The EMBO Journal vol.11 no.8 pp.3059-3065, 1992

Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by ^{1}H NMR

Nelly Morellet, Nathalie Jullian, Hugues De Rocquigny, Bernard Maigret[1], Jean-Luc Darlix[2] and Bernard P.Roques[3]

Département de Chimie Organique, U 266 INSERM, UA 498 CNRS, UFR des Sciences Pharmaceutiques et Biologiques, 4 avenue de l'Observatoire, 75006 Paris, [1]Université Nancy I, Laboratoire Chimie Théorique, BP 239, 54506 Vandoeuvre-Les-Nancy Cédex and [2]Ecole Normale Supérieure, Laboratoire Retro-CJF-INSERM, Laboratoire de Biologie Moléculaire et Cellulaire, 46 allée d'Italie, 69364 Lyon Cedex 07, France

Communicated by K.Wüthrich

[3]Corresponding author

The retroviral gag nucleocapsid protein NCp7 (72 amino acids) of HIV-1 (LAV strain), which contains two successive zinc fingers of the Cys-X_2-Cys-X_4-His-X_4-Cys form linked by a stretch of basic residues, promotes viral RNA dimerization and encapsidation and activates annealing of the primer tRNA to the initiation site of reverse transcription. The structure of NCp7 and other shorter fragments was studied by 600 MHz ^{1}H nuclear magnetic resonance (NMR) in aqueous solution to account for its various biological properties. Complete sequence specific ^{1}H NMR assignments of the 13-51 residues of NCp7 encompassing the two zinc fingers was achieved by two-dimensional NMR experiments and the three-dimensional structure of (13-51)NCp7 was deduced from DIANA calculations, using nuclear Overhauser effects as constraints. The structure of the zinc complexed form of NCp7 is characterized by a kink at the Pro31 level in the basic Arg29-Ala-Pro-Arg-Lys-Lys-Gly35 RNA binding linker leading to a proximity of the Lys14-Cys18 to the Gly35-Cys39 sequences, which belong to the folded proximal and distal zinc fingers, respectively. Accordingly, the aromatic residues Phe16 and Trp37 were found to be spatially close. The Lys33 and Lys34 side-chains involved in viral RNA dimerization were solvent exposed. The N- and C-terminal sequences of NCp7 behave as flexible independent domains. The proposed structure of NCp7 might be used to rationally design new anti-viral agents aimed at inhibiting its functions.

Key words: distance geometry/HIV-1/nucleocapsid NCp7/NMR/protein structure

Introduction

The nucleocapsid protein NCp15 of the human immunodeficiency virus type 1 (HIV-1) (Barré-Sinoussi *et al.*, 1983; Wain-Hobson *et al.*, 1985) is derived from the C-terminus of the Pr55gag polyprotein precursor and is ultimately processed into NCp7 and p6 in mature HIV virions (Di Marzo Veronese *et al.*, 1987). In the viral particle, NC protein molecules are a key component of the capsid and are in close contact with the dimeric RNA genome (Darlix *et al.*, 1990). HIV-1 NCp7, like most retroviral NC proteins, is highly basic and contains two motifs of the form Cys-X_2-Cys-X_4-His-X_4-Cys, which is well conserved in retroviruses and retroelements (Berg, 1986; Covey, 1986). This motif has been shown to coordinate a zinc ion and thus is often designated as a zinc finger (Cornille *et al.*, 1990; Green and Berg, 1990). The zinc fingers of the nucleocapsid proteins of HIV-1, Rous sarcoma virus (RSV) and murine leukaemia virus (MuLV) are required for the specific packaging of the dimeric RNA genome into virions (Méric and Spahr, 1986; Gorelick *et al.*, 1988; Méric and Goff, 1989; Aldovini and Young, 1990; Dupraz *et al.*, 1990). *In vitro*, HIV-1 NCp15, like RSV NCp12 and MuLV NCp10, promote both the dimerization of viral RNA and the annealing of the replication primer tRNA onto the viral RNA, two processes that are necessary for the production of infectious virions (Coffin, 1984; Prats *et al.*, 1988, 1990; Barat *et al.*, 1989; Bieth *et al.*, 1990). Moreover, the 72 amino acid NCp7 containing the two zinc fingers linked by a stretch of basic residues with a proline was recently shown to be as active as the intact NCp15 *in vitro* (De Rocquigny *et al.*, 1991) and therefore could be used in biochemical and structural studies.

It has recently been shown that replacement of both the zinc fingers by a Gly-Gly linker did not inhibit the RNA binding and annealing activities of the HIV-1 NCp7 protein (De Rocquigny *et al.*, 1992). However, deletion of the short sequences Val13-Lys and Arg29-Ala-Pro-Arg-Lys-Lys-Gly35 (RAPRKKG) flanking the first finger led to a complete loss of NCp7 functions *in vitro* (De Rocquigny *et al.*, 1992). Thus, studies of the spatial arrangement of the peptide sequences involved in the various functions of NCp7 are of major interest for understanding the mode of action of this small viral protein and will aid in the design of anti-viral agents capable of inhibiting it.

In the present study, NCp7 and NCp7 related fragments (Figure 1) were studied by 600 MHz ^{1}H NMR spectroscopy in aqueous solution. The specific assignment of all resonances belonging to the (13-51) domain of NCp7 was established by the procedures described by Wüthrich (1986). The constraints obtained from NMR were used through distance geometry calculations (DIANA), to generate a three-dimensional model of the (13-51) peptide. This small protein is characterized by a spatial proximity of the two folded zinc fingers induced by the basic linker. This rather globular domain is surrounded by the conformationally independent and flexible N- and C-terminal sequences.

