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

Xiaowei Chen, Pascale Coric, Valéry Larue, Serge Turcaud, Xiao Wang, Sylvie Nonin-Lecomte, Serge Bouaziz

▶ **To cite this version:**

Xiaowei Chen, Pascale Coric, Valéry Larue, Serge Turcaud, Xiao Wang, et al.. The HIV-1 maturation inhibitor, EP39, interferes with the dynamic helix-coil equilibrium of the CA-SP1 junction of Gag. European Journal of Medicinal Chemistry, 2020, 204, pp.112634. 10.1016/j.ejmech.2020.112634 . hal-02914317

HAL Id: hal-02914317
https://hal.science/hal-02914317

Submitted on 11 Aug 2020

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.

Revised Manuscript Click here to view linked References

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.*

***Corresponding Author E-mail:** serge.bouaziz@parisdescartes.fr

Address correspondence to Serge Bouaziz, CiTCoM, CNRS, UMR 8038, Université de Paris, 4 Avenue de l'Observatoire, Paris, 75270, France.

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.

Key words: HIV-1, maturation inhibitor, EP39, CA-SP1 junction, dynamics,
NMR

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 CA^{226} to NC^{248}. Thus, proper folding of

SP1 could depend on the presence of the NC domain [7].

Bevirimat (BVM) (Fig 1*A*), 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 1*B*), 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 1*C*). CA_{CTD}-SP1-NC is in monomer-dimer equilibrium [7]. In $CA_{CTD}^{W184A, M185A}$-SP1-NC, the 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}SP1NC that dimerizes. NC is present because we believe it can ensure the stable folding of SP1.

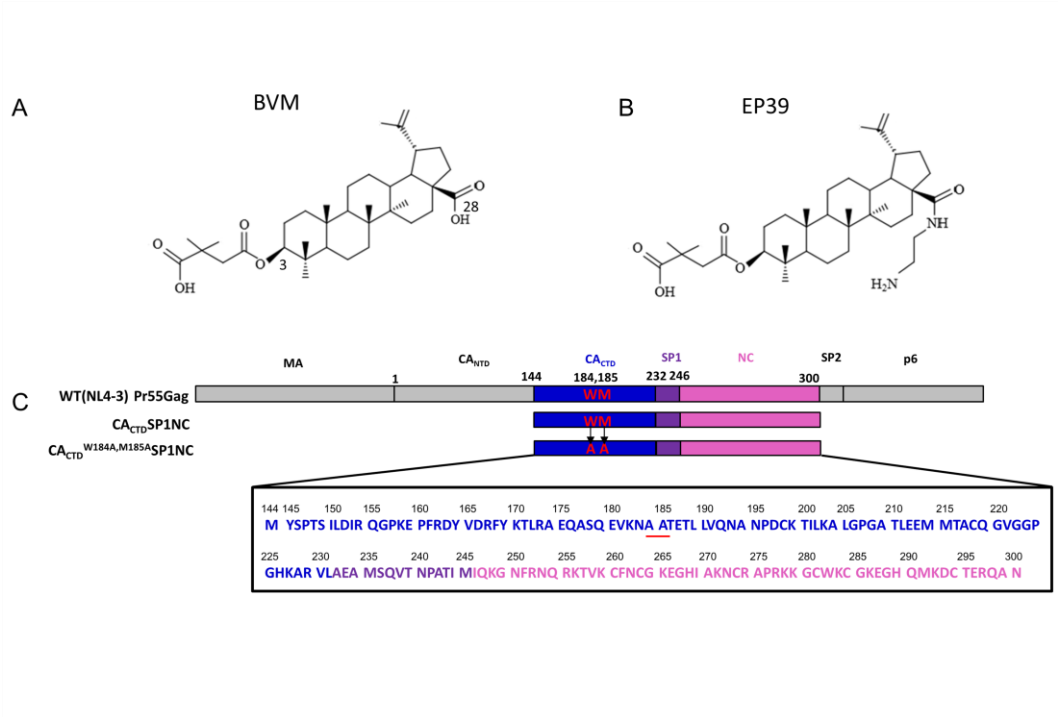


Figure 1. Schematic representations of BVM (A), EP39 (B) and the Gag polyprotein (C). EP39 retains the C3 function of BVM which is essential for the maturation inhibitory effect

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.

