Structure of the zinc finger domain encompassing residues 13–51 of the nucleocapsid protein from simian immunodeficiency virus

Nelly MORELLET^{*1}, Hervé MEUDAL[†], Serge BOUAZIZ^{*} and Bernard Pierre ROQUES^{*}

^{*}Unité de Pharmacologie Chimique et Génétique, INSERM U640 – CNRS UMR 8151, UFR des Sciences Pharmaceutiques et Biologiques, 4, avenue de l'Observatoire, 75270 Paris cedex 06, France, and [†]Centre de Biophysique Moléculaire, UPR 4301/CNRS Rue Charles-Sadron, 45071 Orléans cedex 2, France

The NCps (nucleocapsid proteins) of HIV-1 (HIV type 1), HIV-2 and SIV (simian immunodeficiency virus) are small highly basic proteins, characterized by the presence of two CCHC ZF (zinc finger) domains. NCps, closely associated with the dimeric RNA genome in the core of the virus particle, were shown to promote the specific encapsidation of the viral RNA and are implicated in reverse transcription. Solution structure of the HIV-1 NCp7 and complexes of NCp7 with RNA or DNA showed the critical relationships between the structure and its various functions. HIV-1 and HIV-2 have resulted respectively from transmissions of SIV from chimpanzees and sooty mangabeys. It has been shown that the SIVhoest (SIV from l'Hoeist monkeys) also has the potential to infect human populations. Since monkeys are of great interest for clinical studies of antiviral drugs, the structure of (13–51)NCp8 (zinc finger domain of NCp8, encompassing residues 13–51) from

SIVhoest was determined by NMR to appraise the influence of major differences in the sequence, since Glu²¹, Gly⁴³ and Met⁴⁶ in NCp7 are replaced by Pro, Glu and Phe respectively in this particular NCp8. The structure of (13–51)NCp8 is very well defined, and surprisingly the structure of each ZF is similar in NCp7 and NCp8. Moreover, contrary to NCp7, the two ZFs are strongly locked to each other in this NCp8. This first reported structure of a simian NCp8 compared with that of NCp7 shows that the main structural differences occur at the flexible linker between the two ZFs but the essential residues responsible for the interaction with oligonucleotides adopt the same orientation in the two proteins.

Key words: HIV type 1, NMR, nucleocapsid protein, simian immunodeficiency virus, SIVhoest, zinc finger domain.

INTRODUCTION

Simian immunodeficiency viruses (SIVs), which together with the HIVs [HIV-1 (HIV type 1) and HIV-2] constitute the primate lentivirus family, are harboured naturally in African primates. Currently, SIV infection has been detected in more than 20 different species of African primates. SIVs are morphologically and antigenically related to HIVs. Six distinct types of viruses from non-human primates have been molecularly characterized: SIVsm (SIV from sooty mangabeys) [1] including SIVmac (SIV from macaques) [2], SIVagm (SIV from the African green monkey) [3], SIVsyk (from Sykes monkeys) [4], SIVcpz (SIV from chimpanzees) [5], SIVcol (SIV from Colobus monkeys) [6] and the sixth SIV lineage, SIVhoest (SIV from l'Hoeist monkeys) [7,8], including SIVmnd (SIV from mandrills) [9]. HIV-1 has been traced to three independent transmissions of SIVcpz in the western part of central Africa [5], while HIV-2 has resulted from several transmissions of SIVsm in west Africa [10]. Two members of the sixth SIV lineage, SIVmnd and SIVhoest, were first isolated from animals held in captivity [7,9], and so it is possible that one (or both) species have been the recipient of a recent cross-species transmission. Moreover, it has been shown that l'Hoeist monkeys are the natural reservoir for SIVhoest [8]. The ability of SIVhoest to infect human PBMC (peripheral blood mononuclear cells) *in vitro* at least as efficiently as it infects macaque PBMC suggests that this virus could infect human populations [7].

All HIV and SIV retroviruses encode the canonical retroviral Gag, Pol and Env proteins. Cleavage of Gag precursor produces four major viral proteins: MA (matrix), CA (capsid), NC (nucleocapsid) and the C-terminal protein p6. The functions of the HIV and SIV NCps (NC proteins) are of particular interest because of their vital role in viral replication [11] and their potential as targets for viral inactivation [12–15]. Retroviral NCps are small basic proteins containing two conserved 14-residue peptide segments, -CX₂CX₄HX₄C-, (CCHC), that bind zinc ions, forming structures referred to as ZFs (zinc fingers). The biological roles of retroviral NCps and their ZFs have been reviewed [11,16] and include functions during viral assembly, maturation and infection. NMR studies have shown that NCp7 from HIV-1 presents a folded structure encompassing the two ZFs [17,18]. In contrast, the N- and C-terminal domains of NCp7 are flexible and unstructured [18,19]. Until now no studies have been devoted to the structure of the ZF domain in SIV. These viruses are presently used in many *in vivo* models of HIV-1 disease and in vaccination assays [20–22] and the retroviral NCp is a recognized target for therapeutics [12–15] and vaccine development [20].

A large number of studies have shown that the central regions of NCps, encompassing the two ZFs, are essential for the various activities of the protein [23–25]. Accordingly, mutations on that domain, leading to structural modifications of NCp7, have been shown to abolish virus infectivity [19,26]. We present in this study the NMR structure of the ZF domain of

Abbreviations used: BLV, bovine leukaemia virus; DQF, double-quantum-filtered; HIV-1, HIV type 1; HTLV, human T-cell leukaemia virus; MMLV, Moloney-murine-leukaemia virus; NC, nucleocapsid; NCp, NC protein; (13–51)NCp8, zinc finger domain of NCp8 encompassing residues 13–51; NOE, nuclear Overhauser effect; PBMC, peripheral blood mononuclear cells; R.M.S.D., root mean square deviation; RSV, rous sarcoma virus; SIV, simian immunodeficiency virus; SIVcpz, SIV from chimpanzees; SIVhoest, SIV from l'Hoeist monkeys; SIVmnd, SIV from mandrills; SIVmne, SIV from pig-tailed macaques; SIVsm, SIV from sooty mangabeys; ZF, zinc finger.