Results and discussion

^{1}H NMR spectrum of HIV-1 NCp7 protein with Zn^{2+} ions in aqueous solution

As compared with the spectrum of NCp7 HIV-1 (Bru stain) in the absence of zinc, the addition of two equivalents of

Zn^{2+} ions resulted in a large scattering of chemical shifts, especially evident for the amide and H-2 and H-4 histidine proton resonances belonging to the zinc fingers (not shown). This was caused by the well known folding of the fingers around the zinc ions (Green and Berg, 1989; Summers *et al.*, 1990; South *et al.*, 1991). Due to the large concentration

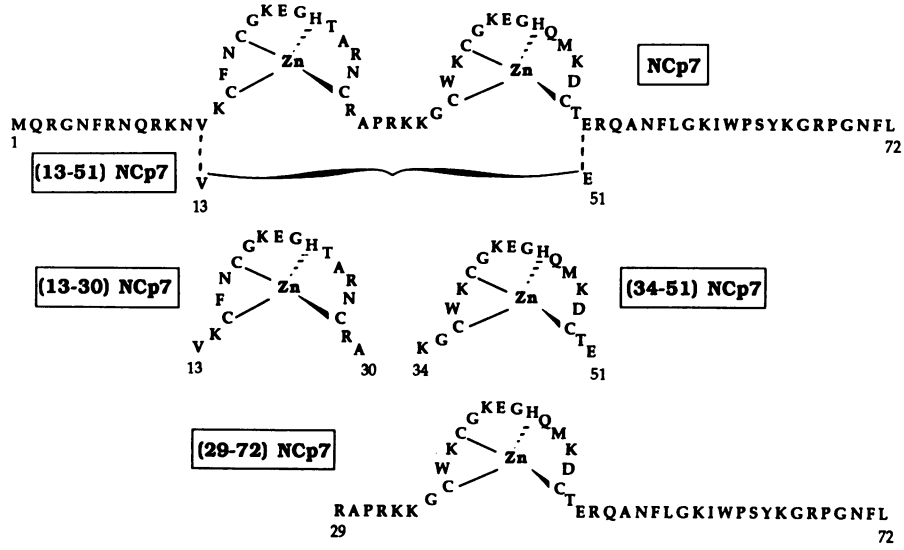


Fig. 1. Primary structures of the NCp7 and related peptides studied.

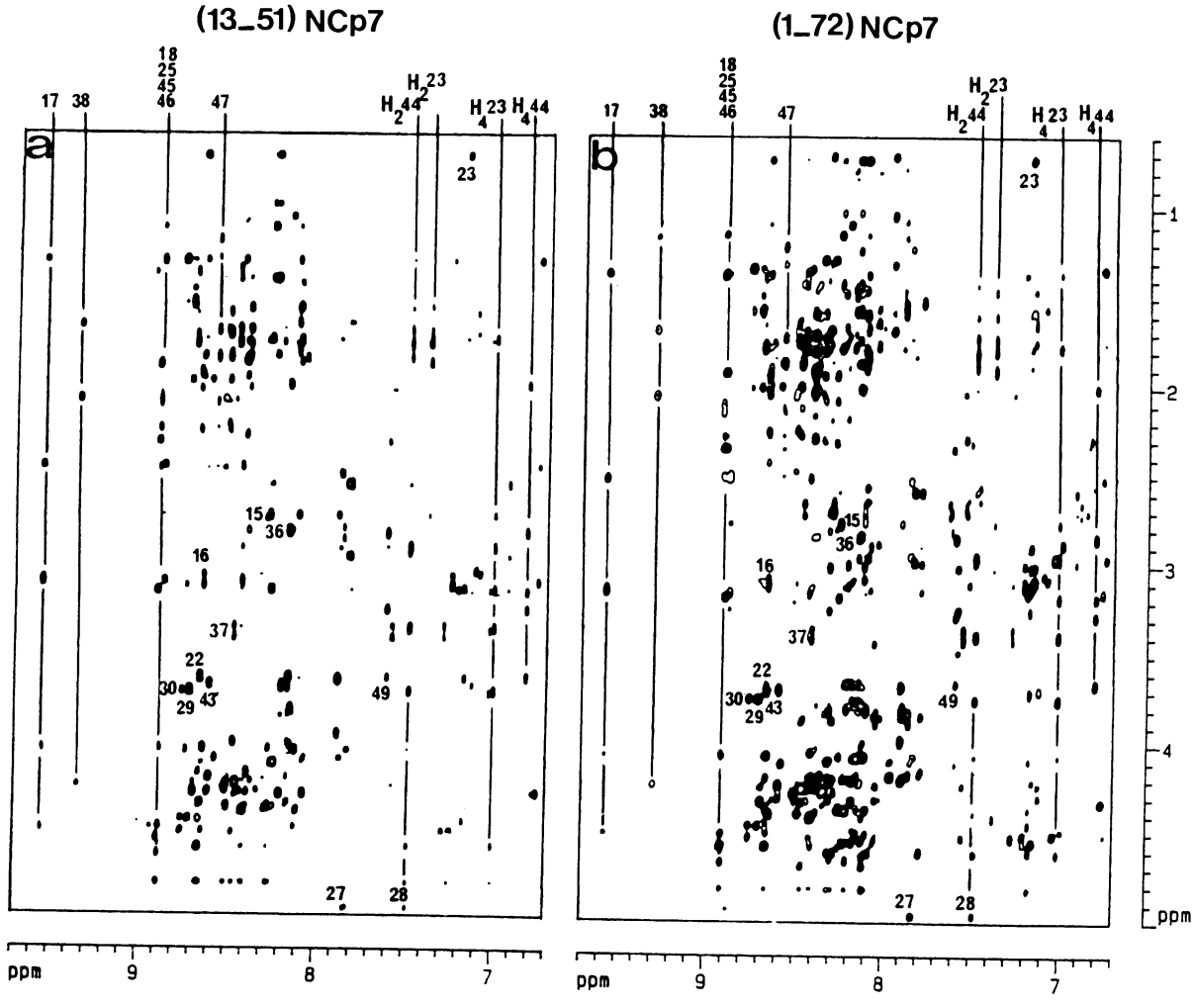


Fig. 2. Part of the two-dimensional NOESY spectra of **(a)** (13−51)NCp7 and **(b)** (1−72)NCp7. Comparison of the 600 MHz NOESY spectra (τm = 200 ms, Te = 293 K, pH = 5.5, in 90% H_2O/10% D_2O) of the two peptides under zinc complexed forms, showing NOE connectivities from signals of amide and aromatic protons to α and side-chain protons. Some groups of resonances which have very similar chemical shifts and similar patterns of NOEs are numbered.