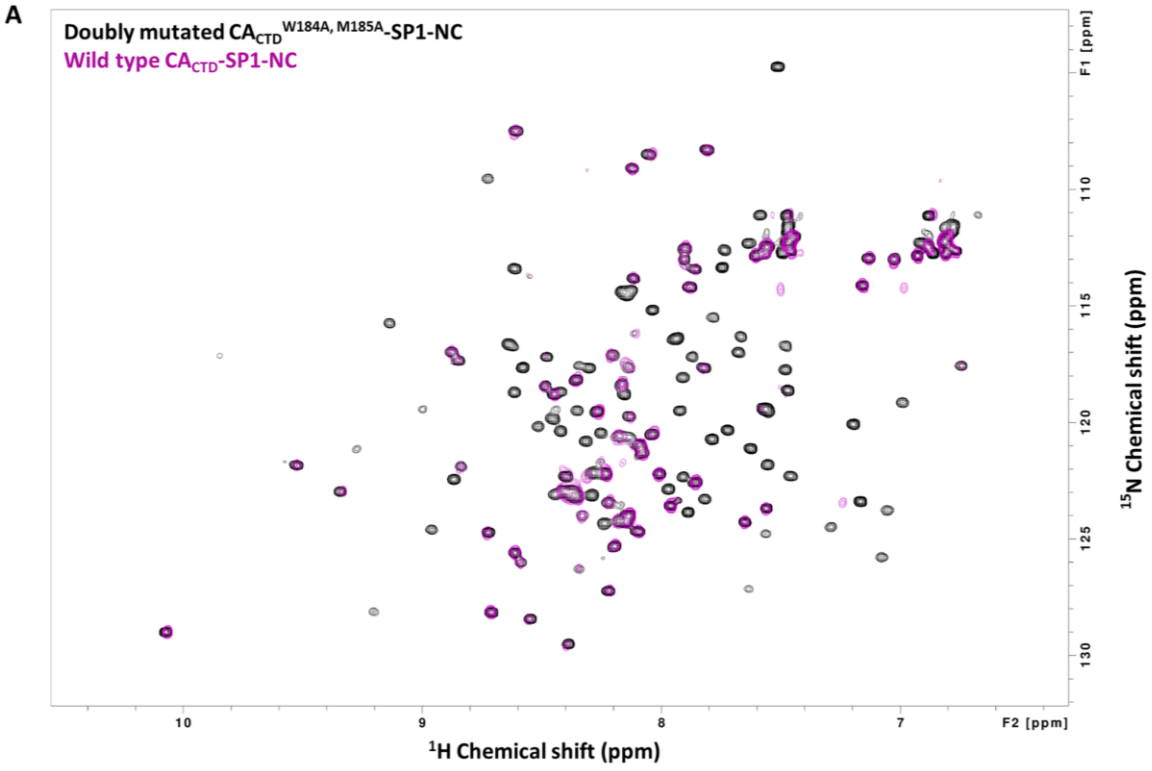
2. Results

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

Comparison between ^{1}H-^{15}N SOFAST-HMQC spectra obtained for the wild type CA_{CTD}-SP1-NC and doubly mutated $CA_{CTD}^{W184A, M185A}$-SP1-NC (Fig 2*A*) shows that their resonances superimposed well, especially the resonance of NC (by reference to the resonances labeled with residue number in Fig 3*A*). 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 2*A*) [7].

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

Comparison of the ^{1}H-^{15}N SOFAST-HMQC spectra of ^{15}N-NC and ^{15}N $CA_{CTD}^{W184A, M185A}$-SP1-NC (Fig 2*B*) 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, $CA_{CTDW184A, M185A}$ or SP1.


