Structural Features in EIAV NCp11: A Lentivirus Nucleocapsid Protein with a Short Linker‡

Pietro Amodeo,§ Maria A. Castiglione Morelli,|| Angela Ostuni,|| Gianantonio Battistuzzi,⊥ and Alfonso Bavoso*,||

Istituto di Chimica Biomolecolare, CNR, via Campi Flegrei 34, Comprensorio A. Olivetti, Ed. 70, 80078 Pozzuoli (NA), Italy, Department of Chemistry, University of Basilicata, via N. Sauro 85, 85100 Potenza, Italy, and Department of Chemistry and Centro SCS, University of Modena and Reggio Emilia, via Campi 183, 41100 Modena, Italy

Received December 7, 2005; Revised Manuscript Received March 10, 2006

ABSTRACT: Lentiviral nucleocapsid proteins are a class of multifunctional proteins that play an essential role in RNA packaging and viral infectivity. They contain two $CX_2CX_4HX_4C$ zinc binding motifs connected by a basic linker of variable length. The 3D structure of a 37-aa peptide corresponding to sequence 22−58 from lentiviral EIAV nucleocapsid protein NCp11, complexed with zinc, has been determined by 2D ^{1}H NMR spectroscopy, simulated annealing, and molecular dynamics. The solution structure consists of two zinc binding domains held together by a five-residue basic linker Arg^{38}-Ala-Pro-Lys-Val^{42} that allows for spatial proximity between the two finger domains. Observed linker folding is stabilized by H bonded secondary structure elements, resulting in an Ω-shaped central region, asymmetrically centered on the linker. The conformational differences and similarities with other NC zinc binding knuckles have been systematically analyzed. The two CCHC motifs, both characterized by a peculiar Pro-Gly sequence preceding the His residue, although preserving Zn-binding geometry and chirality of other known NC proteins, exhibit local fold differences both between each other and in comparison with other previously characterized retroviral CCHC motifs.

Lentiviruses constitute a genus of the retroviridae family, which includes several viruses (HIV-1 and HIV-2, SIV, visna virus, FIV, EIAV, CAEV, BIV, and others) infecting a diverse array of species. Lentiviruses are host-specific viruses that cause slowly progressive diseases in their hosts, characterized by cytopathic changes in host cells (*1*).

All lentiviruses encode a *gag* product that is processed by retroviral protease to generate different proteins. One of these, nucleocapsid protein (NC[1]), is the major structural protein of the ribonucleoprotein complex bound to genomic RNA in mature virions. NC proteins of all lentiviruses and most retroviruses share a high percentage of basic residues and a zinc-binding motif composed of a highly conserved amino acid sequence with invariably spaced Cys and His residues of the form -Cys-X_2-Cys-X_4-His-X_4-Cys- (CCHC motif), where X indicates a variable amino acid residue. These sequences (zinc knuckle) are found either once or twice, depending on the retroviral species. NC proteins of different retrovirus species not only share conserved amino acid sequences but also appear to have similar biological functions (*2*). Because of its essential role in the virus life cycle, NC protein is viewed as an attractive target for the development of antiretroviral drugs (*3*).

A limited number of structural studies on NC proteins of different retroviruses, such as HIV-1 (*4*−*7*), HIV-2 (*8*), MMTV (*9*), MPMV(*10*), and MoMuLV (*11*), have been performed, focusing on either the entire protein or the relevant peptide fragments. Among NC proteins of the lentivirus genus, HIV-1 NCp7 is the only NC protein that has been structurally characterized both in its bound (*12*−*14*) and unbound states to nucleic acid fragments (*5*−*7*), allowing the identification of key residues for the interaction with linear RNA and RNA stem loop elements (*15*).

A reliable identification of sequence-dependent versus sequence-invariant structural features within the lentiviral NC

‡ The coordinates have been deposited in the Protein Data Bank with accession number 2bl6.

* Corresponding author. Tel: +39 971 202208. Fax: +39 971 202223. E-mail: bavoso@unibas.it.

§ Istituto di Chimica Biomolecolare.

|| University of Basilicata.

⊥ University of Modena and Reggio Emilia.

[1] Abbreviations: HIV, human immunodeficience virus; SIV, simian immunodeficience virus; CIV, chimpanzee variant of simian immunodeficience virus; FIV, feline immunodeficience virus; EIAV, equine infectious anaemia virus; CAEV, caprine arthritis encephalitis virus; BIV, bovine immunodeficience virus; NC, nucleocapsid; CCHC, -Cys-X_2-Cys-X_4-His-X_4-Cys- residue pattern; MMTV, murine mammary tumour virus; MPMV, Mason−Pfizer monkey virus; Mo-MuLV, Moloney murine leukemia virus; CD, circular dichroism; NMR, nuclear magnetic resonance; ZF18, 18-residue peptide corresponding to residues 22−39 of the Ncp11 protein; ZF20, 20-residue peptide corresponding to residues 39−58 of the Ncp11 protein; ZF37, 37-residue peptide corresponding to residues 22−58 of the Ncp11 protein; EDTA, ethylene diamine tetraacetic acid; DFQ-COSY, double-quantum filter-correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear overhauser effect spectroscopy; NOE, nuclear overhauser effect; SA, simulated annealing; RIWSA, restrained, implicit water, simulated annealing; REM, restrained energy minimization; GBSA, generalized born with surface area contribution; rmsd, root mean square deviation; REWMD, restrained, explicit water, molecular dynamics; UEWMD, unrestrained, explicit water, molecular dynamics; EM, energy minimization; Fmoc, *N*-fuorenylmethoxycarbonyl; HMP, *P*-hydroxymethylphenoxy methylpolystyrene; TFA, trifluoroacetic acid; EDT, 1,2-ethane dithiole; HPLC, high-performance liquid chromatography; TCEP, tris (2-CarboxyEthyl) phosphine; UEM, unrestrained energy minimization; MD, molecular dynamics.

10.1021/bi0524924 CCC: $33.50
Published on Web 04/07/2006

protein family could provide invaluable information on their functional features and the ability to design antiretroviral drugs against such a variable and elusive virus class. A compared sequence inspection of these proteins showed that EIAV NCp11 exhibits several sequence peculiarities with respect to other NC proteins of known structure that could lead to interesting structural findings.

EIAV(*16*), a prototype of nonprimate lentiviruses, causes equine infectious anemia, the most important infectious disease in horses. It has also been considered a much safer vector for gene therapy than primate retroviruses (*17*−*19*) because it transduces both dividing and nondividing cells without replicating in human cells. In the case of EIAV, the *gag* (*20*) products are matrix (MA, p15), capsid (CA, p26), nucleocapsid (NC, p11), and p9 proteins. NCp11 of EIAV, like other NC proteins, is a small basic nucleoprotein containing two copies of a CCHC motif (*20*). NC proteins bind tightly to nucleic acids (*21*) including double stranded supercoiled DNA (*22*). In this respect, EIAV NC protein has unwinding activity resembling that of topoisomerase I (*23*); this activity is required for the topological changes that occur during integration into the host cell genome and expression of the proviral DNA, when EIAV, like other retroviruses, establishes persistent infection. EIAV NC protein is also presumably involved in RNA packaging during virus assembly in a mature retroviral particle and plays the other biological roles of lentivirus NC proteins (*24*−*26*).

Because the *N*- and *C*-terminal tails have been generally found unstructured in other NC proteins, we have focused the present article on the metal binding properties of the fragment peptides corresponding to the two zinc knuckles of EIAV NCp11 and on the solution structure of a 37-aa peptide corresponding to the two CCHC motifs and the linker region between them (residues 22−58 of the NCp11, 381−417 of the *gag* polyprotein isolate Wyoming), from UV−vis, CD, and ^{1}H NMR spectroscopy data. A detailed comparison with other known retroviral NCp structures is also reported.