(5 mM) and the high degree of purity (99%) of NCp7 and derivatives used in this study, sharp resonances were observed for all protons, including the exchangeable amide protons.

The complete specific assignment of all protons corresponding to the (13−51) domain of NCp7 (Figure 1) was obtained using the procedures described by Wüthrich (1986). First, the complete spin systems of each amino acid were identified using DQ-filtered COSY (Rance *et al.*, 1983) and TOCSY spectra (Braunschhweiler and Ernst, 1983; Davis and Bax, 1985). The next step was to establish the connectivities across the peptide bonds from the results of NOE experiments (Kumar *et al.*, 1980). In the presence of zinc, the position of the cross peaks of corresponding protons found in the NOESY spectra of the complete NCp7 and in the (13−51)NC fragment were found to be almost identical (Figure 2), suggesting that the amino acids located on either side of the (13−51)NCp7 sequence do not significantly modify the solvated conformation of this domain in native NCp7. This conclusion is supported by the almost unchanged chemical shifts observed for the corresponding amide protons in the spectra of NCp7 (except Phe16 and Trp37, as discussed further), its (13−30) proximal finger and the (29−72) C-terminal fragment, containing the distal finger at pH 5.5 and pH 8.0 (Figure 3). Under both conditions the three peptides were under zinc complexed forms as illustrated by the almost identical position of the His23 and His44 aromatic protons in (13−30)NCp7 (Figures 3a1 and 3a2).

Despite the known increase in the rate of amide proton exchange at basic pH, several slowly exchanging resonances were observed at pH 8.0 in the spectra of the (13−30) zinc finger, the (29−72) NC fragment and NCp7 (Figures 3a2, 3b and 3c) under conditions of selective irradiation of the water resonance during the relaxation period. These amide protons are therefore most probably involved in hydrogen bonds or at least buried from the bulk of the water. The same pattern of hydrogen bonded or buried amide protons was observed in the two fingers of NCp7, suggesting that the same type of folding should occur in both fingers as already suggested for the fingers of the (1−55)NCp7 of HIV-1 MN strain studied alone (South *et al.*, 1991).

The region of slowly exchanging amide protons of NCp7 (Figure 3c) corresponded to the addition of the amide regions of the N-terminal finger and of the (29−72) extended C-terminal domain (Figures 3a2 and b). Nevertheless, there were two exceptions since an additional signal corresponding to the NH of Glu21 was observed in the spectrum of NCp7 (Figure 3c), whereas the signal of the NH of Ala30 present in the first zinc finger (Figure 3a2) had disappeared. Moreover, in NCp7, only the NH of Glu21, but not that of Glu42, was involved in a hydrogen bond despite the presence of both residues in the same (18−23) and (39−44) Cys-Gly-Lys-Glu-Gly-His (CGKEGH) sequences. An important observation was that no additional slowly exchanging protons arising from the (1−12) and (51−72) sequences surrounding the two fingers were observed in the spectrum of NCp7 (Figure 3c), suggesting that the N and C-terminal sequences were less structured than the central (13−51) domain of NCp7. In the latter, the backbone of RAPRKKG did not appear to be stabilized by strong hydrogen bonds in spite of a folding induced by Pro31 (discussed below).

The presence of a *cis−trans* equilibrium around the Trp61-Pro62 bond was indicated by a set of signals from

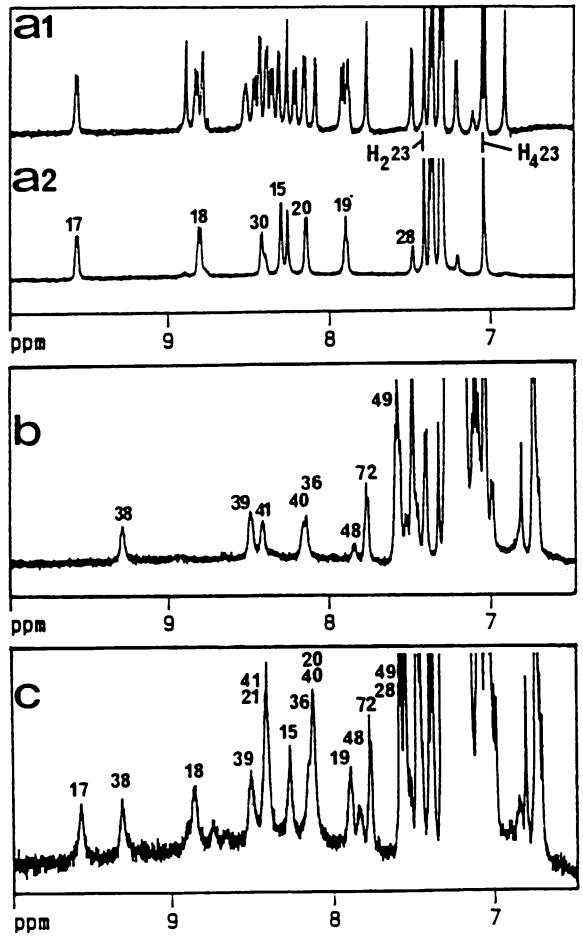


Fig. 3. Slowly exchanging amide protons in zinc complexed (13−30)NCp7 **(a2)**, (29−72)NCp7 **(b)** and NCp7 **(c)**. The spectra were obtained with presaturation of the water signal during the relaxation period in 90% H_2O/10% D_2O (Te = 293 K) at pH 5.5 **(a1)** or pH 8.0 (a2, b and c). The exchanging amide protons present at pH 8.0 are very probably involved in hydrogen bonds or buried from the bulk water.