A
Doubly mutated CA_CTD W184A, M185A-SP1-NC
Wild type CA_CTD-SP1-NC
F1 [ppm]
110
115
120
125
130
15N Chemical shift (ppm)
10
9
8
7
F2 [ppm]
1H Chemical shift (ppm)

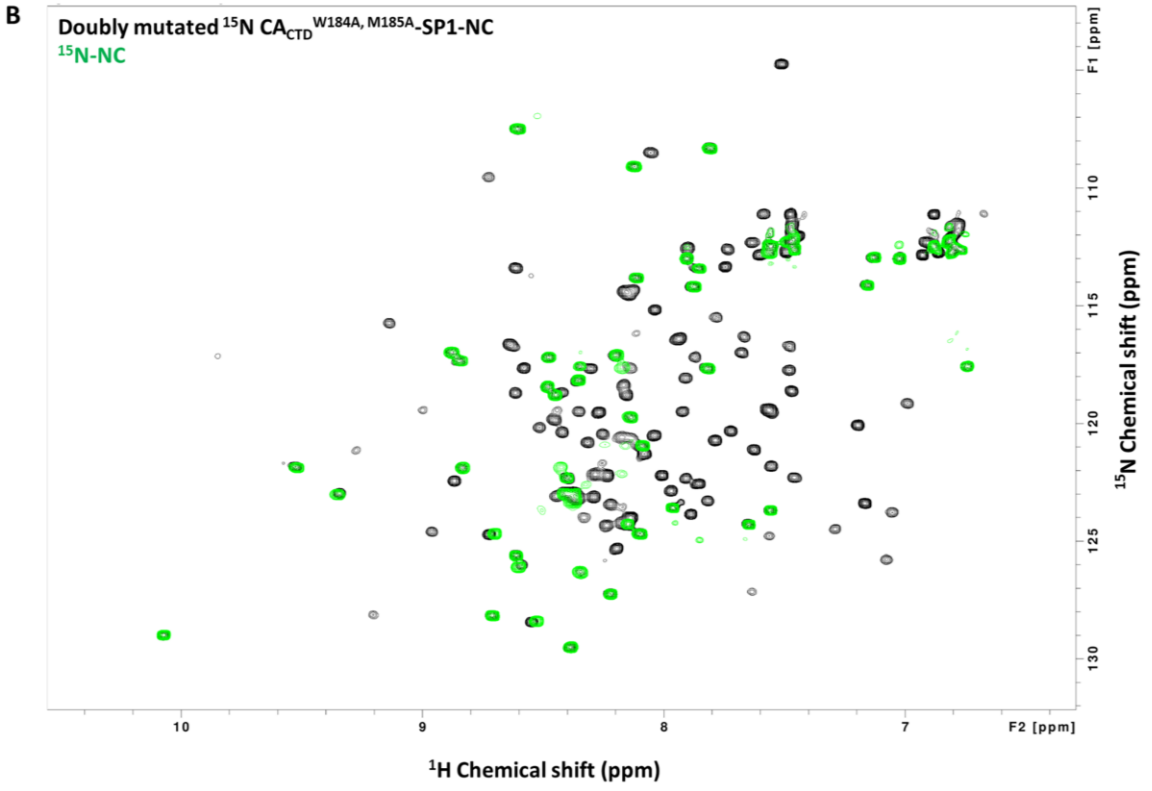

B
Doubly mutated 15N CA_CTD W184A, M185A-SP1-NC
15N-NC
F1 [ppm]
110
115
120
125
130
15N Chemical shift (ppm)
10
9
8
7
F2 [ppm]
1H Chemical shift (ppm)

Figure 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 ^{1}H-^{15}N SOFASTHMQC 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 ^{1}H-^{15}N SOFAST-HMQC spectra of the isolated ^{15}N labeled NC (green spectrum) and doubly mutated ^{15}N-$CA_{CTD}^{W184A, M185A}$-SP1NC (black spectrum).

2.3. EP39 stabilizes the SP1 domain on both sides

^{1}H-^{15}N SOFAST-HMQC spectra were recorded on the monomer $CA_{CTD}^{W184A, M185A}$-SP1-NC in the absence and in the presence of EP39 (Fig 3*A*). In the absence of EP39, the amide resonances of residues CA^{218} to NC^{257} (encompassing the fourteen C-terminal residues of CA, SP1 and the twelve Nterminal 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 3*A* 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 (CA^{218}-NC^{257}) (Fig 3*B*) 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-T^{210} 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-K^{158} (in major homology region MHR) and CA-G^{222} 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-K^{158} [11].

