51>.
- [193] Y. Koh, H. Nakata, K. Maeda, H. Ogata, G. Bilcer, T. Devasamudram, J.F. Kincaid, P. Boross, Y.-F. Wang, Y. Tie, P. Volarath, L. Gaddis, R.W. Harrison, I.T. Weber, A.K. Ghosh, H. Mitsuya, Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro, *Antimicrob. Agents Chemother.* 47 (2003) 3123–3129. <https://doi.org/10.1128/aac.47.10.3123-3129.2003>.
- [194] A.K. Ghosh, H.L. Osswald, G. Prato, Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS, *J. Med. Chem.* 59 (2016) 5172–5208. <https://doi.org/10.1021/acs.jmedchem.5b01697>.
- [195] 90–90–90 - An ambitious treatment target to help end the AIDS epidemic, (n.d.). <https://www.unaids.org/en/resources/documents/2017/90-90-90> (accessed August 12, 2020).
- [196] M.R. Loutfy, T. Antoniou, S. Shen, C. Diong, M. Vlaicu, R. Halpenny, C. Kovacs, D. Fletcher, J.M. Raboud, Virologic and immunologic impact and durability of enfuvirtide-based antiretroviral

- therapy in HIV-infected treatment-experienced patients in a clinical setting, *HIV Clin. Trials*. 8 (2007) 36–44. <https://doi.org/10.1310/hct0801-36>.
- [197] P. Dorr, M. Westby, S. Dobbs, P. Griffin, B. Irvine, M. Macartney, J. Mori, G. Rickett, C. Smith-Burchnell, C. Napier, R. Webster, D. Armour, D. Price, B. Stammen, A. Wood, M. Perros, Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity, *Antimicrob. Agents Chemother.* 49 (2005) 4721–4732. <https://doi.org/10.1128/AAC.49.11.4721-4732.2005>.
- [198] M. Potempa, S.-K. Lee, N. KurtYilmaz, E.A. Nalivaika, A. Rogers, E. Spielvogel, C.W. Carter, C.A. Schiffer, R. Swanstrom, HIV-1 Protease Uses Bi-Specific S2/S2' Subsites To Optimize Cleavage of Two Classes of Target Sites, *J. Mol. Biol.* 430 (2018) 5182–5195. <https://doi.org/10.1016/j.jmb.2018.10.022>.
- [199] J.-Y. Wang, H. Ling, W. Yang, R. Craigie, Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein, *EMBO J.* 20 (2001) 7333–7343. <https://doi.org/10.1093/emboj/20.24.7333>.
- [200] D.J. Hazuda, Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells, *Science*. 287 (2000) 646–650. <https://doi.org/10.1126/science.287.5453.646>.
- [201] K. Salzwedel, D.E. Martin, M. Sakalian, Maturation inhibitors: a new therapeutic class targets the virus structure, *AIDS Rev.* 9 (2007) 162–172.
- [202] K. Van Baelen, K. Salzwedel, E. Rondelez, V. Van Eygen, S. De Vos, A. Verheyen, K. Steegen, Y. Verlinden, G.P. Allaway, L.J. Stuyver, Susceptibility of Human Immunodeficiency Virus Type 1 to the Maturation Inhibitor Bevirimat Is Modulated by Baseline Polymorphisms in Gag Spacer Peptide 1, *Antimicrob. Agents Chemother.* 53 (2009) 2185–2188. <https://doi.org/10.1128/AAC.01650-08>.
- [203] T. Fujioka, Y. Kashiwada, R.E. Kilkuskie, L.M. Cosentino, L.M. Ballas, J.B. Jiang, W.P. Janzen, I.-S. Chen, K.-H. Lee, Anti-AIDS Agents, 11. Betulinic Acid and Platonic Acid as Anti-HIV Principles from *Syzigium claviflorum*, and the Anti-HIV Activity of Structurally Related Triterpenoids, *J. Nat. Prod.* 57 (1994) 243–247. <https://doi.org/10.1021/np50104a008>.
- [204] Y. Kashiwada, F. Hashimoto, L.M. Cosentino, C.-H. Chen, P.E. Garrett, K.-H. Lee, Betulinic Acid and Dihydrobetulinic Acid Derivatives as Potent Anti-HIV Agents, *J. Med. Chem.* 39 (1996) 1016–1017. <https://doi.org/10.1021/jm950922q>.
- [205] T. Kanamoto, Y. Kashiwada, K. Kanbara, K. Gotoh, M. Yoshimori, T. Goto, K. Sano, H. Nakashima, Anti-human immunodeficiency virus activity of YK-FH312 (a betulinic acid derivative), a novel compound blocking viral maturation, *Antimicrob. Agents Chemother.* 45 (2001) 1225–1230. <https://doi.org/10.1128/AAC.45.4.1225-1230.2001>.
- [206] F. Li, R. Goila-Gaur, K. Salzwedel, N.R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D.E. Martin, J.M. Orenstein, G.P. Allaway, E.O. Freed, C.T. Wild, PA-457: A potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing, *Proc. Natl. Acad. Sci.* 100 (2003) 13555–13560. <https://doi.org/10.1073/pnas.2234683100>.
- [207] P.F. Smith, A. Ogundele, A. Forrest, J. Wilton, K. Salzwedel, J. Doto, G.P. Allaway, D.E. Martin, Phase I and II Study of the Safety, Virologic Effect, and Pharmacokinetics/Pharmacodynamics of Single-Dose 3-O-(3',3'-Dimethylsuccinyl)Betulinic Acid (Bevirimat) against Human Immunodeficiency Virus Infection, *Antimicrob. Agents Chemother.* 51 (2007) 3574–3581. <https://doi.org/10.1128/AAC.00152-07>.
- [208] J. Verheyen, C. Verhofstede, E. Knops, L. Vandekerckhove, A. Fun, D. Brunen, K. Dauwe, A. Wensing, H. Pfister, R. Kaiser, M. Nijhuis, High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates, *Aids*. 24 (2010) 669–673. <https://doi.org/10.1097/QAD.0b013e32833160fa>.
- [209] W. Lu, K. Salzwedel, D. Wang, S. Chakravarty, E.O. Freed, C.T. Wild, F. Li, A Single Polymorphism in HIV-1 Subtype C SP1 Is Sufficient To Confer Natural Resistance to the

- Maturation Inhibitor Bevirimat ▽, *Antimicrob. Agents Chemother.* 55 (2011) 3324–3329. <https://doi.org/10.1128/AAC.01435-10>.
- [210] C.S. Adamson, M. Sakalian, K. Salzwedel, E.O. Freed, Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV-1 maturation inhibitor bevirimat, *Retrovirology*. 7 (2010) 36. <https://doi.org/10.1186/1742-4690-7-36>.
- [211] P.W. Keller, C.S. Adamson, J.B. Heymann, E.O. Freed, A.C. Steven, HIV-1 Maturation Inhibitor Bevirimat Stabilizes the Immature Gag Lattice, *J. Virol.* 85 (2011) 1420–1428. <https://doi.org/10.1128/JVI.01926-10>.
- [212] C.S. Adamson, S.D. Ablan, I. Boeras, R. Goila-Gaur, F. Soheilian, K. Nagashima, F. Li, K. Salzwedel, M. Sakalian, C.T. Wild, E.O. Freed, In vitro resistance to the human immunodeficiency virus type 1 maturation inhibitor PA-457 (Beverimat), *J. Virol.* 80 (2006) 10957–10971. <https://doi.org/10.1128/JVI.01369-06>.
- [213] M.D. Purdy, D. Shi, J. Chrustowicz, J. Hattne, T. Gonen, M. Yeager, MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat, *Proc. Natl. Acad. Sci.* 115 (2018) 13258–13263. <https://doi.org/10.1073/pnas.1806806115>.
- [214] S. Gupta, J.M. Louis, R. Tycko, Effects of an HIV-1 maturation inhibitor on the structure and dynamics of CA-SP1 junction helices in virus-like particles, *Proc. Natl. Acad. Sci.* (2020). <https://doi.org/10.1073/pnas.1917755117>.
- [215] W.S. Blair, J. Cao, J. Fok-Seang, P. Griffin, J. Isaacson, R.L. Jackson, E. Murray, A.K. Patick, Q. Peng, M. Perros, C. Pickford, H. Wu, S.L. Butler, New Small-Molecule Inhibitor Class Targeting Human Immunodeficiency Virus Type 1 Virion Maturation, *Antimicrob. Agents Chemother.* 53 (2009) 5080–5087. <https://doi.org/10.1128/AAC.00759-09>.
- [216] K. Qian, I.D. Bori, C.-H. Chen, L. Huang, K.-H. Lee, Anti-AIDS Agents 90. Novel C-28 Modified Bevirimat Analogs as Potent HIV Maturation Inhibitors, *J. Med. Chem.* 55 (2012) 8128–8136. <https://doi.org/10.1021/jm301040s>.
- [217] P. Coric, S. Turcaud, F. Souquet, L. Briant, B. Gay, J. Royer, N. Chazal, S. Bouaziz, Synthesis and biological evaluation of a new derivative of bevirimat that targets the Gag CA-SP1 cleavage site, *Eur. J. Med. Chem.* 62 (2013) 453–465. <https://doi.org/10.1016/j.ejmech.2013.01.013>.
- [218] Z. Dang, P. Ho, L. Zhu, K. Qian, K.-H. Lee, L. Huang, C.-H. Chen, New betulinic acid derivatives for bevirimat-resistant human immunodeficiency virus type-1, *J. Med. Chem.* 56 (2013) 2029–2037. <https://doi.org/10.1021/jm3016969>.
- [219] J. Tang, S.A. Jones, J.L. Jeffery, S.R. Miranda, C.M. Galardi, D.M. Irlbeck, K.W. Brown, C.B. McDanal, N. Han, D. Gao, Y. Wu, B. Shen, C. Liu, C. Xi, H. Yang, R. Li, Y. Yu, Y. Sun, Z. Jin, E. Wang, B.A. Johns, Synthesis and Biological Evaluation of Macrocyclized Betulin Derivatives as a Novel Class of Anti-HIV-1 Maturation Inhibitors, *Open Med. Chem. J.* 8 (2014) 23–27. <https://doi.org/10.2174/1874104501408010023>.
- [220] Z. Liu, J.J. Swidorski, B. Nowicka-Sans, B. Terry, T. Protack, Z. Lin, H. Samanta, S. Zhang, Z. Li, D.D. Parker, S. Rahematpura, S. Jenkins, B.R. Beno, M. Krystal, N.A. Meanwell, I.B. Dicker, A. Regueiro-Ren, C-3 benzoic acid derivatives of C-3 deoxybetulinic acid and deoxybetulin as HIV-1 maturation inhibitors, *Bioorg. Med. Chem.* 24 (2016) 1757–1770. <https://doi.org/10.1016/j.bmc.2016.03.001>.
- [221] B. Nowicka-Sans, T. Protack, Z. Lin, Z. Li, S. Zhang, Y. Sun, H. Samanta, B. Terry, Z. Liu, Y. Chen, N. Sin, S.-Y. Sit, J.J. Swidorski, J. Chen, B.L. Venables, M. Healy, N.A. Meanwell, M. Cockett, U. Hanumegowda, A. Regueiro-Ren, M. Krystal, I.B. Dicker, Identification and Characterization of BMS-955176, a Second-Generation HIV-1 Maturation Inhibitor with Improved Potency, Antiviral Spectrum, and Gag Polymorphic Coverage, *Antimicrob. Agents Chemother.* 60 (2016) 3956–3969. <https://doi.org/10.1128/AAC.02560-15>.
- [222] E. Urano, S.D. Ablan, R. Mandt, G.T. Pauly, D.M. Sigano, J.P. Schneider, D.E. Martin, T.J. Nitz, C.T. Wild, E.O. Freed, Alkyl Amine Bevirimat Derivatives Are Potent and Broadly Active HIV-1 Maturation Inhibitors, *Antimicrob. Agents Chemother.* 60 (2016) 190–197. <https://doi.org/10.1128/AAC.02121-15>.

- [223] E. Chrobak, K. Marciniak, A. Dąbrowska, P. Pęczak, E. Bębenek, M. Kadela-Tomanek, A. Bak, M. Jastrzębska, S. Boryczka, New Phosphorus Analogs of Bevirimat: Synthesis, Evaluation of Anti-HIV-1 Activity and Molecular Docking Study, *Int. J. Mol. Sci.* 20 (2019) 5209. <https://doi.org/10.3390/ijms20205209>.
- [224] J. Cao, J. Isaacson, A.K. Patick, W.S. Blair, High-throughput human immunodeficiency virus type 1 (HIV-1) full replication assay that includes HIV-1 Vif as an antiviral target, *Antimicrob. Agents Chemother.* 49 (2005) 3833–3841. <https://doi.org/10.1128/AAC.49.9.3833-3841.2005>.
- [225] K. Waki, S.R. Durell, F. Soheilian, K. Nagashima, S.L. Butler, E.O. Freed, Structural and Functional Insights into the HIV-1 Maturation Inhibitor Binding Pocket, *PLoS Pathog.* 8 (2012). <https://doi.org/10.1371/journal.ppat.1002997>.
- [226] J. Fontana, P.W. Keller, E. Urano, S.D. Ablan, A.C. Steven, E.O. Freed, Identification of an HIV-1 Mutation in Spacer Peptide 1 That Stabilizes the Immature CA-SP1 Lattice, *J. Virol.* 90 (2015) 972–978. <https://doi.org/10.1128/JVI.02204-15>.
- [227] Insights into the activity of maturation inhibitor PF-46396 on HIV-1 clade C, (n.d.). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333120/> (accessed December 6, 2019).
- [228] A. Neyret, B. Gay, A. Cransac, L. Briant, P. Coric, S. Turcaud, P. Laugâa, S. Bouaziz, N. Chazal, Insight into the mechanism of action of EP-39, a bevirimat derivative that inhibits HIV-1 maturation, *Antiviral Res.* 164 (2019) 162–175. <https://doi.org/10.1016/j.antiviral.2019.02.014>.
- [229] A. Regueiro-Ren, Z. Liu, Y. Chen, N. Sin, S.-Y. Sit, J.J. Swidorski, J. Chen, B.L. Venables, J. Zhu, B. Nowicka-Sans, T. Protack, Z. Lin, B. Terry, H. Samanta, S. Zhang, Z. Li, B.R. Beno, X.S. Huang, S. Rahematpura, D.D. Parker, R. Haskell, S. Jenkins, K.S. Santone, M.I. Cockett, M. Krystal, N.A. Meanwell, U. Hanumegowda, I.B. Dicker, Discovery of BMS-955176, a Second Generation HIV-1 Maturation Inhibitor with Broad Spectrum Antiviral Activity, *ACS Med. Chem. Lett.* 7 (2016) 568–572. <https://doi.org/10.1021/acsmchemlett.6b00010>.
- [230] J. Morales-Ramirez, J.R. Bogner, J.-M. Molina, J. Lombaard, I.B. Dicker, D.A. Stock, M. DeGrosky, M. Gartland, T.P. Dumitrescu, S. Min, C. Llamoso, S.R. Joshi, M. Lataillade, Safety, efficacy, and dose response of the maturation inhibitor GSK3532795 (formerly known as BMS-955176) plus tenofovir/emtricitabine once daily in treatment-naive HIV-1-infected adults: Week 24 primary analysis from a randomized Phase IIb trial, *PLOS ONE.* 13 (2018) e0205368. <https://doi.org/10.1371/journal.pone.0205368>.
- [231] U. Timilsina, D. Ghimire, B. Timalisina, T.J. Nitz, C.T. Wild, E.O. Freed, R. Gaur, Identification of potent maturation inhibitors against HIV-1 clade C, *Sci. Rep.* 6 (2016) 1–6. <https://doi.org/10.1038/srep27403>.
- [232] E. Urano, U. Timilsina, J.A. Kaplan, S. Ablan, D. Ghimire, P. Pham, N. Kuruppu, R. Mandt, S.R. Durell, T.J. Nitz, D.E. Martin, C.T. Wild, R. Gaur, E.O. Freed, Resistance to Second-Generation HIV-1 Maturation Inhibitors, *J. Virol.* 93 (2019). <https://doi.org/10.1128/JVI.02017-18>.
- [233] I.R. Kleckner, M.P. Foster, An introduction to NMR-based approaches for measuring protein dynamics, *Biochim. Biophys. Acta.* 1814 (2011) 942–968. <https://doi.org/10.1016/j.bbapap.2010.10.012>.
- [234] V. Kanelis, J.D. Forman-Kay, L.E. Kay, Multidimensional NMR Methods for Protein Structure Determination, *IUBMB Life.* 52 (2001) 291–302. <https://doi.org/10.1080/152165401317291147>.
- [235] G. Bodenhausen, D.J. Ruben, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chem. Phys. Lett.* 69 (1980) 185–189. [https://doi.org/10.1016/0009-2614\(80\)80041-8](https://doi.org/10.1016/0009-2614(80)80041-8).
- [236] Protein NMR Spectroscopy - 2nd Edition, (n.d.). <https://www.elsevier.com/books/protein-nmr-spectroscopy/cavanagh/978-0-12-164491-8> (accessed May 22, 2020).
- [237] P. Schanda, B. Brutscher, 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 (2005) 8014–8015. <https://doi.org/10.1021/ja051306e>.

- [238] P. Schanda, Ě. Kupĉe, B. Brutscher, SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds, *J. Biomol. NMR.* 33 (2005) 199–211. <https://doi.org/10.1007/s10858-005-4425-x>.
- [239] A. Ross, M. Salzmänn, H. Senn, Fast-HMQC using Ernst angle pulses: An efficient tool for screening of ligand binding to target proteins, (n.d.) 8.
- [240] M. S, A. C, J. Mo, van Z. Pc, Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation, *J. Magn. Reson. B.* 108 (1995). <https://doi.org/10.1006/jmrb.1995.1109>.
- [241] B.T. Farmer, R.A. Venters, L.D. Spicer, M.G. Wittekind, L. Müller, A refocused and optimized HNCA: Increased sensitivity and resolution in large macromolecules, *J. Biomol. NMR.* 2 (1992) 195–202. <https://doi.org/10.1007/BF01875530>.
- [242] S. Grzesiek, A. Bax, *JOURNAL OF MAGNETIC RESONANCE* 96,432-440 (1992) Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein, n.d.
- [243] D. Marion, L.E. Kay, S.W. Sparks, D.A. Torchia, A. Bax, Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins, *J. Am. Chem. Soc.* 111 (1989) 1515–1517. <https://doi.org/10.1021/ja00186a066>.
- [244] A. Bax, M. Ikura, An efficient 3D NMR technique for correlating the proton and¹⁵N backbone amide resonances with the α -carbon of the preceding residue in uniformly¹⁵N/¹³C enriched proteins, *J. Biomol. NMR.* 1 (1991) 99–104. <https://doi.org/10.1007/BF01874573>.
- [245] D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn, G.M. Clore, Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.β., *Biochemistry.* 28 (1989) 6150–6156. <https://doi.org/10.1021/bi00441a004>.
- [246] 3D Manual. HNHA. Introduction, (n.d.). <http://triton.iqfr.csic.es/guide/man/3d/chap5-1.html> (accessed May 21, 2020).
- [247] F.A.A. Mulder, D. Schipper, R. Bott, R. Boelens, Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilis*ins11Edited by P. E. Wright, *J. Mol. Biol.* 292 (1999) 111–123. <https://doi.org/10.1006/jmbi.1999.3034>.
- [248] M. Nilges, Calculation of Protein Structures with Ambiguous Distance Restraints. Automated Assignment of Ambiguous NOE Crosspeaks and Disulphide Connectivities, *J. Mol. Biol.* 245 (1995) 645–660. <https://doi.org/10.1006/jmbi.1994.0053>.
- [249] M. Nilges, M.J. Macias, S.I. O'Donoghue, H. Oschkinat, Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β -spectrin11Edited by P. E. Wright, *J. Mol. Biol.* 269 (1997) 408–422. <https://doi.org/10.1006/jmbi.1997.1044>.
- [250] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr.* 54 (1998) 905–921.
- [251] A.T. Brünger, Version 1.2 of the Crystallography and NMR system, *Nat. Protoc.* 2 (2007) 2728–2733. <https://doi.org/10.1038/nprot.2007.406>.
- [252] D.F. Hansen, P. Vallurupalli, L.E. Kay, An Improved ¹⁵N Relaxation Dispersion Experiment for the Measurement of Millisecond Time-Scale Dynamics in Proteins, *J. Phys. Chem. B.* 112 (2008) 5898–5904. <https://doi.org/10.1021/jp074793o>.
- [253] J.P. Loria, M. Rance, A.G. Palmer, A Relaxation-Compensated Carr–Purcell–Meiboom–Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy, *J. Am. Chem. Soc.* 121 (1999) 2331–2332. <https://doi.org/10.1021/ja983961a>.
- [254] F.A.A. Mulder, N.R. Skrynnikov, B. Hon, F.W. Dahlquist, L.E. Kay, Measurement of Slow (μ s–ms) Time Scale Dynamics in Protein Side Chains by ¹⁵N Relaxation Dispersion NMR

- Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme, *J. Am. Chem. Soc.* 123 (2001) 967–975. <https://doi.org/10.1021/ja003447g>.
- [255] G. Bouvignies, L.E. Kay, Measurement of Proton Chemical Shifts in Invisible States of Slowly Exchanging Protein Systems by Chemical Exchange Saturation Transfer, *J. Phys. Chem. B.* 116 (2012) 14311–14317. <https://doi.org/10.1021/jp311109u>.
- [256] P. Vallurupalli, G. Bouvignies, L.E. Kay, Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation, *J. Am. Chem. Soc.* 134 (2012) 8148–8161. <https://doi.org/10.1021/ja3001419>.
- [257] L.E. Kay, D.A. Torchia, Backbone Dynamics of Proteins As Studied by ¹⁵N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease?, (n.d.) 8.
- [258] D.S. Wishart, B.D. Sykes, F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry.* 31 (1992) 1647–1651. <https://doi.org/10.1021/bi00121a010>.
- [259] H. Zhang, S. Neal, D.S. Wishart, RefDB: a database of uniformly referenced protein chemical shifts, *J. Biomol. NMR.* 25 (2003) 173–195. <https://doi.org/10.1023/a:1022836027055>.
- [260] S. Schwarzingler, G.J. Kroon, T.R. Foss, J. Chung, P.E. Wright, H.J. Dyson, Sequence-dependent correction of random coil NMR chemical shifts, *J. Am. Chem. Soc.* 123 (2001) 2970–2978. <https://doi.org/10.1021/ja003760i>.
- [261] M.-S. Cheung, M.L. Maguire, T.J. Stevens, R.W. Broadhurst, DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure, *J. Magn. Reson.* 202 (2010) 223–233. <https://doi.org/10.1016/j.jmr.2009.11.008>.
- [262] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, ARIA2: Automated NOE assignment and data integration in NMR structure calculation, *Bioinformatics.* 23 (2007) 381–382. <https://doi.org/10.1093/bioinformatics/btl589>.
- [263] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR.* 44 (2009) 213–223. <https://doi.org/10.1007/s10858-009-9333-z>.
- [264] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry.* 28 (1989) 8972–8979. <https://doi.org/10.1021/bi00449a003>.
- [265] HIV-1 Gag Polymorphisms Determine Treatment Response to Bevirimat (PA-457), (n.d.). http://www.natap.org/2008/ResisWksp/ResisWksp_23.htm (accessed March 5, 2020).
- [266] E. Urano, U. Timilsina, J.A. Kaplan, S. Ablan, D. Ghimire, P. Pham, N. Kuruppu, R. Mandt, S.R. Durell, T.J. Nitz, D.E. Martin, C.T. Wild, R. Gaur, E.O. Freed, Resistance to Second-Generation HIV-1 Maturation Inhibitors, *J. Virol.* 93 (2019). <https://doi.org/10.1128/JVI.02017-18>.
- [267] U.K. von Schwedler, K.M. Stray, J.E. Garrus, W.I. Sundquist, Functional Surfaces of the Human Immunodeficiency Virus Type 1 Capsid Protein, *J. Virol.* 77 (2003) 5439–5450. <https://doi.org/10.1128/JVI.77.9.5439-5450.2003>.
- [268] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ -RNA packaging signal. implications for genome recognition 11 Edited by P. Wright, *J. Mol. Biol.* 301 (2000) 491–511. <https://doi.org/10.1006/jmbi.2000.3979>.
- [269] R.N.D. Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 Ψ -RNA Recognition Element, *Science.* 279 (1998) 384–388. <https://doi.org/10.1126/science.279.5349.384>.