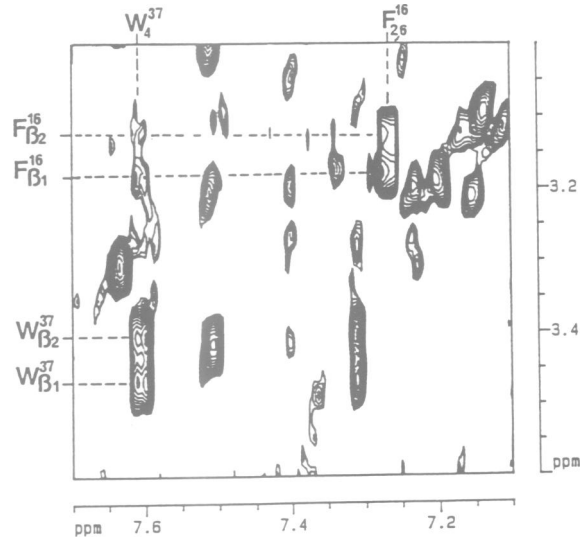


Fig. 4. Part of the two-dimensional NOESY spectra of (13−51)NCp7. (τm = 200 ms) showing the NOE effects between the β protons of Phe16 and H-4 of Trp37 residues belonging to the proximal and distal zinc fingers, respectively.

the 4 and 7 protons of Trp61, and *ortho* and *meta* protons of Tyr64. The Pro62 preferentially adopted a *trans* bond conformation and the position and intensity of Trp61 and Tyr64 signals were almost identical in the (51−72) C-terminal fragment and in NCp7, again indicating the absence of major changes in the conformation of the C-terminal domain of NCp7 in the presence of the complexed zinc fingers.

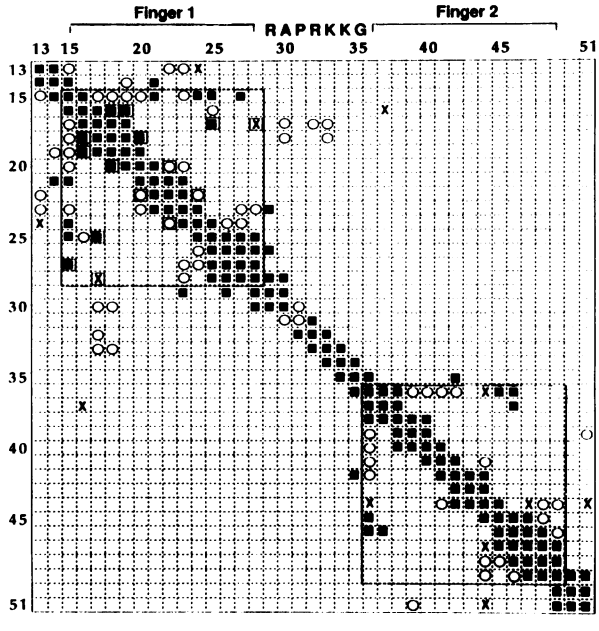


Fig. 5. Diagonal plot of the NOEs observed in (13−51)NCp7. Both axes represent the amino acid sequence. The filled black squares (■) indicate NOEs between backbone protons, the circles (○) indicate NOEs between backbone and side-chain protons and the crosses (X) indicate NOEs between side-chain protons. When two residues are connected by more than one NOE, only the one that involves the largest number of backbone protons is shown (Williamson *et al.*, 1985). Frame squares, circles or crosses indicate NOEs present in first finger only.

Spatial contacts between the two zinc fingers of NCp7

Spatial relationships between the two zinc fingers of NCp7 were indicated by the following data. A comparison of the chemical shifts of the protons of all aromatic residues in NCp7 (not shown) with those of the corresponding residues in the (13−30) proximal, the (34−51) distal zinc finger and the (51−72) C-terminal fragment showed that the signals from the Phe16 and Trp37 residues and the proton H-4 of His44 located in the zinc fingers underwent a slight upfield shift in NCp7. The values ranged from 0.05−0.07 p.p.m. for Trp37 and 0.07−0.08 p.p.m. for Phe16, while the chemical shifts of the other aromatic residues remained almost unmodified (0.02 p.p.m.). This weak shielding effect was most probably due to the mutual ring current effect of Phe16 and Trp37, suggesting that they were closer when the two zinc fingers were linked together. This is supported by the specific changes observed at the level of the Phe16 and Trp37 amide protons, which were shifted upfield by 0.15 and 0.12 p.p.m. respectively in NCp7 while conversely the Hα of Trp37 underwent a deshielding of 0.11 p.p.m. as a result of the spatial proximity of the two aromatic rings. These results suggested that the RAPRKKG linker had induced a specific spatial arrangement of the two zinc fingers in NCp7, an assumption confirmed by the observation of a significant long range NOE between the β protons of Phe16 and the H-4 proton of Trp37 located in the N- and C-terminal zinc fingers, respectively (Figure 4).