¹ To whom correspondence should be addressed (email morellet@pharmacie.univ-paris5.fr).

The co-ordinates of the 20 best calculated simulated annealing structures of (12–51)NCp8 and the NMR restraints have been deposited in the Brookhaven Protein Data Bank under the accession number 2A51.

		n	n+7		n'	n'+7	
		13	15	28 29	35 36	49 51	
NCp7	HIV-1	VK	Cf n Cg k Eg H T a R n C	R A P R R K G	C w K C g K e G H Q m K D C	T E	
NCp7	HIV-2kr	IK	C w n C g K d G H S a R a R C	R A P R R Q G	C w K C g K s G H V m A n C	P E	
NCp7	HIV-2cam2	IK	C w n C g K e G H S a R a R C	R A P R R Q G	C w K C g K p G H I m T n C	P D	
NCp8	SIVlhoest	LT	C f n C g K p G H T a R n C	R Q P R Q E G	C w n C s K s E H R f A Q C	P K	
NCp8	SIVcpz	IK	C f n C g K e G H L a R n C	K A P R R K G	C w R C g Q e G H Q m K D C	T G	
NCp8	SIVmne	IK	C w n C g K e G H S a R a R C	R T P R R Q G	C w K C g Q m G H V m A K C	P D	
Consensus			C f n C g K X G H X a R X C	r X P R X X G	C w X C g X X G H X m X X C		

Figure 1 Comparison of the (13-51) domain primary sequences of the HIV-1, SIVcpz, HIV-2kr, SIVsm, HIV-2cam2, SIVlhoest and SIVmne NCps

In boldface are represented identical residues in the sequences. The consensus sequence of the CCHC domain was obtained after comparison of numerous HIV-1, HIV-2 and SIV strains. The residues conserved or homologous in terms of their physicochemical properties are in lower-case letters f: [F (in HIV-1), W (in HIV-2)] and m (in majority M, but can be replaced by L, Q, K, F or I). The residues conserved at a level < 50% are represented by X.

NCp8, (13-51)NCp8 (residues 13–57 of NCp8), from SIVlhoest (Figure 1). The alignment of the HIV-1 and SIVlhoest ZF domains shows that 85.7 and 42.9% of residues are conserved in ZF1 and ZF2 respectively, with three major differences in positions 21, 43 and 46, since Glu, Gly and Met in NCp7 are replaced respectively by Pro, Glu and Phe in NCp8. For these reasons, to appraise the influence of such important modifications in the SIVlhoest compared with the HIV-1 NC, the structure of this (13-51)NCp8 domain has been elucidated and compared with that of NCp7 from HIV-1 [17,18].

The global structure is perfectly well characterized and the structure of each individual ZF is similar to the corresponding one reported for the two HIV-1 ZFs of NCp7 despite the sequence differences. However, the relative orientation of the two ZFs is different in the two proteins [18,19], the main structural change occurring at the level of the flexible linker between the two ZF domains. Interactions between the two ZFs are reinforced in NCp8 from SIVlhoest by comparison with those observed in NCp7 but the essential residues responsible for the interaction with oligonucleotides [27,28] adopt the same orientation in both proteins. The high structural similarity between NCp8 from SIVlhoest and NCp7 from HIV-1 found in the present study is important as NCp8 could potentially be used for designing possible antiviral agents directed towards NCps, which can be tested initially against SIV-induced diseases in monkeys and then against HIV.

MATERIALS AND METHODS

Peptide synthesis

(13-51)NCp8 was obtained by automated solid phase synthesis using the Fmoc (fluoren-9-ylmethoxycarbonyl) strategy and then purified by reversed-phase HPLC, using procedures already reported for the production of other retroviral proteins [29]. The peptide was analysed by MS and confirmed to have purity higher than 98%.

Sample preparation and NMR spectroscopy

The (13-51)NCp8 peptide was dissolved at pH 6.0 in 90%/10% H₂O/²H₂O and in 100% ²H₂O to give a final concentration of approx. 1 mM. Three equivalents of ZnCl₂ with respect to peptide concentration were added. Two-dimensional phase-sensitive ¹H Clean-TOCSY with 70 ms spin lock, DQF (double-quantum-filtered)-COSY and 100 and 200 ms mixing times NOESY experiments, were recorded at 283, 293 and 303 K on an AVANCE Bruker spectrometer operating at 600.14 MHz without sample spinning, with 2 K real points in t₂, a spectral width of 6000 Hz

and 512 t₁-increments. The transmitter frequency was set to the water signal, which was suppressed using a 3-9-19 pulse sequence with gradients. The data were processed using XWINNMR software (Bruker). A $\pi/6$ phase-shifted sine bell window function was applied prior to Fourier transformation in both dimensions (t₁ and t₂). The temperature was externally controlled using a special temperature control system (BCU 05 Bruker).