Appendices

I Assigned chemical shifts

The assigned chemical shifts which correspond to the protein CA_{CTD}^{W184A, M185A}-SP1-NC (in the presence of 6 eq EP39) concentrated to 0.3 mM in a buffer 25 mM Acetate Sodium, 25 mM NaCl, 0.1 mM BME, 0.1 mM ZnCl₂ and pH 6.5. The residue number is set from 1 to 157, corresponding to residue number 144 and 300 in the text.

Residue	Assign Name	Shift	SD
6Ser	H	8.40111	0.00
6Ser	N	118.63268	0.00
6Ser	Ha	4.49797	0.00
6Ser	Hba	3.70431	0.00
6Ser	Hbb	3.85337	0.00
9Asp	H	7.64658	0.00
9Asp	N	116.29609	0.01
9Asp	Ha	4.6894	0.00
9Asp	Hba	2.54624	0.00
9Asp	Hbb	2.86143	0.00
10Ile	H	7.44036	0.00
10Ile	N	122.27981	0.00
10Ile	Ha	4.12478	0.00
10Ile	Hb	2.26334	0.02
10Ile	Hg2*	0.91491	0.00
12Gln	Ha	4.86132	0.00
12Gln	H	7.61493	0.00
12Gln	N	127.13344	0.00
13Gly	Haa	3.91607	0.02
13Gly	H	9.82517	0.00
13Gly	N	117.12003	0.00
13Gly	Hab	4.32698	0.00
16Glu	H	7.05742	0.00
16Glu	N	125.77224	0.00
16Glu	Ha	4.59687	0.00
16Glu	Hba	1.75835	0.01
16Glu	Hga	2.27755	0.00
17Pro	Ha	4.41096	0.00
17Pro	Hba	2.53267	0.00
17Pro	Hga	2.21657	0.00
18Phe	Ha	4.48328	0.00
18Phe	Hba	2.95059	0.02
18Phe	Hbb	3.34417	0.01
18Phe	N	128.10494	0.00
18Phe	H	9.18696	0.00
19Arg	He	7.28228	0.00
19Arg	Hda	3.16301	0.00
19Arg	Hdb	3.22531	0.00
19Arg	Hga	1.66249	0.00
19Arg	Hgb	1.66249	0.00
19Arg	H	9.11965	0.00
19Arg	N	115.71005	0.00
19Arg	Ha	3.75197	0.00
19Arg	Hba	1.7485	0.01
19Arg	Hbb	1.90804	0.01
20Asp	Hba	2.85577	0.00
20Asp	Ha	4.41671	0.01
20Asp	Hbb	2.85756	0.00
20Asp	H	6.97376	0.00

Residue	Assign Name	Shift	SD
21Tyr	H	7.54543	0.00
21Tyr	N	124.74995	0.00
21Tyr	Ha	4.3879	0.01
21Tyr	Hba	3.16518	0.01
22Val	Hga*	0.21167	0.00
22Val	Hb	1.77064	0.00
22Val	Ha	2.71127	0.00
22Val	H	8.23758	0.01
22Val	N	120.42027	0.00
22Val	Hgb*	0.8359	0.00
23Asp	Hbb	2.77861	0.00
23Asp	Ha	4.42895	0.01
23Asp	Hba	2.72318	0.03
23Asp	H	7.46294	0.00
23Asp	N	117.72534	0.00
24Arg	Hbb	1.96949	0.00
24Arg	Hgb	1.5236	0.00
24Arg	Hda	3.15828	0.00
24Arg	Hdb	3.35953	0.00
24Arg	H	7.90455	0.00
24Arg	N	119.47094	0.00
24Arg	Ha	4.00497	0.01
24Arg	Hba	1.88186	0.03
24Arg	Hga	1.52064	0.00
25Phe	Hd*	6.32183	0.00
25Phe	Hbb	2.85378	0.00
25Phe	H	8.9449	0.00
25Phe	N	124.57999	0.00
25Phe	Ha	3.65695	0.00
25Phe	Hba	2.85025	0.00
26Tyr	Hbb	3.00108	0.00
26Tyr	Hd*	7.26184	0.00
26Tyr	He*	6.66856	0.00
26Tyr	H	8.60207	0.00
26Tyr	N	116.76975	0.00
26Tyr	Ha	4.09873	0.00
26Tyr	Hba	2.96316	0.00
27Lys	Hga	1.37515	0.00
27Lys	Hgb	1.37515	0.00
27Lys	Hea	3.15581	0.00
27Lys	Heb	3.15581	0.00
27Lys	Hbb	1.91914	0.00
27Lys	Hda	1.57246	0.00
27Lys	Hdb	1.57246	0.00
27Lys	H	8.29945	0.00
27Lys	N	120.78305	0.00
27Lys	Ha	3.9885	0.01
27Lys	Hba	1.84553	0.03

28Thr	Hb	4.01476	0.01	36Gln	He2a	6.76655	0.00
28Thr	Ha	3.84366	0.00	37Glu	Hba	1.96713	0.00
28Thr	H	7.66087	0.00	37Glu	Hgb	2.22202	0.00
28Thr	N	116.97308	0.00	37Glu	H	8.43796	0.00
28Thr	Hg1	0.93602	0.01	37Glu	N	119.82	0.00
29Leu	Hba	1.40075	0.00	37Glu	Ha	4.1497	0.00
29Leu	Hdb*	0.7308	0.00	37Glu	Hbb	1.96978	0.00
29Leu	H	7.89123	0.00	37Glu	Hga	2.22144	0.00
29Leu	N	122.30088	0.00	38Val	Hgb*	0.91117	0.00
29Leu	Ha	3.78752	0.00	38Val	H	7.77074	0.00
29Leu	Hbb	1.52515	0.00	38Val	N	120.70735	0.00
29Leu	Hg	1.32291	0.00	38Val	Ha	3.94414	0.00
29Leu	Hda*	0.65216	0.02	38Val	Hga*	0.85007	0.00
30Arg	Hbb	1.8912	0.00	38Val	Hb	2.06987	0.00
30Arg	Hda	3.15744	0.00	39Lys	Hba	1.76884	0.00
30Arg	H	7.92559	0.01	39Lys	Hgb	1.45006	0.00
30Arg	N	116.3732	0.00	39Lys	H	8.27407	0.00
30Arg	Ha	4.10046	0.01	39Lys	N	123.10007	0.00
30Arg	Hba	1.86196	0.02	39Lys	Ha	4.18889	0.01
30Arg	Hga	1.64303	0.03	39Lys	Hbb	1.77064	0.00
31Ala	H	7.5379	0.00	39Lys	Hga	1.43245	0.01
31Ala	N	121.80086	0.00	40Asn	H	8.4245	0.00
31Ala	Ha	4.19357	0.00	40Asn	N	119.3997	0.00
31Ala	Hb*	1.4765	0.00	41Ala	Ha	4.30275	0.01
32Glu	Hbb	2.23401	0.00	41Ala	H	8.15315	0.00
32Glu	Hga	2.08082	0.01	41Ala	N	123.53992	0.00
32Glu	Hgb	2.08082	0.01	41Ala	Hb*	1.38342	0.00
32Glu	H	7.89267	0.00	42Ala	H	8.21871	0.00
32Glu	N	118.05087	0.00	42Ala	N	122.18473	0.01
32Glu	Ha	4.20815	0.01	42Ala	Ha	4.26936	0.00
32Glu	Hba	1.98276	0.02	42Ala	Hb*	1.37788	0.00
33Gln	Hba	2.02851	0.00	43Thr	Hb	4.34315	0.00
33Gln	Hga	2.31623	0.00	43Thr	H	8.14511	0.00
33Gln	H	8.13877	0.00	43Thr	N	114.35989	0.00
33Gln	N	118.77094	0.00	43Thr	Hg2*	1.12791	0.00
33Gln	Ha	4.1872	0.01	43Thr	Ha	4.22455	0.00
33Gln	Hbb	2.04582	0.02	44Glu	Hga	2.31488	0.00
33Gln	Hgb	2.33092	0.01	44Glu	Hgb	2.31488	0.00
34Ala	H	7.95346	0.00	44Glu	H	8.49803	0.00
34Ala	N	122.84939	0.00	44Glu	N	120.14131	0.00
34Ala	Ha	4.24668	0.00	44Glu	Ha	3.99626	0.00
34Ala	Hb*	1.34886	0.00	44Glu	Hba	2.06713	0.00
35Ser	H	8.11159	0.00	44Glu	Hbb	2.06713	0.00
35Ser	N	114.29994	0.00	45Thr	Hb	4.04516	0.00
35Ser	Ha	4.37555	0.00	45Thr	Ha	4.19671	0.00
35Ser	Hba	4.00081	0.00	45Thr	H	7.89119	0.00
36Gln	H	8.47537	0.00	45Thr	N	112.97553	0.00
36Gln	N	121.33898	0.00	46Leu	Hbb	1.86814	0.00
36Gln	Ne2	111.05826	0.03	46Leu	Hda*	0.87707	0.00
36Gln	He2b	7.4551	0.00	46Leu	Hdb*	0.87707	0.00

46Leu	N	123.26594	0.00
46Leu	H	7.8023	0.00
46Leu	Ha	4.09682	0.01
46Leu	Hba	1.78639	0.02
47Leu	Hba	1.47193	0.00
47Leu	Hbb	1.55244	0.00
47Leu	Hda*	0.87051	0.00
47Leu	H	7.5435	0.00
47Leu	N	119.40202	0.00
47Leu	Ha	3.97918	0.02
47Leu	Hdb*	0.87538	0.00
48Val	Hga*	0.87325	0.00
48Val	H	7.46132	0.00
48Val	N	116.70485	0.00
48Val	Ha	3.56215	0.00
48Val	Hb	2.04393	0.00
48Val	Hgb*	1.02892	0.00
49Gln	Hbb	2.16518	0.00
49Gln	H	7.85411	0.00
49Gln	N	117.13135	0.00
49Gln	Ha	3.84907	0.00
49Gln	Hba	2.08477	0.02
49Gln	Hga	2.41293	0.00
50Asn	H	7.76705	0.00
50Asn	N	115.51812	0.01
50Asn	Ha	4.86828	0.02
50Asn	Hba	3.29853	0.00
50Asn	Hbb	3.56228	0.00
51Ala	H	7.27504	0.00
51Ala	N	124.47394	0.00
51Ala	Ha	4.20936	0.01
51Ala	Hb*	1.5733	0.00
52Asn	Hba	3.05715	0.00
52Asn	Ha	4.64858	0.01
52Asn	H	8.9816	0.00
52Asn	N	119.41995	0.00
53Pro	Hba	1.85861	0.00
53Pro	Hbb	2.35957	0.00
54Asp	Hbb	2.56369	0.00
54Asp	H	7.91407	0.00
54Asp	N	116.30965	0.00
54Asp	Ha	4.35443	0.01
54Asp	Hba	2.56319	0.00
55Cys	H	8.28574	0.00
55Cys	N	117.6409	0.00
55Cys	Hbb	3.4106	0.01
55Cys	Hba	2.76708	0.01
55Cys	Ha	4.6455	0.01
56Lys	Hbb	1.79734	0.00
56Lys	Hda	1.70779	0.00

56Lys	Hdb	1.70779	0.00
56Lys	N	121.15924	0.01
56Lys	Hba	1.74514	0.01
56Lys	Ha	3.65455	0.00
56Lys	H	9.25698	0.00
57Thr	Ha	3.75464	0.01
57Thr	Hb	4.19363	0.00
57Thr	H	7.72737	0.00
57Thr	N	113.33487	0.00
57Thr	[47]	4.19647	0.00
57Thr	[48]	3.74443	0.00
57Thr	Hg1	1.22747	0.00
58Ile	Hg1a	1.03265	0.00
58Ile	Hg2*	0.87868	0.00
58Ile	H	7.0367	0.00
58Ile	N	123.75013	0.00
58Ile	Ha	3.65567	0.01
58Ile	Hb	2.11384	0.00
58Ile	Hd1*	0.73598	0.00
58Ile	Hg1b	1.69875	0.02
59Leu	Hg	1.81613	0.00
59Leu	Hda*	1.27529	0.00
59Leu	Hdb*	1.27529	0.00
59Leu	Hbb	1.86882	0.00
59Leu	H	8.59721	0.00
59Leu	N	118.67999	0.00
59Leu	Ha	3.86633	0.01
59Leu	Hba	1.86847	0.00
60Lys	Hba	1.82128	0.00
60Lys	Hda	1.571	0.00
60Lys	Hdb	1.571	0.00
60Lys	Hga	1.42629	0.00
60Lys	Hgb	1.46829	0.04
60Lys	H	8.33619	0.00
60Lys	N	119.48	0.00
60Lys	Ha	3.95009	0.00
60Lys	Hbb	1.82192	0.00
61Ala	H	7.17762	0.01
61Ala	N	120.05743	0.00
61Ala	Ha	4.22879	0.00
61Ala	Hb*	1.47397	0.00
62Leu	Hda*	0.61129	0.00
62Leu	Hbb	1.82174	0.00
62Leu	Hdb*	0.73108	0.00
62Leu	H	7.45433	0.00
62Leu	N	118.61778	0.00
62Leu	Ha	4.14424	0.00
62Leu	Hba	1.81684	0.01
62Leu	Hg	1.58457	0.01
63Gly	H	7.49434	0.00

63Gly	N	104.70594	0.00
63Gly	Haa	3.97238	0.00
63Gly	Hab	4.15161	0.01
65Gly	Haa	3.85088	0.00
65Gly	H	8.70917	0.00
65Gly	N	109.51013	0.00
65Gly	Hab	3.85487	0.00
66Ala	H	7.15004	0.00
66Ala	N	123.36403	0.00
66Ala	Ha	4.38681	0.01
66Ala	Hb*	1.32723	0.00
67Thr	Hb	4.38053	0.01
67Thr	Hg2*	1.32657	0.00
67Thr	H	8.59771	0.00
67Thr	N	113.36987	0.00
67Thr	Ha	4.42527	0.00
68Leu	Hbb	1.69076	0.00
68Leu	H	8.8499	0.00
68Leu	N	122.42992	0.00
68Leu	Ha	4.19741	0.00
68Leu	Hba	1.63832	0.01
69Glu	Hbb	2.01648	0.00
69Glu	Hga	2.21909	0.00
69Glu	Hba	1.88872	0.00
69Glu	H	8.5624	0.00
69Glu	N	117.60997	0.00
69Glu	Ha	3.95086	0.00
69Glu	Hgb	2.31262	0.00
70Glu	Hga	2.22466	0.00
70Glu	Hbb	2.0198	0.00
70Glu	H	7.70639	0.00
70Glu	N	120.28632	0.02
70Glu	Ha	3.93382	0.00
70Glu	Hgb	2.37128	0.01
70Glu	Hba	1.87793	0.01
71Met	Hgb	2.81041	0.00
71Met	Hbb	2.40186	0.00
71Met	Hba	2.16054	0.00
71Met	Ha	3.6024	0.00
71Met	Hga	2.81041	0.00
71Met	H	8.40292	0.00
71Met	N	120.34221	0.00
72Met	Hga	2.91052	0.00
72Met	Hba	2.11731	0.00
72Met	Ha	4.14865	0.00
72Met	Hbb	2.34171	0.01
72Met	Hgb	2.96239	0.05
72Met	H	8.62176	0.00
72Met	N	116.61594	0.00
73Thr	H	8.01678	0.01

73Thr	N	115.1501	0.00
73Thr	Ha	3.85581	0.00
73Thr	Hb	4.11828	0.00
73Thr	Hg2*	1.17285	0.00
74Ala	H	7.86969	0.00
74Ala	N	123.76499	0.02
74Ala	Ha	4.0481	0.00
74Ala	Hb*	1.32946	0.00
75Cys	H	7.56816	0.00
75Cys	N	111.0876	0.00
75Cys	Ha	4.59207	0.00
75Cys	Hbb	3.0303	0.01
75Cys	Hba	2.51845	0.00
76Gln	Hbb	2.45462	0.00
76Gln	Hgb	2.53066	0.00
76Gln	Hba	2.2155	0.00
76Gln	Hga	2.34316	0.00
76Gln	H	7.60706	0.00
76Gln	N	121.10321	0.00
76Gln	Ha	4.29761	0.00
78Val	H	7.55596	0.00
78Val	N	119.35042	0.00
78Val	Ha	3.9483	0.00
78Val	Hb	2.07396	0.00
78Val	Hga*	0.88201	0.00
79Gly	Haa	4.04084	0.00
79Gly	H	8.53163	0.00
79Gly	N	113.69075	0.00
80Gly	H	8.03883	0.00
80Gly	N	108.46368	0.00
80Gly	Haa	4.10353	0.00
85Ala	H	8.21964	0.00
85Ala	N	124.30057	0.00
85Ala	Ha	4.84433	0.00
85Ala	Hb*	1.39919	0.00
87Val	Hga*	0.88024	0.00
87Val	Hgb*	0.88024	0.00
87Val	Ha	3.99652	0.00
87Val	N	121.27027	0.00
87Val	Hb	1.99829	0.01
87Val	H	8.06514	0.00
88Leu	Hbb	1.58651	0.01
88Leu	Hg	1.56218	0.01
88Leu	H	8.17681	0.00
88Leu	N	125.29009	0.00
88Leu	Hba	1.58612	0.01
88Leu	Ha	4.24753	0.00
88Leu	Hda*	0.85672	0.02
89Ala	H	8.12251	0.00
89Ala	N	123.99023	0.00

89Ala	Ha	4.20331	0.00
89Ala	Hb*	1.37754	0.00
90Glu	Hba	1.96702	0.00
90Glu	Hgb	2.22599	0.00
90Glu	H	8.25145	0.00
90Glu	N	119.51019	0.00
90Glu	Ha	4.15398	0.00
90Glu	Hbb	1.96835	0.00
90Glu	Hga	2.22394	0.00
91Ala	H	8.15513	0.00
91Ala	N	124.27118	0.02
91Ala	Ha	4.20275	0.00
91Ala	Hb*	1.37681	0.00
92Met	Ha	4.39346	0.00
92Met	Hbb	2.08505	0.00
92Met	Hga	2.52507	0.00
92Met	Hgb	2.56304	0.00
92Met	H	8.1474	0.00
92Met	N	118.34959	0.00
92Met	Hba	2.02841	0.01
93Ser	H	8.09919	0.00
93Ser	N	116.16732	0.00
94Gln	Ha	4.33442	0.01
94Gln	H	8.23875	0.00
94Gln	N	121.6797	0.00
94Gln	Hba	1.97621	0.00
94Gln	Hbb	1.97621	0.00
94Gln	Hga	2.33292	0.01
95Val	H	8.02142	0.00
95Val	N	120.4919	0.00
95Val	Ha	4.11935	0.00
95Val	Hb	2.07195	0.00
95Val	Hga*	0.89236	0.00
96Thr	Ha	4.26657	0.00
96Thr	H	8.11028	0.01
96Thr	N	117.64365	0.00
97Asn	H	8.38538	0.00
97Asn	N	122.29974	0.00
97Asn	Hd2a	7.50378	0.00
97Asn	Hba	2.85419	0.00
97Asn	Hbb	2.85419	0.00
100Thr	Hg2*	1.15532	0.00
100Thr	H	7.88527	0.00
100Thr	N	112.51446	0.01
100Thr	Ha	4.19908	0.00
101Ile	Hg1b	1.42588	0.00
101Ile	Hg1a	1.13013	0.00
101Ile	Hg2*	0.8349	0.00
101Ile	H	7.84	0.00
101Ile	N	122.5404	0.00
101Ile	Ha	4.07902	0.00
101Ile	Hb	1.82428	0.00
102Met	H	8.20599	0.00
102Met	N	123.40028	0.00
102Met	Ha	4.44472	0.00
102Met	Hba	1.97449	0.00
102Met	Hbb	1.97449	0.00
102Met	Hga	2.51584	0.00
102Met	Hgb	2.51584	0.00
103Ile	Hg1b	1.42479	0.00
103Ile	Hg1a	1.1301	0.00
103Ile	Hg2*	0.84277	0.00
103Ile	H	7.99138	0.00
103Ile	N	122.20003	0.00
103Ile	Hb	1.82646	0.00
103Ile	Ha	4.05644	0.01
104Gln	Hga	2.31829	0.00
104Gln	Hgb	2.31829	0.00
104Gln	Ha	4.27374	0.00
104Gln	Hba	1.99229	0.00
104Gln	Hbb	1.99229	0.00
104Gln	H	8.31623	0.00
104Gln	N	123.95894	0.00
114Thr	H	8.12104	0.00
114Thr	N	117.64809	0.00
114Thr	Ha	4.28888	0.01
115Val	Hgb*	0.7324	0.00
115Val	Hb	1.80901	0.02
115Val	H	8.34481	0.00
115Val	N	123.12003	0.02
115Val	Ha	4.10307	0.00
115Val	Hga*	0.72815	0.01
116Lys	Hbb	1.42425	0.00
116Lys	Hga	0.98056	0.00
116Lys	H	8.0822	0.00
116Lys	N	124.6598	0.00
116Lys	Ha	4.44079	0.00
116Lys	Hba	1.42344	0.00
116Lys	Hgb	0.98506	0.00
116Lys	Hz*	7.66242	0.00
117Cys	Hba	1.72078	0.00
117Cys	H	8.20224	0.00
117Cys	N	127.21984	0.00
117Cys	Ha	4.09617	0.00
117Cys	Hbb	2.75747	0.01
118Phe	Hbb	3.15776	0.00
118Phe	H	8.53252	0.00
118Phe	N	128.41467	0.01
118Phe	Ha	4.55104	0.01
118Phe	Hba	3.08421	0.03

119Asn	Hbb	3.08239	0.00
119Asn	H	9.51095	0.01
119Asn	N	121.82005	0.00
119Asn	Ha	4.69258	0.01
119Asn	Hba	2.46653	0.00
120Cys	Hba	2.47253	0.00
120Cys	H	8.83733	0.01
120Cys	N	117.32123	0.00
120Cys	Ha	4.89052	0.00
120Cys	Hbb	3.20924	0.00
121Gly	H	7.84368	0.00
121Gly	N	113.38986	0.00
121Gly	Hab	3.99929	0.00
121Gly	Haa	3.69376	0.01
122Lys	Hbb	1.86959	0.00
122Lys	Hda	1.57537	0.00
122Lys	Hdb	1.57537	0.00
122Lys	Hga	1.47396	0.00
122Lys	H	8.07494	0.00
122Lys	N	120.9411	0.00
122Lys	Ha	4.39705	0.00
122Lys	Hba	1.78249	0.01
122Lys	Hgb	1.47854	0.00
123Glu	Hbb	1.96847	0.00
123Glu	Hga	2.1182	0.00
123Glu	H	8.3417	0.00
123Glu	N	118.15028	0.00
123Glu	Ha	4.58848	0.00
123Glu	Hba	1.88359	0.01
123Glu	Hgb	2.26852	0.00
124Gly	Haa	3.65006	0.00
124Gly	H	8.59163	0.00
124Gly	N	107.44987	0.00
124Gly	Hab	4.44495	0.00
125His	Hba	3.11825	0.01
125His	H	7.14244	0.00
125His	N	114.09692	0.00
125His	Ha	4.69584	0.01
125His	Hbb	3.12074	0.01
126Ile	Hg1a	1.23355	0.01
126Ile	Hg2*	0.87808	0.00
126Ile	H	7.79395	0.00
126Ile	N	108.26014	0.00
126Ile	Ha	4.6931	0.01
126Ile	Hb	2.11501	0.00
127Ala	Ha	4.68867	0.00
127Ala	H	8.70627	0.00
127Ala	N	124.68319	0.00
127Ala	Hb*	1.27975	0.01
128Lys	Hgb	1.5726	0.00
128Lys	H	8.18756	0.00
128Lys	N	117.08013	0.00
128Lys	Ha	4.05127	0.00
128Lys	Hba	1.87443	0.00
128Lys	Hga	1.56656	0.01
129Asn	H	7.86487	0.00
129Asn	N	114.16004	0.00
129Asn	Ha	4.98941	0.00
129Asn	Hba	2.52244	0.01
129Asn	Hbb	3.06459	0.01
130Cys	H	7.63237	0.00
130Cys	N	124.25098	0.00
130Cys	Ha	3.70551	0.01
130Cys	Hba	2.95717	0.00
130Cys	Hbb	3.33123	0.01
131Arg	H	8.6097	0.00
131Arg	Ha	4.45355	0.00
132Ala	H	8.69517	0.00
132Ala	N	128.13288	0.00
132Ala	Ha	4.51707	0.00
132Ala	Hb*	1.32671	0.00
135Lys	Hga	1.3748	0.00
135Lys	Hgb	1.3748	0.00
135Lys	Hba	1.76809	0.00
135Lys	H	8.57298	0.00
135Lys	N	125.99086	0.00
135Lys	Ha	4.24436	0.00
135Lys	Hbb	1.77113	0.00
136Lys	Hga	1.42448	0.00
136Lys	Hgb	1.42448	0.00
136Lys	Hda	1.57569	0.00
136Lys	Hdb	1.57569	0.00
136Lys	Hba	1.81995	0.00
136Lys	H	8.38424	0.00
136Lys	N	122.29998	0.00
136Lys	Ha	4.38618	0.00
136Lys	Hbb	1.82139	0.00
137Gly	Haa	3.69913	0.00
137Gly	H	8.10099	0.00
137Gly	N	109.06349	0.00
137Gly	Hab	3.70536	0.01
138Cys	Hba	2.02118	0.00
138Cys	H	8.12771	0.00
138Cys	N	124.24141	0.00
138Cys	Ha	4.03339	0.01
138Cys	Hbb	2.85976	0.00
139Trp	H2	7.46012	0.00
139Trp	Hbb	3.41913	0.00
139Trp	He1	10.05244	0.00
139Trp	Ne1	128.97937	0.00