Structure of the (13−51) sequence of NCp7

Nuclear Overhauser effects were used as constraints to generate conformations for the (13−51) domain using the DIANA programme. Seven solution structures compatible with NMR data, in particular without violation of long range constraints including those between the β protons of Phe16 and the H-4 aromatic proton of Trp37, were obtained. Superposition of the backbone atoms of the seven structures led to root mean square deviations (r.m.s.d.) of 0.6 ± 0.1

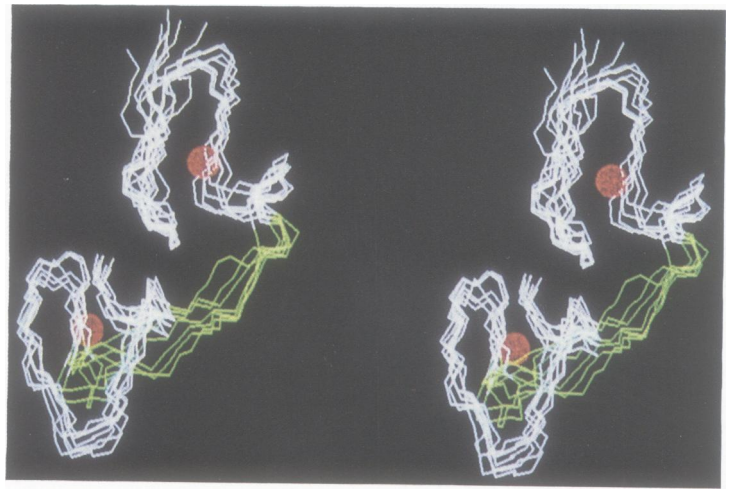
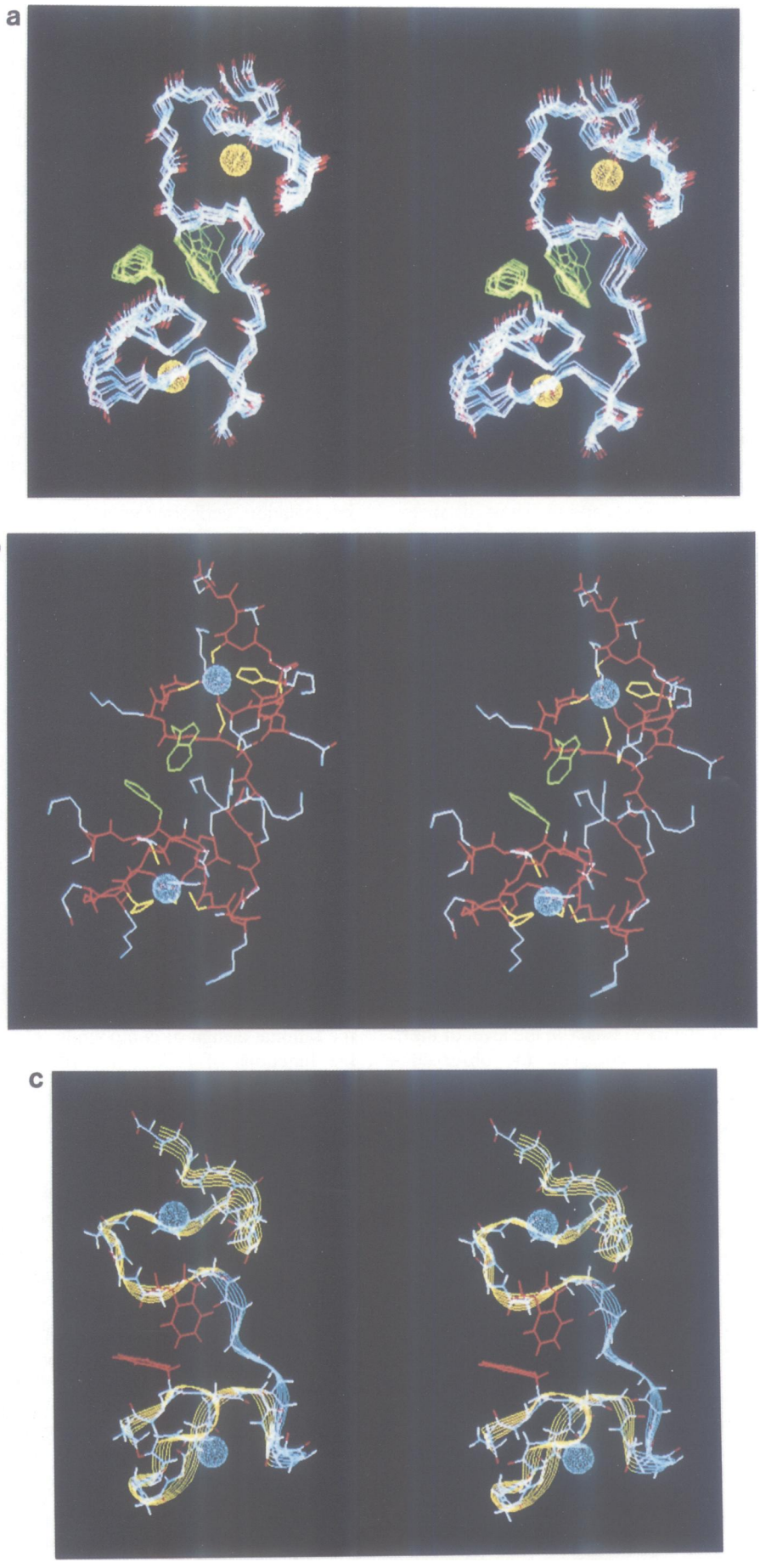


Fig. 6. Stereo view of a family of seven superimposed DIANA structures. The linker region RAPRKKG is in green and the zinc atoms are shown as red Van der Waals spheres. The N-terminal finger is situated below. This is the same for all other representations.

Fig. 7. Conformational behaviour of (13−51)NCp7. **(a)** Superimposition of dynamic conformers. Only backbone atoms and side-chains of Trp37 and Phe16 are shown. Zinc atoms are pictured as yellow van der Waals spheres. **(b)** Stereo view of the (13−51)NCp7 domain showing the spatial proximity of both folded zinc fingers. Backbone atoms are red, aromatics are green and residues involved in zinc binding are yellow. **(c)** Stereo view of a ribbon representation of the (13−51)NCp7. Only side-chain atoms of aromatic residues (red) and of Lys14 and Lys38 (green) are shown.


a
b
c

Å and 0.5 ± 0.2 Å for the proximal and the distal zinc fingers, respectively, whereas the r.m.s.d. of the Arg29-Ala-Pro-Arg-Lys-Lys-Gly35 linker (1.5 ± 0.3 Å) was significantly higher therefore indicating a greater flexibility of the spacer. This could account for the independent behaviour of the two folded domains reported in a recent preliminary NMR study of the (13−51)NCp7 (Omichinski *et al.*, 1991). Interestingly enough, the distance between the aromatic rings of Phe16 and Trp37 was in the range of 4.4 to 6.9 Å. After energy minimization using a zinc force field derived *ab initio* (Jacob, 1990; O.Jacob and B.Maigret, in preparation) and parameterized in AMBER (Singh *et al.*, 1986), all the structures had similar energies (302 ± 10 Kcal/mol).