EXPERIMENTAL PROCEDURES

Peptide Synthesis. A 37-residue peptide QT**C**YN**C**GK-PG**H**LSSQ**C**RAPKV**C**FK**C**KQPG**H**FSKQ**C**RS (ZF37), an 18-residue peptide QT**C**YN**C**GKPG**H**LSSQ**C**RA (ZF18), and a 20-residue peptide APKV**C**FK**C**KQPG**H**FSKQ**C**RS (ZF20) were synthesized using standard solid-phase peptide synthesis techniques as reported in Supporting Information.

Circular Dichroism Spectroscopy. CD spectra were recorded on a Jasco J-600 spectropolarimeter using a 0.1-cm path-length cylindrical quartz cell. Peptide samples (0.1 mg/mL) were dissolved in a sodium phosphate buffer (10 mM) at pH 7.5. $ZnCl_2$ was dissolved in water at a concentration of 1 mM and added to peptides in different molar ratios 30 min before the spectra were recorded. Data were acquired at 20 °C and were baseline corrected by subtracting the buffer/TCEP and/or buffer/TCEP/$ZnCl_2$. The final spectrum in each case is the sum of 16 separate spectra with a step resolution of 0.1 nm, a 2 s time constant, a 50 nm min^{-1} scan speed, a 1 nm bandwidth, and a 20 mdg sensitivity.

Metal Binding Studies. The dissociation constant for the Co^{2+} adduct was determined by titrating a TCEP-reduced peptide (0.1 mM) in a 10 mM phosphate buffer at pH 7.5 with a metal ion and monitoring the absorption spectrum (*27*). Additions were performed in a sequential manner with negligible changes in total solution volume. After each metal addition, each sample was mixed thoroughly and allowed to equilibrate for 1 min prior to recording UV−vis spectra. The Zn^{2+} binding constant was obtained by following the bleaching in the absorption spectrum of Co^{2+} adduct upon metal addition, as reported elsewhere (*27*, *28*). UV−vis absorption spectra were recorded from 200 to 800 nm on a Cary 5 double beam spectrophotometer with 1-cm quartz cuvettes. Data were recorded at a speed of 600 nm min^{-1} with a signal averaging time of 0.1 s.

NMR Spectroscopy. ZF37 and related peptides were dissolved in either 90% H_2O/10% D_2O or 100% D_2O at concentrations of approximately 2.5 mM at pH 5.8. $ZnCl_2$ was added in slight excess (3 equivalents of Zn^{2+} for ZF37 and 1.5 equivalents for ZF18 and ZF20) with respect to zinc finger concentration. The reducing agent, TCEP, was added to prevent the oxidation of cysteine residues.

NMR data were collected using a VARIAN Unity Inova 500 MHz or a BRUKER DRX-500 spectrometer and with samples at 288, 293, 298, and 303 K. Proton chemical shift assignments for peptides were obtained from DQF-COSY-(*29*), TOCSY (*30*, *31*), and NOESY (*32*) experiments. In TOCSY and NOESY experiments, selective excitation was applied to suppress water resonance (WATERGATE pulse sequence (*33*)). Two dimensional experiments were recorded in phase sensitive mode using the States method (*34*) or the time-proportional phase incrementation scheme (*35*). The mixing times for TOCSY spectra were 50 and 80 ms and 100, 150, and 200 ms for NOESY spectra. Weak NOEs observed only at a mixing time of 200 ms but not at 100 ms were considered to be spin diffusion artifacts and were not used in the calculations.$^3J_{NH\alpha}$ coupling constants were determined from an analysis of high-resolution DQF-COSY spectra. Prior to Fourier transformation and baseline correction with polynomial functions, all data were zero-filled and multiplied with shifted sine-bell window functions in both dimensions.

Structure Calculations. Calculations were performed with the Sander module of AMBER6 (*36*), using AMBER all-atom 1994 parametrization (*37*). Conformational sampling was obtained by restrained simulated annealing (SA), including an implicit solvation model (leading to the so-called RIWSA structure set), with proton−proton distance restraints derived from experimental NOE intensities.

Final structures were energy minimized (EM) both with the same penalty function used for SA (RIWSA/REM structure set) and under totally unrestrained conditions (RIWSA/UEM). The final selected EM structure was solvated and underwent a molecular dynamics (MD) simulation with NMR-derived and Zn-coordination restraints (REWMD structures), followed by MD with application of only the Zn-coordination restraints (UEWMD structures). The details of computational protocols and parameters are reported in Supporting Information. The analysis modules of the AMBER 6.0 package and the MOLMOL (*38*) program were used for structural analyses.

RESULTS

Confirmation of Zinc Binding. UV−vis and CD spectra were preliminary measured to confirm zinc binding to

NCp11. The characterization of peptide-metal complexes requires the coordination of an optically active ion such as Co^{2+} (*27*). Metal coordination occurs via nitrogen and sulfur atoms of histidine and cysteine residues, respectively. The addition of Co^{2+} to reduced ZF18 at pH 7.5 originates absorption bands at 307, 348 (shoulder), 598 (shoulder), 644, and 697 nm (Supporting Information). Very similar spectra were also obtained for Co^{2+}−ZF20 and Co^{2+}−ZF37 complexes**.** The titrations of peptides with Co^{2+} were spectrophotometrically followed. The band intensities increase linearly with increasing metal concentration, and data fitted by nonlinear least-squares methods yielded dissociation constants of 1.4×10^{-7} and 3.6×10^{-6} M for the Co^{2+} adducts of ZF18 and ZF20, respectively. The formation of zinc-fingers by these synthetic peptides was confirmed by the disappearance of characteristic charge−transfer and d−d transition bands in visible spectra upon the addition of $ZnCl_2$ to Co^{2+} complexes; these results indicate that Zn^{2+} ions replace Co^{2+} ions in tetrahedral complexes. The dissociation constants for the Zn^{2+} derivatives formed by peptides were roughly measured by titrating the Co^{2+} adducts with Zn^{2+} in the presence of a slight excess of Co^{2+}(*27*). The dissociation constants were 4×10^{-12} and 5×10^{-10} M for ZF18 and ZF20, respectively. These results confirm the zinc binding properties of EIAV NC protein (*39*).

The effects of Zn^{2+} on the secondary structures of these peptides were monitored by CD and NMR spectroscopies. The CD spectra of ZF18, ZF20, and ZF37 (Supporting Information) in the absence of zinc are characteristic of random coil conformations. In the presence of zinc (molar ratio zinc/Zn-finger 1.2:1), a positive maximum is observed at ~220 and ~213 nm for ZF20 and ZF37, respectively; moreover, a significant decrease in negative ellipticity was observed. The CD spectrum of the ZF18−Zn complex shows a positive peak at ~195 nm as well as at ~215 nm, probably due to the Y^{25} residue. The effects of zinc on these peptides were completely reversed by the addition of 10 mol of EDTA/mol of zinc. Altogether, CD data indicate that these peptides are folded upon complexation with zinc.

Three-Dimensional NMR Structure Determination and Validation of ZF37. 1D ^{1}H NMR spectra of ZF37 peptide in the presence of three equivalents of Zn^{2+} exhibited a marked dispersion of chemical shifts in comparison to those in the spectra in the absence of zinc, which are characteristic of a peptide in a random conformation. Upon interaction with a metal ion, as already observed for other retroviral finger peptides (*7*, *40*), ZF37 adopts a folded metal-complexed structure.