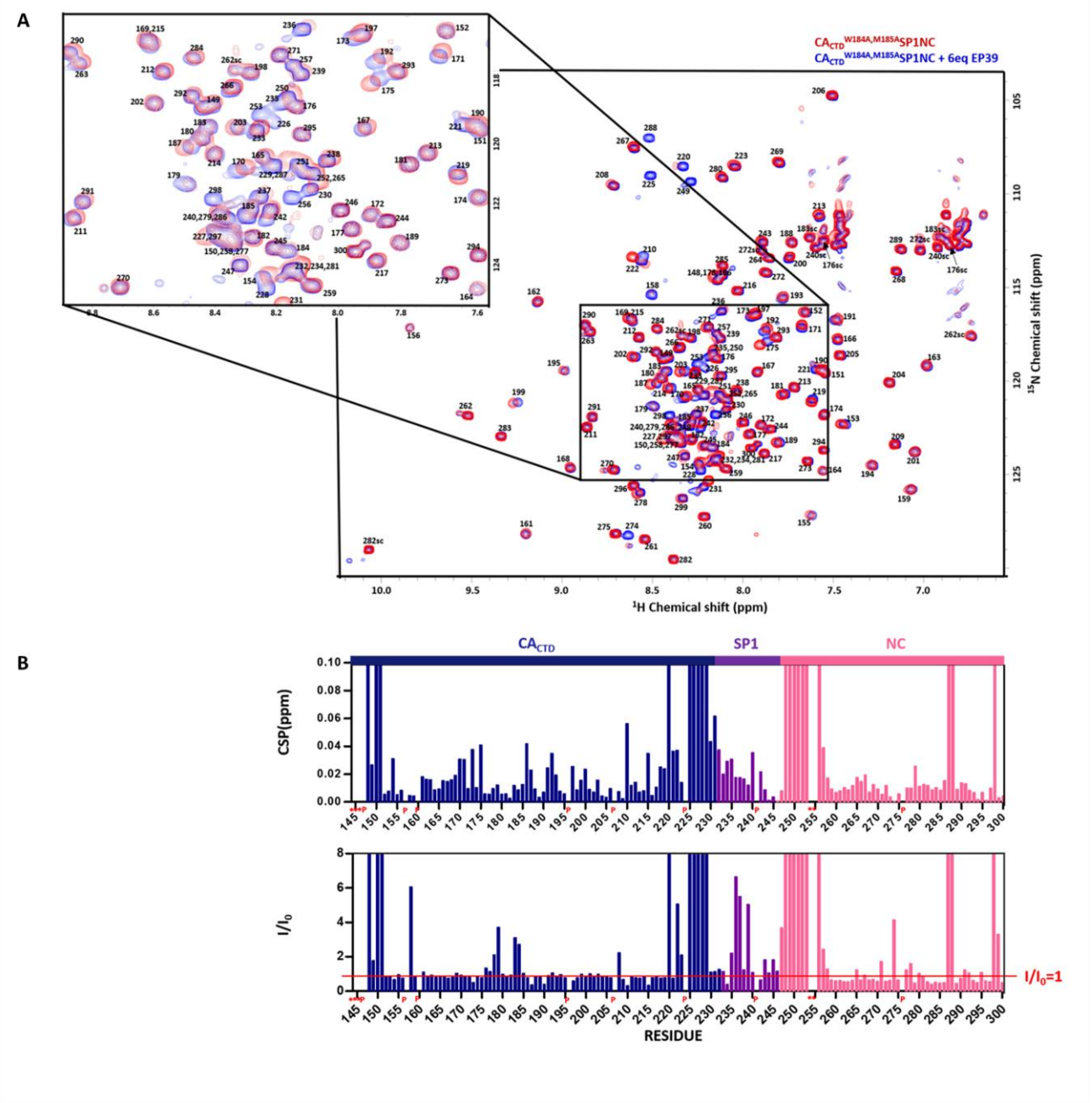


Figure 3. Solution NMR of EP39 binding to $CA_{CTD}^{W184A, M185A}$-SP1-NC. (*A*) Superposition of ^{1}H-^{15}N SOFAST-HMQC spectra recorded on ^{15}N labeled $CA_{CTD}^{W184A, M185A}$-SP1-NC in the absence (red) and in the presence of 6 equivalents of EP39 (blue) at 303K.