Structure calculation of (13-51)NCp8

NOE (nuclear Overhauser effect) cross-peak volumes (NOESY, 200 ms mixing time and 293 K) were converted into distances, semi-quantitatively, by counting contour levels. Using the Asn- β H geminal protons as calibration peak, NOE signals were classified into six categories with upper distance limits ranging from 2.5 to 5 Å (1 Å = 10⁻¹⁰ m). Pseudoatom corrections were added when necessary. Calculations were performed with the Discover/NMRchitect software package from Accelrys with the Amber forcefield. The dielectric constant was set to be distance dependent ($\epsilon = 4r$), in order to diminish *in vacuo* electrostatic effects. A total of 50 initial structures were generated using simulated annealing. The calculation starts with a minimization of the randomly distributed atoms with force constant of the covalent terms, NOE terms (k_{NOE}) and repulsive (k_{repel}) terms reduced to 0.001 kcal · mol⁻¹ · Å⁻² (1 cal = 4.184 J). After the preparation phase, molecular dynamics are initiated such that the internal force constants are adjusted to values where the potential energy of the system is approximately equal to the kinetic energy at 1000 K during 1 ps. The force constants of the NOE restraints are scaled up to their full value during this phase, while the covalent terms are scaled up more slowly. The repulsion force constant remains at its initial low value until k_{NOE} is at its full value and the molecule is loosely folded with covalency established (1 ps). The folding stage of the simulated annealing procedure is completed when all of the force constants have been increased to their full value at 1000 K (except the non-bond term which is at 25% of its full value) (2 ps). The structure is refined in a second stage of dynamics during which the temperature is decreased to 300 K and the non-bond term is increased to its full value in four steps of 2, 1.7, 1.7 and 5 ps each. The NMR constraints were maintained by a force constant of 32 kcal · mol⁻¹ · Å⁻² throughout the refinement process. All peptide bonds were forced into the *trans* orientation ($\omega = 180^\circ$) by torsional constraints with a force constant of 200 kcal · mol⁻¹ · rad⁻² operating for deviations higher than 10° from the *trans* form. The simulated annealing was followed by energy minimization until a maximum gradient value of 0.01 kcal · mol⁻¹ · Å⁻² was obtained. The 20 structures that had the lowest total energy and number of NOE restraint violations were selected for the final structural analysis (Table 1).

RESULTS

NMR study of the (13-51)NCp8 peptide

The one-dimensional ¹H NMR spectrum of the (13-51)NCp8 domain, in the presence of three equivalents of Zn²⁺, exhibits marked dispersion of chemical shifts, reflecting the presence of a folded metal-complexed structure in comparison with the spectrum in the absence of zinc, which is characteristic of a peptide in a random coil conformation. ¹H NMR assignments were obtained using the strategies developed by Wüthrich [30]. Clean-TOCSY and DQF-COSY experiments were used for spin system identification, and NOESY cross-signals connecting HN, H α and H β of residue *i* with residue *i* + 1 were used for sequential assignment of (13-51)NCp8 in solution at pH 6.0. The sequential and medium-range NOEs [*i*/*i* + 2 and *i*/*i* + 3 NOEs (Figure 2)]

Table 1 NMR restraints and structural statistics

The statistics are done on the final 20 simulated annealing structures. The numbers of terms for the various restraints are given in parentheses. NOE NMR-constant energy calculated using square-well potentials with a force constant of $32 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$; E_{L-J} is the value of the Lennard-Jones equation, which was not included in the simulated annealing step but included in the final energy-minimization steps.

Parameter	Value
NMR-derived distance restraints	474
Intraresidue	125
Sequential ($ i - j = 1$)	155
Medium-range ($1 < i - j \leq 4$)	78
Long-range ($ i - j > 4$)	116
Distance violations	
> 0.2	0
R.M.S.D. from experimental	
Distance restraints (Å)	0.038 ± 0.004 (474)
R.M.S.D. from idealized geometry	
Bonds (Å)	0.0063 ± 0.0005 (613)
Angles (deg)	2.18 ± 0.07 (1108)
Dihedral ω (deg)*	4.27 ± 0.13 (38)
Discover potential energies†	
E_{NOE} (kcal $\cdot$ mol $^{-1}$)	22.20 ± 4.43
E_{L-J} (kcal $\cdot$ mol $^{-1}$)	-126.43 ± 4.46
E_{total} (kcal $\cdot$ mol $^{-1}$)	-66.69 ± 8.52
Ramachandran analysis of residues	
In favoured region (%)	83.7
In additional allowed regions (%)	16.3
In generously allowed regions (%)	0
In disallowed regions (%)	0
Pairwise R.M.S.D. (Å)	
Backbone atoms	
(15-28)NCp8	0.21 ± 0.08
(36-49)NCp8	0.24 ± 0.08
(15-49)NCp8	0.53 ± 0.16
Heavy atoms	
(15-28)NCp8	1.06 ± 0.28
(36-49)NCp8	1.25 ± 0.31
(15-49)NCp8	1.64 ± 0.34
R.M.S.D. between backbone atoms NCp8 and NCp7 (Å)	
(15-28)NCp8/(15-28)NCp7	0.67 ± 0.09
(36-49)NCp8/(36-49)NCp7	0.73 ± 0.06
(15-49)NCp8/(15-49)NCp7	2.96 ± 0.18
(15-28)NCp8/(36-49)NCp8	0.84 ± 0.04
(15-28)NCp7/(36-49)NCp7	0.98 ± 0.09

* The conformation of the peptide bond was maintained as *trans* ($\omega = 180^\circ$) with a force constant of $200 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{rad}^{-2}$ ($1 \text{ kcal} = 4.184 \text{ kJ}$).

indicate that despite differences in amino acid content, the two ZFs adopt almost the same secondary structure with two turns involving the Cys¹⁵ → Lys²⁰ and Cys³⁶ → Lys⁴¹ domains on one hand and the Pro²¹ → His²³ and Lys⁴² → His⁴⁴ fragments on the other hand. The C-terminal part of each ZF, Ala²⁵ → Cys²⁸ and Phe⁴⁶ → Cys⁴⁹, exhibited $H\alpha i/NH(i+3)$ and $H\alpha i/H\beta(i+3)$ NOE patterns characteristic of helical segments (Figure 2). Two $H\alpha i/NH(i+2)$ connectivities Cys²⁸/Gln³⁰ and Cys⁴⁹/Lys⁵¹ indicated the presence of two additional turns in the vicinity of each ZF (Figure 2). Comparison of the long-range NOEs observed between the side chains inside each ZF indicates that they fold in an almost similar manner (Figure 3). Some differences are observed since the following side chain NOEs, Phe¹⁶/Thr²⁴, on one hand, and Trp³⁷/Cys³⁹ and Ser⁴¹/Glu⁴³ on the other, are observed in ZF1 and ZF2 respectively, while the corresponding NOEs were not observed in the other ZF. The Lys²⁰–Pro²¹ and Gln³⁰–Pro³¹ bonds did not show any *cis-trans* equilibrium and both are in *trans* conformation.