A complete sequential assignment of ZF37 peptide−Zn^{2+} complex was accomplished using a conventional 2D NMR methodology (*41*). The spin systems of each amino acid were identified using DQF-COSY (*29*) and TOCSY (*30*, *31*) spectra; then, connectivities across the peptide bonds were established from NOESY (*32*) experiments. When necessary, to resolve ambiguities arising from chemical shift degeneracy, spectra were recorded at different temperatures (288, 293, 298, and 303 K). In the presence of zinc, identified ^{1}H chemical shifts of corresponding residues in complete ZF37 and in isolated ZF18 and ZF20 peptides exhibited chemical shift differences ($\Delta\delta$) variable from 0.02 to 0.28 ppm. The aromatic protons of residues Tyr^{25} and Phe^{44} give rise to $\Delta\delta$ values of 0.03 ppm or less from corresponding protons in

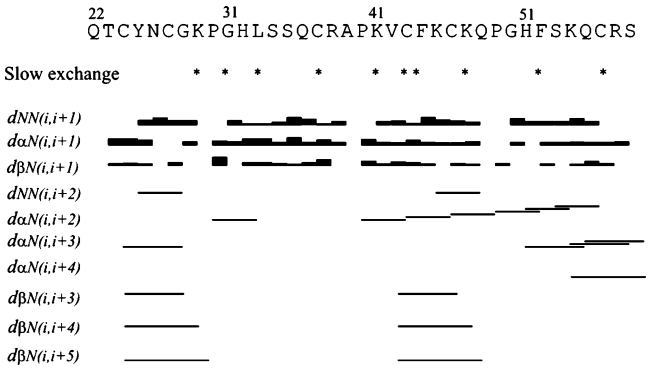

FIGURE 1: Summary of short-range NOEs involving NH, αCH, and *β*CH protons observed for the ZF37 peptide−Zn^{2+} complex. NOEs are classified as strong, medium, and weak according to the thickness of the bars. Asterisks label residues for which the slow exchange of NH protons with deuterons was observed.

ZF18/ZF20. These two residues are involved in a spatial interaction between the zinc-binding domains of ZF37, as evidenced by the presence of weak, long-range NOEs between (1) aromatic ring protons of Phe^{44} and the two *β*CH protons of Tyr^{25} and (2) aromatic ring protons of Tyr^{25} and the αCH proton of Phe^{44}. However, consistent with the lack of ring current effect, we did not observe mutual NOEs between the aromatic protons of Tyr^{25} and Phe^{44}. The observation of almost unchanged chemical shifts suggests the absence of major changes in the conformations of the *N*- and *C*-termini of ZF37 peptide in comparison to those in the isolated domains (data not shown), as already found for NCp7 and its (13−30) proximal finger and the (29−72) *C*-terminal fragment (*7*). However, in NCp7, two residues exhibited more consistent variations in the chemical shifts, and they were the interacting Phe^{16} and Trp^{37} because of a ring current effect of their aromatic protons (*7*).

The secondary structure of the ZF37 peptide was delineated from the qualitative analysis of sequential (αCH_i-NH_{i+1} and NH_i-NH_{i+1}) and medium range (αCH_i-NH_{i+n}, $1 < n < 4$) NOE connectivities from slowly exchanging amide protons and from $^3J_{HN\alpha}$ coupling constants. Figure 1 shows a summary of observed NOEs and the relative exchange rates of amide protons for ZF37 at 303 K and pH 5.8. Although three Pro residues (Pro^{30}, Pro^{40}, and Pro^{49}) were present, no evidence of cis−trans equilibrium for any of Xxx-Pro peptide bonds was observed, and a trans bond conformation was always preferred.

As already observed for another zinc finger peptide (*5*), Cys^{24} from the *N*-terminal domain and Cys^{43} from the *C*-terminal domain show characteristic βCH_i-NH_{i+n}, (n = 3,4,5) connectivities. Besides, the two zinc binding histidines, His^{32} and His^{51}, exhibit several connectivities from the imidazole ring protons to residues i - 3, i + 4, and i + 5. Besides the aforementioned interknuckle NOEs between Tyr^{25} and Phe^{44}, another 25 long-range NOEs were observed between residues in the *N*-terminal finger and the linker region and 22 long-range NOEs between the residues in the *C*-terminal finger and the linker. In the *C*-terminal finger, we observed long-range NOEs between (1) αCH protons of Cys^{43} and Phe^{52} and (2) *β*CH protons of Cys^{43} and HN proton of Phe^{52}.

Preliminary 3D structures of ZF37 were generated with the DYANA program (*42*) starting with 267 manually

assigned NOEs. The iterative back assignment from calculated structures allowed the stereospecific assignment of 23 out of 31 nondegenerate β-methylene proton pairs and the assignment of 218 stereo-ambiguous cross-peaks. Both initial and final datasets were checked against restraints deposited for corresponding residues in NCp7(12−53) (pdb accession number 1esk.mr). Already, after a visual inspection, NOE patterns observed for the two molecules exhibit substantial discrepancies. Also, a refinement of both molecules after interchanging corresponding restraint sets and a new refinement of NCp7(12−53) with its own dataset and our computational protocol (data not shown) demonstrated that the differences detected in conformational behavior between the two molecules are derived from experimental datasets and not from different structural refinement protocols.

The final set of NOE-derived interproton distance restraints (Table 1) for all residues includes a total of 683 distance restraints, which were subsequently used to generate 3D structures of ZF37 by the simulated annealing (SA) approach described in Experimental Procedures, where the different structure bundles that were analyzed are also defined. The final RIWSA/REM bundle includes 100 structures exhibiting good convergence of all parameters reported in Table 1. Particularly noteworthy are the values of Procheck-NMR (*43*) G-factors and the residue distribution in the Ramachandran map, although the structures lack an extensive helix/sheet secondary structure and contain noncanonical conformations induced by zinc binding. The final bundle was also checked against the whole ensemble of experimental information from NOESY spectra by calculating all short (i.e., <5 Å) distances involving proton pairs whose chemical shift could be assigned and looking for their counterparts in all processed spectra. The results reported in Table 1 show that no truly unexpected or inconsistent distances or any strong to weak spectral peaks without a counterpart in the short distance list, was observed. In particular, no other inter-knuckle short distances are predicted, apart from those experimentally detected. The final set (683 restraints) corresponds to 52 and 81% of total and theoretically observable short distance sets, respectively, associated to our bundle.

Structural Features. The solution structure of the ZF37−zinc complex consists of two zinc binding domains held together by a five-residue basic linker Arg[38]-Ala-Pro-Lys-Val[42] (Figure 2) that allows a spatial proximity between the two finger domains and, after a backbone atom superposition (Table 1), shows a remarkably small rmsd value (0.09 ± 0.04 Å on the Arg[38]-Val[42] region).

Although lacking extensive canonical secondary structures, like other viral CCHC zinc-binding NC proteins, the ZF37 bundle exhibits a definitively folded structure, as witnessed both by the identification of several turns/structural motifs and from H-bond patterns (Tables 2and 3). In particular, in the final bundle, the PROMOTIF 3.0 program (*44*) identified the following: (a) a short 3_{10} helix, spanning residues His[51]-Lys[54], stabilized by His[51]-Lys[54] and Phe[52]-Gln[55] strong O_i-H_{i+3} H bonds; (b) from 8 to 13 fused β-turns; (c) a *C*-terminal 5-turn locked by a very strong Phe[52]-Arg[57] O_i-H_{i+5} H bond stabilized by other interactions from the 3_{10} helix and fused β-turns, thus giving rise to one of the most convergent ZF37 regions (rmsd of 0.13 Å for backbone atoms in the His[51]-Arg[57] range); (d) two bridge-1 structures: Cys[24]-Pro[40], locked by a strong Thr[23]-Lys[41] a weaker Tyr[25]-Ala[39] H bond, and Thr[23]-Pro[30], locked by a medium Gln[22]-His[32] and a very strong Cys[24]-Lys[29] H bond.