Resonances are labeled with amino acids numbers. (*B*) Chemical shift perturbations (CSP) and resonance intensity ratios (I/I_0) of the amide groups ($^{1}H/^{15}N$ combined chemical shift) in the presence and in the absence of EP39. The chemical shift perturbations and resonance intensity ratios of the residues whose ^{1}H-^{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 are colored as Fig 1*C.*

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 isolated in the spectrum. 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. In these ten residues, SP1-T^{239} has the highest binding affinity (52.4 μM); SP1-S^{236}, SP1-Q^{237} and SP1-V^{238} have relatively high binding affinities (179.6 μM, 154,5 μM and 204.4 μM respectively); SP1-E^{233}, SP1-M^{235} and NC-T^{257} have medium binding affinities (237.0 μM, 235,4 μM, 308.4 μM respectively); CA-V^{221}, CA-L^{231} and SP1-A^{232} have weak binding affinities (669.6 μM, 678.4 μM and 691.9 μM respectively) (Fig 4). The higher binding affinity at SP1-$Q^{237}V^{238}T^{239}$ is consistent with previous study which shows that SP1-$Q^{237}V^{238}T^{239}$ motif polymorphism confers intrinsic resistance to BVM [13].

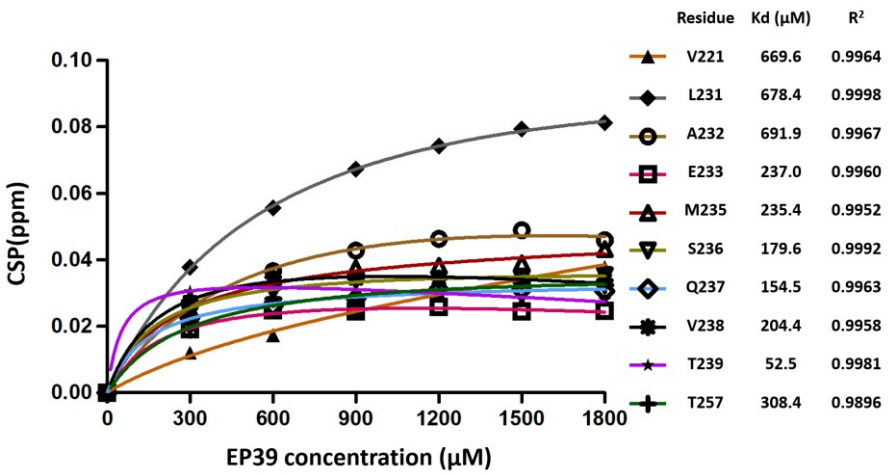


Figure 4. Variations of chemical shift perturbations (CSP) of ten residues (CAV^{221}, CA-L^{231}, SP1-A^{232}, SP1-E^{233}, SP1-M^{235}, SP1-S^{236}, SP1-Q^{237}, SP1-V^{238}, SP1-T^{239}, NC-T^{257}) isolated in spectrum as a function of EP39 whose concentration increase from one to six equivalents of protein. Kds were computed as described in Materials and Methods.

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

The structure of the monomeric mutant $CA_{CTD}^{W184A,M185A}$-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 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}-$SP1_{239}$) and H6 (SP1^{241}-NC^{251}); two zinc fingers: ZF1 (NC$^{260-273}$) and ZF2 (NC$^{281-294}$) connected by seven amino acids (Fig 5*A*).

Table 1. Statistics for the top 10 NMR structures of monomeric mutant

HIV-1 $CA_{CTDW184A, M185A}$-SP1-NC

Restraints

Total number of NMR restraints	2121
Total number of NOE distance restraints	1898
Long range (|i-j| > 5)	90
Medium range (2≤|i-j|) ≤ 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

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}-NC^{251}, 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 5*B*). 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 5*B*). Six amino acids, $CA^{221-226}$, starting the second domain are very flexible and display different orientations (middle Fig 5*B*). A flexible elbow between H5 and H6, corresponding to the QVT motif, allows these two helices to adopt different relative orientations. (middle Fig 5*B*). 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 5*B*).