As suggested by preliminary NMR results for the (1−55) NC fragment from HIV-1 (MN strain) (South *et al.*, 1990, 1991), the two zinc fingers folded in a quite similar manner, as shown by the satisfactory r.m.s.d. (2.3 ± 0.1 Å) obtained after superimposition of the backbone atoms of the proximal (Cys15 to Cys28) and the distal (Cys36 to Cys49) fingers. Moreover, all of the hydrogen bonds and the short distance contacts already found in the zinc fingers alone (South *et al.*, 1990, 1991) were observed. No hydrogen bond was found in the RAPRKKG sequence, in agreement with the experimental data. It is interesting to note that molecular dynamics studies performed in the absence of NMR derived constraints showed that the distance between the two aromatic rings fluctuated between 4.8 and 7.1 Å (Figure 7a), in accordance with the range of distances observed in DIANA derived structures. As expected, dynamic studies performed under constraints led both to an improvement of the r.m.s.d. (0.5 ± 0.1 versus 1.0 ± 0.6) and to structures (not shown here) very close to those derived from DIANA. This clearly shows that zinc coordination is the main driving force ensuring the stability of the fingers while aromatic ring interactions could participate in bringing the fingers together.

Structural characteristics of the (13−51) RNA binding domain of NCp7 and functional relevance

The Lys14-Cys18 and Gly35-Cys39 sequences of NCp7 were drawn together (Figure 7) by the kink occurring in the Arg29-Ala-Pro-Arg-Lys-Lys-Gly35 linker at the level of the Ala30-Pro31 residues in agreement with the observed dihedral angles: $\Psi_{30} = 157°$ and $\Phi_{31} = -57°$, which were close to those found for corresponding residues in the kink of protein chain ($\Psi \sim 180°$, $\Phi < 60°$) (Creighton, 1984). This structural arrangement shortens the distance between the two aromatic residues Phe16 and Trp37 (Figures 7b and c) to ~6 Å which is compatible with the experimentally observed slight shielding of Phe16 and Trp37 proton resonances and accounts for the NOE effect observed between Phe16 and Trp37 side-chain protons. The weakness of the shielding effects was most probably due both to the almost perpendicular orientation of the two aromatic rings, also found in a recently described zinc finger peptide (Krizek *et al.*, 1991), and to the degree of freedom around the χ_2 dihedral angles, averaging the current shifts.

The basic linker region was well exposed to the solvent. This supports the idea that the compact globular zinc fingers may function independently from the basic RAPRKKG sequence. In fact, deletion of the two zinc fingers did not affect the RNA annealing activities of the NC protein, provided that the basic domain was linked by glycine residues to the N- and C-terminal domains of NCp7 (De Rocquigny *et al.*, 1992). Therefore the role of the zinc fingers might be to direct the spatial recognition by the basic VK and RAPRKKG sequences of specific sites on the viral RNA and replication primer tRNA (Barat *et al.*, 1989; Darlix *et al.*, 1990). In agreement with this possibility, mutations of the two successive lysine residues corresponding to Lys33 and Lys34 in NCp7 resulted in a drastic decrease in the RNA annealing activities of MuLV NCp10 (Prats *et al.*, 1991).

The proximity of the Phe16 and Trp37 aromatic rings and the presence of Lys14 and Lys38 side-chains in their proximity probably has important functional implications in NC protein viral RNA recognition. In fact, retroviral NC proteins were characterized by the presence of two aromatic rings located either in one zinc finger as in MuLV, or in two successive zinc fingers as in RSV and HIV (Van Beveren *et al.*, 1984; Wain-Hobson *et al.*, 1985). Replacement of one aromatic ring, even with a hydrophobic moiety, led to a large decrease in genomic RNA packaging (Méric and Goff, 1989; Dupraz *et al.*, 1990). It is therefore tempting to speculate that the aromatic rings could, through RNA intercalation, cooperate with the basic domain to selectively recognize the packaging signal of the viral RNA *in vivo*. This possibility is presently under study.

The three-dimensional structure of the (13−51)NCp7 domain indicated that the zinc ions were accessible to reagents while the Cys-Phe16 and Cys-Trp37 bonds were buried in the zinc fingers. This may explain why the recently observed cleavage of the NC protein of the HIV related equine infectious anaemia virus (EIAV) occurring at the Cys-Tyr25 and Cys-Phe44 bonds required a prior dissociation of zinc ions (Roberts *et al.*, 1991). Moreover, the fewer NOEs in the distal finger compared with the proximal finger of NCp7 (Figure 5) suggested that the former was less structured. This is in agreement with the higher temperature-dependent exchange rate of the NH signals occurring in the C-terminal finger (unpublished results). This could explain the lower affinity of the distal finger for the metal ion in NCp7 (De Rocquigny *et al.*, 1991; Fitzgerald and Coleman, 1991; Y.Mély, E.Piémont, H.Rocquigny, N.Jullian, N.Morellet, B.P.Roques and D.Gerard, in preparation).

Finally, the structural study reported here should facilitate the rational design of compounds able to selectively inhibit the functions of NCp7; this is now in progress in our laboratories.