139Trp	Hd1	7.26501	0.00
139Trp	H	8.37221	0.00
139Trp	N	129.49938	0.01
139Trp	Ha	4.54168	0.00
139Trp	Hba	3.39799	0.02
140Lys	H	9.32882	0.00
140Lys	N	122.93537	0.01
140Lys	Ha	4.25046	0.00
140Lys	Hba	2.06637	0.01
140Lys	Hga	1.23499	0.00
140Lys	Hda	1.67294	0.01
141Cys	H	8.46513	0.01
141Cys	N	117.1692	0.00
141Cys	Ha	4.89448	0.00
141Cys	Hbb	3.18282	0.00
141Cys	Hba	2.48023	0.01
142Gly	H	8.10205	0.00
142Gly	N	113.7958	0.01
142Gly	Haa	3.80163	0.00
142Gly	Hab	4.04055	0.00
143Lys	H	8.38036	0.00
143Lys	N	122.30036	0.00
143Lys	Ha	4.37274	0.00
143Lys	Hba	1.72083	0.00
143Lys	Hbb	1.72083	0.00
143Lys	Hga	1.40791	0.00
143Lys	Hgb	1.40791	0.00
143Lys	Hda	1.59097	0.00
143Lys	Hdb	1.59097	0.00
145Gly	Hab	4.27458	0.03
145Gly	Haa	3.71788	0.01
146His	Hba	3.15779	0.00
146His	H	7.11206	0.00
146His	N	112.92434	0.00
146His	Ha	4.77528	0.01
146His	Hbb	3.17333	0.02
147Gln	Hbb	1.91485	0.00
147Gln	Hga	2.31568	0.00
147Gln	N	116.96219	0.00
147Gln	Ha	4.54959	0.00
147Gln	Hba	1.91404	0.00
147Gln	Hgb	2.31988	0.00
147Gln	H	8.86202	0.00
148Met	H	8.82241	0.00
148Met	N	121.86043	0.00
148Met	Ha	4.74024	0.05
148Met	Hba	2.13014	0.01
149Lys	Hea	3.2034	0.00
149Lys	Heb	3.2034	0.00
149Lys	Hda	1.71804	0.00

149Lys	Hdb	1.71804	0.00
149Lys	Ha	4.09692	0.00
149Lys	H	8.46635	0.00
149Lys	N	118.42022	0.00
149Lys	Hba	1.86121	0.02
150Asp	H	7.80398	0.00
150Asp	N	117.62835	0.00
150Asp	Ha	4.83635	0.00
150Asp	Hba	2.51489	0.00
150Asp	Hbb	2.91595	0.00
151Cys	H	7.54308	0.00
151Cys	N	123.67013	0.00
151Cys	Ha	3.64601	0.01
151Cys	Hbb	3.24255	0.00
151Cys	Hba	2.84422	0.01
152Thr	H	8.1226	0.00
152Thr	N	119.70135	0.00
152Thr	Ha	4.35494	0.00
153Glu	Hbb	2.01806	0.00
153Glu	Hgb	2.38705	0.00
153Glu	H	8.59584	0.00
153Glu	N	125.5727	0.00
153Glu	Ha	4.19019	0.00
153Glu	Hba	1.9286	0.01
153Glu	Hga	2.21629	0.00
156Ala	H	8.33122	0.00
156Ala	N	126.26093	0.00
156Ala	Hb*	1.3275	0.00
156Ala	Ha	4.29966	0.00
157Asn	Hba	2.62934	0.00
157Asn	Hbb	2.72354	0.00
157Asn	Ha	4.44087	0.00
157Asn	H	7.94551	0.00
157Asn	N	123.55103	0.00

The assigned chemical shifts which correspond to the protein CA_{CTD}^{W184A, M185A}-SP1-NC concentrated to 0.3 mM in a buffer 25 mM Acetate Sodium, 25 mM NaCl, 0.1 mM BME, 0.1 mM ZnCl₂ and pH 6.5. The residue number is set from 1 to 157, corresponding to residue number 144 and 300 in the text.

Residue	Assign Name	Shift	SD
4Pro	Ha	4.35354	0.00
4Pro	Hda	3.61186	0.00
4Pro	Hdb	3.61186	0.00
4Pro	Hba	2.22242	0.00
4Pro	Hbb	2.22242	0.00
4Pro	Hgb	1.92795	0.00
4Pro	Hga	1.92795	0.00
5Thr	H	8.12425	0.00
5Thr	N	114.33992	0.00
5Thr	Hg2*	1.13753	0.00
5Thr	Ha	4.2052	0.00
5Thr	Hb	4.10533	0.00
6Ser	H	8.41521	0.00
6Ser	N	118.7881	0.00
6Ser	Ha	4.50017	0.00
6Ser	Hba	3.70781	0.00
6Ser	Hbb	3.85752	0.00
7Ile	Hd1*	0.69244	0.00
7Ile	Hg1a	1.37908	0.00
7Ile	Hb	1.72947	0.00
7Ile	H	8.34639	0.00
7Ile	N	123.00988	0.00
7Ile	Ha	4.10289	0.00
7Ile	Hg2*	0.73973	0.00
8Leu	Hba	1.53378	0.00
8Leu	Hda*	0.83988	0.00
8Leu	Hdb*	0.84047	0.00
8Leu	H	7.54534	0.00
8Leu	N	119.40984	0.02
8Leu	Ha	4.00398	0.00
8Leu	Hbb	1.53382	0.00
9Asp	H	7.65115	0.00
9Asp	N	116.32604	0.00
9Asp	Ha	4.69751	0.00
9Asp	Hba	2.57062	0.00
9Asp	Hbb	2.84381	0.00
10Ile	Hd1*	0.63923	0.00
10Ile	Hg2*	0.93802	0.00
10Ile	Hg1a	1.53214	0.00
10Ile	Ha	4.15377	0.00
10Ile	Hb	2.27294	0.00
10Ile	N	122.29997	0.00
10Ile	H	7.43492	0.00
11Arg	Hbb	1.78118	0.00
11Arg	Hga	1.48289	0.00
11Arg	Hgb	1.48299	0.00
11Arg	Hda	3.11477	0.00
11Arg	Hdb	3.16494	0.00
11Arg	H	8.24632	0.00
11Arg	N	124.41994	0.00
11Arg	Ha	4.8737	0.00
11Arg	Hba	1.63085	0.00
12Gln	Ha	3.78188	0.00
12Gln	Hgb	2.29862	0.01
12Gln	Hbb	2.02944	0.00
12Gln	Hba	1.78359	0.00
12Gln	Hga	2.2981	0.01
12Gln	H	7.61474	0.00
12Gln	N	127.15214	0.00
13Gly	Hab	4.35659	0.00
13Gly	Haa	3.94504	0.00
13Gly	H	9.84374	0.00
13Gly	N	117.15	0.00
14Pro	Hda	3.76102	0.00
14Pro	Ha	4.24988	0.00
14Pro	Hbb	2.4215	0.00
14Pro	Hba	2.08124	0.00
14Pro	Hdb	3.7631	0.00
15Lys	Hbb	2.07768	0.00
15Lys	Hba	2.07706	0.00
15Lys	Hgb	1.33543	0.00
15Lys	Hda	1.58237	0.00
15Lys	Hdb	1.58262	0.00
15Lys	Ha	4.50204	0.00
15Lys	H	8.51064	0.00
15Lys	N	115.36001	0.00
15Lys	Hga	1.33355	0.00
16Glu	Hbb	2.07468	0.00
16Glu	Ha	4.60291	0.00
16Glu	Hba	1.78216	0.00
16Glu	Hgb	2.27462	0.00
16Glu	Hga	2.27456	0.00
16Glu	H	7.06619	0.00
16Glu	N	125.79892	0.00
17Pro	Hbb	2.53958	0.01
17Pro	Hba	2.52526	0.00
17Pro	Ha	4.41381	0.00
18Phe	He*	7.06915	0.00
18Phe	Hh	6.89815	0.00
18Phe	Hd*	7.2929	0.00
18Phe	Ha	4.49954	0.00
18Phe	Hba	2.96466	0.00
18Phe	Hbb	3.36821	0.00
18Phe	H	9.20222	0.00
18Phe	N	128.18999	0.00
19Arg	Hda	3.21602	0.00
19Arg	Hdb	3.21625	0.00
19Arg	Hba	1.73287	0.00
19Arg	Hbb	1.92822	0.00
19Arg	Hga	1.68473	0.00

19Arg	Hgb	1.68497	0.00
19Arg	Ha	3.75955	0.00
19Arg	H	9.13852	0.00
19Arg	N	115.77997	0.00
20Asp	Hbb	2.8666	0.00
20Asp	Hba	2.86646	0.00
20Asp	Ha	4.40298	0.00
20Asp	H	6.98048	0.00
20Asp	N	119.20311	0.00
21Tyr	Hd*	6.82405	0.00
21Tyr	Hbb	2.8647	0.00
21Tyr	Ha	4.3954	0.01
21Tyr	Hba	2.8646	0.00
21Tyr	H	7.55286	0.00
21Tyr	N	124.79998	0.00
22Val	Hga*	0.23675	0.01
22Val	Ha	2.72054	0.00
22Val	H	8.24495	0.00
22Val	N	120.4952	0.00
22Val	Hgb*	0.8423	0.00
22Val	Hb	1.78213	0.00
23Asp	Hbb	2.78502	0.00
23Asp	H	7.47145	0.00
23Asp	N	117.80998	0.00
23Asp	Ha	4.4261	0.00
23Asp	Hba	2.69202	0.00
24Arg	Hga	1.53182	0.00
24Arg	Hda	3.16789	0.00
24Arg	Hdb	3.3663	0.00
24Arg	Hbb	1.99671	0.00
24Arg	Hgb	1.82094	0.00
24Arg	H	7.92067	0.00
24Arg	N	119.52998	0.00
24Arg	Ha	4.00649	0.00
24Arg	Hba	1.87897	0.00
25Phe	Hd*	6.33156	0.00
25Phe	He*	6.83983	0.00
25Phe	Hz	6.98272	0.01
25Phe	Hbb	2.86745	0.00
25Phe	Ha	3.65979	0.00
25Phe	Hba	2.86706	0.00
25Phe	H	8.95516	0.00
25Phe	N	124.66001	0.00
26Tyr	Hd*	7.27214	0.00
26Tyr	He*	6.67821	0.00
26Tyr	Ha	4.09862	0.01
26Tyr	Hbb	3.01717	0.00
26Tyr	H	8.6137	0.00
26Tyr	N	116.68103	0.00
26Tyr	Hba	2.96685	0.00
27Lys	Hda	1.3844	0.00
27Lys	Hdb	1.38466	0.00
27Lys	Hba	1.82997	0.00
27Lys	Hga	1.58139	0.00
27Lys	H	8.32549	0.01
27Lys	N	120.89993	0.00
27Lys	Ha	3.98162	0.00
27Lys	[468]	1.8932	0.00
27Lys	Hbb	1.9302	0.00
27Lys	Hgb	1.5823	0.00
28Thr	Ha	3.84285	0.00
28Thr	Hb	4.00652	0.00
28Thr	Hg2*	0.93992	0.00
28Thr	H	7.67382	0.00
28Thr	N	117.15007	0.00
29Leu	Hba	1.33648	0.00
29Leu	Hbb	1.53211	0.00
29Leu	Hda*	0.64164	0.00
29Leu	Hdb*	0.74093	0.00
29Leu	H	7.89871	0.00
29Leu	N	122.351	0.00
29Leu	Ha	3.80871	0.00
30Arg	Hda	3.16488	0.00
30Arg	Hgb	1.63266	0.00
30Arg	Hga	1.63255	0.00
30Arg	Hba	1.82678	0.00
30Arg	H	7.9535	0.00
30Arg	N	116.53206	0.00
30Arg	Ha	4.10558	0.00
30Arg	Hbb	1.87949	0.00
31Ala	Ha	4.20398	0.00
31Ala	Hb*	1.48352	0.00
31Ala	H	7.54828	0.00
31Ala	N	121.74893	0.00
32Glu	Hgb	2.27634	0.00
32Glu	Hbb	2.07638	0.00
32Glu	H	7.87046	0.00
32Glu	N	117.86147	0.00
32Glu	Ha	4.20385	0.00
32Glu	Hba	1.97861	0.00
32Glu	Hga	2.27586	0.00
33Gln	He2b	7.55344	0.00
33Gln	Ne2	112.51534	0.00
33Gln	He2a	6.87193	0.00
33Gln	Hbb	2.07548	0.00
33Gln	Hgb	2.32543	0.00
33Gln	H	8.13838	0.00
33Gln	N	118.79008	0.00
33Gln	Ha	4.14808	0.00
33Gln	Hba	2.0754	0.00

33Gln	Hga	2.32484	0.00
34Ala	N	122.81004	0.00
34Ala	Ha	4.2536	0.00
34Ala	Hb*	1.33579	0.00
34Ala	H	7.95937	0.00
35Ser	Hbb	4.05422	0.00
35Ser	H	8.10982	0.00
35Ser	N	114.34011	0.00
35Ser	Ha	4.40077	0.00
35Ser	Hba	3.90645	0.00
36Gln	Hba	2.07642	0.00
36Gln	Hgb	2.37412	0.00
36Gln	Ha	4.15152	0.00
36Gln	Hbb	2.07649	0.00
36Gln	Hga	2.37394	0.00
36Gln	H	8.49227	0.00
36Gln	N	121.34111	0.00
37Glu	Hba	1.9782	0.00
37Glu	Hgb	2.22745	0.00
37Glu	H	8.44906	0.00
37Glu	N	119.83992	0.00
37Glu	Ha	4.15471	0.00
37Glu	Hbb	1.97833	0.00
37Glu	Hga	2.22684	0.00
38Val	Hgb*	0.91614	0.00
38Val	H	7.76927	0.00
38Val	N	120.68107	0.00
38Val	Ha	3.95625	0.00
38Val	Hb	2.07691	0.00
38Val	Hga*	0.86085	0.00
39Lys	Hea	2.91776	0.00
39Lys	Heb	2.91776	0.00
39Lys	Hga	1.33293	0.00
39Lys	Hbb	1.78264	0.00
39Lys	Hda	1.5829	0.00
39Lys	Hdb	1.58298	0.00
39Lys	Hgb	1.44755	0.00
39Lys	H	8.27968	0.00
39Lys	N	123.09001	0.00
39Lys	Ha	4.19423	0.00
39Lys	Hba	1.78219	0.00
40Asn	Hd2a	6.89663	0.00
40Asn	Nd2	112.36023	0.00
40Asn	Hbb	2.87353	0.00
40Asn	H	8.42259	0.01
40Asn	N	119.46069	0.00
40Asn	Ha	4.59964	0.00
40Asn	Hba	2.83107	0.00
40Asn	Hd2b	7.62902	0.00
41Ala	H	8.16256	0.00
41Ala	N	123.54203	0.00
41Ala	Ha	4.25303	0.00
41Ala	Hb*	1.38694	0.00
42Ala	H	8.28072	0.01
42Ala	N	122.11955	0.00
42Ala	Ha	4.2983	0.00
42Ala	Hb*	1.43778	0.00
43Thr	Hb	4.35268	0.00
43Thr	Ha	4.00484	0.00
43Thr	Hg2*	1.28535	0.00
43Thr	H	8.13199	0.01
43Thr	N	114.59999	0.00
44Glu	Hbb	2.08003	0.00
44Glu	Hga	2.32521	0.00
44Glu	Hgb	2.32535	0.00
44Glu	H	8.47441	0.01
44Glu	N	120.11992	0.00
44Glu	Ha	4.00598	0.00
44Glu	Hba	2.07988	0.00
45Thr	Hb	4.20562	0.00
45Thr	Ha	4.06904	0.00
45Thr	Hg2*	1.2372	0.00
45Thr	H	7.7269	0.00
45Thr	N	112.59995	0.00
46Leu	Hbb	1.88227	0.00
46Leu	Hda*	0.88837	0.00
46Leu	H	7.80003	0.00
46Leu	N	123.26801	0.00
46Leu	Ha	4.10513	0.00
46Leu	Hba	1.78134	0.00
46Leu	Hdb*	0.88908	0.00
47Leu	Hda*	0.93599	0.00
47Leu	Hbb	1.5345	0.00
47Leu	H	7.56216	0.00
47Leu	N	119.51379	0.02
47Leu	Ha	4.05658	0.00
47Leu	Hba	1.53449	0.00
48Val	Hgb*	1.03713	0.00
48Val	Ha	3.5661	0.00
48Val	Hb	2.05056	0.00
48Val	Hga*	0.87826	0.01
48Val	H	7.4876	0.00
48Val	N	116.72995	0.00
49Gln	Hba	2.07947	0.00
49Gln	Hbb	2.17184	0.00
49Gln	Hga	2.41774	0.00
49Gln	Hgb	2.41801	0.00
49Gln	H	7.87168	0.00
49Gln	N	117.301	0.00
49Gln	Ha	3.85711	0.00

50Asn	Nd2	109.45444	0.02
50Asn	Hd2b	7.62006	0.00
50Asn	Hd2a	6.42846	0.00
50Asn	N	115.58982	0.00
50Asn	Ha	4.89171	0.01
50Asn	Hba	3.30966	0.00
50Asn	Hbb	3.60552	0.01
50Asn	H	7.77828	0.01
51Ala	Ha	4.20606	0.00
51Ala	Hb*	1.58273	0.00
51Ala	H	7.27941	0.00
51Ala	N	124.51095	0.00
52Asn	Hd2a	6.66975	0.00
52Asn	Nd2	111.15324	0.00
52Asn	Hd2b	7.41674	0.00
52Asn	Hba	3.03015	0.01
52Asn	Hbb	3.03926	0.01
52Asn	Ha	4.64598	0.00
52Asn	H	8.98613	0.00
52Asn	N	119.45756	0.00
53Pro	Hba	1.88406	0.00
53Pro	Hbb	2.37289	0.00
53Pro	Hdb	3.83153	0.00
53Pro	Hda	3.80834	0.00
53Pro	Ha	4.10636	0.00
54Asp	Hbb	2.57255	0.00
54Asp	H	7.92155	0.00
54Asp	N	116.45897	0.00
54Asp	Ha	4.35312	0.00
54Asp	Hba	2.5721	0.00
55Cys	H	8.28652	0.00
55Cys	N	117.7	0.00
55Cys	Ha	4.649	0.00
55Cys	Hba	2.77069	0.00
55Cys	Hbb	3.41358	0.00
56Lys	Hgb	1.4785	0.00
56Lys	Hga	1.47842	0.00
56Lys	Hbb	1.77967	0.00
56Lys	Hdb	1.68229	0.00
56Lys	Hda	1.68224	0.00
56Lys	Ha	3.65824	0.00
56Lys	Hba	1.77952	0.00
56Lys	H	9.24714	0.00
56Lys	N	121.12997	0.00
57Thr	Hb	4.20381	0.00
57Thr	H	7.74364	0.00
57Thr	N	113.4499	0.00
57Thr	Ha	3.75979	0.00
57Thr	Hg2*	1.23541	0.00
58Ile	Hg1a	1.03984	0.00
58Ile	Hg2*	0.8378	0.00
58Ile	Hg1b	1.68425	0.00
58Ile	Hd1*	0.74273	0.00
58Ile	Ha	3.66041	0.00
58Ile	Hb	2.12292	0.00
58Ile	H	7.04356	0.01
58Ile	N	123.78999	0.00
59Leu	Hbb	1.86941	0.00
59Leu	Hg	1.83103	0.00
59Leu	Hdb*	1.2861	0.00
59Leu	Hda*	0.74136	0.00
59Leu	H	8.59759	0.00
59Leu	N	118.70105	0.00
59Leu	Ha	3.85807	0.00
59Leu	Hba	1.86834	0.00
60Lys	Hba	1.82925	0.00
60Lys	Hga	1.43361	0.00
60Lys	Hgb	1.43383	0.00
60Lys	Hda	1.58253	0.00
60Lys	Hdb	1.58263	0.00
60Lys	H	8.3222	0.01
60Lys	N	119.46894	0.00
60Lys	Ha	3.95648	0.00
60Lys	Hbb	1.83022	0.00
61Ala	Ha	4.25111	0.00
61Ala	Hb*	1.48215	0.00
61Ala	H	7.1855	0.01
61Ala	N	120.08996	0.00
62Leu	Hda*	0.74092	0.00
62Leu	Hbb	1.83209	0.00
62Leu	Hdb*	0.74146	0.00
62Leu	Ha	4.15264	0.00
62Leu	Hg	1.58301	0.00
62Leu	Hba	1.78286	0.00
62Leu	H	7.46048	0.00
62Leu	N	118.62005	0.00
63Gly	Haa	3.98668	0.00
63Gly	Hab	4.15462	0.00
63Gly	H	7.50202	0.00
63Gly	N	104.75103	0.00
64Pro	Hdb	3.63795	0.00
64Pro	Hga	1.92782	0.00
64Pro	Hgb	1.92853	0.00
64Pro	Hda	3.60487	0.00
64Pro	Ha	4.41346	0.00
64Pro	Hba	1.97545	0.00
64Pro	Hbb	2.26267	0.00
65Gly	Hab	3.87908	0.00
65Gly	Haa	3.844	0.00
65Gly	H	8.71218	0.00