The complete NMR spectrum analyses of (13-51)NCp8 indicate the presence of a spatial interaction between the two ZFs as

Figure 2 Comparison of the sequential and medium-range NOEs showing that the same local secondary structures occur in the two ZFs (ZF1 and ZF2) in (13-51)NCp8, in spite of differences in the amino acid sequence

Two turns involving the Cys¹⁵ → Lys²⁰ and Cys³⁶ → Lys⁴¹ domains on one hand and the Pro²¹ → His²³ and Lys⁴² → His⁴⁴ fragments on the other hand can be described. The $H\alpha i/NH(i+3)$ and $H\alpha i/H\beta(i+3)$ NOE patterns observed in the C-terminal part of each ZF, Ala²⁵ → Cys²⁸ and Phe⁴⁶ → Cys⁴⁹, are characteristic of helical segments. Moreover, the $H\alpha i/NH(i+2)$ connectivities between Cys²⁸ and Gln³⁰, and between Cys⁴⁹ and Lys⁵¹, indicated the presence of two additional turns close to each ZF.

Figure 3 Diagonal plot of side-chain/side-chain NOEs showing the main differences between the (13-51) domains of NCp8 (lower part) and NCp7 (upper part)

illustrated by numerous NOEs between amino acid side chains belonging to the proximal and distal ZF: Thr¹⁴/Trp³⁷, Thr¹⁴/Phe⁴⁶, Phe¹⁶/Trp³⁷, Phe¹⁶/Phe⁴⁶, Asn¹⁷/Trp³⁷, Gly¹⁹/Trp³⁷ and Ala²⁵/Trp³⁷ (Figures 3 and 4).

NMR-derived three-dimensional structure of the (13-51)NCp8 peptide

A total of 474 interproton distance restraints identified from the NOESY spectrum were used for the structure calculation using standard simulated annealing methods. Of these distance restraints, 125 were intra-residual, 155 sequential, 78 medium-range

Table 3 NMR chemical shifts of the NCp8 and NCp7 peptides at 293 K and pH 6.0, calibrated to H₂O

Residue	Chemical shift			
	NH	α H	β H	Others
Phe ¹⁶ (NCp8)	8.71	4.58	3.20, 3.35	2,6H 7.25; 3,5H 7.37; 4H 7.37
Phe ¹⁶ (NCp7)	8.69	4.52	3.10, 3.12	2,6H 7.24; 3,5H 7.32; 4H 7.32
Trp ³⁷ (NCp8)	7.81	4.57	3.15, 3.36	2H 7.21; 4H 7.24; 5H 7.00; 6H 6.84; 7H 6.59; NH 8.24
Trp ³⁷ (NCp7)	8.44	4.56	3.40, 3.44	2H 7.30; 4H 7.48; 5H 7.12; 6H 6.22; 7H 7.49; NH 10.11
Phe ⁴⁶ (NCp8)	8.96	4.89	2.74, 2.80	2,6H 6.89; 3,5H 7.05; 4H 7.13

sulphur in ZF2. Moreover, the amide protons of Lys²⁰ in ZF1 and Ser⁴¹ in ZF2 are pointing in a direction consistent with hydrogen-bonding to the sulphur of Cys¹⁸ and Cys³⁹ respectively.

The aromatic proton resonances of Trp³⁷ and Phe⁴⁶ belonging to ZF2 underwent a superior upfield shift to Phe¹⁶, by comparison with the standard chemical shifts of Trp and Phe protons in random coil peptide [30] (Table 3). These particular chemical shifts result from the partial stacking of the Trp³⁷ and Phe⁴⁶ aromatic cycles (Figures 5 and 7).

The two CCHC domains form a single folded protein domain characterized by spatial contacts between residues belonging to ZF1 and ZF2 (Figures 3–5 and 7). This relative orientation is locked by hydrophobic and aromatic interactions between Thr¹⁴, Phe¹⁶, Thr²⁴ and Ala²⁵ on one side and Trp³⁷, Phe⁴⁶ and Ala⁴⁷ on the other side, belonging to ZF1 and ZF2 respectively.