Materials and methods

Peptide synthesis

Solid phase synthesis of large quantities (≥100 mg) of NCp7 and related peptides (Figure 1) was carried out using the stepwise solid phase method and Fmoc amino acids on an automatic reprogrammed Applied Biosystems 431 A synthesizer, as already described (De Rocquigny *et al.*, 1991). To preserve the highly oxidizable cysteine residues of the peptides (purity ≥99%), all manipulations were carried out under an argon atmosphere.

NMR sample preparation and NMR measurements

NCp7 and related peptides were dissolved in 100% D_2O or 90% H_2O/10% D_2O argon-degassed solutions at concentrations around 5.0 mM. $ZnCl_2$ was added in slight excess with respect to the concentration of the zinc finger(s). The pH was adjusted to 5.5 ± 0.2 using 1 M NaOD or DCl. The proton assignments for the various peptides and NCp7 (Figure 1) were obtained at 600 MHz ^1H frequency on a Brüker AMX 600 spectrometer from COSY, NOESY and TOCSY, experiments recorded in D_2O and H_2O solution at 293 K. In each case, the spectra were recorded with irradiation of the solvent resonance during the relaxation delay and, for the NOE spectra, during the mixing time. All two-dimensional NMR spectra were acquired in the phase-sensitive mode using the time proportional phase incrementation method (Marion and Wüthrich, 1983). Two-dimensional spectra were

generally recorded with 2048 data points and 512 t_1 increments. The number of scans per t_1 increment varied between 64 and 128. Two-dimensional NMR data were collected without sample spinning. All two-dimensional spectra were processed using shifted sine-bell window functions in both dimensions. After zero filling and double Fourier transformation, base-line corrections were performed in both dimensions. Digital resolution in the final transformed spectrum was usually 3.53 or 7.07 Hz/point and 7.07 Hz/point in the ω_2 and the ω_1 dimensions. NOESY spectra were recorded with mixing times of 100 and 200 ms. No zero quantum suppression technique was applied. Weak NOEs observed at a mixing time of 200 ms, but not 100 ms, were generally regarded as spin diffusion effects and not used in distance geometry calculations (Kochoyan *et al.*, 1991).

Structure calculations

Experimental NOE intensities were converted into proton–proton distance constraints classified in three categories: 2.0–2.5 Å, 2.0–3.5 Å and 2.0–4.5 Å. Pseudoatoms were introduced when necessary and interproton distances were corrected accordingly (Wüthrich *et al.*, 1983). Supplementary structural constraints were added to impose tetrahedral coordination around the zinc ion. As previously described in the case of cadmium complexed metallothionein (Arseniev *et al.*, 1988), a pseudoatom, QZn, was attached to S^{γ} of each Cys residue with a bond length of 2.3 Å. The distance between QZn and Nϵ2 of His residue in the fingers, was imposed to be ~2.0 Å. Since the six QZn pseudoatoms must represent two metal ions, we imposed an upper limit of 0.1 Å between any two QZn that corresponded to the same zinc ion. Tetrahedral geometry was ensured by imposing a fixed distance between each pair of atoms bound to the same zinc ion. Preliminary calculations, with the distance geometry program DIANA (Güntert *et al.*, 1991), generated a first set of structures. A structural analysis of the best conformers using PRO-EXPLORE, resulted in suplementary stereospecific assignments of 14 methylene protons which consisted in correlating the distances obtained from the calculated structures with NOE intensities (Wüthrich, 1986). All these data yielded 320 distance constraints that were used in DIANA to generate a set of 100 conformers. The seven best conformations, i.e. those with the lowest target function value (Braun and Go, 1985) are superimposed as depicted in Figure 6. The degree of refinement is shown by the weak value of the r.m.s.d. (1.6 ± 0.6 Å) when the backbone atoms of residues from Cys15 to Cys49 of the seven structures were superimposed. The agreement with the experimental data was good as no long range constraint violations >0.15 Å were observed. For final structure calculation, the seven structures were then subjected to energy refinement using the AMBER modelling package. During the minimization and molecular dynamics procedure, all NMR-derived constraints including the distance between the Phe16 side-chain and the Trp37 aromatic ring were excluded. A zinc force field obtained from *ab initio* calculations (Jacob, 1990; O.Jacob and B.Maigret, in preparation) and parameterized in the AMBER framework was used in the refinement procedures to account for metal–peptide interactions. Moreover, a dielectric constant of 80 was used to minimize ion pair interactions overestimated in *in vacuo* calculations. After 3000 gradient minimization steps, one conformer presenting the lowest energy (292 kcal/mol) and no long-range constraint violation, was selected and then submitted to 10 ps of molecular dynamics refinement at 300 K, using AMBER. Ten conformers were extracted at 1 ps intervals and superimposed after additional energy minimization (Figure 7a) essentially to analyse the mobility of the aromatic rings. Computer graphic representations were obtained using the program BIO-EXPLORE (BIOSTRUCTURE S.A., Strasbourg, France) on a Personal Iris 4D35 workstation (Silicon Graphics Inc.).

Acknowledgements

We thank P.Güntert, W.Braun and K.Wüthrich for the gift of the DIANA program. We acknowledge G.Perrot for helpful discussions and technical assistance and C.Dupuis for typing the manuscript. Comments and stylistic revision of the paper by M.Lapadat are acknowledged. N.Jullian is grateful to BIOSTRUCTURE for financial support. Gratitude is expressed to the 'Groupement Scientifique' IBM-CNRS. This work was supported by the French Program against AIDS (ANRS).

References

Aldovini,A. and Young,R.A. (1990) *J. Virol.*, **64**, 1920–1926.

Arseniev,A., Schultze,P., Wörgötter,T., Braun,W., Wagner,G. Vasak,M., Kägi,J.H.R. and Wüthrich,K. (1988) *J. Mol. Biol.*, **201**, 637–657.