65Gly	N	109.55095	0.00
66Ala	Hb*	1.32899	0.01
66Ala	Ha	4.37986	0.00
66Ala	H	7.15405	0.00
66Ala	N	123.36993	0.00
67Thr	Hb	4.44816	0.00
67Thr	Hg2*	1.33533	0.00
67Thr	N	113.26994	0.00
67Thr	Ha	4.40394	0.00
67Thr	H	8.55939	0.00
68Leu	Hba	1.64777	0.00
68Leu	Hda*	0.84438	0.00
68Leu	Hdb*	0.9898	0.00
68Leu	Ha	4.20398	0.00
68Leu	Hbb	1.67177	0.00
68Leu	H	8.86155	0.00
68Leu	N	122.47979	0.00
69Glu	Hga	2.22756	0.00
69Glu	Hgb	2.32609	0.00
69Glu	Hbb	2.01463	0.00
69Glu	H	8.57434	0.00
69Glu	N	117.67992	0.00
69Glu	Ha	3.95731	0.00
69Glu	Hba	1.92539	0.00
70Glu	Hba	1.87861	0.00
70Glu	Hbb	2.02819	0.00
70Glu	Hga	2.25412	0.00
70Glu	Hgb	2.374	0.00
70Glu	H	7.71434	0.00
70Glu	N	120.33108	0.00
70Glu	Ha	3.94199	0.00
71Met	Hbb	2.12888	0.00
71Met	Hga	2.81995	0.00
71Met	Hgb	2.82009	0.00
71Met	H	8.40163	0.00
71Met	N	120.38001	0.00
71Met	Ha	3.61047	0.00
71Met	Hba	1.92855	0.00
72Met	Hga	2.51457	0.01
72Met	He*	2.21427	0.01
72Met	Hba	2.12745	0.00
72Met	H	8.61568	0.00
72Met	N	116.82002	0.00
72Met	Ha	4.15874	0.00
72Met	Hbb	2.32422	0.00
72Met	Hgb	2.51861	0.01
73Thr	H	8.02205	0.00
73Thr	N	115.18	0.00
73Thr	Ha	3.8582	0.00
73Thr	Hb	4.10207	0.00
73Thr	Hg2*	1.18576	0.00
74Ala	H	7.87803	0.00
74Ala	N	123.87212	0.00
74Ala	Ha	4.0562	0.00
74Ala	Hb*	1.33563	0.00
75Cys	H	7.57212	0.00
75Cys	N	111.22002	0.00
75Cys	Ha	4.59963	0.00
75Cys	Hba	2.52147	0.00
75Cys	Hbb	3.0166	0.00
76Gln	He2a	6.87356	0.00
76Gln	Ne2	111.14817	0.00
76Gln	He2b	7.46346	0.00
76Gln	Hgb	2.52317	0.00
76Gln	Hbb	2.32759	0.00
76Gln	H	7.6136	0.00
76Gln	N	120.95896	0.00
76Gln	Ha	4.30248	0.00
76Gln	Hba	2.22437	0.00
76Gln	Hga	2.42463	0.00
77Gly	H	8.33384	0.01
77Gly	N	108.49927	0.00
77Gly	Haa	3.90773	0.00
77Gly	Hab	4.02946	0.00
78Val	Hga*	0.89015	0.00
78Val	H	7.59149	0.00
78Val	N	119.38831	0.00
78Val	Ha	4.0056	0.00
78Val	Hb	2.07828	0.00
78Val	Hgb*	0.90024	0.01
79Gly	H	8.56072	0.00
79Gly	N	113.55106	0.00
79Gly	Haa	3.85644	0.00
79Gly	Hab	4.00666	0.00
80Gly	Haa	4.0531	0.00
80Gly	Hab	4.10499	0.00
80Gly	H	8.044	0.00
80Gly	N	108.54104	0.00
81Pro	Ha	4.39959	0.00
81Pro	Hba	1.92828	0.00
81Pro	Hbb	2.02551	0.00
81Pro	Hda	3.61162	0.00
81Pro	Hdb	3.65928	0.00
82Gly	H	8.51086	0.00
82Gly	N	108.99037	0.00
82Gly	Haa	3.85404	0.00
82Gly	Hab	3.89078	0.00
83His	Ha	4.45767	0.00
83His	Hba	3.08615	0.00
83His	Hbb	3.16	0.00

83His	H	8.20236	0.00
83His	N	119.36911	0.00
84Lys	Hga	1.38422	0.00
84Lys	H	8.38616	0.00
84Lys	N	123.12098	0.00
84Lys	Ha	4.29856	0.00
84Lys	Hba	1.78005	0.00
85Ala	Ha	4.20241	0.00
85Ala	Hb*	1.36286	0.00
85Ala	H	8.22457	0.00
85Ala	N	124.71906	0.00
86Arg	Hba	1.77767	0.00
86Arg	N	120.74996	0.00
86Arg	H	8.21248	0.00
86Arg	Ha	4.25637	0.00
86Arg	Hbb	1.77791	0.00
87Val	Hga*	0.89005	0.00
87Val	H	8.08927	0.00
87Val	N	121.49	0.00
87Val	Hgb*	0.89046	0.00
87Val	Hb	2.02747	0.00
87Val	Ha	4.00617	0.00
88Leu	Hg	1.5857	0.00
88Leu	Hda*	0.83792	0.00
88Leu	Hdb*	0.87999	0.00
88Leu	Hbb	1.58506	0.00
88Leu	H	8.21269	0.00
88Leu	N	125.60064	0.00
88Leu	Ha	4.25155	0.00
88Leu	Hba	1.58476	0.00
89Ala	H	8.15153	0.00
89Ala	N	124.14516	0.02
89Ala	Ha	4.20416	0.00
89Ala	Hb*	1.38551	0.00
90Glu	Hbb	2.02167	0.00
90Glu	Hgb	2.22906	0.00
90Glu	H	8.26525	0.00
90Glu	N	119.59993	0.00
90Glu	Ha	4.16611	0.00
90Glu	Hba	1.93117	0.00
90Glu	Hga	2.22891	0.00
91Ala	H	8.13919	0.00
91Ala	N	124.08393	0.00
91Ala	Ha	4.2474	0.01
91Ala	Hb*	1.38561	0.00
92Met	Hbb	2.07944	0.00
92Met	Hgb	2.61965	0.00
92Met	H	8.16355	0.00
92Met	N	118.50992	0.00
92Met	Ha	4.42523	0.00

92Met	Hba	2.01941	0.00
92Met	Hga	2.53059	0.00
93Ser	Hbb	3.85582	0.00
93Ser	H	8.11364	0.00
93Ser	N	116.25037	0.00
93Ser	Ha	4.40296	0.00
93Ser	Hba	3.85574	0.00
94Gln	Hba	1.95633	0.00
94Gln	Hbb	2.12728	0.00
94Gln	Hga	2.3243	0.00
94Gln	Hgb	2.32443	0.00
94Gln	H	8.25264	0.00
94Gln	N	121.74811	0.00
94Gln	Ha	4.35173	0.00
95Val	Hga*	0.90308	0.00
95Val	H	8.03356	0.01
95Val	N	120.4582	0.03
95Val	Hb	2.07767	0.00
95Val	Hgb*	0.9036	0.00
95Val	Ha	4.12693	0.00
96Thr	Hg2*	1.13716	0.00
96Thr	Hb	4.1065	0.00
96Thr	Ha	4.26405	0.00
96Thr	H	8.12111	0.00
96Thr	N	117.71895	0.00
97Asn	Hd2b	7.59532	0.00
97Asn	Hd2a	6.91915	0.00
97Asn	Nd2	112.89208	0.00
97Asn	Hba	2.70136	0.00
97Asn	Hbb	2.86	0.00
97Asn	H	8.35647	0.00
97Asn	N	122.44997	0.00
97Asn	Ha	4.94988	0.00
98Pro	Hda	3.73077	0.00
98Pro	Hdb	3.7983	0.00
98Pro	Hba	1.97603	0.00
98Pro	Hbb	2.2773	0.00
98Pro	Ha	4.30165	0.00
98Pro	Hga	1.93268	0.00
98Pro	Hgb	1.93287	0.00
99Ala	H	8.22227	0.00
99Ala	N	122.34001	0.00
99Ala	Ha	4.25935	0.00
99Ala	Hb*	1.38509	0.00
100Thr	Hb	4.2226	0.00
100Thr	Ha	4.20372	0.00
100Thr	Hg2*	1.13904	0.00
100Thr	H	7.88425	0.01
100Thr	N	112.57108	0.00
101Ile	Hd1*	0.84185	0.00

101Ile	Hg1b	1.43787	0.00	109Arg	N	120.95106	0.00
101Ile	Hg1a	1.18808	0.00	109Arg	Ha	4.64345	0.00
101Ile	Ha	4.10259	0.00	110Asn	Hba	2.70668	0.00
101Ile	Hb	1.83005	0.00	110Asn	Hbb	2.85513	0.00
101Ile	Hg2*	0.88948	0.00	110Asn	Ha	4.5994	0.00
101Ile	H	7.84167	0.00	110Asn	H	8.24561	0.00
101Ile	N	122.54901	0.00	110Asn	N	119.05201	0.00
102Met	Hgb	2.55312	0.00	112Arg	Ha	4.55077	0.00
102Met	Hbb	2.01187	0.00	113Lys	Hea	3.11529	0.00
102Met	H	8.20795	0.00	113Lys	Ha	4.20262	0.00
102Met	N	123.42004	0.00	113Lys	Hbb	1.77858	0.00
102Met	Hba	1.97891	0.00	113Lys	Hga	1.48318	0.00
102Met	Hga	2.47858	0.00	113Lys	Hgb	1.48335	0.00
102Met	Ha	4.43928	0.00	113Lys	H	8.1516	0.00
103Ile	Hd1*	0.81842	0.00	113Lys	N	121.77097	0.00
103Ile	Hg1b	1.4739	0.00	113Lys	Hba	1.68073	0.00
103Ile	H	7.9917	0.00	114Thr	H	8.14179	0.00
103Ile	N	122.20083	0.00	114Thr	N	117.45048	0.00
103Ile	Ha	4.05803	0.00	114Thr	Ha	4.32763	0.00
103Ile	Hb	1.83039	0.00	114Thr	Hb	4.0075	0.00
103Ile	Hg2*	0.85577	0.00	114Thr	Hg2*	1.13807	0.00
103Ile	Hg1a	1.13678	0.00	115Val	Hgb*	0.74138	0.00
104Gln	He2a	6.7774	0.00	115Val	H	8.34435	0.00
104Gln	Hgb	2.32482	0.00	115Val	N	123.20884	0.00
104Gln	Hbb	2.05434	0.00	115Val	Ha	4.10705	0.00
104Gln	Hga	2.32458	0.00	115Val	Hb	1.86151	0.00
104Gln	H	8.31803	0.00	115Val	Hga*	0.73491	0.01
104Gln	N	124.00005	0.00	116Lys	Hbb	1.43445	0.00
104Gln	Ha	4.25823	0.00	116Lys	Hgb	0.99203	0.00
104Gln	Hba	1.93204	0.00	116Lys	H	8.09082	0.01
105Lys	Hbb	1.78183	0.00	116Lys	N	124.69103	0.00
105Lys	Hba	1.7815	0.00	116Lys	Ha	4.45023	0.00
105Lys	N	122.33999	0.00	116Lys	Hba	1.43443	0.00
105Lys	Ha	4.25223	0.00	116Lys	Hga	0.99203	0.00
105Lys	H	8.28966	0.00	117Cys	Ha	4.10439	0.00
106Gly	Hab	3.85831	0.00	117Cys	Hbb	2.76829	0.00
106Gly	H	8.28777	0.00	117Cys	Hba	1.7307	0.00
106Gly	N	109.32008	0.00	117Cys	H	8.2095	0.00
106Gly	Haa	3.85791	0.00	117Cys	N	127.24215	0.00
107Asn	H	8.18491	0.00	118Phe	Hd*	7.22942	0.00
107Asn	N	118.68059	0.00	118Phe	Hbb	3.16663	0.01
107Asn	Ha	4.64246	0.00	118Phe	Ha	4.55156	0.00
107Asn	Hba	2.62357	0.00	118Phe	Hba	3.06692	0.00
107Asn	Hbb	2.72014	0.00	118Phe	H	8.53986	0.00
108Phe	Hbb	3.11768	0.00	118Phe	N	128.46102	0.00
108Phe	Ha	4.55821	0.00	119Asn	Hd2a	6.72952	0.00
108Phe	H	8.12605	0.00	119Asn	Nd2	117.48163	0.01
108Phe	N	120.73204	0.00	119Asn	Hd2b	8.34179	0.00
108Phe	Hba	3.01326	0.00	119Asn	Ha	4.69924	0.00
109Arg	H	8.10815	0.00	119Asn	Hbb	3.08439	0.01

119Asn	Hba	2.47301	0.00
119Asn	H	9.522	0.01
119Asn	N	121.85004	0.00
120Cys	Hbb	3.21436	0.00
120Cys	Hba	2.50628	0.00
120Cys	H	8.85546	0.00
120Cys	N	117.38004	0.00
120Cys	Ha	4.89707	0.00
121Gly	H	7.85327	0.01
121Gly	N	113.44993	0.00
121Gly	Haa	3.70787	0.00
121Gly	Hab	4.00722	0.00
122Lys	Hbb	1.87084	0.00
122Lys	Hgb	1.4813	0.00
122Lys	Hda	1.58462	0.00
122Lys	Hdb	1.58462	0.00
122Lys	H	8.08618	0.01
122Lys	N	120.99384	0.00
122Lys	Ha	4.40189	0.00
122Lys	Hba	1.77544	0.00
122Lys	Hga	1.48067	0.00
123Glu	Hbb	1.97586	0.00
123Glu	Hgb	2.2757	0.00
123Glu	H	8.35358	0.00
123Glu	N	118.22901	0.00
123Glu	Ha	4.60044	0.00
123Glu	Hba	1.88214	0.00
123Glu	Hga	2.12652	0.00
124Gly	Hab	4.45186	0.00
124Gly	Haa	3.65908	0.00
124Gly	H	8.60325	0.00
124Gly	N	107.55105	0.00
125His	Hd2	6.81885	0.01
125His	He1	6.83714	0.00
125His	[978]	7.37216	0.00
125His	Hbb	3.14284	0.00
125His	H	7.14947	0.00
125His	N	114.13089	0.00
125His	Ha	4.69814	0.00
125His	Hba	3.14206	0.00
126Ile	Ha	4.70076	0.00
126Ile	Hb	2.12875	0.00
126Ile	Hg1a	1.23469	0.00
126Ile	Hg2*	0.89225	0.00
126Ile	H	7.7973	0.00
126Ile	N	108.33086	0.00
127Ala	Hb*	1.2919	0.01
127Ala	Ha	4.10628	0.00
127Ala	H	8.71019	0.00
127Ala	N	124.749	0.00
128Lys	Hga	1.38258	0.00
128Lys	Ha	4.05552	0.00
128Lys	H	8.1898	0.00
128Lys	N	117.11904	0.00
128Lys	Hgb	1.38487	0.00
128Lys	Hba	1.58088	0.00
128Lys	Hbb	1.87951	0.00
129Asn	Hd2b	7.88824	0.00
129Asn	Nd2	113.03951	0.00
129Asn	Hd2a	7.01597	0.00
129Asn	H	7.87158	0.00
129Asn	N	114.23003	0.00
129Asn	Ha	4.99399	0.00
129Asn	Hbb	3.06763	0.01
129Asn	Hba	2.52643	0.01
130Cys	H	7.63681	0.01
130Cys	N	124.26911	0.00
130Cys	Ha	3.71114	0.00
130Cys	Hbb	3.35649	0.01
130Cys	Hba	2.96805	0.00
131Arg	N	128.76943	0.00
131Arg	Hda	3.16174	0.00
131Arg	Hdb	3.16346	0.00
131Arg	Hga	1.48207	0.00
131Arg	Hgb	1.48239	0.00
131Arg	Hbb	1.58284	0.00
131Arg	Hba	1.5825	0.00
131Arg	H	8.62614	0.00
131Arg	Ha	4.46165	0.00
132Ala	Ha	4.52288	0.00
132Ala	Hb*	1.33425	0.00
132Ala	H	8.70335	0.00
132Ala	N	128.12902	0.00
133Pro	Hda	3.50998	0.00
133Pro	Hdb	3.50998	0.00
133Pro	Hga	1.98027	0.00
133Pro	Hgb	1.98027	0.00
133Pro	Ha	4.30023	0.00
134Arg	Hga	1.38627	0.00
134Arg	Hgb	1.43395	0.00
134Arg	Hba	1.61542	0.00
134Arg	H	8.35168	0.00
134Arg	N	123.17075	0.00
134Arg	Hbb	1.73377	0.00
134Arg	Ha	4.30243	0.00
135Lys	Hgb	1.38591	0.00
135Lys	Hbb	1.77905	0.00
135Lys	Hdb	1.58506	0.00
135Lys	Hda	1.58506	0.00
135Lys	Ha	4.24852	0.00

135Lys	Hba	1.77822	0.00
135Lys	Hga	1.38544	0.00
135Lys	H	8.56823	0.00
135Lys	N	125.95587	0.00
136Lys	Hda	1.63054	0.00
136Lys	Hdb	1.63064	0.00
136Lys	Hga	1.38469	0.00
136Lys	Hbb	1.7345	0.00
136Lys	Hgb	1.38509	0.00
136Lys	H	8.3952	0.00
136Lys	N	122.45931	0.00
136Lys	Ha	4.30479	0.00
136Lys	Hba	1.7342	0.00
137Gly	Haa	3.67285	0.03
137Gly	Hab	3.70814	0.00
137Gly	H	8.10889	0.00
137Gly	N	109.13088	0.00
138Cys	Hbb	2.87009	0.00
138Cys	H	8.13027	0.01
138Cys	N	124.15892	0.00
138Cys	Ha	4.05388	0.00
138Cys	Hba	2.86953	0.00
139Trp	Hh2	7.17584	0.00
139Trp	Hbb	3.41285	0.00
139Trp	Ha	4.5262	0.00
139Trp	Hba	3.41259	0.00
139Trp	Hd1	7.27596	0.01
139Trp	H2	7.46758	0.00
139Trp	H	8.37873	0.01
139Trp	He1	10.0678	0.00
139Trp	Ne1	129.01795	0.00
139Trp	N	129.55003	0.00
140Lys	Hbb	2.12485	0.00
140Lys	Hda	1.71957	0.00
140Lys	Hgb	1.27441	0.00
140Lys	Ha	4.25469	0.00
140Lys	Hba	2.07454	0.00
140Lys	Hdb	1.71968	0.00
140Lys	Hga	1.27399	0.00
140Lys	H	9.34109	0.00
140Lys	N	122.91998	0.00
141Cys	H	8.47621	0.01
141Cys	N	117.20898	0.00
141Cys	Ha	4.89523	0.00
141Cys	Hba	2.5002	0.00
141Cys	Hbb	3.21221	0.00
142Gly	H	8.11178	0.00
142Gly	N	113.83096	0.00
142Gly	Haa	3.80886	0.00
142Gly	Hab	4.05766	0.00
143Lys	Hbb	1.77949	0.00
143Lys	Hga	1.36291	0.00
143Lys	Hgb	1.36314	0.00
143Lys	Hda	1.63583	0.00
143Lys	Hdb	1.63584	0.00
143Lys	H	8.39443	0.00
143Lys	N	122.36002	0.00
143Lys	Ha	4.29758	0.00
143Lys	Hba	1.77843	0.00
144Glu	Hga	2.32579	0.00
144Glu	Hgb	2.32602	0.00
144Glu	Hba	2.07348	0.00
144Glu	Hbb	2.0736	0.00
144Glu	H	8.21409	0.00
144Glu	N	120.7491	0.00
144Glu	Ha	4.24213	0.00
145Gly	Haa	3.70903	0.00
145Gly	Hab	4.25503	0.00
145Gly	H	8.52133	0.00
145Gly	N	106.9809	0.00
146His	Hd2	6.82455	0.00
146His	Hbb	3.16926	0.00
146His	H	7.12115	0.00
146His	N	112.97836	0.00
146His	Hba	3.1677	0.00
146His	Ha	4.79367	0.01
147Gln	Hba	1.92918	0.00
147Gln	Hga	2.27931	0.00
147Gln	Ha	4.55029	0.00
147Gln	H	8.87293	0.01
147Gln	N	117.02058	0.00
147Gln	Hbb	1.92966	0.00
147Gln	Hgb	2.35117	0.00
148Met	Hga	2.52264	0.00
148Met	Hbb	2.27326	0.00
148Met	Ha	4.79767	0.00
148Met	Hba	2.12733	0.00
148Met	Hgb	2.5231	0.00
148Met	H	8.83321	0.00
148Met	N	121.93092	0.00
149Lys	Hbb	1.87884	0.00
149Lys	Hda	1.73101	0.00
149Lys	Hdb	1.73117	0.00
149Lys	Hgb	1.33518	0.00
149Lys	Hga	1.23728	0.00
149Lys	H	8.47534	0.00
149Lys	N	118.45089	0.00
149Lys	Ha	4.10556	0.00
149Lys	Hba	1.87839	0.00
150Asp	H	7.81511	0.00

150Asp	N	117.64313	0.00
150Asp	Ha	4.83566	0.00
150Asp	Hba	2.52495	0.00
150Asp	Hbb	2.9177	0.00
151Cys	H	7.54666	0.00
151Cys	N	123.68003	0.00
151Cys	Ha	3.65824	0.00
151Cys	Hba	2.86166	0.00
151Cys	Hbb	3.26277	0.00
152Thr	H	8.1154	0.00
152Thr	N	119.71806	0.00
152Thr	Ha	4.35568	0.00
152Thr	Hg2*	1.09021	0.00
153Glu	Hba	1.92858	0.00
153Glu	Hbb	2.02953	0.00
153Glu	Hga	2.2266	0.00
153Glu	Hgb	2.42181	0.00
153Glu	Ha	4.19921	0.00
153Glu	H	8.6008	0.00
153Glu	N	125.56996	0.00
154Arg	H	8.35578	0.00
154Arg	N	122.86948	0.00
154Arg	Ha	4.30282	0.00
155Gln	N	121.83699	0.00
155Gln	Ha	4.27249	0.00
155Gln	Hba	1.95802	0.00
155Gln	H	8.40534	0.00
156Ala	H	8.33125	0.00
156Ala	N	126.25987	0.00
156Ala	Ha	4.30329	0.00
156Ala	Hb*	1.33384	0.00
157Asn	Ha	4.40163	0.00
157Asn	H	7.94044	0.00
157Asn	N	123.55908	0.00

II Titration of ¹⁵N-CA_{CTD}^{W184A, M185A}-SP1-NC by EP39

Residue	Reference			6 eq EP39			Δδ ¹ H ¹⁵ N	I/I ₀
	δ ¹ H (ppm)	δ ¹⁵ N (ppm)	Intensity	δ ¹ H (ppm)	δ ¹⁵ N (ppm)	Intensity		
144Met								
145Tyr								
146Ser								
147Pro								
148Thr				8.123	114.338	4.58E+05		
149Ser	8.402	118.634	1.29E+05	8.412	118.781	2.31E+05	0.027	1.784
150Ile				8.348	123.007	2.54E+05		
151Leu				7.545	119.525	1.68E+05		
152Asp	7.650	116.297	1.52E+05	7.649	116.328	1.23E+05	0.005	0.809
153Ile	7.443	122.277	1.99E+05	7.436	122.298	1.62E+05	0.008	0.815
154Arg	8.221	124.300	1.93E+05	8.245	124.417	1.30E+05	0.031	0.671
155Gln	7.617	127.131	6.12E+04	7.615	127.160	5.87E+04	0.005	0.959
156Gly	9.833	117.120	4.83E+04	9.839	117.150	3.66E+04	0.008	0.758
157Pro								
158Lys	8.501	115.333	2.61E+04	8.502	115.361	1.58E+05	0.005	6.055
159Glu	7.062	125.784	1.09E+05	7.066	125.794	9.72E+04	0.004	0.889
160Pro								
161Phe	9.189	128.109	7.74E+04	9.201	128.190	8.51E+04	0.018	1.099
162Arg	9.122	115.712	1.23E+05	9.134	115.777	1.05E+05	0.016	0.856
163Asp	6.974	119.120	1.30E+05	6.978	119.213	1.19E+05	0.016	0.917
164Tyr	7.549	124.749	8.33E+04	7.547	124.799	6.67E+04	0.009	0.801
165Val	8.238	120.426	1.74E+05	8.244	120.472	1.42E+05	0.009	0.815
166Asp	7.466	117.724	1.96E+05	7.471	117.809	1.60E+05	0.015	0.815
167Arg	7.906	119.477	2.01E+05	7.918	119.529	1.52E+05	0.015	0.753
168Phe	8.947	124.580	1.27E+05	8.955	124.661	1.01E+05	0.016	0.798
169Tyr	8.601	116.768	1.61E+05	8.614	116.684	1.69E+05	0.019	1.051
170Lys	8.299	120.796	2.02E+05	8.325	120.896	1.88E+05	0.031	0.930
171Thr	7.664	116.977	1.98E+05	7.673	117.153	1.64E+05	0.031	0.828
172Leu	7.892	122.307	1.86E+05	7.896	122.356	1.49E+05	0.009	0.802
173Arg	7.923	116.389	3.91E+05	7.951	116.542	1.96E+05	0.038	0.502
174Ala	7.540	121.795	2.07E+05	7.546	121.746	1.68E+05	0.010	0.811
175Glu	7.893	118.055	1.80E+05	7.867	117.867	1.39E+05	0.041	0.776
176Gln	8.138	118.775	1.86E+05	8.133	118.793	2.49E+05	0.006	1.338
177Ala	7.957	122.843	2.18E+05	7.959	122.811	2.51E+05	0.006	1.153
178Ser	8.114	114.299	2.00E+05	8.108	114.343	4.24E+05	0.010	2.115
179Gln	8.480	121.339	4.21E+04	8.493	121.346	1.56E+05	0.012	3.703
180Glu	8.438	119.820	2.21E+05	8.443	119.837	2.18E+05	0.005	0.988
181Val	7.771	120.701	2.71E+05	7.766	120.685	2.25E+05	0.006	0.830
182Lys	8.275	123.102	2.62E+05	8.277	123.090	2.47E+05	0.003	0.942
183Asn	8.425	119.398	7.01E+04	8.421	119.465	2.18E+05	0.012	3.108