An analysis of side chain atom interactions shows only two relevant backbone-side chain H bonds, both anchoring

Table 1: Structural Statistics for the Bundle of 100 Accepted ZF37 Structures

Experimental Restraints[a]	
intraresidue NOEs	330 dupl
inter-residue sequential NOEs ($\|i-j\|=1$)	187 dupl
inter-residue medium-range NOEs ($1<\|i-j\|\leq 4$)	105 dupl
inter-residue long-range NOEs ($\|i-j\|>4$)	61 dupl
total NOEs	683 dupl
Zn-coordination upper limit/range restraints	8 dupl, 12 arange
Zn-coordination lower limit restraints	95 dlol
total restraints	798 (691 dupl, 95 dlol, 12 arange)
Total and Average per Residue (in Parentheses) Restraint Number within ZF37 Regions	
N-terminal CCHC(Cys[24]-Cys[37])	236 (25.6)
linker (Arg[38]-Val[42])	71 (18.4)
C-terminal CCHC(Cys[43]-Cys[56])	267 (27.8)
Predicted Number and Percent (in Parentheses) of Short Distances vs Observed NOE Peaks	
total predicted short distances (<5 Å)	1304
unambiguously assigned	485 (36%)
initially ambiguous but finally assignable	198 (15%)
strongly overlapped	463 (36%)
not clearly identified	158 (12%)
not clearly identified with corresponding distance <4 Å[b]	60 (4%)
Restraints Violations[c]	
sum of restraint violations:	0.15 ± 0.18 Å (min 0.0, max 0.9)
NOE distances with violations >0.1 Å:	0.4 ± 0.8 (min 0, max 3)
NOE distances with violations >0.2 Å:	0.1 ± 0.4 (min 0, max 2)
NOE distances with violations >0.3 Å:	0.01 ± 0.1 (min 0 max 1)
AMBER94 GBSA energy (kcal mol^{-1})	−1658.5 ± 3.4
rmsd from Ideal Covalent Geometry	
bonds (Å)	0.0115 ± 0.0004
angles (°)	2.46 ± 0.02
rmsd from Average Structure	
N-terminal CCHC(Cys[24]-Cys[37]) (Å)[d]	0.29 ± 0.14
linker (Arg[38]-Val[42]) (Å)[d]	0.09 ± 0.04
C-terminal CCHC(Cys[43]-Cys[56]) (Å)[d]	0.17 ± 0.06
all selected (Cys[24]-Cys[56]) (Å)[d]	0.38 ± 0.13
all backbone (Å)[e]	0.46 ± 0.17
Procheck_NMR (G-factor and Ramachandran Analysis)	
G-factors	
overall	−0.69
dihedral	−0.97
$\chi_1-\chi_2$	−0.73
χ_1-only	0.17
residues in the favored region (%)	66.4
residues in the additional allowed region (%)	30.7
residues in the generously allowed region (%)	2.9
residues in the disallowed region (%)	0.0

[a] dupl = upper limit of the distance; dlol = lower limit of the distance; arange = angle range. [b] All of these peaks involve either flexible side chains or atoms providing weaker than average peaks. [c] No restraint violation larger than 0.38 Å was detected. [d] Calculated on all backbone atoms plus side chain heavy atoms of Cys/His residues and Zn ions included in the reported range. [e] Calculated on backbone atoms only.

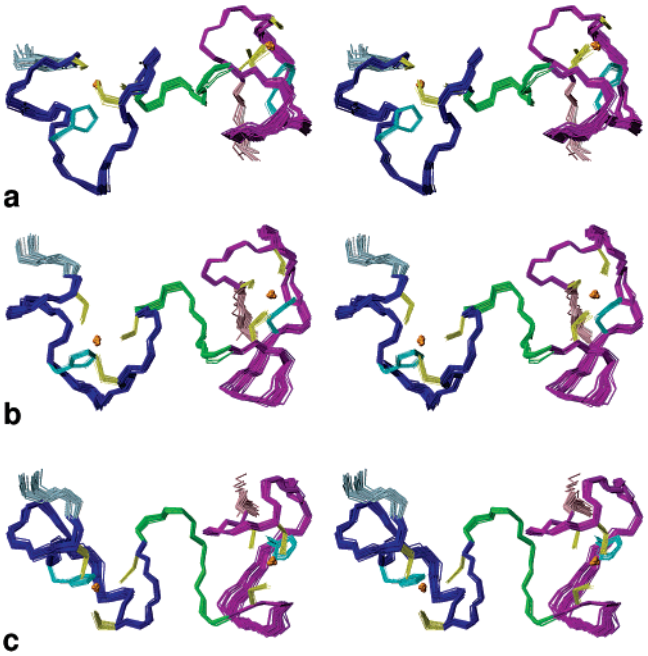

FIGURE 2: Stereo stick representation of the 50 lowest-energy structures from the final ZF37 bundle with the backbone colored by sequence domain. The first two residues, the *N*-terminal CCHC motif, the central linker, the *C*-terminal CCHC motif, and the two last residues are colored pink, magenta, green, blue, and light blue, respectively. Zn-binding side chains are also represented, with yellow and cyan thinner sticks for Cys and His, respectively. Orange spheres indicate Zn ions. The three views are obtained by 45° rotation around the horizontal axis of the figure.

His rings through their $H^{\delta}1$ protons to amide O atoms from the backbone. The two H bonds exhibit different residue patterns and stabilities between the two domains: His^{32} *N*-terminal $H^{\delta 1}$ is H bonded in 84% of the final structures, forming either $i + 2$ (with Pro^{30} O, 26%) or $i + 3$ (with Lys^{29} O, 58%) interactions, whereas *C*-terminal His^{51} $H^{\delta 1}$ forms a stable $i + 3$ H bond with Gln^{48} O in 100% of the bundle structures. This finding could explain how critical this His position is within CCHC motifs for nucleic acid-binding affinities and conformational stability, as observed in the NMR structure and RNA-affinity characterization of CCCC and CCHH point mutants of HIV-1 Ncp7 protein (*45*).

Thus, the overall conformational features of ZF37 can be summarized as follows. (1) Linker folding is provided by H-bonded secondary structure elements, which also ensure the anchoring of the two CCHC domains (bridge-1 involving *N*-terminal CCHC motif and type I/III β-turn, including the first Cys of *C*-terminal CCHC), resulting in an Ω-shaped central region, asymmetrically centered on the linker, spanning residues Arg^{38}−Lys^{45} (Figure 3).

(2) The two CCHC motifs exhibit some local fold differences, resulting in a rmsd value of 1.93 ± 0.08 Å on backbone atoms from residues Cys^{24}−Cys^{37} and Cys^{43}−Cys^{56} of the final structures (the side chains of Cys and His residues and zinc atoms included) and in the different secondary structures and H-bond patterns: one bridge-1 and two type I β-turns, altogether spanning the Thr^{23}−Ser^{35} range, in the *N*-terminal domain versus a short helix, strongly locked at both ends by two distorted β-turns and a *C*-terminal 5-turn, globally involving the Pro^{49}−Arg^{57} region in the *C*-terminal domain. These differences may be a result of either the intrinsic conformational preferences or the interaction with the linker and the other CCHC motif. The observed similarity in the chemical shifts of the two isolated CCHC motifs in ZF18 and ZF20 with the corresponding domains in ZF37 suggests a preeminent role for the intrinsic conformational properties of each CCHC motif, the different observed patterns of Pro and Gly residues possibly providing a substantial contribution to these differences. In particular, Lys^{47}, replacing the Gly residue observed in the ZF37 *N*-terminal finger and in many other CCHC motifs, exhibits ϕ, ψ values rather unfavorable for a L-amino acid, inducing some local structural distortion.