Barat,C., Lullien,V., Schatz,O., Keith,G., Nugeyre,M.T., Grüninger-Leitch,F., Barré-Sinoussi,F., Le Grice,S.F.J. and Darlix,J.L. (1989) *EMBO J.*, **8**, 3279–3285.

Barré-Sinoussi,F. *et al.* (1983) *Science*, **220**, 868–870.

Berg,J.M. (1986) *Science*, **232**, 485–487.

Bieth,E., Gabus,C. and Darlix,J.L. (1990) *Nucleic Acids Res.*, **18**, 119–127.

Braun,W. and Go,N. (1985) *J. Mol. Biol.*, **186**, 611–626.

Braunschweiler,L. and Ernst,R.R. (1983) *J. Magn. Reson.*, **53**, 521–528.

Coffin,J.M. (1984) In Weiss,R., Teich,N., Varmus,H. and Coffin,J.M. (eds), *RNA Tumor Viruses*, Vol. I. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY, pp. 261–368.

Cornille,F., Mély,Y., Ficheux,D., Salvignol,I., Gérard,D., Darlix,J.L., Fournié-Zaluski,M.C. and Roques,B.P. (1990) *Int. J. Pep. Prot. Res.*, **36**, 551–558.

Covey,S.N. (1986) *Nucleic Acids Res.*, **14**, 623–632.

Creighton,T.E. (1984) *Proteins*. W.H.Freeman and Company, New York.

Darlix,J.L., Gabus,C., Nugeyre,M.T., Clavel,F. and Barré-Sinoussi,F. (1990) *J. Mol. Biol.*, **216**, 689–699.

Davis,D.G. and Bax,A. (1985) *J. Am. Chem. Soc.*, **107**, 2821–2822.

De Rocquigny,H., Ficheux,D., Gabus,C., Fournié-Zaluski,M.C., Darlix,J.L. and Roques,B.P. (1991) *Biochem. Biophys. Res. Commun.*, **180**, 1010–1018.

De Rocquigny,H., Gabus,C., Vincent,A., Fournié-Zaluski,M.C. and Roques,B.P. (1992) *Proc. Natl. Acad. Sci. USA*, in press.

Di Marzo Veronese,F., Rahman,R., Copeland,T., Oroszlan,S., Gallo,R.C. and Sarngadharan,M.G. (1987) *AIDS Res. Human Retrov.*, **3**, 253–264.

Dupraz,P., Oertle,S., Méric,C., Damay,P. and Spahr,P.F. (1990) *J. Virol.*, **64**, 4978–4987.

Fitzgerald,D.W. and Coleman,J.E. (1991) *Biochemistry*, **30**, 5192–5201.

Gorelick,R.J., Henderson,L.E., Hanser,J.P. and Rein,A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8420–8424.

Green,L.M. and Berg,J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 4047–4051.

Green,L.M. and Berg,J.M. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6403–6407.

Güntert,P., Braun,W. and Wüthrich,K. (1991) *J. Mol. Biol.*, 217, 517–530.

Jacob,O. (1990) Ph.D. Thesis, Université Louis Pasteur, Strasbourg, France.

Kochoyan,M., Havel,T.F., Nguyen,D., Dahl,C.E., Keutmann,H.T. and Weiss,M.A. (1991) *Biochemistry*, **30**, 3371–3386.

Krizek,B.A., Amann,B.T., Kilfoil,V.J., Merkle,D.L. and Berg,J.M. (1991) *J. Am. Chem. Soc.*, **113**, 4518–4523.

Kumar,A., Ernst,R.R. and K. Wüthrich (1980) *Biochem. Biophys. Res. Commun.*, **95**, 1–6.

Marion,D. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

Méric,C. and Goff,S.P. (1989) *J. Virol.*, **63**, 1558–1568.

Méric,C. and Spahr,P.F. (1986) *J. Virol.*, **60**, 450–459.

Omichinski,J.G., Clore,G.M., Sakaguchi,K., Appella,E. and Gronenborn,A.M. (1991) *FEBS Lett.*, **292**, 25–30.

Prats,A.C., Sarih,L., Gabus,C., Litvak,S., Keith,G. and Darlix,J.L. (1988) *EMBO J.*, **7**, 1777–1783.

Prats,A.C., Roy,C., Wang,P., Erard,M., Housset,V., Gabus,C., Paoletti,C. and Darlix,J.L. (1990) *J. Virol.*, **64**, 774–783.

Prats,A.C., Housset,V., De Billy,G., Cornille,F., Prats,H., Roques,B.P. and Darlix,J.L., (1991) *Nucleic Acids Res.*, **19**, 3533–3541.

Rance,M., Sorensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Roberts,M.M., Copeland,T.D. and Oroszlan,S. (1991) *Protein Engng*, **4**, 695–700.

Singh,U.C., Weiner,P.K., Caldwell,J. and Kollman,P.A. (1986) AMBER (UCSF version 3.0), School of Pharmacy, University of California, San Francisco, CA 94143 .

South,T.L., Blake,P.R., Sowder,R.C., Arthur,L.O., Henderson,L.E. and Summers,M.F. (1990) *Biochemistry*, **29**, 7786–7789.

South,T.L., Blake,P.R., Hare,D.R. and Summers,M.F. (1991) *Biochemistry*, **30**, 6342–6349.

Summers,M.F., South,T.L., Kim,B. and Hare,D.R. (1990) *Biochemistry*, **29**, 329–340.

Van Beveren,C., Coffin,J.M. and Hughes,S. (1984) In Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds), *RNA Tumor Viruses*, Vol. II. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY, pp. 559–1209.

Wain-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M. (1985) *Cell*, **40**, 9–17.

Williamson,M.P., Marion,D. and Wüthrich,K. (1985) *J. Mol. Biol.*, **182**, 295–315.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley and Sons, New York.

Wüthrich,K., Billeter,M. and Braun,W. (1983) *J. Mol. Biol.*, **169**, 949–961.

Received on February 20, 1992; revised on April 22, 1992