184Ala	8.153	123.539	6.60E+04	8.161	123.543	1.80E+05	0.009	2.721
185Ala	8.268	122.152	2.45E+05	8.279	122.113	2.54E+05	0.013	1.033
186Thr	8.146	114.359	3.04E+05	8.133	114.599	2.34E+05	0.042	0.771
187Glu	8.498	120.142	3.35E+05	8.476	120.117	1.23E+05	0.023	0.368
188Thr	7.718	112.585	2.04E+05	7.727	112.597	1.73E+05	0.009	0.848
189Leu	7.803	123.265	1.73E+05	7.800	123.266	1.43E+05	0.004	0.824
190Leu	7.556	119.396	4.74E+05	7.563	119.404	1.88E+05	0.007	0.398
191Val	7.463	116.706	1.37E+05	7.487	116.727	1.23E+05	0.024	0.901
192Gln	7.855	117.142	1.44E+05	7.877	117.304	1.54E+05	0.035	1.073
193Asn	7.768	115.489	1.22E+05	7.779	115.585	1.02E+05	0.019	0.835
194Ala	7.275	124.472	1.05E+05	7.278	124.514	1.02E+05	0.008	0.970
195Asn	8.982	119.419	6.89E+04	8.984	119.452	5.36E+04	0.006	0.778
196Pro								
197Asp	7.916	116.305	3.28E+05	7.921	116.455	1.93E+05	0.026	0.588
198Cys	8.286	117.648	2.22E+05	8.285	117.700	1.71E+05	0.009	0.771
199Lys	9.259	121.173	6.48E+04	9.245	121.128	6.43E+04	0.016	0.993
200Thr	7.730	113.330	2.82E+05	7.743	113.446	2.28E+05	0.024	0.808
201Ile	7.037	123.753	1.24E+05	7.044	123.790	1.25E+05	0.009	1.011
202Leu	8.600	118.680	1.93E+05	8.594	118.704	1.55E+05	0.007	0.804
203Lys	8.337	119.480	1.48E+05	8.321	119.465	1.47E+05	0.016	0.994
204Ala	7.182	120.066	2.72E+05	7.184	120.088	2.39E+05	0.004	0.881
205Leu	7.454	118.605	2.78E+05	7.457	118.622	2.23E+05	0.004	0.803
206Gly	7.496	104.704	3.08E+05	7.500	104.755	2.38E+05	0.010	0.774
207Pro								
208Gly	8.710	109.510	1.01E+05	8.711	109.555	2.26E+05	0.008	2.233
209Ala	7.151	123.368	3.85E+05	7.154	123.368	2.63E+05	0.003	0.684
210Thr	8.598	113.367	2.39E+05	8.544	113.268	7.72E+04	0.056	0.323
211Leu	8.852	122.428	2.36E+05	8.861	122.475	1.91E+05	0.012	0.810
212Glu	8.564	117.609	2.85E+05	8.573	117.676	2.19E+05	0.014	0.769
213Glu	7.708	120.301	2.42E+05	7.711	120.339	1.84E+05	0.007	0.760
214Met	8.404	120.343	1.98E+05	8.399	120.381	1.72E+05	0.008	0.869
215Met	8.621	116.614	2.52E+05	8.615	116.821	8.49E+04	0.035	0.337
216Thr	8.022	115.153	2.65E+05	8.023	115.180	2.09E+05	0.005	0.788
217Ala	7.874	123.816	2.90E+05	7.877	123.878	2.35E+05	0.011	0.809
218Cys	7.569	111.071	2.11E+05	7.573	111.221	1.62E+05	0.025	0.768
219Gln	7.609	121.098	2.89E+05	7.611	120.956	2.26E+05	0.024	0.781
220Gly	8.358	108.553	3.60E+04	8.331	108.499	2.14E+05	0.028	5.931
221Val	7.558	119.354	3.99E+05	7.595	119.386	3.02E+05	0.037	0.756
222Gly	8.532	113.691	5.15E+04	8.561	113.555	2.61E+05	0.037	5.064
223Gly	8.038	108.466	2.01E+05	8.043	108.544	4.23E+05	0.014	2.101
224Pro								
225Gly				8.510	108.993	2.26E+05		
226His				8.201	119.368	9.50E+04		
227Lys				8.387	123.125	2.76E+05		
228Ala				8.227	124.711	3.34E+05		
229Arg				8.212	120.749	6.51E+05		
230Val	8.065	121.272	2.96E+05	8.089	121.490	3.28E+05	0.043	1.107
231Leu	8.179	125.291	2.71E+05	8.210	125.611	3.03E+05	0.062	1.116
232Ala	8.124	123.993	5.01E+05	8.150	124.152	6.34E+05	0.038	1.265
233Glu	8.251	119.512	3.23E+05	8.265	119.597	3.71E+05	0.020	1.148
234Ala	8.154	124.225	2.41E+05	8.138	124.078	9.77E+04	0.029	0.406
235Met	8.150	118.346	1.86E+05	8.165	118.507	4.13E+05	0.031	2.215
236Ser	8.103	116.167	5.26E+04	8.114	116.249	3.50E+05	0.018	6.649

237Gln	8.240	121.679	5.83E+04	8.252	121.750	3.21E+05	0.017	5.505
238Val	8.023	120.494	2.65E+05	8.035	120.565	3.28E+05	0.016	1.237
239Thr	8.117	117.647	7.34E+04	8.120	117.717	3.70E+05	0.012	5.044
240Asn	8.386	122.298	2.94E+05	8.361	122.449	3.20E+05	0.035	1.090
241Pro								
242Ala	8.220	122.212	2.97E+05	8.220	122.343	1.95E+05	0.022	0.655
243Thr	7.884	112.521	2.40E+05	7.884	112.574	4.34E+05	0.009	1.807
244Ile	7.840	122.546	3.72E+05	7.839	122.549	3.98E+05	0.001	1.071
245Met	8.206	123.402	1.87E+05	8.205	123.422	3.41E+05	0.004	1.821
246Ile	7.992	122.200	3.30E+05	7.991	122.203	3.81E+05	0.001	1.156
247Gln	8.318	123.954	9.69E+04	8.318	124.002	3.56E+05	0.008	3.678
248Lys				8.295	122.340	6.35E+05		
249Gly				8.287	109.321	3.73E+05		
250Asn				8.185	118.684	2.77E+05		
251Phe				8.129	120.735	3.98E+05		
252Arg				8.109	120.953	4.57E+05		
253Asn				8.247	119.054	2.39E+05		
254Gln								
255Arg								
256Lys				8.150	121.773	3.31E+05		
257Thr	8.121	117.648	1.18E+05	8.143	117.456	2.89E+05	0.039	2.451
258Val	8.352	123.124	3.18E+05	8.342	123.205	4.07E+05	0.017	1.280
259Lys	8.082	124.657	3.48E+05	8.089	124.695	2.32E+05	0.009	0.666
260Cys	8.204	127.216	2.89E+05	8.207	127.252	1.76E+05	0.007	0.609
261Phe	8.532	128.424	2.32E+05	8.536	128.466	1.44E+05	0.008	0.621
262Asn	9.508	121.821	2.40E+05	9.517	121.851	1.33E+05	0.010	0.555
263Cys	8.848	117.333	1.83E+05	8.845	117.382	1.03E+05	0.009	0.564
264Gly	7.844	113.387	2.45E+05	7.850	113.448	1.50E+05	0.012	0.613
265Lys	8.075	120.942	3.16E+05	8.089	121.004	3.89E+05	0.018	1.230
266Glu	8.339	118.156	3.53E+05	8.348	118.226	2.32E+05	0.015	0.658
267Gly	8.591	107.448	2.31E+05	8.599	107.554	2.13E+05	0.019	0.923
268His	7.143	114.095	4.17E+05	7.146	114.134	2.69E+05	0.007	0.646
269Ile	7.792	108.263	3.01E+05	7.796	108.335	2.07E+05	0.013	0.688
270Ala	8.707	124.692	2.68E+05	8.706	124.748	1.53E+05	0.009	0.569
271Lys	8.188	117.082	1.61E+05	8.187	117.114	2.76E+05	0.005	1.710
271Asn	7.864	114.161	2.83E+05	7.865	114.231	1.63E+05	0.012	0.575
273Cys	7.634	124.250	3.51E+05	7.636	124.266	2.19E+05	0.004	0.624
274Arg	8.622	128.777	3.20E+04	8.622	128.774	1.33E+05	0.001	4.147
275Ala	8.695	128.133	4.01E+05	8.701	128.127	2.54E+05	0.006	0.634
276Pro								
277Arg	8.365	123.156	2.89E+05	8.359	123.175	3.55E+05	0.010	1.225
278Lys	8.573	125.996	1.25E+05	8.567	125.948	2.01E+05	0.010	1.603
279Lys	8.385	122.300	2.65E+05	8.394	122.444	1.28E+05	0.026	0.482
280Gly	8.106	109.067	1.99E+05	8.108	109.134	2.03E+05	0.011	1.022
281Cys	8.132	124.231	5.02E+05	8.133	124.155	4.29E+05	0.013	0.854
282Trp	8.373	129.493	2.66E+05	8.379	129.552	1.43E+05	0.012	0.538
283Lys	9.327	122.939	2.41E+05	9.335	122.919	9.67E+04	0.009	0.400
284Cys	8.465	117.159	1.83E+05	8.472	117.205	9.66E+04	0.010	0.527
285Gly	8.102	113.790	2.52E+05	8.107	113.832	1.22E+05	0.009	0.485
286Lys	8.380	122.303	2.63E+05	8.393	122.361	1.31E+05	0.015	0.497
287Glu				8.212	120.748	6.51E+05		
288Gly				8.517	106.983	1.13E+05		
289His	7.116	112.922	3.25E+05	7.120	112.972	1.51E+05	0.009	0.463

290Gln	8.863	116.963	1.81E+05	8.873	117.023	1.38E+05	0.014	0.766
291Met	8.824	121.863	1.32E+05	8.829	121.935	1.63E+05	0.013	1.236
292Lys	8.469	118.422	1.68E+05	8.475	118.455	1.80E+05	0.008	1.074
293Asp	7.805	117.626	2.57E+05	7.810	117.653	1.58E+05	0.007	0.616
294Cys	7.545	123.673	3.32E+05	7.546	123.681	1.66E+05	0.002	0.500
295Thr	8.122	119.706	1.58E+05	8.116	119.713	1.72E+05	0.007	1.090
296Glu	8.597	125.567	3.20E+05	8.598	125.569	1.90E+05	0.001	0.595
297Arg	8.353	122.901	2.77E+05	8.360	122.866	1.54E+05	0.010	0.557
298Gln				8.405	121.837	2.27E+05		
299Ala	8.333	126.265	8.14E+04	8.330	126.258	2.68E+05	0.003	3.293
300Asn	7.946	123.554	5.60E+05	7.942	123.556	2.76E+05	0.005	0.493

III ¹⁵N relaxation data for CA_{CTD}^{W184A, M185A}-SP1-NC

Residue	Reference		6 eq EP39		Reference		6 eq EP39		Reference		6 eq EP39	
	T1 (s)	Erro	T1 (s)	Erro	T2 (s)	Erro	T2 (s)	Erro	NOE	Erro	NOE	Erro
144Met												
145Tyr												
146Ser												
147Pro												
148Thr			0.6121	0.0055			0.1632	0.0006			0.4012	0.0150
149Ser	0.5485	0.0086	0.5983	0.0078	0.1738	0.0307	0.1483	0.0011	0.4547	0.0192	0.4627	0.0274
150Ile			0.5807	0.0055			0.2599	0.0012			0.1259	0.0139
151Leu			0.5892	0.0069			0.0748	0.0013			0.6891	0.0252
152Asp	0.5715	0.0110	0.6424	0.0146	0.0606	0.0029	0.0624	0.0019	0.6606	0.0302	0.6135	0.0391
153Ile	0.6375	0.0087	0.6226	0.0089	0.1083	0.0022	0.1073	0.0016	0.6819	0.0237	0.6844	0.0368
154Arg	0.6144	0.0078	0.6065	0.0074	0.1179	0.0020	0.1881	0.0233	0.6250	0.0199	0.4210	0.0237
155Gln	0.5846	0.0193	0.6217	0.0193	0.0635	0.0050	0.0831	0.0035	0.6548	0.0405	0.6306	0.0547
156Gly	0.5287	0.0269	0.5693	0.0364	0.0673	0.0081	0.0738	0.0065	0.7060	0.0621	0.7477	0.0834
157Pro												
158Lys			0.6285	0.0170			0.0672	0.0023			0.7292	0.0458
159Glu	0.5908	0.0132	0.5819	0.0134	0.1169	0.0067	0.0966	0.0024	0.7476	0.0357	0.9291	0.0534
160Pro												
161Phe	0.6930	0.0779	0.6214	0.0223	0.0388	0.0015	0.0493	0.0016	0.7836	0.0788	0.7860	0.0588
162Arg	0.6396	0.0135	0.6364	0.0172	0.0654	0.0025	0.0712	0.0021	0.6503	0.0314	0.7925	0.0484
163Asp	0.6578	0.0125	0.6434	0.0136	0.0977	0.0030	0.0921	0.0020	0.7649	0.0263	0.7647	0.0409
164Tyr	0.6238	0.0186	0.6099	0.0194	0.0707	0.0046	0.0726	0.0033	0.7540	0.0352	0.6942	0.0454
165Val	0.6555	0.0081	0.6455	0.0093	0.0947	0.0033	0.1685	0.0018	0.6164	0.0197	0.3513	0.0224
166Asp	0.6390	0.0077	0.6255	0.0098	0.0756	0.0023	0.0702	0.0016	0.8331	0.0257	0.7658	0.0387
167Arg	0.6473	0.0063	0.6336	0.0081	0.0824	0.0018	0.0838	0.0014	0.7259	0.0228	0.7958	0.0346
168Phe	0.6433	0.0116	0.6091	0.0143	0.0752	0.0027	0.0809	0.0026	0.7385	0.0298	0.8037	0.0480
169Tyr	0.6482	0.0078	0.6076	0.0057	0.0899	0.0015	0.0903	0.0010	0.7006	0.0153	0.7777	0.0224
170Lys	0.6260	0.0075	0.6277	0.0075	0.0815	0.0019	0.1226	0.0013	0.7334	0.0230	0.6088	0.0242
171Thr	0.6573	0.0099	0.6154	0.0111	0.0828	0.0021	0.0787	0.0016	0.7645	0.0269	0.7641	0.0391
172Leu	0.6573	0.0082	0.6154	0.0102	0.0931	0.0021	0.0993	0.0017	0.5389	0.0207	0.7306	0.0419
173Arg	0.6200	0.0043	0.6334	0.0077	0.0820	0.0010	0.0754	0.0010	0.6759	0.0124	0.7022	0.0246
174Ala	0.5920	0.0086	0.6462	0.0107	0.0936	0.0022	0.0882	0.0015	0.6198	0.0253	0.6344	0.0369
175Glu	0.6247	0.0109	0.6148	0.0093	0.0920	0.0023	0.0966	0.0015	0.5539	0.0296	0.5684	0.0341
176Gln	0.6076	0.0117	0.5906	0.0063	0.1164	0.0027	0.1355	0.0011	0.4772	0.0240	0.4521	0.0199
177Ala	0.6611	0.0133	0.6038	0.0055	0.1465	0.0030	0.1445	0.0010	0.5361	0.0258	0.4749	0.0251
178Ser	0.6970	0.0856	0.6050	0.0035	0.1296	0.0038	0.1604	0.0006	0.4779	0.0161	0.4347	0.0111
179Gln			0.5891	0.0106			0.1638	0.0016			0.3639	0.0271
180Glu	0.6428	0.0161	0.6028	0.0073	0.1167	0.0034	0.1164	0.0011	0.5370	0.0276	0.4651	0.0204
181Val	0.6129	0.0081	0.6043	0.0070	0.1338	0.0020	0.1268	0.0012	0.4894	0.0201	0.5421	0.0282
182Lys	0.6073	0.0133	0.6063	0.0059	0.1303	0.0030	0.1509	0.0011	0.3546	0.0240	0.3763	0.0237
183Asn	0.6236	0.0569	0.5793	0.0077	0.1230	0.0180	0.1308	0.0013	0.5848	0.0504	0.4892	0.0252
184Ala			0.5961	0.0109			0.1500	0.0018			0.4293	0.0283
185Ala	0.6367	0.0165	0.6027	0.0036	0.1686	0.0039	0.2418	0.0008	0.3324	0.0219	0.1931	0.0102
186Thr	0.6643	0.0139	0.6139	0.0061	0.1539	0.0030	0.1447	0.0006	0.4199	0.0156	0.4605	0.0177
187Glu	0.5802	0.0122	0.5950	0.0114	0.0836	0.0027	0.0977	0.0017	0.6281	0.0302	0.5736	0.0280
188Thr	0.6276	0.0121	0.6145	0.0098	0.0732	0.0029	0.0764	0.0015	0.6034	0.0279	0.6951	0.0357
189Leu	0.5910	0.0081	0.6042	0.0104	0.1055	0.0022	0.1025	0.0016	0.7239	0.0263	0.8374	0.0466
190Leu	0.6215	0.0046	0.6011	0.0077	0.0947	0.0011	0.0924	0.0012	0.5861	0.0099	0.6094	0.0213
191Val	0.6154	0.0108	0.5902	0.0123	0.0753	0.0030	0.0766	0.0023	0.8074	0.0376	0.7564	0.0494
192Gln	0.5960	0.0097	0.5731	0.0181	0.0530	0.0025	0.1685	0.0017	0.6777	0.0293	0.4744	0.0265
193Asn	0.5868	0.0114	0.5795	0.0122	0.0738	0.0027	0.0792	0.0020	0.6668	0.0295	0.8202	0.0446
194Ala	0.5645	0.0105	0.5480	0.0105	0.0978	0.0032	0.0958	0.0019	0.6326	0.0286	0.1817	0.0243
195Asn	0.5867	0.0176	0.5666	0.0179	0.1033	0.0045	0.1017	0.0034	0.7660	0.0395	0.7025	0.0522
196Pro												
197Asp	0.6191	0.0052	0.6525	0.0073	0.0821	0.0012	0.0868	0.0013	0.7034	0.0134	0.7334	0.0243
198Cys	0.6128	0.0076	0.6169	0.0087	0.0951	0.0020	0.0967	0.0014	0.6904	0.0224	0.7169	0.0314
199Lys	0.6201	0.0190	0.5608	0.0179	0.0849	0.0050	0.0867	0.0032	0.8108	0.0518	0.8145	0.0629
200Thr	0.5798	0.0055	0.5718	0.0075	0.0971	0.0016	0.0974	0.0011	0.6806	0.0197	0.8003	0.0319
201Ile	0.6072	0.0109	0.5969	0.0135	0.1055	0.0031	0.1045	0.0022	0.7408	0.0341	0.6876	0.0459

202Leu	0.592513	0.006718	0.579635	0.008299	0.09671	0.002295	0.093754	0.001521	0.754875	0.025664	0.804871	0.041097
203Lys	0.584751	0.007397	0.588991	0.007831	0.102317	0.002384	0.119771	0.001424	0.712354	0.028557	0.633333	0.028982
204Ala	0.623574	0.005271	0.646367	0.005747	0.105251	0.001611	0.155889	0.001142	0.772875	0.020119	0.722538	0.032209
205Leu	0.6266	0.005257	0.665354	0.009287	0.106191	0.001449	0.104838	0.001317	0.699513	0.019796	0.713513	0.03247
206Gly	0.573407	0.005235	0.58828	0.008043	0.102507	0.001409	0.096716	0.001306	0.626309	0.017659	0.706579	0.035112
207Pro												
208Gly	0.734314	0.106026	0.620423	0.008281	0.131729	0.019651	0.13729	0.001409	0.466325	0.09165	0.715857	0.03307
209Ala	0.638475	0.004715	0.622631	0.005441	0.129425	0.001295	0.123033	0.000933	0.663548	0.015867	0.534486	0.024305
210Thr	0.701967	0.008438	0.648882	0.010174	0.099677	0.00184	0.114113	0.001457	0.636409	0.021812	0.538055	0.026144
211Leu	0.610405	0.008567	0.62868	0.00855	0.102882	0.002275	0.103579	0.001476	0.697155	0.024036	0.756041	0.033829
212Glu	0.626401	0.005628	0.61071	0.006934	0.100092	0.001467	0.097892	0.00121	0.678968	0.020758	0.772021	0.036498
213Glu	0.636796	0.006571	0.624567	0.008603	0.098119	0.001604	0.098732	0.001319	0.710249	0.020719	0.775831	0.034275
214Met	0.640653	0.008706	0.636426	0.009691	0.08696	0.002045	0.100799	0.001431	0.766219	0.024002	0.750029	0.035075
215Met	0.614813	0.005121	0.596933	0.006879	0.093804	0.00147	0.084264	0.001298	0.730632	0.015055	0.769433	0.023147
216Thr	0.638753	0.006276	0.610109	0.00914	0.097618	0.00152	0.099457	0.00122	0.717478	0.019749	0.722305	0.02843
217Ala	0.621287	0.00441	0.609493	0.005853	0.097569	0.001333	0.090491	0.000984	0.789395	0.019209	0.781764	0.028817
218Cys	0.661484	0.00825	0.62862	0.011943	0.101304	0.00188	0.09753	0.001449	0.638053	0.019774	0.621734	0.029554
219Gln	0.579495	0.006185	0.573754	0.006964	0.105187	0.001643	0.106714	0.001129	0.680757	0.021186	0.617809	0.026663
220Gly			0.559232	0.012338			0.137839	0.001748			0.578003	0.037271
221Val	0.635606	0.005621	0.610395	0.005208	0.09467	0.00106	0.12332	0.000908	0.597037	0.009723	0.490066	0.01977
222Gly	0.847114	0.161167	0.609527	0.007951	0.156133	0.043504	0.134606	0.001316	0.107352	0.180995	0.482373	0.025393
223Gly	0.716187	0.035257	0.646696	0.005945	0.243134	0.015784	0.284991	0.001236	0.200368	0.035358	0.179126	0.019416
224Pro												
225Gly			0.642326	0.015377			0.267539	0.021754			0.031	0.042579
226His			0.680532	0.043651			0.173992	0.005352			0.048619	0.061889
227Lys			0.615648	0.008826			0.171134	0.001637			0.425002	0.018752
228Ala			0.652704	0.007093			0.227598	0.032021			0.144596	0.018182
229Arg			0.646819	0.004675			0.284712	0.000932			0.018654	0.01305
230Val	0.668321	0.028118	0.628653	0.005593	0.224942	0.003906	0.231641	0.014507	0.088458	0.12731	0.150404	0.01691
231Leu	0.659639	0.015389	0.638174	0.006086	0.157428	0.003541	0.184488	0.001172	-0.00686	0.024156	-0.02734	0.026735
232Ala	0.701948	0.054016	0.650233	0.003652	0.224578	0.030529	0.241327	0.000725	0.077735	0.012279	0.012596	0.014359
233Glu	0.707253	0.05814	0.6771	0.004981	0.241798	0.003886	0.260266	0.025476	-0.03193	0.021149	-0.09458	0.018153
234Ala	0.57937	0.01843	0.666738	0.004169	0.20231	0.003896	0.242645	0.019866	0.215311	0.016371	0.05728	0.01331
235Met	0.70255	0.05267	0.652023	0.005358	0.215752	0.025448	0.272372	0.001205	0.210375	0.023447	-0.05319	0.014983
236Ser	0.254337	0.053045	0.678722	0.00826	0.186868	0.067845	0.273391	0.014916	0.022655	0.123319	-0.06787	0.027537
237Gln			0.659965	0.008701			0.316231	0.00212			-0.13375	0.023979
238Val	0.762409	0.022763	0.685827	0.007235	0.229042	0.019658	0.245247	0.001448	-0.16381	0.024125	-0.15614	0.0262
239Thr	0.618052	0.125606	0.652679	0.007261	0.258054	0.029479	0.305494	0.001654	-0.3239	0.083501	-0.16359	0.022397
240Asn			0.649341	0.008628			0.321178	0.002109			-0.07802	0.018288
241Pro												
242Ala	0.666221	0.018779	0.632535	0.003446	0.280148	0.004692	0.312706	0.000951	0.001709	0.020412	-0.04877	0.013159
243Thr	0.676499	0.031113	0.617856	0.004355	0.233398	0.008992	0.290948	0.001035	0.063152	0.028299	-0.00682	0.018161
244Ile	0.642362	0.011293	0.640363	0.004135	0.326419	0.004656	0.330707	0.001381	-0.07317	0.019859	-0.08949	0.02308
245Met	0.607001	0.036883	0.617839	0.006436	0.215011	0.014136	0.255764	0.035188	-0.11983	0.037814	-0.0027	0.023062
246Ile	0.672064	0.013264	0.650353	0.004962	0.319493	0.005092	0.322468	0.001424	0.002222	0.021682	-0.06617	0.020009
247Gln			0.604398	0.010785			0.323123	0.001601			-0.17684	0.024845
248Lys			0.605378	0.00341			0.296886	0.000923			0.058333	0.008692
249Gly			0.65636	0.008806			0.321868	0.001889			-0.13236	0.019092
250Asn			0.667991	0.011289			0.324248	0.002699			-0.17114	0.025764
251Phe			0.654741	0.006523			0.365846	0.001748			-0.05393	0.015825
252Arg			0.660922	0.004882			0.308302	0.001044			-0.10127	0.011931
253Asn			0.611487	0.010419			0.292484	0.00282			-0.01835	0.027699
254Gln												
255Arg												
256Lys			0.612289	0.008375			0.326795	0.002154			-0.14926	0.027938
257Thr	0.551821	0.122127	0.620349	0.013713	0.283367	0.049962	0.267948	0.015811	-0.27907	0.080285	-0.01672	0.036012
258Val	0.637911	0.017522	0.596259	0.005019	0.225558	0.004395	0.267491	0.001205	0.251815	0.017164	0.099354	0.013456
259Lys	0.549011	0.006273	0.513617	0.008915	0.193144	0.001921	0.202205	0.031441	0.416062	0.016239	0.471708	0.045829
260Cys	0.50977	0.004496	0.499174	0.011487	0.105297	0.001274	0.132669	0.002337	0.538809	0.016539	0.505902	0.051443
261Phe	0.516989	0.010947	0.475635	0.019987	0.137371	0.003221	0.160553	0.004682	0.444487	0.030708	0.668883	0.155446