(3) The two zinc knuckles show differences in intrinsic flexibility as suggested by the unrestrained MD simulation (UEWMD) whose results, however, correlate well with the observed dispersion in the RIWSA/REM structures (Figure 3). In particular, the residues in the third loop (HX_4C) of the first knuckle appear to be more flexible than the corresponding residues of the second knuckle.

(4) The appreciable convergence of the UEWMD structures, showing no substantial relative reorientation of the two CCHC motifs even after 3 ns of unrestrained MD, their similarity to those from both RIWSA/REM and REWMD ensembles (rmsd on backbone atoms plus the side chains of Cys and His and Zn atoms of 1.13 and 1.09 Å, respectively), the total compliance of predicted distances with full spectral

Table 2: Comparison of Turn/Helix Distribution in ZF37, 1ESK, 1MFS, and 1NC8 Structures[a]

residue range		ZF37		1ESK		1MFS		1NC8	
start	end	turn	%	turn	%	turn	%	turn	%
I(−1)	I(1)							γ^{i}	33
I(1)	***I(4)***	***β^{IV}***	***100***	***β^{IV}***	***100***	***β^{IV}***	***100***	***β^{IV}***	***100***
I(2)	***I(5)***	***β^{IV}***	***35***	***β^{IV}***	***100***	***β^{IV}***	***100***	***β^{IV}***	***66***
I(5)	I(7)							γ^{i}	27
I(6)	I(9)	β^{I}/β^{IV}	67/21			β^{IV}	10		
I(7)	I(9)	γ^{i}	27						
I(10)	I(13)	β^{I}	100	β^{I}/β^{IV}	22/78			β^{IV}	47
I(11)	*I(14)*	*β^{IV}*	*100*	*β^{IV}*	*89*	*β^{III}*	*100*	*β^{I}*	*100*
I(12)	I(15)	β^{IV}	37						
I(13)	I(15)							γ^{i}	33
I(14)	I(17)	β^{IV}	9			β^{IV}	30		
II(−2)	II(1)	β^{IV}	100						
II(−1)	II(2)	β^{IV}	36						
II(1)	***II(4)***	***β^{IV}***	***100***	***β^{IV}***	***100***	***β^{IV}***	***100***		
II(2)	***II(5)***	***β^{IV}***	***35***	***β^{I}/β^{IV}***	***11/89***	***β^{IV}***	***100***		
II(6)	II(8)					γ^{i}	57		
II(6)	***II(9)***	***β^{IV}***	***100***	***β^{IV}***	***44***	***β^{IV}***	***80***		
II(7)	II(9)			γ^{i}	56				
II(7)	II(10)			β^{IV}	100				
II(8)	II(10)					β^{IV}	23		
II(9)	II(12)	β^{III}	100						
II(10)	II(13)				β^{IV}	100	β^{III}	100	
II(11)	II(14)				β^{IV}	100			
I(12)	I(15)	β^{IV}	72						
I(13)	I(16)	β^{IV}	100						
II(14)	II(17)				β^{IV}	44	β^{IV}	43	

[a] Types of turn/helix, alongside with corresponding percentages of observed secondary structure within each bundle are reported. To allow comparison among structures, donor and acceptor residues are labeled according to their position within CCHC motifs. A roman numeral indicates the *N*-terminal (I) or *C*-terminal (II) motif, whereas the number in parentheses corresponds to the residue position within its CCHC motif (assigning 1 to the first Cys residue). Only elements including at least one residue in a CCHC motif are shown. Boldface italic type indicates secondary structures conserved in all analyzed structures, whereas lightface italic type is used when only turn types are conserved but subtypes are different.

Table 3: Comparison of Main H-Bond Distribution in ZF37, 1ESK, 1MFS, and 1NC8 Structures[a]

H bond		ZF37	1ESK	1MFS	1NC8
H- - - - - - - - - -	-O	%	%	%	%
I(1)	*I(−1)*				87
I(1)	*I(6)*	***91***	***66***	***100***	***60***
I(1)	*I(9)*				40
I(2)	*I(0)*			90	
I(2)	*I(16)*	*66*			
I(5)	*I(1)*			100	100
I(6)	*I(1)*				40
I(7)	*I(5)*				33
I(8)	*I(−1)*	***15***	***100***	***100***	***100***
I(8)	*I(6)*				33
I(9)	*I(−1)*	*26*			100
I(9)	*I(6)*	*5*	56	63	
I(9)	*I(6)*	*58*			
I(9)	*I(7)*	*71*			
I(9)	*I(7)*	*26*			
I(13)	*I(10)*	99		100	100
I(14)	*I(10)*	*85*	100		100
I(14)	*I(11)*			100	33
I(15)	*I(12)*			93	
I(15)	*I(13)*				80
I(18)	*I(0)*	100			
I(19)	*I(3)*		66	-	-
II(1)	*I(−2)*	100			
II(1)	*II(6)*		100	100	
II(1)	*II(9)*		44	30	
II(2)	*II(0)*			40	
II(5)	*II(1)*		78	100	
II(7)	*II(5)*			66	
II(8)	*II(6)*			77	
II(9)	*II(6)*		100		
II(9)	*II(7)*			56	
II(11)	*II(0)*			66	
II(12)	*II(9)*		100		
II(13)	*II(10)*		100		100
II(14)	*II(10)*	*45*	00		
II(14)	*II(11)*			100	
II(15)	*II(10)*	*100*			
II(15)	*II(13)*			87	

Percentages of observed H bonds within each bundle are reported. To allow for comparison among structures, donor and acceptor residues are labeled according to their position within the CCHC motifs. A roman numeral indicates the *N*-terminal (I) or *C*-terminal (II) motif, whereas the number in parentheses corresponds to the residue position within its CCHC motif (assigning 1 to the first Cys residue). Only H-bonds including at least one residue in a CCHC motif are shown. Boldface italic type indicates H-bonds conserved in all analyzed structures. All acceptors and donors are backbone amide O and H atoms, respectively, except for underlined residues, where the donor is $H^{\delta 1}$ from the His side chain.

information, together strongly suggests that the final RIWSA/REM bundle is representative of a largely predominant ZF37 conformer in solution. This result, the observed inter knuckle interaction pattern and the fact that the chemical shifts found in ZF37 are quite similar to those of the corresponding atoms in the isolated zinc fingers suggest that in the experimental conditions used the two zinc fingers are weakly in contact, mainly by interactions mediated by the linker rather than by direct contact between the two zinc-fingers, but in a definite relative orientation to each other.

Comparison with Other Nucleocapsid Proteins. The sequence of the ZF37 peptide was preliminary compared using PSI-BLAST (*46*) with other known NC protein sequences. It appears to be highly conserved among EIAV strains and is quite similar (about 50−60% of identity) to sequences of corresponding proteins of other lentiviruses.