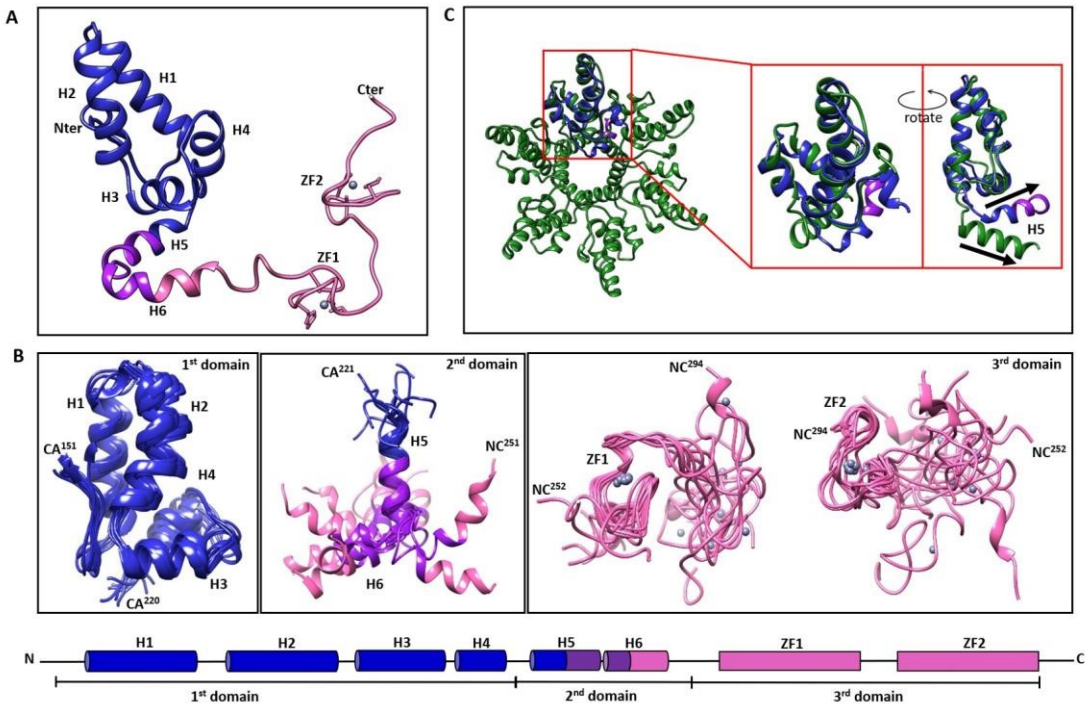


Figure 5. NMR structure of $CA_{CTD}^{W184A,\ M185A}$-SP1-NC in the presence of EP39. The secondary structure of the protein is indicated at the bottom. H: helix; ZF: zinc finger. Residues are colored as in Fig 1*C*. *(A*) the structure which is closest to the average structure is represented. From N-terminal 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}-$SP1^{239}$), H6 (residue $SP1^{241}$-NC^{251}) and two zinc fingers: ZF1 (residue $NC^{260-273}$) and ZF2 (residue $NC^{281-294}$). (*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, $CA^{151-220}$, shows a RMSD of 1.06 Å calculated on the backbone atoms. Middle: Superimposition on the second domain, CA^{221}-NC^{251}, 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}-

SP1^{238} between the NMR structure (Fig.5*A*) 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.

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 CA^{148}-SP1^{239} backbone atoms between our structure and the chain G of the hexamer is 1.15 Å (Fig 5*C*). 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$^{221\text{-}226}$) relative to H4 (CA$^{212\text{-}218}$) between our structure and the hexameric one as shown in Fig 5*C*, demonstrating again the flexibility of the residues between H4 and H5. H6 which is missing in the crystal and cryoEM structures is well folded in our NMR structure (Fig 5*A*) and adopts different orientations compared to H5.

2.7. EP39 stabilizes the H5 and H6 helices of the second domain CA^{221}-NC^{251} ^{15}N 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-amplitude motions.

In the absence of EP39, the N- and C-terminal parts of $CA_{CTD}^{W184A, M185A}$-SP1NC 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 ^{1}H-^{15}N 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 (CA^{225}, CA^{226}, CA^{227}, CA^{228}, CA^{229}, NC^{248}, NC^{249}, NC^{250}, NC^{251}) of the second domain appear in $^{1}H^{15}N$ 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-Q^{237}, SP1-V^{238} and SP1-T^{239}, 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-T^{239} with lowest HetNOE value. After adding EP39, HetNOE value of SP1-T^{239} 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.