262Asn	0.53674	0.006303	0.501794	0.01988	0.150386	0.00197	0.155051	0.004307	0.521851	0.02267	0.569776	0.104103
263Cys	0.534126	0.007274	0.500011	0.024533	0.156928	0.002261	0.156752	0.0052	0.553275	0.018527	0.53211	0.085104
264Gly	0.486338	0.004613	0.460805	0.017601	0.147513	0.001653	0.156876	0.003722	0.617055	0.019494	0.754035	0.082504
265Lys	0.55335	0.004255	0.669042	0.004529	0.160332	0.001377	0.3107	0.00117	0.387515	0.009823	-0.06661	0.011938
266Glu	0.544097	0.004713	0.541282	0.009004	0.164355	0.001435	0.169438	0.001783	0.575845	0.017714	0.47443	0.046342
267Gly	0.541693	0.016631	0.491857	0.012002	0.165606	0.004512	0.144161	0.002162	0.444381	0.029024	0.522823	0.057066
268His	0.508813	0.00267	0.495299	0.006487	0.156548	0.001152	0.162461	0.00154	0.629796	0.013969	0.616164	0.039067
269Ile	0.511268	0.004379	0.49405	0.010558	0.155563	0.001621	0.156003	0.002084	0.507083	0.016543	0.575306	0.051574
270Ala	0.501946	0.005416	0.513545	0.01982	0.159432	0.001993	0.171986	0.004743	0.576414	0.023673	0.683072	0.128792
271Lys	0.52359	0.035869	0.504838	0.007981	0.136021	0.008016	0.146644	0.001645	0.72653	0.055187	0.605855	0.04338
271Asn	0.519051	0.005937	0.513336	0.010232	0.160995	0.001863	0.160949	0.002163	0.517203	0.017994	0.57438	0.05555
273Cys	0.520955	0.003088	0.485714	0.007631	0.180157	0.016211	0.186885	0.022969	0.595968	0.014274	0.55372	0.040002
274Arg	0.551097	0.053491	0.456143	0.016924	0.121435	0.012685	0.146222	0.004462	0.604171	0.101803	0.491133	0.091749
275Ala	0.553962	0.004457	0.547447	0.008137	0.118312	0.001567	0.154997	0.001833	0.561497	0.016332	0.63712	0.043939
276Pro												
277Arg	0.580327	0.014288	0.570913	0.006297	0.194514	0.004005	0.251812	0.001585	0.392691	0.019139	0.1865	0.015414
278Lys	0.514834	0.015273	0.547157	0.016832	0.163854	0.012524	0.169248	0.003405	0.414118	0.031864	0.372091	0.108091
279Lys	0.491455	0.006769	0.594627	0.009243	0	0	0.299981	0.002542	0.523513	0.018689	0.034869	0.020271
280Gly	0.620832	0.031156	0.56	0.014547	0.276764	0.010306	0.270692	0.003833	0.307758	0.042371	0.496641	0.091763
281Cys	0.529613	0.004108	0.64208	0.003212	0.203293	0.001556	0.258109	0.000714	0.272957	0.00841	0.035095	0.011775
282Trp	0.491933	0.007773	0.459919	0.017698	0.101974	0.002062	0.102685	0.003193	0.490959	0.027329	0.660372	0.136226
283Lys	0.502445	0.007966	0.482125	0.023371	0.182329	0.002807	0.183393	0.031217	0.533178	0.024989	0.448531	0.125117
284Cys	0.521993	0.007208	0.504592	0.034932	0.144155	0.002375	0.12755	0.005486	0.520468	0.025846	0.535196	0.167371
285Gly	0.484333	0.006093	0.591927	0.028137	0.144863	0.002158	0.138737	0.002962	0.511133	0.018721	0.3992	0.022664
286Lys	0.515074	0.006148	0.631512	0.011732	0.167394	0.00203	0.335445	0.003358	0.499576	0.017882	-0.0707	0.018769
287Glu			0.629808	0.003815			0.286331	0.000929			0.000517	0.012348
288Gly			0.503223	0.057821			0.138278	0.009391			0.591857	0.212767
289His	0.511081	0.00814	0.472404	0.020507	0.12425	0.002657	0.109582	0.003693	0.552839	0.028402	0.461836	0.100162
290Gln	0.526534	0.02352	0.540064	0.024311	0.136619	0.006625	0.150007	0.004691	0.627974	0.025385	0.513492	0.067886
291Met	0.573662	0.050321	0.463489	0.016437	0.165183	0.013532	0.162864	0.004113	0.499734	0.048745	0.635616	0.100593
292Lys	0.537993	0.02784	0.484113	0.016599	0.152099	0.006587	0.139519	0.003027	0.564863	0.031041	0.599684	0.077136
293Asp	0.491393	0.005693	0.53574	0.011705	0.169567	0.001898	0.180748	0.002219	0.497281	0.020292	0.451564	0.043607
294Cys	0.502211	0.003944	0.475276	0.014457	0.190313	0.001639	0.182296	0.003533	0.506638	0.01827	0.483076	0.087637
295Thr			0.503398	0.02607			0.175521	0.004023			0.724278	0.146804
296Glu	0.518021	0.009228	0.508163	0.014362	0.187915	0.003033	0.199762	0.003559	0.363326	0.023436	0.614084	0.101617
297Arg	0.622887	0.016077			0.233798	0.011141			0.227058	0.016128		
298Gln			0.615934	0.020011			0.397613	0.006444			0.011362	0.027531
299Ala			0.724658	0.01757			0.288476	0.019869			-0.82571	0.085903
300Asn	1.0077	0.006666	1.070422	0.006558	0.785554	0.005827	0.787994	0.00479	-1.12576	0.019741	-1.23394	0.029854

IV Published scientific articles

Xiaowei Chen, Pascale Coric, Valery Larue, Serge Turcaud, Xiao Wang, Sylvie Nonin-Lecomte, Serge Bouaziz*. The HIV-1 maturation inhibitor, EP39, interferes with the dynamic helix-coil equilibrium of the CA-SP1 junction of Gag. **Eur J Med Chem** (2020) Jul 18;204:112634. doi: 10.1016/j.ejmech.2020.112634



The HIV-1 maturation inhibitor, EP39, interferes with the dynamic helix-coil equilibrium of the CA-SP1 junction of Gag

Xiaowei Chen^a, Pascale Coric^a, Valery Larue^a, Serge Turcaud^b, Xiao Wang^a, Sylvie Nonin-Lecomte^a, Serge Bouaziz^{a,*}

^a CITCoM, CNRS, UMR 8038, Université de Paris, 4 Avenue de L'Observatoire, Paris, 75270, France

^b LCBPT, CNRS, UMR 8601, Université de Paris, Paris, 45 Rue des Saints Pères, 75270, France

ARTICLE INFO

Article history:

Received 7 May 2020

Received in revised form

29 June 2020

Accepted 29 June 2020

Available online 18 July 2020

Keywords:

HIV-1

Maturation inhibitor

EP39

CA-SP1 junction

Dynamics

NMR

ABSTRACT

During the maturation of HIV-1 particle, the Gag polyprotein is cleaved into several proteins by the HIV-1 protease. These proteins rearrange to form infectious virus particles. In this study, the solution structure and dynamics of a monomeric mutated domain encompassing the C-terminal of capsid, the spacer peptide SP1 and the nucleocapsid from Gag was characterized by Nuclear Magnetic Resonance in the presence of maturation inhibitor EP39, a more hydro-soluble derivative of BVM. We show that the binding of EP39 decreases the dynamics of CA-SP1 junction, especially the QVT motif in SP1, and perturbs the natural coil-helix equilibrium on both sides of the SP1 domain by stabilizing the transient alpha helical structure. Our results provide new insight into the structure and dynamics of the SP1 domain and how HIV-1 maturation inhibitors interfere with this domain. They offer additional clues for the development of new second generation inhibitors targeting HIV-1 maturation.

© 2020 Elsevier Masson SAS. All rights reserved.

1. Introduction

Maturation of human immunodeficiency virus type 1 (HIV-1) particle is an essential step for viral infectivity. During maturation, the HIV-1 Pr55^{Gag} (Gag) precursor is cleaved by the HIV-1 protease into several proteins [1]: matrix (MA) which binds to the plasma membrane, capsid (CA) which forms the virus capsid, spacer peptide 1 (SP1), nucleocapsid (NC) which binds to RNA and forms a ribonucleoprotein (RNP) complex, spacer peptide 2 (SP2) and p6 involved in Vpr encapsidation. The fastest to slowest cleavage rate is as follows: SP1-NC > MA-CA/SP2-p6 > CA-SP1/NC-SP2 [1–3]. After cleavage, these proteins rearrange to form the mature HIV-1 particle.

The final cleavage at CA-SP1 junction is the limiting step for HIV-1 maturation. The structure of CA-SP1 junction has been reported in different studies. The crystal structure of CA_{CTD}-SP1 hexamer forms an α -helix at CA_{CTD}-SP1 junction encompassing the nine C-terminal residues of CA_{CTD} and seven N-terminal residues of SP1 [4]. However, in the cryo-EM structure of CA-SP1, no detectable

electron density was present in SP1 which means SP1 is probably highly dynamics [5,6]. In the solution structure of CA_{CTD}-SP1-NC, the residues in SP1, the final thirteen residues of CA_{CTD} and the thirteen N-terminal residues of NC are flexible and have a low propensity of forming α -helix [7]. In 30% Trifluoroethanol (TFE) solution, the short peptide CA^{211–231}-SP1^{232–245}-NC^{246–258} can form a long α -helix extending from CA²²⁶ to NC²⁴⁸. Thus, proper folding of SP1 could depend on the presence of the NC domain [7].

Bevirimat (BVM) (Fig. 1A), the first-in-class maturation inhibitor (MI), can block the final cleavage between CA and SP1 [8]. Recently, it was reported that BVM could stabilize the helical conformation of SP1 [5]. Additional studies on the HIV-1 Gag structure by microED revealed that CA-SP1 junction associates into a six-helix bundle and BVM binds to the center of this bundle [9]. Interestingly, our group found that EP39 (Fig. 1B), a derivative of BVM which shows better antiviral activity and water solubility compared to BVM [10], might share the same binding pocket with BVM, while they interact differently in the hexameric structure of CA_{CTD}-SP1 [11]. Nevertheless, the mechanism by which BVM or EP39 stabilizes the immature lattice is still unclear. Here, we use nuclear magnetic resonance (NMR) to investigate how EP39 impacts the structure and dynamics of CA_{CTD}^{W184A, M185A}-SP1-NC (Fig. 1C). CA_{CTD}-SP1-NC is in monomer-dimer equilibrium [7]. In CA_{CTD}^{W184A, M185A}-SP1-NC, the

* Corresponding author.

E-mail address: serge.bouaziz@parisdescartes.fr (S. Bouaziz).

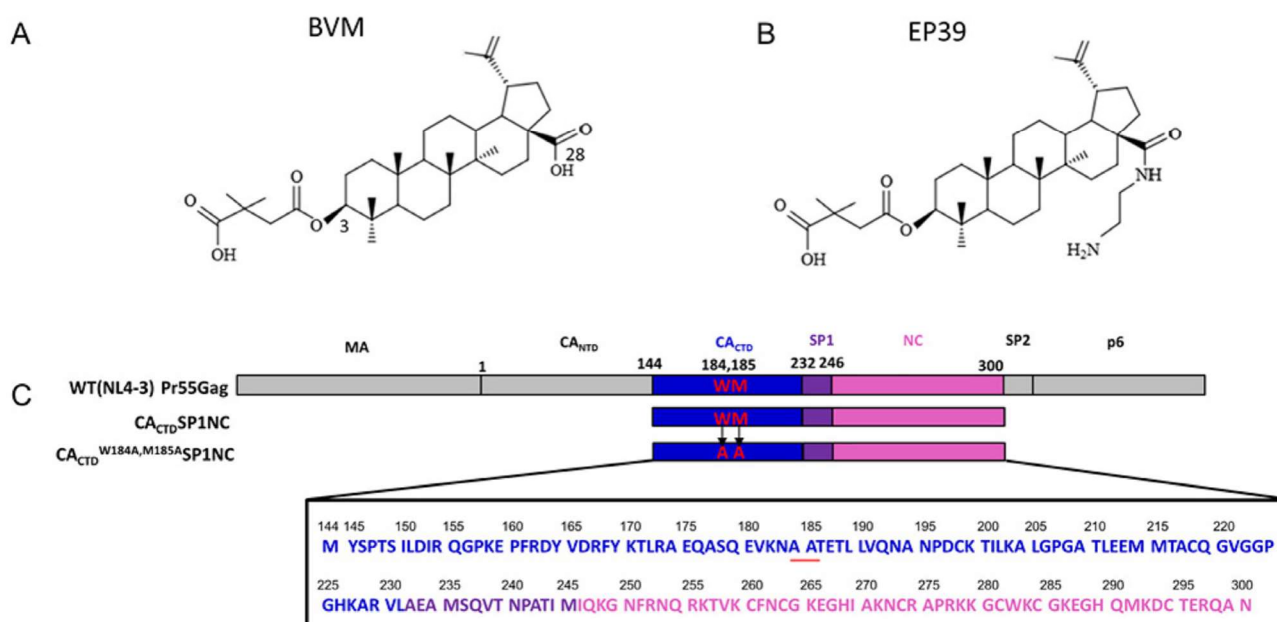


Fig. 1. Schematic representations of BVM (A), EP39 (B) and the Gag polyprotein (C). EP39 retains the C3 function of BVM which is essential for the maturation inhibitory effect and has substituted C28 function to achieve better water solubility [10]. Wild type CA_{CTD}-SP1-NC from the Gag polyprotein contains the C-terminal domain of CA, the spacer peptide SP1 and the nucleocapsid protein NC; CA_{CTD}^{W184A, M185A}-SP1-NC contains two mutations (W184A, M185A) in the C-terminal domain of CA to prevent dimerization. Residues of CA_{CTD}^{W184A, M185A}, SP1 and NC are colored in blue, purple and pink respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

two mutations (W184A, M185A) abolish CA dimerization and do not interfere with the structure of CA_{CTD} [12]. Thus, by NMR in solution, it will be easier to study the structural and dynamic properties of the monomeric CA_{CTD}^{W184A, M185A}-SP1-NC than the wild type CA_{CTD}-SP1-NC that dimerizes. NC is present because we believe it can ensure the stable folding of SP1.

2. Results

2.1. The two mutations, W184A and M185A, abolish CA dimerization

Comparison between ¹H-¹⁵N SOFAST-HMQC spectra obtained for the wild type CA_{CTD}-SP1-NC and doubly mutated CA_{CTD}^{W184A, M185A}-SP1-NC (Fig. 2A) shows that their resonances superimposed well, especially the resonance of NC (by reference to the resonances labeled with residue number in Fig. 3A). However, many resonances of the CA_{CTD} domain (residue 144–231) are missing in wild type CA_{CTD}-SP1-NC compared with the mutated CA_{CTD}^{W184A, M185A}-SP1-NC because of the line broadening due to a monomer-dimer exchange in an intermediate state (Fig. 2A) [7].

2.2. NC does not seem to interact with CA_{CTD}^{W184A, M185A} or SP1

Comparison of the ¹H-¹⁵N SOFAST-HMQC spectra of ¹⁵N-NC and ¹⁵N CA_{CTD}^{W184A, M185A}-SP1-NC (Fig. 2B) shows that the resonances of NC domain of both proteins superimpose well, which indicates that the conformation of NC is preserved in NC alone and in CA_{CTD}^{W184A, M185A}-SP1-NC. We can conclude that in CA_{CTD}^{W184A, M185A}-SP1-NC, NC does not seem to interact with the rest of the protein, CA_{CTD}^{W184A, M185A} or SP1.

2.3. EP39 stabilizes the SP1 domain on both sides

¹H-¹⁵N SOFAST-HMQC spectra were recorded on the monomer CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and in the presence of EP39 (Fig. 3A). In the absence of EP39, the amide resonances of residues CA²¹⁸ to NC²⁵⁷ (encompassing the fourteen C-terminal residues of CA, SP1 and the twelve N-terminal residues of NC) exhibit weak dispersion with narrow linewidths, indicating that this region is highly flexible, in accordance with the structural properties of CA_{CTD}-SP1-NC [7]. Many resonances in this region are missing due to residue dynamics and fast proton exchange with the solvent. Upon EP39 addition, new resonances corresponding to residues CA^{220, 225, 226, 227, 228, 229} and NC^{248, 249, 250, 251, 252, 253, 256} appear as shown in Fig. 3A and Fig. S1 in supplementary material, which means the regions encompassing these residues are stabilized by EP39. The chemical shift perturbations (CSP) and peak intensity changes observed on the domain (CA²¹⁸-NC²⁵⁷) (Fig. 3B) suggest that EP39 modifies the dynamics and/or the conformation of the corresponding domain, consistent with the previously described effect of BVM on CA_{CTD}-SP1 [5]. We also note CA-T²¹⁰ undergoes a relatively high CSP of 0.06 ppm. It could be caused by an allosteric structural change following the MI blocking of the transient interaction between SP1 and the first α -helix of CA_{CTD} [5]. The resonance intensity of residue CA-K¹⁵⁸ (in major homology region MHR) and CA-G²²² is six and five times stronger after EP39 addition consistent with our previously docking results of EP39 on CA_{CTD}-SP1 showing that the carbonyl in position C-28 of EP39 has a close contact with CA-K¹⁵⁸ [11].

2.4. Kd evaluation between EP39 and CA_{CTD}^{W184A, M185A}-SP1-NC

We monitored the dissociation constants (Kd) of EP39 on protein by following the CSP of ten residues whose resonances are

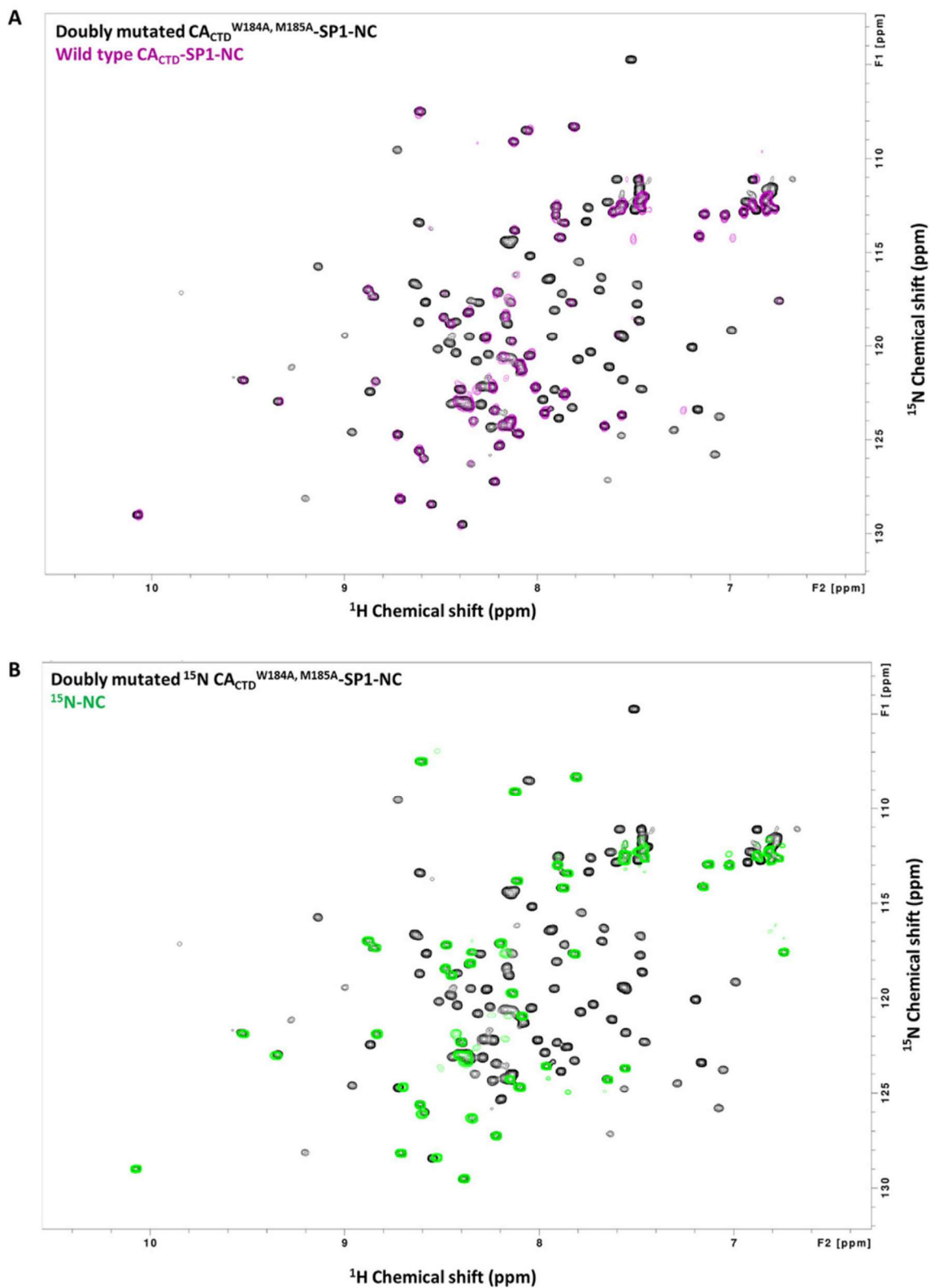


Fig. 2. Solution NMR spectra of CA_{CTD} -SP1-NC, $CA_{CTD}^{W184A, M185A}$ -SP1-NC and NC at 303 K. (A) Superimposition of natural abundance 1H - ^{15}N SOFAST-HMQC NMR spectra recorded on wild type protein CA_{CTD} -SP1-NC (magenta) and doubly mutated protein $CA_{CTD}^{W184A, M185A}$ -SP1-NC (black). (B) Superimposition of the 1H - ^{15}N SOFAST-HMQC spectra of the isolated ^{15}N labeled NC (green spectrum) and doubly mutated ^{15}N - $CA_{CTD}^{W184A, M185A}$ -SP1-NC (black spectrum). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