Comparison of the 13-51 domains of NCp8 from SIVhoest and NCp7 from HIV-1

The (12-53)NCp7 and (13-51)NCp8 structures and ZF NOE patterns comparison gives insight into the similarities and differences between these two zinc-binding domains that could be induced by differences in amino acid sequence, especially by the replacement of Glu²¹, Gly⁴³ and Met⁴⁶ in NCp7 by Pro²¹, Glu⁴³ and Phe⁴⁶ in NCp8 of almost all the SIVhoests, except in SIVhoest524 where Met is present at the 46th position (Figure 1). The major differences appeared in the number of medium- and long-range NOEs observed for NCp8 by comparison with NCp7. A total of 30 medium-range NOEs characteristic of regular secondary structure (Figure 2) were found for NCp8, instead of 22 for NCp7 [19]. This could explain the differences observed for the backbone atom pairwise R.M.S.D. values, since values of 0.21 ± 0.08 and 0.34 ± 0.22 Å were calculated for the (15-28) domains of NCp8 and NCp7, and 0.24 ± 0.08 and 0.49 ± 0.23 Å for the (36-49) domains of NCp8 and NCp7 respectively. Interestingly, in spite of the important differences in amino acid composition, the same local secondary structures and the same three-dimensional structures (Figure 3) are observed for ZF1 and ZF2 in NCp8 and NCp7. Thus the pairwise backbone R.M.S.D.s between the (15-28) N-terminal and (36-49) C-terminal ZF domains of NCp7 and NCp8 were respectively 0.67 ± 0.09 and 0.73 ± 0.06 Å. The number of hydrogen bonds observed after simulated annealing, stabilizing the local secondary structures of each ZF, is greater in NCp8 compared with NCp7, but all those found in NCp7 are also observed in NCp8 (Table 2).

Moreover, NOEs between seven pairs of residues located on ZF1 and ZF2 (Thr¹⁴/Trp³⁷, Thr¹⁴/Phe⁴⁶, Phe¹⁶/Trp³⁷, Phe¹⁶/Phe⁴⁶, Asn¹⁷/Trp³⁷, Gly¹⁹/Trp³⁷ and Ala²⁵/Trp³⁷) were found for NCp8

Figure 7 Comparison of the (13-51)NCp8 and (13-51)NCp7 stereoview structures after superposition of the (36-49) backbone domains, showing the relative proximity of the two ZFs in each structure

Only the side chain of residues involved in oligonucleotide interactions are represented.

(Figures 3 and 4) instead of two for NCp7 (Phe¹⁶/Trp³⁷ and Ala²⁵/Trp³⁷) [19]. The number of NOEs, which account for the structure of each ZF and for their spatial proximity in NCp8, is greater than that found in NCp7 and explain the better convergence of the NCp8 structures since the pairwise R.M.S.D. values calculated on the backbone atoms are 0.89 ± 0.37 and 0.53 ± 0.16 Å for the (15-49) domains of NCp7 and NCp8 respectively (Table 1). The better definition of the (15-49)NCp8 domain could also be due to an increased rigidity of each ZF in NCp8 by comparison with NCp7 since Glu²¹ and Gly⁴³ in NCp7 are replaced by Pro²¹ and Glu⁴³ in NCp8. This convergence is also increased by the number of hydrophobic residues responsible for the proximity of the two ZFs, which is higher in NCp8 (Thr¹⁴, Phe¹⁶, Thr²⁴, Ala²⁵, Trp³⁷, Phe⁴⁶ and Ala⁴⁷) than in NCp7 (Phe¹⁶, Ala²⁵, Trp³⁷ and Met⁴⁶), and to interaction between the three aromatic residues (Phe¹⁶, Trp³⁷ and Phe⁴⁶) in NCp8, instead of two (Phe¹⁶ and Trp³⁷) in NCp7.

The relative orientation of the two ZFs is different in the two proteins since the pairwise backbone R.M.S.D. corresponding to the (15-49) domain of NCp8 and NCp7 is 2.96 ± 0.18 Å (Table 1). The main structural change occurs at the level of the flexible linker between the two ZFs (residues 29-35) (R.M.S.D. = 1.30 ± 0.10 Å), while the structural distortions undergone by both ZFs are weaker (Table 1). Nevertheless, these values are of the same order of magnitude as those calculated from the backbone of the (12-53)NCp7 peptide in its free and complexed state with ACGCC [27], the main structural change in this case also occurring at the level of the linker between the two ZFs.

Thus, in spite of the high R.M.S.D. calculated on the backbone (15-49) domain between NCp7 and NCp8 due to the flexibility of the linker domain (R²⁹QPRQEG³⁵ in NCp8 and R²⁹APRKKG³⁵ in NCp7), we can observe in Figure 7 that the residues at positions 13, 16, 24, 25, 26, 32, 37 and 46, responsible for the formation of NCp7/nucleic acid complexes [27,28], are accessible for interactions with DNA in the same manner in the two proteins.

DISCUSSION

The SIVs are used as models for HIV-1 disease in many *in vivo* models and for the development of vaccination treatment [20-22]. Moreover, the retroviral NCp is presently a recognized target for therapeutic [12-15,32] and vaccine [20,33] approaches. Among the simian viruses, SIVlhoest is one of few SIVs capable of infecting human PBMC *in vitro*, a prerequisite for cross-species transmission to humans [7]. Such cross-species transmission to humans has already been observed for the SIVsm isolate that is now circulating among human populations as HIV-2 [10], and for HIV-1 which has been shown to originate from SIVcpz [5].

In spite of a great similarity in the biological functions of the retroviral NCps, amino acid composition can vary in an important way. The analysis of multiple alignments of the ZF domains of the HIV-1, HIV-2 and SIV gag proteins gives a consensus sequence, which presents at least 64.3 and 42.9 % of conserved residues for the first and the second ZFs respectively (Figure 1). The value for ZF1 climbs to 78.6 % if the amino acids which have homologous physicochemical properties are included (Figure 1). HIV-1, HIV-2 and SIV strains show 57.1 % sequence conservation in the linker situated between the two ZFs. All the conserved or similar residues are implicated either in the structure of the NCps [17,18] (Cys¹⁵, Cys¹⁸, Gly¹⁹, Gly²², His²³ and Cys²⁸ in ZF1 and Cys³⁶, Cys³⁹, Gly⁴⁰, His⁴⁴ and Cys⁴⁸ in ZF2 and Pro³¹ for the linker), or in nucleic acid interactions [27,28] [(Val,

Ile, Leu)¹³, (Phe, Trp)¹⁶, Ala²⁵, (Arg, Lys)²⁶ in ZF1, Trp³⁷, Met⁴⁶ in ZF2 and Arg³² in the linker (Figure 1)]. NCp8 proteins from SIVcpz and SIVsm do not show important variation in their amino acid sequence by comparison with NCp7 from HIV-1 and -2 (Figure 1), suggesting that their structures are probably not very different from that of NCp7 [19]. On the contrary, NCp8 from SIVlhoest shows significant modifications in amino acid sequence as compared with NCp7, leading possibly to important structural changes in each ZF structure as regards to those of NCp7. In spite of marked similarity, NCp8 from SIVlhoest shows interesting differences with the other HIV and SIV strains, since Pro, Glu and Phe residues located at the 21st, 43rd and 46th positions are replaced by Glu, Gly and Met respectively, in the majority of the HIV and SIV NCps (Figure 1). The structural effects of the proline and glycine residues on the protein structures are well known. Glycine residues have unique conformational flexibility, and the cyclic ring of proline imposes rigid constraints on rotation about the N- α bond of the backbone. The replacement of Glu²¹ by Pro and Gly⁴³ by Glu in NCp8 from SIVlhoest by comparison with the other NCps of HIV and SIV leads to a higher rigidity of the individual ZF in NCp8. Nevertheless, the PGH motif found in NCp8 ZF1 (Figure 1) is also found in other viruses such as in the first ZF of RSV (rouv sarcoma virus) [34], in the second ZF of some HIV-2 strains such as HIV-2cam2 [35] (Figure 1) and BIV (bovine immunodeficiency virus) [36], or in the two ZFs of FIV (feline immunodeficiency virus) [37] and EIAV (equine infectious anaemia virus) [38]. In BLV (bovine leukaemia virus) [39] and in HTLV (human T-cell leukaemia virus) [40], the glycine residue in the PGH motif is replaced respectively by a threonine (PTH motif) and a serine (PSH motif) residue.

Moreover, it has been shown that in spite of the replacement in NCp7 of the acidic residue, Glu²¹ (in the E²¹GH²³ motif), by a hydrophobic one such as alanine, the HIV-1 is still infectious. It seems then that HIV-1 replication is insensitive to amino acid type at this position of the NCp sequence.

A conserved glycine residue at position $n + 7$ (Figure 1), which is always present in NCps constituted by a single CCHC box such as that of MMLV (Moloney-murine-leukaemia virus) [41], is also present in the proximal ZF of retroviruses with two CCHC boxes [41]. On the contrary, Gly at position $n' + 7$ (Figure 1) in the distal ZF is not well conserved since it is replaced by Thr in BLV [39], Ser in HTLV [40], Glu in SIVlhoest [7,8] and Asp in several SIVmn strains [42]. The high degree of conservation of position $n + 7$ (Gly²² in NCp7 and in NCp8), which is invariant among proximal cysteine-histidine boxes of almost all retroviruses, suggests that this residue might be essential to give flexibility to the main chain at this position. However, replication is not significantly affected by the G22A mutation in HIV-1 [43] and the equivalent mutation, G28A, in RSV [44]. On the contrary, substitution of Gly by Val results in a severe replication defect in RSV and MMLV, probably due to steric hindrance caused by the increased size of the amino acid side chain [44,45]. The Gly $\rightarrow$ Asp mutations at positions $n + 7$ in ZF1 and $n' + 7$ in ZF2 of NCp8 of SIV from pig-tailed macaques (SIVmne) were studied [46] because several SIV proviral clones have an aspartic residue adjacent to the conserved His in the second ZF of SIV NC [42]. The study showed that the Gly $\rightarrow$ Asp change in ZF2 of SIVmne did not modify the infectivity of the mutant consistent with the occurrence of this replacement in native SIVmm, whereas this mutation was detrimental in ZF1 [46].

Our study shows that in spite of several modifications in the sequence of SIVlhoest NCp8 compared with HIV-1 NCp7, the structure of each ZF is similar in the two proteins. It seems that NCps can accept modifications in their amino acid sequence, as long as the intrinsic structure of the ZFs is not modified. Moreover,

the magnitude of the difference measured between the NCp8 and NCp7 flexible linker conformation, which induces the relative orientation of the two ZFs in the two proteins, is of the same size order as that observed when we compared the free NCp7 and NCp7 in interaction with ACGCC [27]. This is consistent with the critical role played by the particular structure of the central folded domain of NCps, regarding their biological and toxicological properties [19,26,47,48].

Consistent with these results, it has been shown that SIVmne NCp8 (Figure 1) was able to recognize and bind preferentially the same single-stranded oligonucleotide sequence as HIV-1 NCp7, suggesting that two contiguous TG bases constitute the main determinant for the high-affinity binding site in NCp8 [49], as has already been demonstrated for NCp7 [50]. The similarity between the experimental (12-53)NCp7/ACGCC [27] complex and the computer-derived (12-53)NCp8/GUGUG structure provides additional support for the functional similarity observed between these two proteins [49]. The solution structure of the complex formed between the (13-51) domain of NCp8 and an oligonucleotide is under investigation in our laboratory.

Since similar structural and biological properties have been reported for HIV-1 NCp7 and numerous SIV NCp8 proteins, the NCp8 structure determined in the present study could be used to better understand the HIV-1 NCp7 functions. Moreover, this structural study could be used for the rational design of new molecules able to selectively inhibit the critical functions of NCps before putative *in vivo* trials in monkeys.

This work was supported by ANRS (Agence Nationale de la Recherche sur le SIDA, Paris, France) and SIDACTION (Paris, France), the French programmes against AIDS.

REFERENCES

- Hirsch, V. M., Olmsted, R. A., Murphy-Corb, M., Purcell, R. H. and Johnson, P. R. (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature (London)* **339**, 389–392
- Novembre, F. J., Hirsch, V. M., McClure, H. M., Fultz, P. N. and Johnson, P. R. (1992) SIV from stump-tailed macaques: molecular characterization of a highly transmissible primate lentivirus. *Virology* **186**, 783–787
- Fukasawa, M., Miura, T., Hasegawa, A., Morikawa, S., Tsujimoto, H., Miki, K., Kitamura, T. and Hayami, M. (1988) Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. *Nature (London)* **333**, 457–461
- Emau, P., McClure, H. M., Isahakia, M., Else, J. G. and Fultz, P. N. (1991) Isolation from African Sykes' monkeys (*Cercopithecus mitis*) of a lentivirus related to human and simian immunodeficiency viruses. *J. Virol.* **65**, 2135–2140
- Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M. et al. (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature (London)* **397**, 436–441
- Courgnaud, V., Pourrut, X., Bibollet-Ruche, F., Mpoudi-Ngole, E., Bourgeois, A., Delaporte, E. and Peeters, M. (2001) Characterization of a novel simian immunodeficiency virus from guereza colobus monkeys (*Colobus guereza*) in Cameroon: a new lineage in the nonhuman primate lentivirus family. *J. Virol.* **75**, 857–866
- Hirsch, V. M., Campbell, B. J., Bailes, E., Goeken, R., Brown, C., Elkins, W. R., Axthelm, M., Murphy-Corb, M. and Sharp, P. M. (1999) Characterization of a novel simian immunodeficiency virus (SIV) from L'Hoeist monkeys (*Cercopithecus lhoesti*): implications for the origins of SIVmnd and other primate lentiviruses. *J. Virol.* **73**, 1036–1045
- Beer, B. E., Bailes, E., Dapolito, G., Campbell, B. J., Goeken, R. M., Axthelm, M. K., Markham, P. D., Bernard, J., Zagury, D., Franchini, G. et al. (2000) Patterns of genomic sequence diversity among their simian immunodeficiency viruses suggest that L'Hoeist monkeys (*Cercopithecus lhoesti*) are a natural lentivirus reservoir. *J. Virol.* **74**, 3892–3898
- Tsujimoto, H., Hasegawa, A., Maki, N., Fukasawa, M., Miura, T., Speidel, S., Cooper, R. W., Moriyama, E. N., Gojibori, T. and Hayami, M. (1989) Sequence of a novel simian immunodeficiency virus from a wild-caught African mandrill. *Nature (London)* **341**, 539–541
- Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M. and Hahn, B. H. (1992) Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature (London)* **358**, 495–499
- Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H. and Roques, B. P. (1995) First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–537
- Rice, W. G., Supko, J. G., Malspeis, L., Buckheit, Jr, R. W., Clanton, D., Bu, M., Graham, L., Schaeffer, C. A., Turpin, J. A., Domagala, J. et al. (1995) Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS. *Science* **270**, 1194–1197
- Musah, R. A. (2004) The HIV-1 nucleocapsid zinc finger protein as a target of antiretroviral therapy. *Curr. Top. Med. Chem.* **4**, 1605–1622
- Huang, M., Maynard, A., Turpin, J. A., Graham, L., Janini, G. M., Covell, D. G. and Rice, W. G. (1998) Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* **41**, 1371–1381
- Berthou, L., Pechoux, C. and Darlix, J. L. (1999) Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication. *J. Virol.* **73**, 10000–10009
- Freed, E. O. (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* **251**, 1–15
- Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J. L. and Roques, B. P. (1992) Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ¹H NMR. *EMBO J.* **11**, 3059–3065
- Summers, M. F., Henderson, L. E., Chance, M. R., Bess, Jr, J. W., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, III, R. C., Hare, D. R. et al. (1992) 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.* **1**, 563–574
- Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J. L., Fournie-Zaluski, M. C. and Roques, B. P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NCp7 of HIV-1 studied by ¹H NMR. *J. Mol. Biol.* **235**, 287–301
- Arthur, L. O., Bess, Jr, J. W., Chertova, E. N., Rossio, J. L., Esser, M. T., Benveniste, R. E., Henderson, L. E. and Lifson, J. D. (1998) Chemical inactivation of retroviral infectivity by targeting nucleocapsid protein zinc fingers: a candidate SIV vaccine. *AIDS Res. Hum. Retroviruses* **14** (Suppl. 3), S311–S319
- Hanke, T., Samuel, R. V., Blanchard, T. J., Neumann, V. C., Allen, T. M., Boyson, J. E., Sharpe, S. A., Cook, N., Smith, G. L., Watkins, D. I. et al. (1999) Effective induction of simian immunodeficiency virus-specific cytotoxic T lymphocytes in macaques by using a multiepitope gene and DNA prime-modified vaccinia virus Ankara boost vaccination regimen. *J. Virol.* **73**, 7524–7532
- Wang, S. W., Kozlowski, P. A., Schmelz, G., Manson, K., Wyand, M. S., Glickman, R., Montefiori, D., Lifson, J. D., Johnson, R. P., Neutra, M. R. et al. (2000) Effective induction of simian immunodeficiency virus-specific systemic and mucosal immune responses in primates by vaccination with proviral DNA producing intact but noninfectious virions. *J. Virol.* **74**, 10514–10522
- Zhang, Y. and Barklis, E. (1995) Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J. Virol.* **69**, 5716–5722
- Dupraz, P., Oertle, S., Meric, C., Damay, P. and Spahr, P. F. (1990) Point mutations in the proximal Cys-His box of Rous sarcoma virus nucleocapsid protein. *J. Virol.* **64**, 4978–4987
- Gorelick, R. J., Gagliardi, T. D., Bosche, W. J., Wiltout, T. A., Coren, L. V., Chabot, D. J., Lifson, J. D., Henderson, L. E. and Arthur, L. O. (1999) Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences. *Virology* **256**, 92–104
- Ramboarina, S., Druillennec, S., Morellet, N., Bouaziz, S. and Roques, B. P. (2004) Target specificity of human immunodeficiency virus type 1 NCp7 requires an intact conformation of its CCHC N-terminal zinc finger. *J. Virol.* **78**, 6682–6687
- Morellet, N., Demene, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournie-Zaluski, M. C. and Roques, B. P. (1998) Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J. Mol. Biol.* **283**, 419–434
- De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**, 384–388
- de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M. C., Darlix, J. L. and Roques, B. P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form. *Biochem. Biophys. Res. Commun.* **180**, 1010–1018

- 30 Wüthrich, K. (1986) NMR of Protein and Nucleic Acids, Wiley and Sons, New York
- 31 Adman, E., Watenpaugh, K. D. and Jensen, L. H. (1975) NH-S hydrogen bonds in *Peptococcus aerogenes* ferredoxin, *Clostridium pasteurianum* rubredoxin, and Chromatium high potential iron protein. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4854–4858
- 32 Druillennec, S., Dong, C. Z., Escaich, S., Gresh, N., Bousseau, A., Roques, B. P. and Fournie-Zaluski, M. C. (1999) A mimic of HIV-1 nucleocapsid protein impairs reverse transcription and displays antiviral activity. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4886–4891
- 33 Rossio, J. L., Esser, M. T., Suryanarayana, K., Schneider, D. K., Bess, Jr, J. W., Vasquez, G. M., Wiltout, T. A., Chertova, E., Grimes, M. K., Sattentau, Q. et al. (1998) Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins. *J. Virol.* **72**, 7992–8001
- 34 Schwartz, D. E., Tizard, R. and Gilbert, W. (1983) Nucleotide sequence of Rous sarcoma virus. *Cell (Cambridge, Mass.)* **32**, 853–869
- 35 Tristem, M., Hill, F. and Karpas, A. (1991) Nucleotide sequence of a Guinea-Bissau-derived human immunodeficiency virus type 2 proviral clone (HIV-2CAM2). *J. Gen. Virol.* **72**, 721–724
- 36 Garvey, K. J., Oberste, M. S., Elser, J. E., Braun, M. J. and Gonda, M. A. (1990) Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus. *Virology* **175**, 391–409
- 37 Olmsted, R. A., Hirsch, V. M., Purcell, R. H. and Johnson, P. R. (1989) Nucleotide sequence analysis of feline immunodeficiency virus: genome organization and relationship to other lentiviruses. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8088–8092
- 38 Stephens, R. M., Casey, J. W. and Rice, N. R. (1986) Equine infectious anemia virus gag and pol genes: relatedness to visna and AIDS virus. *Science* **231**, 589–594
- 39 Oroszlan, S., Copeland, T. D., Henderson, L. E., Stephenson, J. R. and Gilden, R. V. (1979) Amino-terminal sequence of bovine leukemia virus major internal protein: homology with mammalian type C virus p30 structural proteins. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2996–3000
- 40 Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S., Miwa, M. and Sugimura, T. (1985) Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3101–3105
- 41 Green, L. M. and Berg, J. M. (1989) A retroviral Cys-Xaa2-Cys-Xaa4-His-Xaa4-Cys peptide binds metal ions: spectroscopic studies and a proposed three-dimensional structure. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4047–4051
- 42 Korber, B., Foley, B., Leitner, T., McCutchan, F., Hahn, B., Mellors, J. W., Myers, G. and Kuiken, C. (1997) Human Retrovirus and AIDS 1997: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences, Los Alamos National Laboratory, Los Alamos, New Mexico
- 43 Dorfman, T., Luban, J., Goff, S. P., Haseltine, W. A. and Gottlinger, H. G. (1993) Mapping of functionally important residues of a cysteine-histidine box in the human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **67**, 6159–6169
- 44 Dupraz, P. and Spahr, P. F. (1992) Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *J. Virol.* **66**, 4662–4670
- 45 Meric, C. and Goff, S. P. (1989) Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J. Virol.* **63**, 1558–1568
- 46 Gorelick, R. J., Benveniste, R. E., Gagliardi, T. D., Wiltout, T. A., Busch, L. K., Bosche, W. J., Coren, L. V., Lifson, J. D., Bradley, P. J., Henderson, L. E. et al. (1999) Nucleocapsid protein zinc-finger mutants of simian immunodeficiency virus strain mne produce virions that are replication defective *in vitro* and *in vivo*. *Virology* **253**, 259–270
- 47 Guo, J., Wu, T., Kane, B. F., Johnson, D. G., Henderson, L. E., Gorelick, R. J. and Levin, J. G. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein. *J. Virol.* **76**, 4370–4378
- 48 Yovandich, J. L., Chertova, E. N., Kane, B. P., Gagliardi, T. D., Bess, Jr, J. W., Sowder, II, R. C., Henderson, L. E. and Gorelick, R. J. (2001) Alteration of zinc-binding residues of simian immunodeficiency virus p8(NC) results in subtle differences in gag processing and virion maturation associated with degradative loss of mutant NC. *J. Virol.* **75**, 115–124
- 49 Urbaneja, M. A., McGrath, C. F., Kane, B. P., Henderson, L. E. and Casas-Finet, J. R. (2000) Nucleic acid binding properties of the simian immunodeficiency virus nucleocapsid protein NCp8. *J. Biol. Chem.* **275**, 10394–10404
- 50 Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M. and Henderson, L. E. (1998) Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.* **72**, 1902–1909

Received 25 July 2005/7 October 2005; accepted 17 October 2005

Published as BJ Immediate Publication 17 October 2005, doi:10.1042/BJ20051203