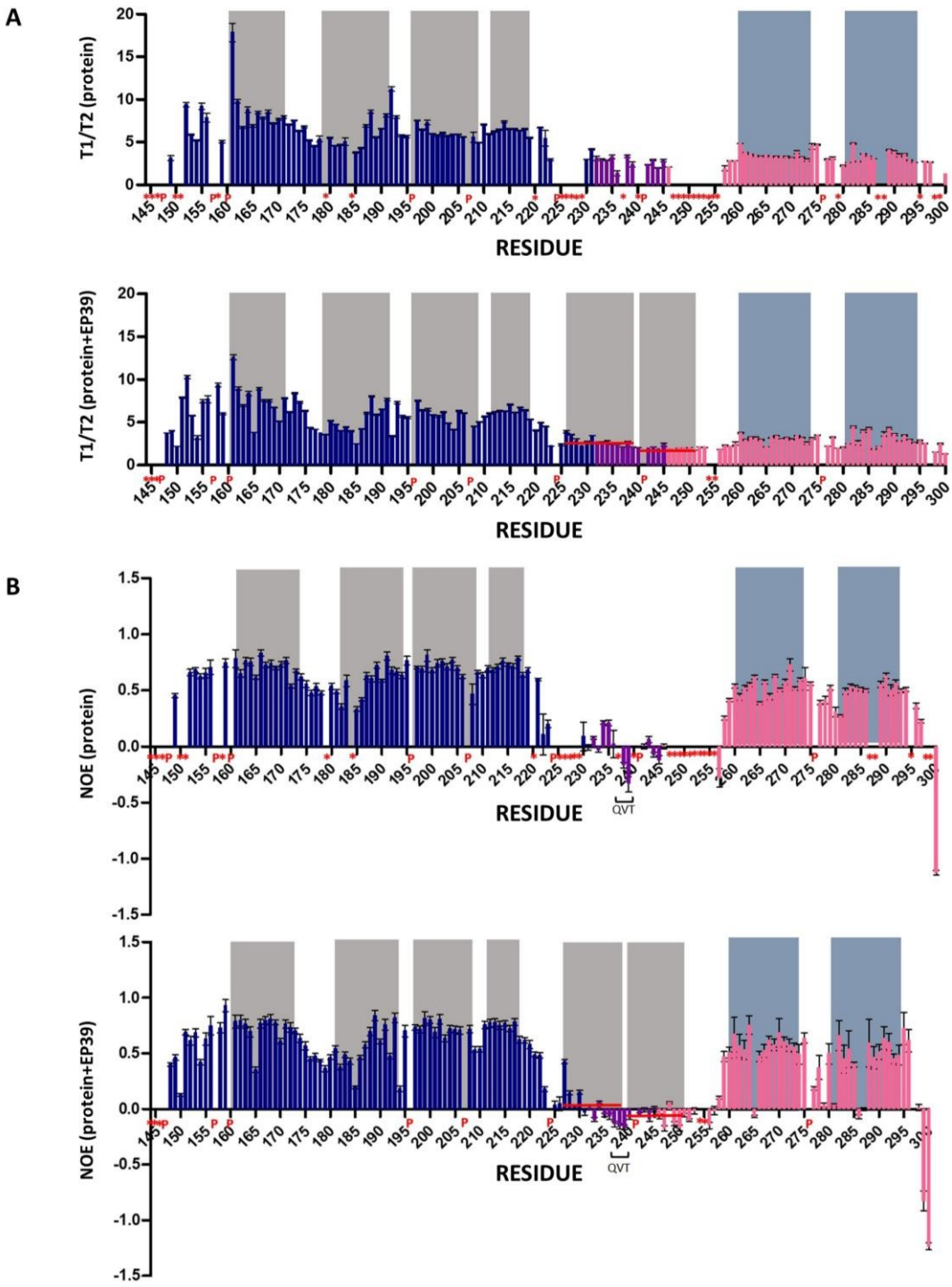


Figure 6. Dynamic characterization of $CA_{CTD}^{W184A,\ M185A}$-SP1-NC in the absence and presence of EP39. Experimental NMR ^{15}N relaxation data T1/T2 (*A*) and HetNOE (*B*) values were obtained at 600 MHz for protein in the absence

(top) and presence (bottom) of EP39 at 303K. 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 1*C*. 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-$Q^{237}V^{238}T^{239}$ are marked in the HetNOE value.