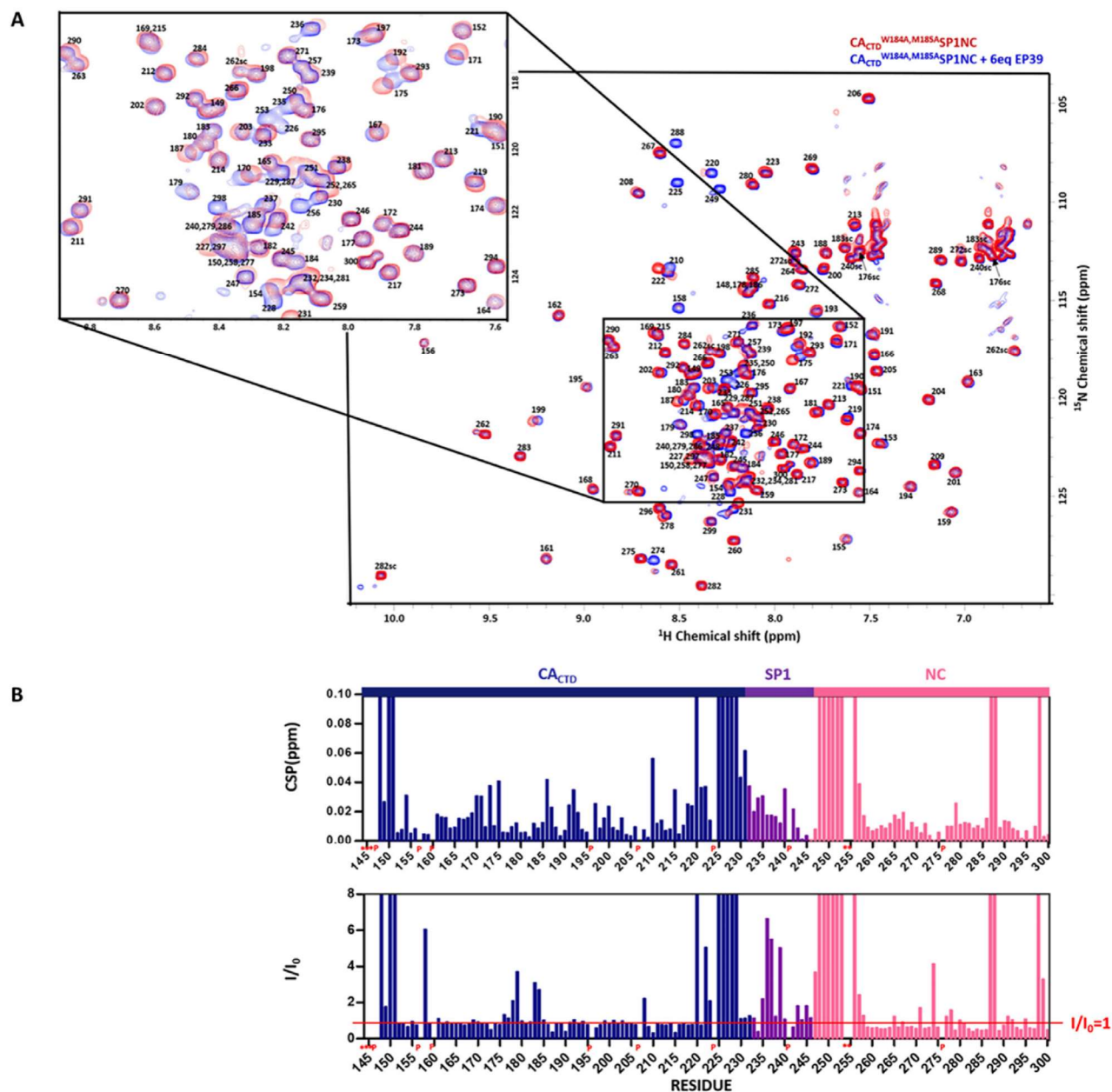


Fig. 3. Solution NMR of EP39 binding to CA^{W184A, M185A}-SP1-NC. (A) Superposition of ¹H-¹⁵N SOFAST-HMQC spectra recorded on ¹⁵N labeled CA^{W184A, M185A}-SP1-NC in the absence (red) and in the presence of 6 equivalents of EP39 (blue) at 303 K. Resonances are labeled with amino acids numbers. (B) Chemical shift perturbations (CSP) and resonance intensity ratios (I/I_0) of the amide groups (¹H/¹⁵N combined chemical shift) in the presence and in the absence of EP39. The chemical shift perturbations and resonance intensity ratios of the residues whose ¹H-¹⁵N resonance are seen only in the presence of EP39 were set at 0.1 and 8 respectively. The red line indicates the resonance intensity ratio of 1. Proline residues are labeled as P and unassigned residues are labeled as *. Residues are colored as Fig. 1C. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

isolated in the spectrum. These K_ds could be approximate because of the slight precipitation of EP39 at high concentration during the experiment, but they give us ideas about the binding affinity of EP39 to different residues of SP1. In these ten residues, SP1-T²³⁹ has the highest binding affinity (52.4 μM); SP1-S²³⁶, SP1-Q²³⁷ and SP1-V²³⁸ have relatively high binding affinities (179.6 μM, 154.5 μM and 204.4 μM respectively); SP1-E²³³, SP1-M²³⁵ and NC-T²⁵⁷ have medium binding affinities (237.0 μM, 235.4 μM, 308.4 μM respectively); CA-V²²¹, CA-L²³¹ and SP1-A²³² have weak binding affinities (669.6 μM, 678.4 μM and 691.9 μM respectively) (Fig. 4). The higher

binding affinity at SP1-Q²³⁷V²³⁸T²³⁹ is consistent with previous study which shows that SP1-Q²³⁷V²³⁸T²³⁹ motif polymorphism confers intrinsic resistance to BVM [13].

2.5. Solution structure of monomeric mutant HIV-1 CA^{W184A, M185A}-SP1-NC in the presence of EP39

The structure of the monomeric mutant CA^{W184A, M185A}-SP1-NC in the presence of EP39 was determined using the program ARIA 2.3.2 [14] based on the experimental NMR restraints summarized in

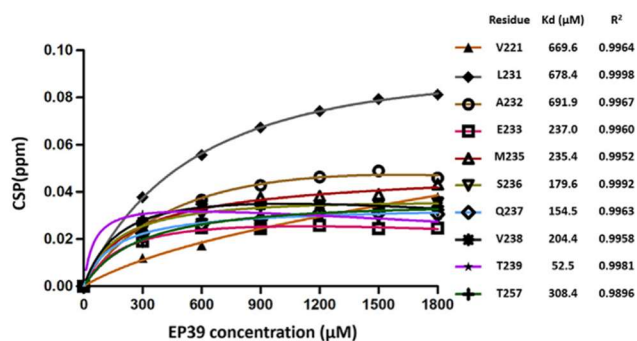


Fig. 4. Variations of chemical shift perturbations (CSP) of ten residues (CA-V²²¹, CA-L²³¹, SP1-A²³², SP1-E²³³, SP1-M²³⁵, SP1-S²³⁶, SP1-Q²³⁷, SP1-V²³⁸, SP1-T²³⁹, NC-T²⁵⁷) isolated in spectrum as a function of EP39 whose concentration increase from one to six equivalents of protein. Kds were computed as described in Materials and Methods.

Table 1. The coordinates of the 10 lowest energy structures have been deposited to the PDB under accession number 6RWG. CA_{CTD}^{W184A,M185A}-SP1-NC contains six alpha helices: H1 (CA^{161–172}), H2 (CA^{179–192}), H3 (CA^{196–208}), H4 (CA^{212–218}), H5 (CA^{227–SP1²³⁹}) and H6 (SP1^{241–NC²⁵¹}); two zinc fingers: ZF1 (NC^{260–273}) and ZF2 (NC^{281–294}) connected by seven amino acids (Fig. 5A).

To better analyze the structure of CA_{CTD}^{W184A,M185A}-SP1-NC, we structurally divided the protein into three domains: the first one, CA^{151–220}, encompassing the four α -helices in CA; the second one, CA^{221–NC²⁵¹}, constituted by CA-SP1 (H5) and SP1-NC (H6) helical junctions and finally the third one, NC^{252–294}, formed by the two zinc fingers connected by seven amino acids (Fig. 5B). The N and C ends of the protein were excluded because of their high flexibility. Then, each domain of the 10 structures was superimposed to evaluate their convergence. The first domain CA^{151–220} displays a good convergence with a backbone atoms RMSD of 1.06 Å (left Fig. 5B). Six amino acids, CA^{221–226}, starting the second domain are very flexible and display different orientations (middle Fig. 5B). A flexible elbow between H5 and H6, corresponding to the QVT motif,

allows these two helices to adopt different relative orientations. (middle Fig. 5B). The backbone atoms RMSDs for H5 and H6 are 0.74 Å and 1.34 Å respectively. The last domain contains two zinc fingers, ZF1 and ZF2, connected by seven residues linker (NC^{274–280}). The flexibility of this linker allows the two zinc fingers to adopt variable orientations with respect to each other. This relative motion can modulate the nucleic acid binding property [15]. RMSDs for ZF1 and ZF2 are 0.92 Å and 1.19 Å respectively on the backbone atoms for the 10 structures (right Fig. 5B).

2.6. The CA_{CTD}-SP1 domain adopts a similar structure in the monomeric or hexameric form

The structure which is closest to the average structure calculated on the 10 best structures was superimposed on the hexameric crystal structure of CA_{CTD}SP1 (PDB code: 5I4T). The RMSD calculated on CA¹⁴⁸-SP1²³⁹ backbone atoms between our structure and the chain G of the hexamer is 1.15 Å (Fig. 5C). This result demonstrates that the structure of CA_{CTD} in monomeric form is consistent with the structure in hexameric form and the two mutations (W184A and M185A) do not modify the structure of CA_{CTD}. We also observed two different orientations for H5 (CA^{221–226}) relative to H4 (CA^{212–218}) between our structure and the hexameric one as shown in Fig. 5C, demonstrating again the flexibility of the residues between H4 and H5. H6 which is missing in the crystal and cryo-EM structures is well folded in our NMR structure (Fig. 5A) and adopts different orientations compared to H5.

2.7. EP39 stabilizes the H5 and H6 helices of the second domain CA^{221–NC²⁵¹}

¹⁵N T1 and T2 relaxation times have been determined and are presented in Fig. S2, in Supporting Information. The T1/T2 and heteronuclear NOE (HetNOE) values are presented in Fig. 6. T1/T2 ratio provides a good estimation of the rate at which each N-H vector reorients with global tumbling. HetNOE values are indicative of the magnitude of local sub-nanosecond motions with high values corresponding to restricted motions and low values to high-

Table 1
Statistics for the top 10 NMR structures of monomeric mutant HIV-1 CA_{CTD}^{W184A,M185A}-SP1-NC.

Restrains	
Total number of NMR restraints	2121
Total number of NOE distance restraints	1898
Long range ($ i-j > 5$)	90
Medium range ($2 \leq i-j \leq 4$)	548
Sequential range $ i-j = 1$	718
Intra-residue NOEs	542
Total number of dihedral restraints	223
Restraint statistics	
Distance violations per structure	
0.2–0.3 Å	2.90
0.3–0.5 Å	0.20
>0.5 Å	0.00
R.M.S. on distance violations per constraint (Å)	0.02
Maximum distance violation (Å)	0.42
Dihedral angle violations per structure	
5–10°	0.80
>10°	0.00
R.M.S. on dihedral violations per constraint (°)	0.73
Maximum dihedral angle violation (°)	8.78
Ramachandran analysis of residues	
Most favored regions (%)	87
Allowed regions (%)	10
Disallowed regions (%)	3

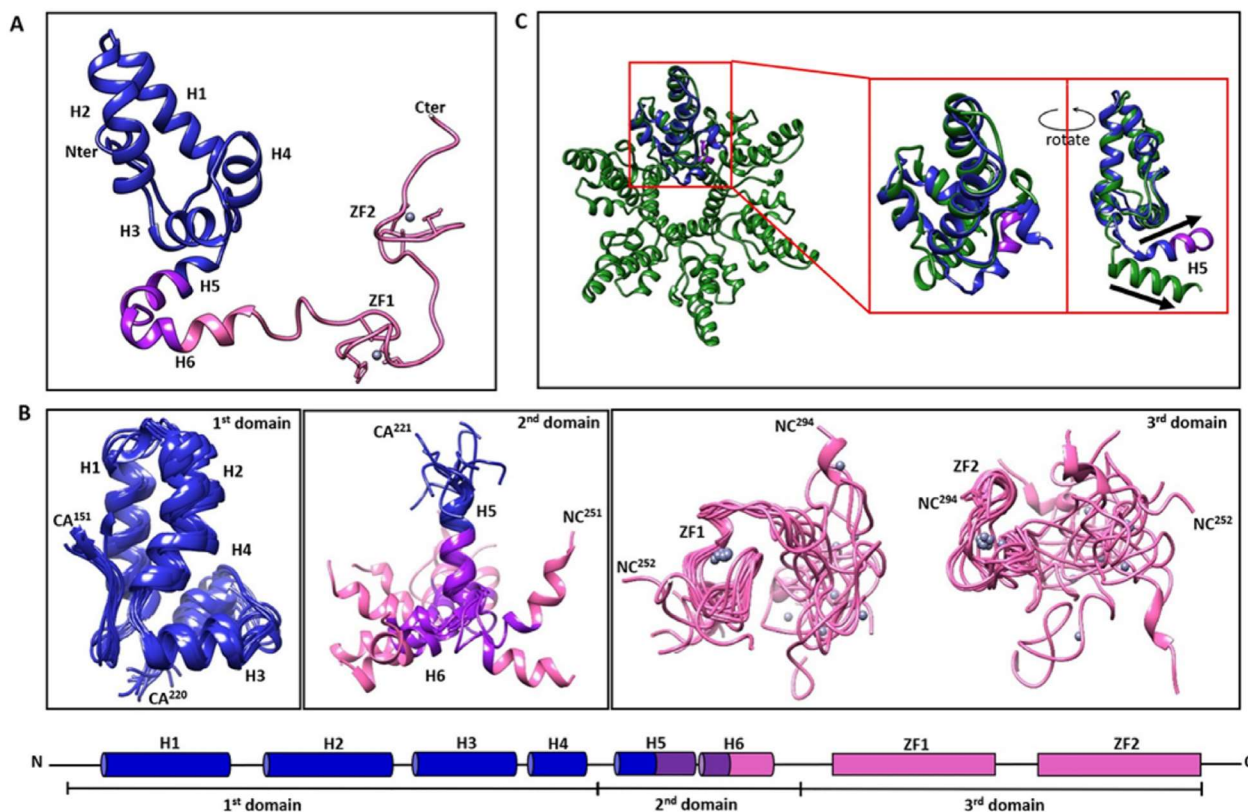


Fig. 5. NMR structure of CA^{W184A, M185A}-SP1-NC in the presence of EP39. The secondary structure of the protein is indicated at the bottom. H: helix; ZF: zinc finger. Residues are colored as in Fig. 1C. (A) the structure which is closest to the average structure is represented. From N-terminal to C-terminal, this structure contains six helices: H1 (residue CA^{161–172}), H2 (residue CA^{179–192}), H3 (residue CA^{196–208}), H4 (residue CA^{212–218}), H5 (residue CA^{227–SP1²³⁹}), H6 (residue SP1^{241–NC²⁵¹}) and two zinc fingers: ZF1 (residue NC^{260–273}) and ZF2 (residue NC^{281–294}). (B) CA^{W184A, M185A}-SP1-NC was divided into three domains. The ten best structures are superimposed based on each domain. Left: superimposition on the first domain, CA^{151–220}, shows a RMSD of 1.06 Å calculated on the backbone atoms. Middle: Superimposition on the second domain, CA^{221–NC²⁵¹}, performed on H5 shows a RMSD of 0.74 Å. Right: superimpositions on the third domain, NC^{252–294}, performed on ZF1 and ZF2 respectively show RMSD of 0.92 Å and 1.19 Å respectively. (C) Superimposition on the region CA^{148–SP1²³⁸} between the NMR structure (Fig. 5A) and chain G (in green) in CA_{CTD}-SP1 hexamer (PDB code: 5I4T). H5 in our structure and in chain G of the crystal structure has different orientations as indicated by black arrows. PDB code for CA^{W184A, M185A}-SP1-NC in the presence of EP39 is 6RWG. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

amplitude motions.

In the absence of EP39, the N- and C-terminal parts of CA^{W184A, M185A}-SP1-NC are highly flexible. The first domain containing four α -helices is the most stable part in CA^{W184A, M185A}-SP1-NC. The second domain is the most flexible part and many amino acids ¹H-¹⁵N correlation peaks are even undetectable. The third domain containing the two stable zinc fingers connected by seven amino acids more flexible than the first domain but is more ordered than SP1.

In the presence of EP39, changes are observed in the second domain. New correlation peaks corresponding to the residues (CA²²⁵, CA²²⁶, CA²²⁷, CA²²⁸, CA²²⁹, NC²⁴⁸, NC²⁴⁹, NC²⁵⁰, NC²⁵¹) of the second domain appear in ¹H-¹⁵N SOFAST-HMQC spectrum. Besides, the average values of T1/T2 and HetNOE for H5 are slightly higher than that of H6 suggesting H5 is less dynamic than H6. The residues SP1-Q²³⁷, SP1-V²³⁸ and SP1-T²³⁹, forming the QVT motif located at the elbow between H5 and H6, have lowest HetNOE values compared to other residues in the second domain, especially SP1-T²³⁹ with lowest HetNOE value. After adding EP39, HetNOE value of SP1-T²³⁹ increases from -0.324 to -0.164 which means this residue is less dynamic in the presence of EP39. Conversely, the first and third domains do not seem to be affected by the addition of EP39.

3. Discussion and conclusions

Several structural studies have been carried out on the CA-SP1 junction and different results have been obtained depending on the conditions of the study. The crystal structure of the hexamer CA_{CTD}-SP1 contains an α -helix at the CA-SP1 junction encompassing the C-terminal nine residues of CA and the N-terminal seven residues of SP1 [4]. In this crystal structure, the C-terminal seven residues of SP1 are undetectable because of their high dynamics. On the other hand, in 30% TFE containing solution, the whole SP1 forms a long α -helix [16]. However, TFE is known to stabilize the alpha-helical structure in proteins and their fragments and the long helix in SP1 could be structurally stabilized by TFE.

In this work, we solved the first monomeric solution structure of CA^{W184A, M185A}-SP1-NC in the presence of the maturation inhibitor EP39 by NMR in physiological conditions. In this structure, from N-terminus to C-terminus, four α -helices are formed at the CA_{CTD} domain, as in the hexameric crystal structure of CA_{CTD}-SP1. The two junctions CA-SP1 and SP1-NC form two α -helices (H5 and H6 respectively) separated by an elbow constituted by the QVT motif. This is the first time that the whole SP1 is demonstrated to be in helical structure at physiological condition. The NC folds into two zinc fingers separated by a flexible linker.

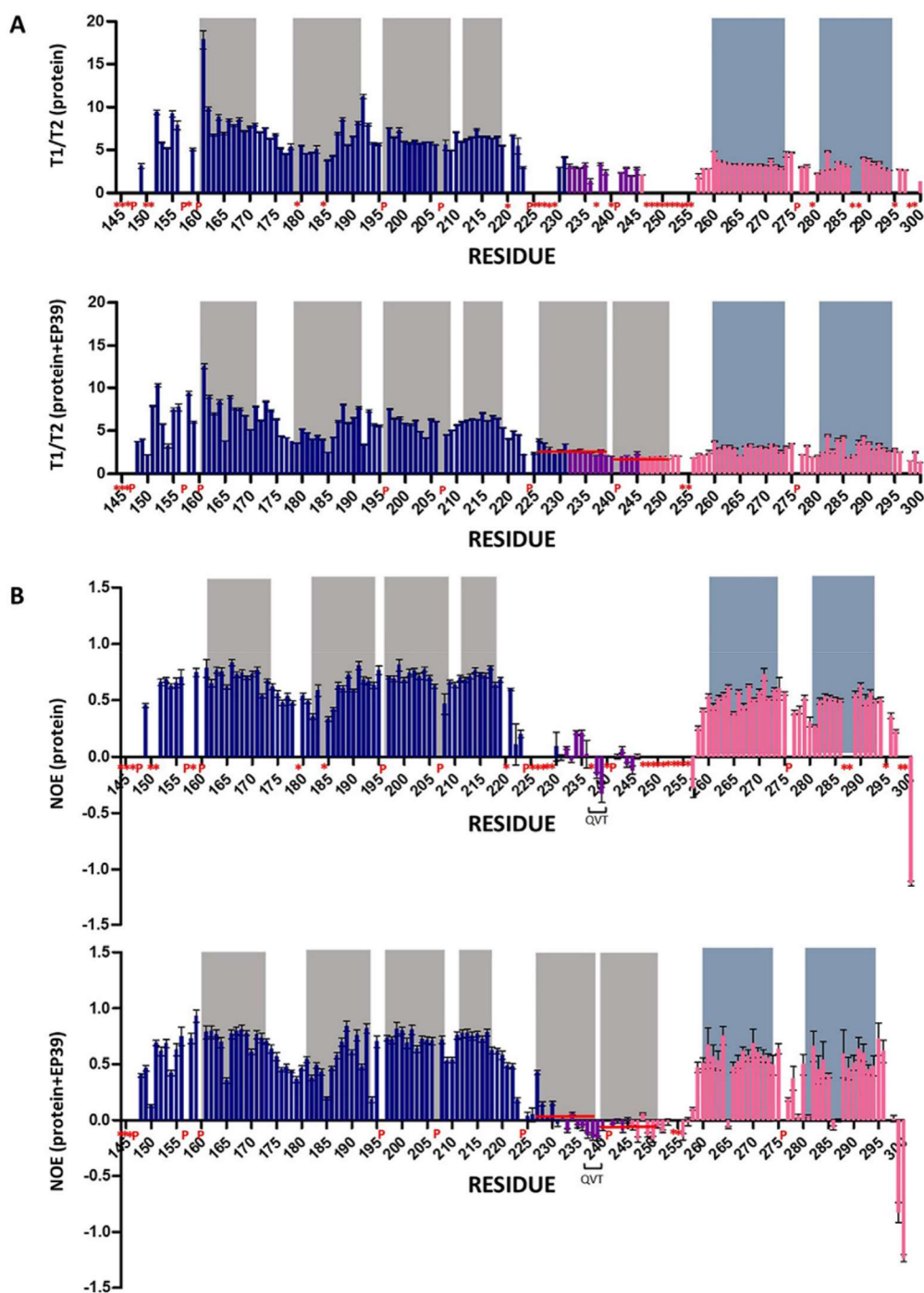


Fig. 6. Dynamic characterization of CA^{W184A, M185A}-SP1-NC in the absence and presence of EP39. Experimental NMR ¹⁵N relaxation data T1/T2 (A) and HetNOE (B) values were obtained at 600 MHz for protein in the absence (top) and presence (bottom) of EP39 at 303 K. The average T1/T2 values for H5 and H6 are 2.68 and 2.03 respectively. The average HetNOE values for helix 5 and helix 6 are 0.0086 and -0.06899 respectively. These average values are indicated as a red bold line. The color code for residues is the same as Fig. 1C. The boxes in grey indicates the four α -helices in the first domain and two α -helices in the second domain after the addition of EP39. The boxes in blue indicates the two zinc fingers in the third domain. Proline residues are labeled as P and unassigned residues are labeled as *. Residues SP1-Q²³⁷/V²³⁸-I²³⁹ are marked in the HetNOE value. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

BVM is known to stabilize the immature Gag lattice [17]. In the microED structure of CA_{CTD}-SP1 complexed with BVM, the inhibitor binds to the hydrophobic center of CA_{CTD}-SP1 hexamer [9]. In our study, EP39 seems to interact with CA_{CTD}^{W184A, M185A}-SP1-NC, mainly with the two alpha helices of the second domain (H5 and H6). The second domain CA²²¹-NC²⁵¹ has been reported to be highly dynamic and presents an equilibrium between coil and alpha helix [7,18]. Our NMR data shows that many resonances corresponding to the second domain CA²²¹-NC²⁵¹ cannot be detected in ¹H-¹⁵N SOFAST-HMQC spectrum in the absence of EP39 (Fig. 3A), because of the high dynamics of this domain. However, in the presence of EP39, this domain seems to be stabilized and showed both a CSP of existing peaks and the appearance of new peaks due to decrease in the rate of exchange with the solvent and the stabilization of this domain (Fig. 3B) that makes this domain to adopt a helical structure in the presence of EP39. Thus, we have shown that the stabilization of this domain by EP39 leads to the formation of two alpha helices, H5 and H6, with a short flexible elbow between them containing the very important QVT motif.

Polymorphism of the SP1 domain, especially the QVT motif forming an elbow between H5 and H6 (SP1-Q²³⁷V²³⁸T²³⁹), has been associated with reduced BVM activity in vivo [13,19] suggesting this motif is essential for BVM binding. The mutation SP1^{T239I} which mimics the effect of BVM binding could attenuate dynamics and increase helicity of residues CA²²⁶-SP1²³⁷ [5]. In this work, residues SP1-Q²³⁷V²³⁸T²³⁹ are the most dynamic amino acids in SP1 and have high binding affinities to EP39 compared to other residues (Fig. 4), especially residue SP1-T²³⁹ whose dynamics is decreased after adding EP39. These results suggest that EP39 binds to protein firstly at the SP1-QVT motif and that on each side of the SP1-QVT motif, H5 and H6 are stabilized to helical structure. These results could explain why the polymorphism of the QVT motif leads to a defect in binding of EP39/BVM and therefore to a high dynamic of the CA-SP1 junction allowing optimal activity of the protease.

In the crystal and microED structures of CA_{CTD}-SP1, the helix formed at the CA-SP1 junction starts at CA²²⁵ and stop at SP1²³⁸ [9]. As we have shown by our relaxation experiments, H6 is more dynamic than H5 (Fig. 6). The residue SP1-T²³⁹ at the end of H5 is the most dynamic residue in the second domain which could be the reason why the CA-SP1 junction helix stops at SP1-V²³⁸ and why H6 is not detectable in the crystal and microED structures of CA_{CTD}-SP1.

In the latest published article about the effect of BVM on the structure and dynamics of CA-SP1 junction, the assigned signals of SP1 ends at SP1-A²³² (A²⁸² in their numbering scheme) due to the high dynamics of SP1. Their data show that BVM causes subtle changes in the structure and dynamics of CA-SP1 junction [20]. However, the lost of information in the SP1 domain makes it impossible to clarify the essential role of SP1-QVT motif in the binding of the maturation inhibitor.

MHR plays a critical role in HIV-1 assembly, maturation and infectivity [21]. There are frequent, short lived contacts between SP1 and MHR [5]. In the docking result between EP39/BVM and CA_{CTD}-SP1, one common hydrogen bond involving the carbonyl in position C-28 of EP39/BVM and the ε-amino of CA-K¹⁵⁸ is observed [11]. The microED structure of CA_{CTD}-SP1 complexed with BVM indicates that the negatively charged dimethylsuccinyl carboxyl group on BVM is located at the center of the positive charges generated by the double ring of lysines (K¹⁵⁸ and K²²⁷) [9]. These results can easily be extrapolated to demonstrate the essential role of CA-K¹⁵⁸ in EP39 binding since EP39 and BVM have the same dimethylsuccinyl carboxyl group. In our NMR experiments, the intensity of the resonance of residue CA-K¹⁵⁸ in MHR is much stronger after EP39 addition (Fig. 3B), which means this residue could be an essential target for EP39 binding.

The SP1-NC junction (residue SP1²⁴¹-NC²⁵¹) in the C-terminal of

the second domain also adopted an α-helix in the presence of EP39. This finding is surprising and can be explained firstly by the presence of NC and secondly by the effect of EP39 binding. The N-terminal of NC (residues 246–257) is known to form a ₃₁₀ helix by electrostatic contacts with SL2/SL3 [22,23]. In 30% TFE solution, the peptide CA^{211–231}SP1^{232–245}NC^{246–258} adopts a helical structure spanning domain CA²²⁶-NC²⁴⁸ [16]. In the present work, the α-helix in the second domain starts from CA²²¹ and stops at NC²⁵¹ with a kink at SP1²⁴⁰ which could be caused by the presence of a proline at this position. This suggests that the N-terminal domain of NC has the propensity to form an alpha helix even in the absence of nucleic acids. Similar with the conformational stabilization of CA-SP1 junction, the helix formed at the SP1-NC junction results from the decrease in the dynamics of the QVT motif caused by EP39 binding. The microED structure of CA_{CTD}-SP1 complexed with BVM contains a helical structure in the N-terminal of SP1 and the C-terminal part of SP1 is invisible [9]. Here, we observe that the C-terminal domain of SP1 and the N-terminal domain of NC adopt an alpha helical structure (H6). This demonstrates that, in the presence of NC, EP39 causes the formation of a helical structure within the SP1-NC junction. Various NMR studies, including ours, have shown that SP1 is poorly structured in aqueous solution but adopts a helical structure if the polarity of the solvent is reduced. This property of the SP1 domain has also been demonstrated in an aqueous medium at high concentrations of the SP1 peptide [24]. These results suggest that SP1 could act as a switch during the assembly of Gag. It seems that we observe the same phenomenon in the presence of EP39 which would stabilize the association of Gag and its stabilization thus disturbing the binding of the protease and then the maturation.

We have observed the effect of EP39 on the structure and dynamics of the protein CA_{CTD}^{W184A, M185A}-SP1-NC, but unfortunately it was not possible to observe any intermolecular correlations between the protein and EP39 in the NOESY experiments which could be caused by the low water solubility of EP39 and high dynamics of the second domain even in the presence of EP39.

To conclude, the binding of EP39 to the second domain appears to particularly disturb the dynamic balance of the helix-coil of CA-SP1 and SP1-NC junctions. Indeed, by binding to the QVT motif of the SP1 domain, the dynamics of this motif is reduced and allows the alpha-helix H5 and H6 to form at the CA-SP1 and SP1-NC junctions respectively. We assert that the decrease in dynamics and the change in conformation of these domains caused by EP39 make it difficult for the protease to cleave the CA-SP1 junction. This therefore results in an inhibition of the maturation of HIV particles by the protease.

4. Experimental section

4.1. Chemical synthesis of EP39

EP39 was prepared as described previously [10] and solubilized in DMSO-*d*₆ at the concentration of about 60 mM. To verify that the structure of the protein was not modified by the addition of DMSO-*d*₆, a second sample containing the protein was prepared at the same concentration in which the same quantity of DMSO-*d*₆ was added.

4.2. Expression and purification of wild type HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC

The wild type HIV-1 CA_{CTD}-SP1-NC, mutated CA_{CTD}^{W184A, M185A}-SP1-NC and NC constructions were kindly provided by Dr. Michael F. Summers. Proteins were overexpressed and purified as previously described [23].

4.3. Binding of EP39 to CA_{CTD}^{W184A, M185A}-SP1-NC

A 300 μM ^{15}N labeled CA_{CTD}^{W184A, M185A}-SP1-NC was prepared in 200 μL acetate buffer (25 mM sodium acetate, pH 6.5, 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM 2-mercaptoethanol, 10% D₂O) in a 3 mm NMR tube. A ^1H - ^{15}N SOFAST-HMQC reference spectrum was recorded.

The addition of 1 μL DMSO-*d*₆ 0.3% solution containing EP39 (for 1 equivalent of protein) in CA_{CTD}^{W184A, M185A}-SP1-NC was carried out and allowed to record a ^1H - ^{15}N SOFAST-HMQC spectrum without significant CSP compared to the spectrum of the protein alone. This procedure was repeated 5 times, allowing 6 equivalents of EP39 to be added to the protein. Due to the low water solubility of EP39, a visible precipitate was observed in the NMR tube after adding 6 equivalents of EP39. Stop the addition of EP39 and collect the NMR data.

The chemical shift perturbations of 10 residues (CA-V²²¹, CA-L²³¹, SP1-A²³², SP1-E²³³, SP1-M²³⁵, SP1-S²³⁶, SP1-Q²³⁷, SP1-V²³⁸, SP1-T²³⁹, NC-T²⁵⁷) during the binding of EP39 to protein from 1 to 6 equivalents were followed. K_ds were calculated using one-site binding model. Non-linear fitting was carried out with Graphpad Prism software with the equation: $Y = B_{\text{max}} * X / (K_d + X) + NS * X$. X is the ligand concentration, Y is the chemical shift perturbation, B_{max} is the maximum specific binding in the same units as Y. K_d is the equilibrium binding constant, in the same units as X. NS is the slope of non-specific binding in Y units divided by X units.

4.4. NMR experiments

NMR spectra of wild type CA_{CTD}-SP1-NC alone (300 μM) and mutated CA_{CTD}^{W184A, M185A}-SP1-NC (300 μM) in the absence and presence of EP39 were recorded in acetate buffer (25 mM sodium acetate, pH 6.5, 25 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM 2-mercaptoethanol, 10% D₂O) at 303 K on a Bruker Avance III 600 MHz spectrometer equipped with a cryogenic probe. The CSPs of the amide moiety were normalized according to the equation: $\Delta\delta = [(\Delta\delta_{\text{N}}/6.5)^2 + (\Delta\delta_{\text{HN}})^2]^{1/2}$ where $\Delta\delta_{\text{N}}$ and $\Delta\delta_{\text{HN}}$ are the chemical shift changes in the nitrogen and proton dimensions respectively [26]. ^1H and ^{15}N resonance assignments were carried out using the following set of spectra ^1H - ^{15}N SOFAST-HMQC [27], NOESY-HSQC [28], TOCSY-HSQC [29], HNHA [30]. Spectra were analyzed with CCPNMR Analysis 2.4.2 [31].

4.5. Structure determination

The NMR structures were calculated with ARIA2.3.2 software [14] coupled to CNS 1.2.1 [32] based on the inter-proton distance restraints from NOESY spectrum and the dihedral angle restraints from TALOS+ program [33]. Two zinc coordination were added (C²⁶⁰-C²⁶³-H²⁶⁸-C²⁷³ and C²⁸¹-C²⁸⁴-H²⁸⁹-C²⁹⁴) for structure calculation for NC. ARIA was run with explicit water refinement using standard protocols. The ten lowest total energy conformers were selected out of the 500 structures calculated in the final run. The average RMSD was calculated on the backbone atoms between the best structure which has the lowest energy, and each of the remaining nine structures on the first, second and third domain respectively.

The coordinates, the chemical shifts, the constraints and the list of peaks have been deposited in the Protein Data Bank under accession code 6RWG.

4.6. ^{15}N relaxation experiments

The ^{15}N backbone relaxation experiments were performed on 300 μM CA_{CTD}^{W184A, M185A}-SP1-NC in the absence and presence of 6 equivalents of EP39. T1, T2 and HetNOE ^{15}N experiments used

standard Bruker pulse sequences [34]. The T1 experiments were recorded with 5, 10, 20, 30, 40 (repeated), 60, 80, 100 (repeated), 200, 300, 400, 600, 800 and 1200 ms for recovery delay. T2 experiments were recorded with 1, 2, 4, 8, 10 (repeated), 12, 14, 16, 18, 20 (repeated), 22, 24, 28 and 32 ms for recovery delay. The 15N-{1H} NOE (HetNOE) values were taken as the ratio between the intensities recorded with and without saturation of the amide protons. The experiments were processed with Dynamic Center 2.3 in Topspin 3.5 pl 7. Select T1 fit function $f(t) = I_0 * [e^{-t/T1}]$, T2 fit function $f(t) = I_0 * [e^{-t/T2}]$, NOE: intensity ratio ($I_{\text{sat}}/I_{\text{un-sat}}$).

Accession codes

PDB code for CA_{CTD}^{W184A, M185A}-SP1-NC in the presence of EP39 is 6RWG.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully acknowledge Michael F. Summers for kindly providing us the plasmids of wild type CA_{CTD}-SP1-NC and mutated CA_{CTD}^{W184A, M185A}-SP1-NC. Xiaowei CHEN was granted by China Scholarship Council [No 201603250053]. This work was supported by CNRS and the University of Paris.

Abbreviations

NMR	nuclear magnetic resonance
HIV-1	human immunodeficiency virus type 1
MA	matrix
CA	capsid
SP1	spacer peptide 1
NC	nucleocapsid
RNP	ribonucleoprotein
SP2	spacer peptide 2
TFE	trifluoroethanol
BVM	bevirimat
MI	maturation inhibitor
CSP	chemical shift perturbations
MHR	major homology region
ZF	zinc finger
RMSD	Root mean square deviation

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmech.2020.112634>.

References

- [1] K. Wieggers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, H.-G. Kräusslich, Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites, *J. Virol.* 72 (1998) 2846–2854.
- [2] L.V. Coren, J.A. Thomas, E. Chertova, R.C. Sowder, T.D. Gagliardi, R.J. Gorelick, D.E. Ott, Mutational Analysis of the C-terminal Gag cleavage sites in human immunodeficiency virus type 1, *J. Virol.* 81 (2007) 10047–10054.
- [3] S.-K. Lee, M. Potempa, R. Swanstrom, The choreography of HIV-1 proteolytic processing and virion assembly, *J. Biol. Chem.* 287 (2012) 40867–40874.
- [4] J.M. Wagner, K.K. Zadrozny, J. Chrustowicz, M.D. Purdy, M. Yeager, B.K. Ganser-Pornillos, O. Pornillos, Crystal structure of an HIV assembly and maturation switch, *ELife* 5 (2016), e17063.
- [5] M. Wang, C.M. Quinn, J.R. Perilla, H. Zhang, R. Shirra Jr., G. Hou, I.-J. Byeon,

- C.L. Suiter, S. Ablan, E. Urano, T.J. Nitz, C. Aiken, E.O. Freed, P. Zhang, K. Schulten, A.M. Gronenborn, T. Polenova, Quenching protein dynamics interferes with HIV capsid maturation, *Nat. Commun.* 8 (2017) 1779.
- [6] T.A.M. Bharat, L.R.C. Menendez, W.J.H. Hagen, V. Lux, S. Igonet, M. Schorb, F.K.M. Schur, H.-G. Kräusslich, J.A.G. Briggs, Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly, *Proc. Natl. Acad. Sci.* 111 (2014) 8233–8238.
- [7] J.L. Newman, E.W. Butcher, D.T. Patel, Y. Mikhaylenko, M.F. Summers, Flexibility in the P2 domain of the HIV-1 Gag polyprotein, *Protein Sci. Publ. Protein Soc.* 13 (2004) 2101–2107.
- [8] F. Li, R. Goila-Gaur, K. Salzwedel, N.R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D.E. Martin, J.M. Orenstein, G.P. Allaway, E.O. Freed, C.T. Wild, PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing, *Proc. Natl. Acad. Sci.* 100 (2003) 13555–13560.
- [9] M.D. Purdy, D. Shi, J. Chrustowicz, J. Hattne, T. Gonen, M. Yeager, MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat, *Proc. Natl. Acad. Sci.* 115 (2018) 13258–13263.
- [10] P. Coric, S. Turcaud, F. Souquet, L. Briant, B. Gay, J. Royer, N. Chazal, S. Bouaziz, Synthesis and biological evaluation of a new derivative of bevirimat that targets the Gag CA-SP1 cleavage site, *Eur. J. Med. Chem.* 62 (2013) 453–465.
- [11] A. Neyret, B. Gay, A. Cransac, L. Briant, P. Coric, S. Turcaud, P. Laugaa, S. Bouaziz, N. Chazal, Insight into the mechanism of action of EP-39, a bevirimat derivative that inhibits HIV-1 maturation, *Antivir. Res.* 164 (2019) 162–175.
- [12] L. Deshmukh, C.D. Schwieters, A. Grishaev, R. Ghirlando, J.L. Baber, G.M. Clore, Structure and dynamics of full length HIV-1 capsid protein in solution, *J. Am. Chem. Soc.* 135 (2013) 16133–16147.
- [13] W. Lu, K. Salzwedel, D. Wang, S. Chakravarty, E.O. Freed, C.T. Wild, F. Li, A single polymorphism in HIV-1 subtype C SP1 is sufficient to confer natural resistance to the maturation inhibitor bevirimat ∇ , *Antimicrob. Agents Chemother.* 55 (2011) 3324–3329.
- [14] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, ARIA2: automated NOE assignment and data integration in NMR structure calculation, *Bioinformatics* 23 (2007) 381–382.
- [15] L. Zargarian, C. Tisné, P. Barraud, X. Xu, N. Morellet, B. René, Y. Mély, P. Fossé, O. Mauffret, Dynamics of linker residues modulate the nucleic acid binding properties of the HIV-1 nucleocapsid protein zinc fingers, *PLoS One* 9 (2014), e102150.
- [16] N. Morellet, S. Druillennec, C. Lenoir, S. Bouaziz, B.P. Roques, Helical structure determined by NMR of the HIV-1 (345–392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging, *Protein Sci. Publ. Protein Soc.* 14 (2005) 375–386.
- [17] P.W. Keller, C.S. Adamson, J.B. Heymann, E.O. Freed, A.C. Steven, HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice, *J. Virol.* 85 (2011) 1420–1428.
- [18] Y. Han, G. Hou, C.L. Suiter, J. Ahn, I.-J.L. Byeon, A.S. Lipton, S. Burton, I. Hung, P.L. Gor'kov, Z. Gan, W. Brey, D. Rice, A.M. Gronenborn, T. Polenova, Magic angle spinning NMR reveals sequence-dependent structural plasticity, dynamics, and the spacer peptide 1 conformation in HIV-1 capsid protein assemblies, *J. Am. Chem. Soc.* 135 (2013) 17793–17803.
- [19] E. Urano, U. Timilsina, J.A. Kaplan, S. Ablan, D. Ghimire, P. Pham, N. Kuruppu, R. Mandt, S.R. Durell, T.J. Nitz, D.E. Martin, C.T. Wild, R. Gaur, E.O. Freed, Resistance to second-generation HIV-1 maturation inhibitors, *J. Virol.* 93 (2019) e02017–e2018.
- [20] S. Gupta, J.M. Louis, R. Tycko, Effects of an HIV-1 maturation inhibitor on the structure and dynamics of CA-SP1 junction helices in virus-like particles, *Proc. Natl. Acad. Sci.* 117 (2020) 10286–10293.
- [21] U.K. von Schwedler, K.M. Stray, J.E. Garrus, W.I. Sundquist, Functional surfaces of the human immunodeficiency virus type 1 capsid protein, *J. Virol.* 77 (2003) 5439–5450.
- [22] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ -RNA packaging signal. implications for genome recognition 11Edited by P. Wright, *J. Mol. Biol.* 301 (2000) 491–511.
- [23] R.N.D. Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ -RNA recognition element, *Science* 279 (1998) 384–388.
- [24] S.A.K. Datta, L.G. Temeselew, R.M. Crist, F. Soheilian, A. Kamata, J. Mirro, D. Harvin, K. Nagashima, R.E. Cachau, A. Rein, On the role of the SP1 domain in HIV-1 particle assembly: a molecular Switch? ∇ , *J. Virol.* 85 (2011) 4111–4121.
- [26] F.A.A. Mulder, D. Schipper, R. Bott, R. Boelens, Altered flexibility in the substrate-binding site of related native and engineered high-alkaline *Bacillus subtilis*11Edited by P. E. Wright, *J. Mol. Biol.* 292 (1999) 111–123.
- [27] P. Schanda, E. Kupce, B. Brutscher, SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, *J. Biomol. NMR* 33 (2005) 199–211.
- [28] O. Zhang, L.E. Kay, J.P. Olivier, J.D. Forman-Kay, Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques, *J. Biomol. NMR* 4 (1994) 845–858.
- [29] D. Marion, P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn, G.M. Clore, Overcoming the overlap problem in the assignment of ^1H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^1H - ^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta, *Biochemistry* 28 (1989) 6150–6156.
- [30] G.W. Vuister, A. Bax, Measurement of four-bond HN-H alpha J-couplings in staphylococcal nuclease, *J. Biomol. NMR* 4 (1994) 193–200.
- [31] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins Struct. Funct. Bioinf.* 59 (2005) 687–696.
- [32] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54 (1998) 905–921.
- [33] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR* 44 (2009) 213–223.
- [34] L.E. Kay, D.A. Torchia, A. Bax, Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry* 28 (1989) 8972–8979.

Résumé

Le traitement du précurseur Gag du VIH-1 (MA-CA-SP1-NC-SP2-P6) par la protéase virale, après bourgeonnement de la particule virale, permet la maturation du virion, étape essentielle et ultime du cycle de réplication. Le clivage final se produit entre la capsid (CA) et le peptide espaceur 1 (SP1). Le bevirimat (BVM), premier inhibiteur de maturation du VIH-1, bloque le clivage de la jonction CA-SP1. Son mécanisme d'action n'est pas clair. Dans ce travail, la structure en solution d'un domaine muté de Gag contenant 157 acides aminés, CACTDW184A, M185A-SP1-NC, englobant le domaine C-ter de CA, SP1 et la nucléocapsid, NCp7, a été déterminée par RMN en présence d'EP39, un dérivé découvert dans notre laboratoire, plus soluble que le BVM. Les propriétés structurales et dynamiques de la protéine en l'absence et en présence d'EP39 ont été caractérisées. Les résultats montrent que la liaison de EP39 diminue la dynamique de la jonction CA-SP1, en particulier du motif QVT de SP1 reconnu par EP39 et perturbe l'équilibre hélice-coil naturel aux deux extrémités du domaine SP1. Pour comprendre le mécanisme d'action des inhibiteurs de maturation du VIH, nous avons étudié par RMN l'effet d'une autre molécule, le PF46396 dérivé de la pyridone, bloquant également le clivage entre CA et SP1. Les résultats montrent que, bien que EP39 et PF46396 aient des structures différentes et des affinités différentes pour la protéine, leur mode d'action par stabilisation du domaine SP1 en hélice est comparable. Nous pensons que l'interaction hydrophobe entre les inhibiteurs et le motif QVT de la jonction CA-SP1 est un déterminant majeur pour stabiliser la structure hélicoïdale de la jonction CA-SP1. Cet effet empêche l'accès de la protéase virale au site de clivage et finalement inhibe la maturation de la particule. Nos résultats fournissent de nouvelles informations sur la structure et la dynamique du domaine SP1 et suggèrent un mécanisme par lequel les inhibiteurs de maturation interfèrent avec le traitement final des particules virales. Ils ouvrent une nouvelle voie pour le développement de nouveaux inhibiteurs ciblant la maturation du VIH-1.

Mots-clés

VIH-1, Gag, maturation, BVM, EP39, PF46396, inhibiteur de maturation, RMN, dynamique des protéines

Abstract

Treatment of the HIV-1 Gag precursor, MA-CA-SP1-NC-SP2-p6, by the viral protease after release of the viral particle allows the virion to mature, an essential step in the replication cycle. The final cleavage occurs between the capsid (CA) and the spacer peptide 1 (SP1). The bevirimat (BVM), the first HIV-1 maturation inhibitor, blocks the cleavage of the CA-SP1 junction. The mechanism by which BVM inhibits the maturation of HIV-1 is unclear. In this work, the solution structure of a mutated domain of Gag, CA_{CTD}^{W184A, M185A}-SP1-NC encompassing the C-ter of CA, SP1 and the nucleocapsid NCp7 (157 aa) was determined by NMR in the presence of EP39, a more soluble BVM derivative discovered in our laboratory. The structural and dynamic properties of the protein in the absence and in the presence of EP39 were characterized by NMR. The results show that the binding of EP39 decreases the dynamics of the CA-SP1 junction, in particular the QVT motif in SP1, recognized first by EP39, and disturbs the natural coil-helix equilibrium on both ends of the SP1 domain. To understand the mechanism of action of HIV maturation inhibitors, we studied the effect of PF46396, a pyridone-based compound, blocking the cleavage between CA and SP1 and inhibiting the maturation by NMR. The results show that, although EP39 and PF46396 are different in structure and have different affinities for the protein, they have the same effect in stabilizing the SP1 domain in helix. We believe that the hydrophobic interaction between the inhibitors and the CA-SP1 junction is the major determinant for stabilizing the helical structure of the CA-SP1 junction, which disrupts access of the viral protease to the cleavage site and ultimately inhibits the maturation of HIV-1 particle. Our results provide new information on the structure and dynamics of the SP1 domain and suggest a mechanism by which maturation inhibitors interfere with the final processing of HIV-1 particles. They offer additional clues for the development of new inhibitors targeting the maturation of HIV-1.

Keywords

HIV-1, Gag, maturation, BVM, EP39, PF46396, maturation inhibitor, NMR, protein dynamic