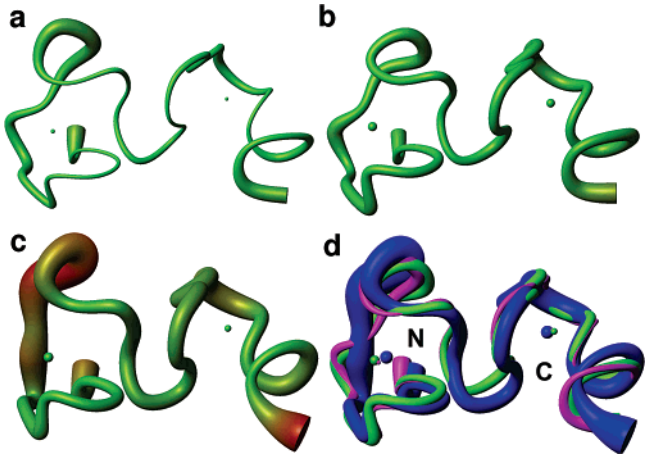

FIGURE 3: Representation of positional fluctuations in RIWSA/REM (100 structures), REWMD (50 structures), and UEWMD (50 structures) bundles. Tubes and spheres, whose radii correspond to positional rmsd values, represent the ZF37 backbone and Zn ions, respectively, of RIWSA/REM (a), REWMD (b), and UEWMD (c). Colors vary from green to red for increasing local fluctuation amplitudes. A superposition of the tubes/spheres (magenta: RIWSA/REM, green: REWMD, blue: UEWMD) is also shown (d) with *N*- and *C*-terminal domains labeled accordingly.

We compared the ZF37 sequence and structure with those of available NC proteins belonging to the same CCHC family as ZF37 in the pdb database (*47*), whose structures were determined in the absence of nucleic acids to investigate sequence/structure relationships corresponding to intrinsic conformational preferences of viral NC proteins/peptides. To ease comparisons among different sequences, the position within or adjacent to the CCHC motif will be named *N(x)*, where *N* is a roman number indicating *N*- (*I*) or *C*-terminal (*II*) CCHC motif, and *x* is a number assuming the value of 1 for the first Cys residue of each motif and a values less than 1 for residues preceding the first Cys.

Figure 4 shows a manual sequence alignment of our sequence (labeled ZF37) with three different strands of HIV-1 (1MFS (*48*), 1AAF, and 1ESK); HIV-2 (1NC8); Mason-Pfizer monkey virus (MPMV) (1CL4); and murine mammary tumor virus (MMTV) (corresponding to two different pdb entries, 1DSQ, containing the *N*-terminal zinc-finger domain and a part of the linker, and 1DSV, containing the rest of the linker plus the *C*-terminal zinc finger domain).

ZF37 includes the shortest linker among structurally determined retroviral NC proteins because the length of the linkers ranges from 5 (EIAV) to 15 (MPMV) residues. A higher conservation is observed in the *N*-terminal than in the *C*-terminal domains, paralleled (see below) by a higher conservation in 3D structures. ZF37 also shows a peculiar distribution of charged and, to a lesser extent, hydrophobic residues. In fact, the ZF37 *C*-terminal Zn-finger domain contains two aromatic residues, *II(2)* and *II(10)*, whereas only one aromatic residue is detected in both HIV-1 (*II(2)*) and MPMV or MMTV (*II(10)*). Even larger differences occur in the distribution of charged residues, with ZF37 showing by far the lowest number of charged residues within the aligned set and, in particular, no Asp or Glu residues and only one and three basic residues within *N*- and *C*-terminal Zn-fingers, respectively, with one and two more basic residues in the two-residue sequences at both ends of each domain, respectively. In addition to obvious zinc-binding

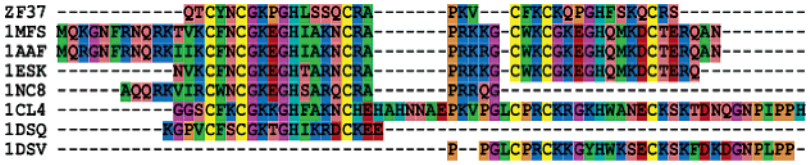

FIGURE 4: Manual alignment of sequences from viral CCHC NC proteins whose 3D structure has been determined in the absence of the nucleic acid. PDB codes are reported on the left, and the following color code for conserved residues was used: yellow, C; cyan, H; blue, K,R; red, D,E; pink, N,Q,S,T; violet, G; orange, P; green, A,F,I,L,M,V,W,Y.

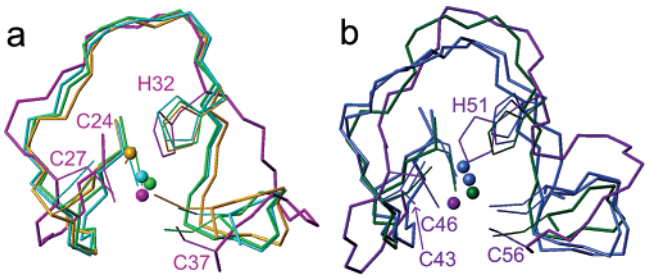

FIGURE 5: Best fit superposition on all shown atoms of *N*-terminal (a) and *C*-terminal (b) CCHC motifs from several retroviral NC proteins. The *C*-terminal motifs are colored in shade that are darker than those of corresponding *N*-terminal ones. According to this scheme, EIAV (ZF37, magenta/dark violet), HIV-1 (1MFS, green/dark green), HIV-2 (1NC8, orange/-), and other retroviruses (1DSQ for *N*-terminal, 1DSV and 1CL4 for *C*-terminal, cyan/blue). Backbone bonds of CCHC residues, side chain heavy atom bonds of Zn-binding residues, and Zn atoms are shows as sticks, thin sticks, and spheres, respectively. Zinc-binding residues in ZF37 are labeled in magenta/dark violet.

residues, a strong conservation is also observed for the Gly-Lys *I(5)*−*I(6)* sequence and for the Gly *I(8)* residue. Also in this view, the *N*-terminal Zn-binding domains appear by far more conserved than the *C*-terminal ones. In particular, the two aforementioned Gly residues in the *N*-terminal domain are absolutely conserved within the aligned sequence set, although the number and position of Gly residues vary in the *C*-terminal regions.

ZF37 considerably extends the structural characterization of the C-X_2-C-X_4-H-X_4-C class of Zn-binding motifs. In fact, ZF37 represents the first resolved structure in this class including at least these novel features: (a) in general, the first case of a Pro residue inside a *N*-terminal CCHC motif; (b) a Pro-Gly *I/II(7)*−*I/II(8)* sequence in both CCHC domains; and (c) in particular, the whole Gly-Lys-Pro-Gly (*I(5)* to *I(8)*) sequence. The latter, while occurring rather frequently in either the *N*- or the *C*-terminal or both knuckles of other nonprimate lentiviral NC proteins, is also interestingly detected in NC proteins from several strains of both HIV-2 and SIV (especially in its SIV(cpz) (also known as CIV) variant) as well as in several human and simian proteins.

A systematic analysis of the positional rmsd values of both ZF37 fingers versus several NC proteins shows that ZF37 exhibits new CCHC motif spatial arrangements (Figure 5) as witnessed by rmsd values against other motifs, on average apporximately 2 Å (for *N*-terminal) and 3 Å (for *C*-terminal ZF37), to be compared with about 0.6 and 1.5 Å observed within the HIV−NC subfamily and among other different NC protein subfamilies, respectively.

A PROMOTIF comparison among ZF37, 1ESK, 1MFS, and 1NC8 allowed the characterization of these differences as well as the identification of common secondary structural submotifs (Tables 2 and 3). The bridge-1 involving the *I-(0)*−*I(7)* range (corresponding to the second bridge-1 motif in ZF37) occurs in all analyzed structures as well as in the *I(10)*−*I(14)* 4-turn. No similar conservation is observed in *C*-terminal motifs, although the 5-turn in the ZF37 *II(10)*−*II(15)* range also occurs in 1NC8 *I(10)*−*I(15)*, probably favored by a high local sequence similarity (FSKQCR in ZF37 vs SARQCR in 1NC8).

Observed rmsd values mainly result from short- to medium-range secondary structures and H-bond patterns, appreciably differing even in *N*- and *C*-terminal structures of 1ESK versus 1MFS (Tables 2 and 3), which just exhibit minor differences in sequence and experimental conditions.

H-bond distribution (Table 3) includes only two absolutely conserved interactions, both occurring in the *N*-terminal bridge-1 structure CCHC motif. The H bonds associated with well-defined and stable β-turns (namely β^{I} and β^{III} types in our ensembles) are also detected, although, at best, in three out of four or two out of three molecules in *N*- and *C*-terminal domains, respectively. Stable *I(0)-I(7)* bridge-1 corresponds to a rather conserved sequence. However, in the rest of the CCHC motif, local secondary structure conservation appears to depend more upon the balancing of environmental conditions, the length of analyzed peptides, or the elongation of *N*- and/or *C*-termini than upon mere homology as shown, for example, by the six H bonds and three turn types for which 1ESK or 1MFS exhibit a higher similarity to ZF37 or 1NC8 than to each other.

DISCUSSION

All structural studies on retroviral NC proteins evidenced few common major points, such as the role of each zinc knuckle in binding interactions and the dynamical behavior of these proteins.

Earlier studies on HIV-1 NCp7 led to the conclusion that the two zinc knuckles are not functionally equivalent and that *N*-terminal zinc finger has a more prominent role in RNA selection and packaging (*49*). However, recent structural data indicate that both knuckles have an important role in specific recognition of different stem-loops (*13*, *14*). Even considering the reported differences between corresponding knuckles of NCp7 and NCp11, linker folding and spatial proximity of the two zinc binding domains in NCp11 suggest that both zinc knuckles should contact nucleic acids and have both an important role in binding nucleic acids as in NCp7.

As for Zn binding, it is noteworthy that the simultaneously higher rigidity and lower zinc affinity in the *C*-terminal rather than in the *N*-terminal zinc knuckle determined for EIAV NCp11, in comparison with the lower rigidity/lower affinity reported for HIV-1 NCp7 *C*- terminal versus *N*-terminal knuckle. The observed combination of higher conformational rigidity/lower zinc affinity in the ZF37 *C*-terminal region seems to suggest either a less optimal orientation or a more

strained arrangement for Zn ligands in this domain rather than in the *N*-terminal region. This result can be related to the lack of a Gly residue in the ZF37 *C*-terminal domain, which is replaced by a Lys moiety, exhibiting a rather unfavorable conformation for a bulky L-amino acid.

The combination of the shortest linker region among NC proteins so far investigated and the experimental conditions used for NMR spectra allowed the identification of a preferred conformation in solution for EIAV NCp11 ZF37 peptide, characterized by a folded structure for the basic linker. In fact, a type-I β-turn is present with a hydrogen bond between Pro^{40} C═O and the amide proton of Cys^{43}. Other hydrogen bonds are formed between the residues of the *C*-terminal zinc knuckle and the residues in the linker. Linker folding is additionally stabilized by inter-knuckle side chain van der Waals contacts. The substantial stability of this conformation in water is also supported by the results of unrestrained MD simulations. This finding is in agreement with the results of Ramboarina et al. (*50*) on NCp7 dynamics reporting the predominance of structures with a folded linker below the transition temperature of 303 K. Because other lentiviral NC proteins with a 5- or 4-residue linker are known, we suggest that interactions between the knuckles could occur in these proteins too. However, such inter-knuckle interactions seem to only marginally affect the conformational preferences of each CCHC motif. Our results, in any case, do not exclude the fact that only transient interactions could be observed in different experimental conditions.

The conformational behavior of the different CCHC motifs appears to depend on a fine tuning of different intrinsic (i.e., sequence-dependent) versus environmental factors. Despite this, some emerging conformational properties within structurally determined retroviral NC proteins as well as the critical role of some conformationally peculiar residues such as Gly and Pro can be evidenced. In this view, ZF37 exhibits new patterns of these types of residues in both CCHC motifs, which results in the differences in the local fold within the -Cys-X_2-Cys-X_4-His-X_4-Cys- subfamily of CCHC motifs. Thus, the emerging overall picture for the conformational behavior of the CCHC motifs in NC proteins is that of a loose conservation of structural submotifs. In fact, zinc-binding restraints and homology in CCHC motifs determine a substantial preservation within the analyzed structure set of coarse-grained structural parameters, such as bends or turns, especially in *N*-terminal CCHC motifs, and some less trivial medium-long and long-range secondary structure patterns. The overall result is a sequence of short elements whose exact type and position vary within a rather large set of combinations, restricted by both the Zn coordination pattern and, to an extent still to be exactly determined, few medium-range secondary structures, such as the bridge-1 and 4-turn identified in the present analysis. This feature may represent a flexible strategy to (1) accommodate sequence variability of either NC proteins themselves or their complementary nucleic acids, a very critical feature for rapidly mutating lentiviruses; (2) provide the required flexibility for the different biological function and aggregation states in which NC proteins are involved; and (3) still preserve the main, critical conformational features of CCHC motifs. In this view, the distribution of Pro and Gly residues appears to behave as a harder conformational switch than other variable sequence features occurring in structurally determined NC proteins.

Design of molecules capable of specifically inactivating NC protein functions is of relevant pharmaceutical interest. Starting from the known structure and binding properties of HIV-1 Ncp7, molecules with antiretroviral activity have already been derived (*51*, *52*). However, one of the problems experienced with anti-HIV drugs is the viral resistance due to the mutations of retroviral proteins. Therefore, it is of great importance to identify those molecular structures and surfaces that are invariantly conserved among different strains of a certain lentivirus and/or among different lentiviruses. Our findings seem to indicate that conserved structural features can be identified in lentiviral NC proteins, although further comparative and systematic analyses are required.

SUPPORTING INFORMATION AVAILABLE

Peptide synthesis addendum, structure calculation addendum, UV−vis absorption spectrum of ZF18−Co^{2+} and ZF20−Co^{2+}complexes and titration data, CD spectra of peptides and peptide-zinc complexes, fingerprint region of a NOESY spectrum with a 0.150 s mixing time at 298 K of the ZF37 peptide-Zn^{2+} complex, average values (degrees) and fluctuations (rmsd (rmsd, degrees) and angular order parameter (S)) of backbone ϕ and ψ angles for the final ZF37 structure bundle, a summary of PROMOTIF comparative analysis of ZF37, 1ESK, 1MFS, and 1NC8, superposition matrix of *N*- and *C*-termini zinc-finger motifs from different NC proteins, the distribution of acidic and basic residues in CCHC Zn-finger domains of structurally determined viral NC proteins, and ^{1}H NMR chemical shifts of the zinc finger peptide ZF37. This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES

1. Haase, A. T. (1986) Pathogenesis of lentivirus infections, *Nature (London) 322*, 130−136.
2. 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.
3. Roques, B. P., Morellet, N., de Rocquigny, H., Demene, H., Schueler, W., and Jullian, N. (1997) Structure, biological functions and inhibition of the HIV-1 proteins Vpr and NCp7, *Biochimie 79*, 673−680.
4. Demene, H., Jullian, N., Morellet, N., de Rocquigny, H., Cornille, F., Maigret, B., and Roques, B. P. (1992) Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ^{1}H NMR, *EMBO J. 11*, 3059−3065.
5. Omichinski, J. G., Clore, G. M., Sakaguchi, K., Appella, E., and Gronenborn, A. M. (1991) Structural characterization of a 39-residue synthetic peptide containing the two zinc binding domains from the HIV-1 p7 nucleocapsid protein by CD and NMR spectroscopy, *FEBS Lett. 292*, 25−30.
6. Summers, M. F, Henderson, L. E., Chance, M. R., Bess, J. W., Jr., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C., III, Hare, D. R., and Arthur, L. O. (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.
7. 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 ^{1}H NMR, *EMBO J. 11*, 3059−3065.
8. Kodera Y., Sato K., Tsukahara T., Komatsu H., Maeda T., and Kohno T. (1998) High-resolution solution NMR structure of the

minimal active domain of the human immunodeficiency virus type-2 nucleocapsid protein, *Biochemistry 37*, 17704−17713.
9. Klein, D. J., Johnson, P. E., Zollars, E. S., De Guzman, R. N., and Summers, M. F. (2000) The NMR structure of the nucleocapsid protein from the mouse mammary tumor virus reveals unusual folding of the *C*-terminal zinc knuckle, *Biochemistry 39*, 1604−1612.
10. Gao, Y., Kaluarachchi K., and Giedroc D. P. (1998) Solution structure and backbone dynamics of Mason-Pfizer monkey virus (MPMV) nucleocapsid protein, *Protein Sci. 7*, 2265−2280.
11. Demene, H., Jullian, N., Morellet, N., de Rocquigny, H., Cornille, F., Maigret, B., and Roques, B. P. (1994) Three-dimensional ^{1}H NMR structure of the nucleocapsid protein NCp10 of Moloney murine leukemia virus, *J. Biomol. NMR 4*, 153−170.
12. Morellet, N., Demene, H., Teielleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournié-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.
13. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element, *Science 279*, 384−388.
14. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the ψ-RNA packaging signal. Implications for genome recognition, *J. Mol. Biol. 301*, 491−511.
15. Jewell, N. A., and Mansky, L. M. (2000) In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly, *J. Gen. Virol. 81*, 1889−1899.
16. Leroux, C., Cadore, J. L., and Montelaro, R. C. (2004) Equine infectious anemia virus (EIAV): what has HIV's country cousin got to tell us?, *Vet. Res. 35*, 485−512.
17. Olsen, J. C. (2001) EIAV, CAEV and other lentivirus vector systems, *Somatic Cell Mol. Genet. 26*, 131−145.
18. Azzouz, M., and Mazarakis, N. (2004) Non-primate EIAV-based lentiviral vectors as gene delivery system for motor neuron diseases, *Curr. Gene Ther. 4*, 277−286.
19. O'Rourke, J. P., Olsen, J. C., and Bunnell, B. A. (2005) Optimization of equine infectious anemia derived vectors for hematopoietic cell lineage gene transfer, *Gene Ther. 12*, 22−29.
20. 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.
21. Karpel, R. L., Henderson, L. E., and Oroszlan, S. (1987) Interactions of retroviral structural proteins with single-stranded nucleic acids, *J. Biol. Chem. 262*, 4961−4967.
22. Lapadat-Tapolsky, M., De Rocquigny, H., Van-Gent, D., Roques, B., Plasterk, R., and Darlix, J. C. (1993) Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle, *Nucleic Acids Res. 21*, 831−839.
23. Matsrafi, L., Aflalo, E., Blair, D. G., and Priel, E. (1996) Isolation of an 11-kDa protein associated with the topoisomerase I activity from equine infectious anemia virus, *Biochem. Biophys. Res. Commun. 220*, 1028−1035.
24. Gorelick, R. J., Henderson, L. E., Hamser, J. P., and Rein, A. (1988) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a "zinc finger-like" protein sequence, *Proc. Natl. Acad. Sci. U.S.A. 85*, 8420−8424.
25. Meric, C., and Spahr, R. F. (1986) Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging, *J. Virol. 60*, 450−459.
26. Prats, A. C., Sarih, L., Gabus, C., Litvak, S., Keith, G., and Darlix, J. L. (1988) Small finger protein of avian and murine retroviruses has nucleic acid annealing activity and positions the replication primer tRNA onto genomic RNA, *EMBO J. 7*, 1777−1783.
27. Bavoso, A., Ostuni, A., Battistuzzi, G., Menabue, L., Saladini, M., and Sola, M. (1998) Metal ion binding to a zinc finger peptide containing the Cys-X2-Cys-X4-His-X4-Cys domain of a nucleic acid binding protein encoded by the Drosophila Fw-element, *Biochem. Biophys. Res. Commun. 242*, 385−389.
28. Berg, J. M., and Merkle D. L. (1989) On the metal ion specificity of zinc finger proteins, *J. Am. Chem. Soc. 111*, 3759−3761.
29. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wuetrich, K. (1983) Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering, *Biochem. Biophys. Res. Commun. 117*, 479−485.
30. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy, *J. Magn. Reson. 53*, 521−528.
31. Davis, D. G., and Bax, A. (1985) Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartman-Hahn spectroscopy, *J. Am. Chem. Soc. 107*, 2820−2821.
32. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys. 71*, 4546−4553.
33. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR 2*, 661−665.
34. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants, *J. Magn. Reson. 48*, 286−292.
35. Marion D., and Wüthrich K. (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin−spin coupling constants in proteins, *Biochem. Biophys. Res. Commun. 113*, 967−74.
36. Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer R., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., Kollman, P. A. (1999) *AMBER 6*, University of California, San Francisco, CA.
37. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M. Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *J. Am. Chem. Soc. 117*, 5179−5197.
38. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graph. 14*, 51−55.
39. Chance, M. R., Sagi, I., Wirt, M. D., Frisbie. S. M., Scheuring, E., Chen, E., Bess, J. W. Jr., Henderson, L. E., Arthur, L. O., South, T. L., Perez-Alvarado, G., and Summers M. F. (1992) Extended X-ray absorption fine structure studies of a retrovirus: equine infectious anemia virus cysteine arrays are coordinated to zinc, *Proc. Natl. Acad. Sci. U.S.A. 89*, 10041−10045.
40. South, T. L., Blake, P. R., Sowder, R. C., Arthur, L. O., Henderson, L. E., and Summers, M. F. (1990) The nucleocapsid protein isolated from HIV-1 particles binds zinc and forms retroviral-type zinc fingers, *Biochemistry 29*, 7786−7789.
41. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, J. Wiley, New York.
42. Güntert, P., Mumenthaler, M., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
43. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
44. Hutchinson, E. G., and Thornton, J. M. (1996) "PROMOTIF − A program to identify and analyze structural motifs in proteins", *Protein Sci. 5*, 212−220.
45. 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.
46. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman D. J. (1997) Gapped BLAST and PSI−BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
47. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank, *Nucleic Acids Res. 28*, 235−242.
48. Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) Dynamical behavior of the HIV-1 Nucleocapsid Protein, *J. Mol. Biol. 279*, 633−649.
49. Gorelick, R. J., Chabot, D. J., Rein, A., Henderson, L. E., and Arthur, L. O. (1993) The two zinc fingers in the human immunodeficiency virus type 1 nucleocapsid protein are not functionally equivalent, *J. Virol. 67*, 4027−4036.
50. Ramboarina, S., Srividya, N., Atkinson, R. A., Morellet, N., Roques, B. P., Lefevre, J. F., Mely, Y., and Kieffer, B. (2002) Effects of temperature on the dynamic behaviour of the HIV-1 nucleocapsid NCp7 and its DNA complex, *J. Mol. Biol. 316*, 611−627.

51. 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.

52. Druillennec, S., and Roques, B. P. (2000) HIV-1 NCp7 as a target for the design of novel antiviral agents, *Drug News & Perspect. 13*, 337−349.

BI0524924