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 CASP1 junction encompassing the C-terminal nine residues of CA and the Nterminal 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 αhelix [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 $CA_{CTD}^{W184A, M185A}$-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.

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 CA^{221}-NC^{251} 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 CA^{221}-NC^{251} cannot be detected in ^{1}H-^{15}N SOFASTHMQC spectrum in the absence of EP39 (Fig 3*A*), 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 3*B*) 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-$Q^{237}V^{238}T^{239}$), 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 CA^{226}-$SP1^{237}$ [5]. In this work, residues SP1-$Q^{237}V^{238}T^{239}$ are the most dynamic amino acids in SP1 and have high binding affinities to EP39 compared to other residues (Fig 4), especially residue SP1-T^{239} 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 CASP1 junction starts at CA^{225} and stop at $SP1^{238}$ [9]. As we have shown by our relaxation experiments, H6 is more dynamic than H5 (Fig 6). The residue $SP1T^{239}$ 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-V^{238} 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-A^{232} (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 CASP1 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-K^{158} 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 (K^{158} and K^{227}) [9]. These results can easily be extrapolated to demonstrate the essential role of CA-K^{158} 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-K^{158} in MHR is much stronger after EP39 addition (Fig 3*B*), which means this residue could be an essential target for EP39 binding.

The SP1-NC junction (residue $SP1^{241}$-NC^{251}) 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 [21,22]. In 30% TFE

solution, the peptide $CA^{211-231}SP1^{232-245}NC^{246-258}$ adopts a helical structure spanning domain CA^{226}-NC^{248} [16]. In the present work, the α-helix in the second domain starts from CA^{221} and stops at NC^{251} with a kink at $SP1^{240}$ 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 makes 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 $CA_{CTDW184A, M185A}$-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 [25].

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 $ZnCl_2$, 0.1 mM 2-mercaptoethanol, 10% D_2O) in a 3 mm NMR tube. A ^{1}H-^{15}N SOFASTHMQC 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 ^{1}H-^{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-V^{221}, CA-L^{231}, SP1-A^{232}, SP1-E^{233}, SP1-M^{235}, SP1-S^{236}, SP1-Q^{237}, SP1-V^{238}, SP1-T^{239}, NC-T^{257}) during the binding of EP39 to protein from 1 to 6 equivalents were followed. Kds were calculated using one-site binding model. Non-linear fitting was carried out with Graphpad Prism software with the equation: Y=Bmax*X/(Kd+X) + NS*X. X is the ligand concentration, Y is the chemical shift perturbation, Bmax is the maximum specific binding in the same units as Y. 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.

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 $ZnCl_2$, 0.1 mM 2-mercaptoethanol, 10% D_2O) 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: $\Delta\delta = [(\Delta\delta_N/6.5)^2 + (\Delta\delta_{HN})^2]^{1/2}$ where $\Delta\delta_N$ and $\Delta\delta_{HN}$ are the chemical shift changes in the nitrogen and proton

dimensions respectively [26]. ^{1}H and ^{15}N resonance assignments were carried out using the following set of spectra ^{1}H-^{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 (C^{260}-C^{263}-H^{268}-C^{273} and C^{281}-C^{284}-H^{289}-C^{294}) 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

($I_{sat}/I_{un\text{-}sat}$).

5. Ancillary Information

Supplementary material

Additional figures illustrating the comparison of ^{1}H-^{15}N SOFAST-HMQC spectra recorded on $CA_{CTD}^{W184A, M185A}$-SP1-NC in the absence and presence of EP39

from 1 to 6 equivalents and the experimental T1 and T2 values for $CA_{CTD}^{W184A, M185A}$-SP1-NC in the absence and presence of EP39.

Accession Codes

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

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

Notes

The authors declare no competing interest.

References

[1] K. Wiegers, 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).

[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. 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.

[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, HIV1 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).

[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. (2020).

[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 ΨRNA 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.

[25] 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 ΨRNA Recognition Element, Science. 279 (1998) 384–388.

[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 subtilisins11Edited 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 15N 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-15N 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.

Bioinforma. 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. 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.

Supplementary Material - For Publication Online

Click here to access/download

Supplementary Material - For Publication Online

Supplementary material.pdf

*Declaration of Interest Statement

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: