NMR Structure of the Complex between the Zinc Finger Protein NCp10 of Moloney Murine Leukemia Virus and the Single-Stranded Pentanucleotide d(ACGCC): Comparison with HIV−NCp7 Complexes[†,‡]

W. Schüler, C.-Z. Dong, K. Wecker, and B. P. Roques*

Département de Pharmacochimie Moléculaire et Structurale, INSERM U266−CNRS UMR 8600, UFR des Sciences Pharmaceutiques et Biologiques, 4, Avenue de l'Observatoire, 75270 Paris Cedex 06, France

Received February 16, 1999; Revised Manuscript Received June 9, 1999

ABSTRACT: The structure of the 56 amino acid nucleocapsid protein NCp10 of retrovirus MoMuLV, which contains a single $CX_2CX_4HX_4C$-type zinc finger, has been determined previously by NMR. The important role of NCp10 (or NCp7 for HIV-1) in the retroviral life cycle seems mainly related to their preferential binding to single-stranded nucleic acids. We report here the structure of the complex formed between the biologically active (14−53)NCp10 and the oligonucleotide d(ACGCC) in aqueous solution determined by 2D ^{1}H NMR based methods. The aromatic residue Trp^{35} of NCp10 directs nucleic acid complexation as shown by its complete fluorescence quenching upon addition of d(ACGCC). ^{1}H and ^{31}P NMR studies support the insertion of Trp^{35} between the G^3 and C^4 bases. A total of 577 NOE distance restraints, of which 40 were intermolecular, were used for the structure determination. The zinc finger provides a well-defined surface for the binding of d(ACGCC) through hydrophobic interactions and tryptophan stacking on the guanine. This latter interaction was also observed in the NMR-derived structures of the complexes between NCp7, which contains two successive zinc fingers, and single-stranded DNA and RNA, supporting the proposal for a major role played by aromatic residues of NCp proteins in nucleic acid recognition. Upon binding to the nucleotide a new loop in NCp10 that participates in the intermolecular interaction is formed. Additional interactions provided by positively charged residues surrounding the zinc finger appear necessary for tight binding. The structure of the complex NCp10−d(ACGCC) gives a structural explanation for the loss of virus infectivity following point mutations in the finger domain.

The retroviral nucleocapsid proteins (NCps)[1] are encoded by the 3′ end of the gag precursor gene. In the virion core they are tightly associated with the viral genome in a nucleoprotein complex containing two identical, unspliced RNA molecules. The NCps are known to be involved in a variety of functions crucial for the viral life cycle (for reviews, see refs 1 and 2). Like NCp7 in HIV-1 (and NC proteins in other retroviruses), the 56 amino acid NCp10 (Figure 1) of Moloney murine leukemia virus potentiates the dimerization of genomic RNA, a process most probably linked to RNA packaging (*3, 4*), and promotes annealing of the replication primer (tRNA^{Pro}) to the initiation site of the reverse transcription on the viral RNA (*5*). In addition, NCps appear to play a critical role in strand transfer (*6, 7*), protection of genomic RNA from nucleases (*8*), and provirus integration (*9, 10*). Moreover, NCp7 was shown to interact with viral proteins such as Vpr and RT (*11, 12*). In contrast to NCp7 of HIV-1 (Figure 1) and several other retroviral NC proteins, which contain two zinc fingers, NCp10 possesses only one conserved $CX_2CX_4HX_4C$ sequence (*3*) which binds one zinc ion with a high affinity constant (1.2×10^{13} M^{-1}) (*13*).

The structures of NCp10 (*14*) and NCp7 (*15, 16*) have been determined by ^{1}H NMR. In both proteins the backbones of the finger domains (Figure 1) are folded around the zinc ion, while the flanking peptide sequences remain flexible. In NCp10, the two aromatic residues Tyr^{28} and Trp^{35} are present in the single zinc finger, whereas in NCp7 a spatial (*15, 17*) and transient (*18*) proximity of the two corresponding residues, Phe^{16} and Trp^{37}, which are located in the N- and the C-terminal finger domains, respectively, was observed. The replacement of one zinc ligand, His^{23}, by Cys in NCp7 was shown to produce a change in the 3D structure of the mutated zinc finger, resulting in a loss of the spatial proximity between the two zinc-complexed domains (*19*). This single mutation was shown to induce a complete loss of HIV-1 infectivity (*19*), a result which was also recently found in the case of NCp10 (*20*). These structure−activity studies have suggested that the typical and highly conserved sequences of the retroviral zinc fingers generating well-

[†] We gratefully acknowledge the grant given to W.S. by the EC (Human Capital and Mobility Program). This work was supported by the French program against AIDS (ANRS) and SIDACTION.

[‡] The coordinates of the complex structure have been deposited in the Protein Data Bank, Brookhaven National Laboratories, Long Island, NY 11973 (accession code 1a6b).

* To whom correspondence should be addressed. Tel: (33)-01-43.25.50.45. Fax: (33)-01-43.26.69.18. E-mail: Roques@pharmacie.univ-paris5.fr.

[1] Abbreviations: HIV, human immunodeficiency virus; MoMuLV, Moloney murine leukemia virus; NC, nucleocapsid protein; NOE, nuclear Overhauser effect; PBS, primer binding site; RSV, rous sarcoma virus; RT, reverse transcriptase; SL, stem loop; tRNA, transfer ribonucleic acid.

10.1021/bi990378d CCC: $18.00
Published on Web 09/15/1999

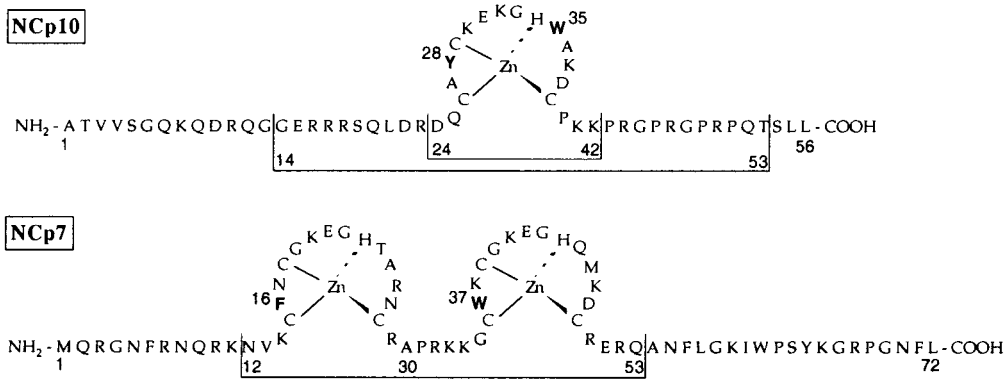

FIGURE 1: Primary sequences of the retroviral nucleocapsid proteins NCp10 and NCp7. The (14−53)NCp10 studied in this paper which retains the nucleic acid binding properties of NCp10 and the inactive zinc finger fragment (24−42)NCp10 are indicated.

structured motifs could be involved in critical steps of the retroviral life cycle (*21*). This assumption is supported by the antiviral activity of compounds capable of destroying the structure of the zinc finger domain (*22*) of to mimicking the main components of the finger domain involved in NCp7 biological functions (*23*).

One general property of NCps is their preferential interaction with single-stranded nucleic acids (*24*−*27*). The high-affinity binding of NCps to nucleic acids is mainly ensured by hydrophobic interactions and binding characteristics of NCps to short DNA molecules apply equally for RNA binding (*26*−*30*).

In this work, we have therefore studied the 3D structure of the complex between the biologically active (14−53)-NCp10 and d(ACGCC) by 2D ^{1}H NMR and molecular modeling. The choice of the same nucleotide as that used in studies with NCp7 (*26*) allowed insights in the common binding characteristics of NC proteins and single-stranded oligonucleotide to be determined. Our results show that the formation of the complex (14−53)NCp10−d(ACGCC) induces a change in the flexible sequences surrounding the zinc finger, resulting in the stabilization of a hydrophobic cluster of amino acids (Leu^{21}, Ala^{27}, Ala^{36}) which participates to the binding of d(ACGCC). The study also supports the key role of the Trp residue for NCp10 interaction with single-stranded nucleic acids. As in the case of the NMR complex formed between (12−53)NCp7 and d(ACGCC) (*26*) or between NCp7 and the RNA stem−loop 3 (SL3) (*27*), the indole ring of NCp10 is stacked on a guanine base. The biological relevance of the Trp interaction has recently been demonstrated using various techniques and with mutants in which Trp was replaced by Leu (*23*, *31*). These structural results account for the partial or complete loss of viral infectivity induced by point mutations in the zinc finger domain. Moreover, they are of great value for the design of new antiviral agents capable of inhibiting the numerous functions of NCps and possibly devoid of resistance effects observed with RT or protease inhibitors (*32*).

MATERIALS AND METHODS

Chemical Synthesis. NCp10, the finger domain 24−42, and the 14−53 fragment of NCp10 were prepared using the automated step-by-step solid-phase synthesis by the classical Fmoc strategy (Applied Biosystems 431A synthesizer) and purified by HPLC as described (*33*). The primary sequence of the peptides was verified by means of mass spectrometry and amino acid analyses. The oligonucleotide d(ACGCC) (generous gift of Dr. T. Huynh-Dinh, Pasteur Institute, France) was prepared by the solid-phase phosphoramidite chemistry on an ABI-380B synthesizer as described (*34*). The final purity of the pentamer was checked with analytical reverse-phase HPLC and capillary electrophoresis (>99%).

Fluorescence Experiments. The fluorescence experiments were performed on a Perkin-Elmer LS 50 by measuring emission at 350 nm (excitation wavelength at 280 nm). The concentrations of the studied peptides, in 5 mM Hepes buffer (pH 7), were 3×10^{-1} M. The respective concentrations of the peptides and the nucleotide were verified by UV measurements using a value of $\epsilon_{mol} = 7000$ M^{-1} cm^{-1} at 280 nm for NCp10 fragments and $\epsilon_{mol} = 51 \times 10^3$ M^{-1} cm^{-1} at 260 nm for d(ACGCC). The binding isotherms were monitored at 20 °C, and the number of sites and the affinity constant were calculated using the Scatchard equation as already reported in detail (*17*).

NMR Spectroscopy. The samples for NMR studies contained 1:2 or 1:1 protein−d(ACGCC) equivalents, respectively, for a nucleotide concentration of 2 mM. The pH was fixed at 6.5 in 90% ^1H_2O/10% ^2H_2O, and 1.5 equiv of Zn^{2+} was added as $ZnCl_2$ as described (*26*).

The NMR experiments were carried out at temperatures of 276, 293, and 313 K on a Bruker AMX600 spectrometer operating at 600 MHz. DQF-COSY (*35*), E.COSY (*36*), NOESY (*37*), and Clean-TOCSY (*38*) methods were used. The spectra were recorded in a phase-sensitive mode, using States−TPPI quadrature detection (*39*). 2K (t_2) × 512(t_1) complex points were recorded (4K × 1K for both DQF-COSY and E.COSY) at a spectral width of 12 ppm, symmetric around the solvent peak. The mixing times employed were 60, 100, and 200 ms for the NOESY experiments. The duration of the TOCSY spin-lock sequence was 50 and 70 ms. The carrier frequency was set to the resonance of H_2O which also served as internal standard to measure the chemical shifts of the proton signals (4.828 ppm at 293 K, 4.628 ppm at 313 K). The solvent signal was suppressed by presaturation during the relaxation delay of 1.6 s (and during the mixing time of the NOESY experiments). TOCSY and NOESY with a jump and return solvent suppression (*40*) were applied, allowing the signals of protons exchanging with water (spectral width of 20 ppm) to be observed.

Prior to Fourier transformation and baseline correction, the data were zero-filled and multiplied in both dimensions

with a sine bell function shifted by $\pi/6$. In the quantification step of the analysis, a squared sine bell function shifted by $\pi/2$ was used. Additional NOE distance restraints were collected from a NOESY experiment in ^2H_2O. Spectra were processed and evaluated using UXNMR (Bruker) and Felix (Biosym). The ^{31}P experiments [162 MHz, 1D and 2D $^1H-^{31}P$ shift correlation (*41*)] were carried out on a Bruker AM400 spectrometer equipped with a selective ^{31}P probe head with a 1:1 sample of (14−53)NCp10−d(ACGCC) in the conditions used for ^1H NMR experiments. The phosphorus chemical shifts were referenced to an external sample of 85% phosphoric acid in $HClO_4$.

Structure Calculations. Distance restraints were obtained from NOESY cross-peak intensities from spectra recorded with a 100 ms mixing time. They were calibrated with respect to an average of a set of geminal β-protons (1.78 Å) of residues 28, 29, 35, 39, and 40 and classified into eight categories of restraints with upper limits between 2.5 and 6.0 Å. Appropriate corrections were added for the upper limits involving pseudoatoms for methylene, methyl, or aromatic groups given by the default values of the software. A total of 537 intramolecular and 40 intermolecular conformational restraints was used as input for the structure calculations. The AMBER force field, which was completed with parameters for zinc as applied in ref *17*, was used for all molecular mechanics simulations with the NMRchitect software (Biosym/MSI).

The system was heated for 1 ps to 1000 K, kept at this temperature for a further 3 ps, and then cooled to 300 K in five steps of 2, 1.7, 1.7, 5, and 5 ps, respectively. During the steps of molecular dynamics the nonbond terms were reduced to a simple quartic van der Waals repulsion form with radii scaled by a factor of 0.85. The force constants for the NOE-derived constraints were maintained at 30 kcal mol^{-1} $Å^{-2}$ throughout the refinement process. All peptide bond angles ω were forced into an E-conformation by torsional constraints with a force constant of 100 kcal mol^{-1} operating for deviations higher than 10° from this conformation. The dielectric constant was set to be distance dependent ($\epsilon = 4r$) in order to reduce "in vacuo" effects. The resulting structures were energy minimized to a maximal gradient of 0.02 kcal mol^{-1} $Å^{-1}$ using a Lennard-Jones nonbond term and a cutoff of 12 Å.

Fifty structures of the DNA-bound (14−53)NCp10 with the correct three-dimensional folds were generated with a dynamical simulated annealing protocol (*42*) using the intramolecular data subset but disregarding the presence of the DNA during the calculation. Two starting structures for d(ACGCC) were generated in standard A and B DNA conformations and prepared with 200 steps of steepest descents and conjugate gradient energy minimization, using the intramolecular NOEs of the nucleotide as described (*26*).

The best protein structure (with respect to total energy and restraint violations) of the above set and each nucleotide structure were used for a docking procedure which applied energy minimization incorporating all NOE restraints of which 40 were intermolecular.

In a second step, the two resulting complex structures were taken as starting structures for dynamical simulated annealing calculations in which the nucleotide conformations were kept fixed. The resulting two sets of 50 structures were energy minimized with all atoms free to move.

To verify that the orientation of the nucleotide in the complex is not biased by the protocol, the best structure, taking into account the minimum of total energy and NOE violations, was dissociated by moving the protein moiety about 20 Å away from the oligonucleotide and turned 180° about the binding axis (*26*). A further dynamical simulated annealing calculation led to the same complex structure with the Trp residue inserted between G^3 and C^4.

The Protein Data Bank accession numbers for (12−53)NCp7−d(ACGCC) and (14−53)NCp10−d(ACGCC) complexes are 1bj6 and 1a6b, respectively.

RESULTS

Fluorescence Studies. The interaction of the fragments (24−42)NCp10 and (14−53)NCp10 with d(ACGCC) has been studied by fluorescence spectroscopy as previously reported for NCp7 (*26*). The addition of increasing concentrations of d(ACGCC) in the range of $(0.5-5) \times 10^{-6}$ M to a 3×10^{-6} M solution of (14−53)NCp10 at 20 °C, pH 6.5, induces drastic quenching of Trp^{35} fluorescence measured at 350 nm (data not shown). An apparent affinity constant of 3×10^7 M $\pm$ 1.1×10^7 with a number of sites of 1.4 ± 0.03 was calculated for this complex by using the Scatchard equation. Similar parameters were obtained for the entire NCp10. Interestingly, under the same conditions, no modification of the fluorescence intensity was observed with the (24−42)NCp10 zinc finger (data not shown).

^{31}P NMR Measurements. We studied the changes of the ^{31}P chemical shifts upon complexation, which are very sensitive to variations of the PO and CO dihedral angles in the nucleotide backbone (*43*) as lengthening of the nucleotide backbone by intercalation of a tryptophan side chain in the nucleotide would result in observable changes. The experiments (data not shown) were carried out under the same experimental conditions as those used for the ^1H spectra. The chemical shifts of the phosphorus signals for d(ACGCC) free (P-I = 1.80 ppm; P-II = 1.55 ppm; P-III = 1.65 ppm; P-IV = 1.80 ppm) and complexed (P-I = 1.53 ppm; P-II = 1.60 ppm; P-III = 1.88 ppm; P-IV = 1.80 ppm) with (14−53)NCp10 (data not shown) were assigned by 2D ($^1H-^{31}P$) shift correlation experiments. The phosphate group P-III, connecting G^3 and C^4, exhibited a low-field shift (−0.23 ppm) upon addition of (14−53)NCp10 to d(ACGCC), while the phosphate group P-I, connecting A^1 and C^2, shows a high-field shift (0.27 ppm). Deshielding effects such as that observed for P-III were found following intercalation of aromatic residues into nucleic acids (*44*).

^1H NMR Studies of the Complex between (14−53)NCp10 and d(ACGCC). The ^1H NMR spectra of (14−53)NCp10 alone or complexed with d(ACGCC) were recorded in aqueous solution at pH 6.5 and 20 °C in the presence of 1.5 equiv of $ZnCl_2$. For the free and the complexed peptide, no sign of aggregation was observed in the concentration range used. Addition of increasing amounts of (14−53)NCp10 to a fixed concentration of d(ACGCC) led to chemical shift modifications of protons belonging to both the peptide and the pentanucleotide (Figure 2) (Table 1 and Figure 1 in Supporting Information). However, only one series of signals was observed during all the titrations, showing that the complex formed between d(ACGCC) and (14−53)NCp10 was in fast exchange on the NMR time scale. No important

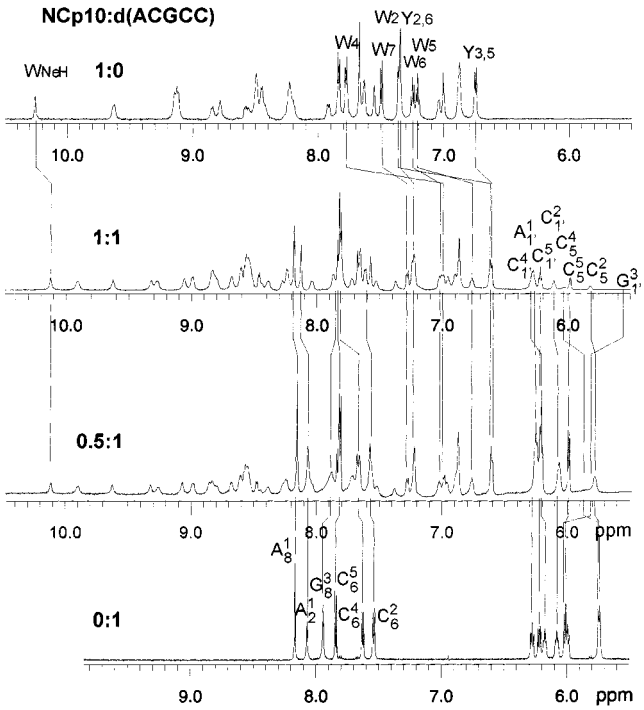

FIGURE 2: Variation of the chemical shifts of (14−53)NCp10 upon binding to d(ACGCC) in a 1:1 complex ratio (concentration 2 mM, 293 K, pH 6.5). The whole zinc finger region is strongly affected upon binding, particularly Trp^{35} aromatic protons, in agreement with the stacking of this residue on G^3. Some important residues located at the interface region of the complex are numbered. The influence of binding upon the oligonucleotide protons is indicated by the variation of various proton resonances as illustrated by the strong shielding of G^3-$H_{1'}$.

changes in the NMR parameters were observed after addition of a peptide amount of more than 1 equiv with respect to d(ACGCC), indicating the formation of a 1:1 complex and suggesting that most of the species in solution were in the complexed form, in agreement with the fluorescence results.

Sequence-specific assignments of the proton resonances were obtained for all residues except for the two N-terminal amino acids. Moreover, due to the cis−trans isomerism of Pro^{49} and Pro^{51}, additional weak signals corresponding to the cis isomers (20%) were found for the five C-terminal residues. In the complex with the nucleotide, the strongest chemical shift changes were observed for Ala^{27}, Ala^{36}, Trp^{35}, and Lys^{37} in the finger domain and for Leu^{21}, Lys^{42}, and Pro^{43} in the vicinity of the CCHC box.

In the (14−53)NCp10−d(ACGCC) complex, several nucleotide protons underwent chemical shift changes and exhibited line broadening relative to the free nucleotide. This is the case for G^3-H_8, $H_{1'}$ (−0.09 and −0.43 ppm), C^4-H_5, H_6, $H_{1'}$ (0.30, 0.19, and 0.13 ppm) and C^2-$H_{2'}$ (−0.40 ppm). The shielding observed on the guanosine protons is due to the strongest ring current shift of the superimposed Trp^{35} as compared to the cytosine only transiently stacked on G^3 in the flexible free d(ACGCC). This is the reverse for C^4, which undergoes the strong shielding effect of G^3 in free d(ACGCC), whereas in the complex, G^3 is replaced by Trp^{35}, which is not in direct contact with the C^4 ring. The strong upfield shift observed for the $H_{2'}$ proton of C^2 is probably due to its spatial proximity to Tyr^{28} in the complex. A line broadening was observed for G^3-H_8, $H_{1'}$ protons and to a lesser extent for C^4-H_5, H_6, $H_{1'}$, which correspond to base protons found

Table 1: Intermolecular NOE Contacts Found for the Complex (14−53)NCp10−d(ACGCC) at pH 6.5

atoms	int[a]	atoms	int[a]	atoms	int[a]
Leu^{21}-Hδ A_1-$H_{1'}$	w	Trp^{35}-4H G_3-$H_{2'}$	w	Trp^{35}-6H C_4-H_5	vw
Leu^{21}-Hδ A_1-$H_{4'}$	w	Trp^{35}-4H G_3-$H_{2''}$	w	Trp^{35}-6H C_4-H_6	w
Leu^{21}-Hδ C_2-$H_{1'}$	*vw*	Trp^{35}-4H G_3-H_8	w	Trp^{35}-7H G_3-$H_{1'}$	w
Leu^{21}-Hδ C_2-H_5	w	Trp^{35}-5H G_3-$H_{2'}$	m	Trp^{35}-7H G_3-$H_{2'}$	w
Leu^{21}-Hδ C_2-H_6	w	Trp^{35}-5H G_3-$H_{2''}$	s	Trp^{35}-7H G_3-$H_{2''}$	w
Asp^{24}-HN C_4-$H_{1'}$	vw	Trp^{35}-5H G_3-H_8	w	Trp^{35}-7H C_4-$H_{1'}$	w
Ala^{27}-Hβ A_1-$H_{1'}$	*vw*	Trp^{35}-5H C_4-$H_{1'}$	vw	Trp^{35}-7H C_4-H_6	w
Ala^{27}-Hβ A_1-$H_{4'}$	w	Trp^{35}-5H C_4-H_5	vw	Ala^{36}-Hβ A_1-$H_{1'}$	w
Ala^{27}-Hβ A_1-$H_{5'}$	*vw*	Trp^{35}-5H C_4-H_6	w	Ala^{36}-Hβ A_1-H_8	m
Ala^{27}-Hβ A_1-$H_{5''}$	*vw*	Trp^{35}-6H G_3-$H_{1'}$	w	Ala^{36}-Hβ C_2-$H_{1'}$	m
Ala^{27}-Hβ C_2-$H_{1'}$	w	Trp^{35}-6H G_3-$H_{2'}$	m	Ala^{36}-Hβ C_2-H_6	w
Ala^{27}-Hβ C_2-H_5	m	Trp^{35}-6H G_3-$H_{2''}$	m	Ala^{36}-Hβ C_2-H_5	w
Ala^{27}-Hβ C_2-H_6	m	Trp^{35}-6H G_3-H_8	w		
Trp^{35}-Hα C_4-H_5	*vw*	Trp^{35}-6H C_4-$H_{1'}$	w		

[a] Classification: vw, very weak; w, weak; m, medium; s, strong.

in the vicinity of the aromatic tryptophan side chain. Moreover, a slight deshielding (0.06−0.08 ppm) was observed for the aromatic protons of C^2 and of A^1-H_2. All aromatic protons of Trp^{35} show high-field shifts in the complex, up to 0.79 ppm, supporting the insertion of the Trp side chain between the G^3 and C^4 nucleotide bases.

The 40 intermolecular NOEs observed allowed the position of the nucleotide relative to the protein in the complex to be established (Table 1). The position of A^1 is defined by a cross signal of its aromatic H_8 proton to the methyl group of Ala^{36} and NOEs between protons of the A^1 ribose moiety in contact with the methyl groups of Leu^{21}, Ala^{27}, and Ala^{36}.

The methyl groups are also in proximity of the aromatic protons H_5 and H_6 of C^2 whose $H_{1'}$ proton is close to Ala^{27} and Ala^{36}. The position of G^3 is defined by 13 contacts with Trp^{35} and one to Arg^{23}. There are intense NOE signals between H_4, H_5, H_6, H_7 of Trp^{35} and $H_{2'}$, $H_{2''}$ of G^3 (Table 2 in Supporting Information). The nine contacts of C^4 to Trp^{35} and to Asp^{24}-HN are weak and affected by line broadening, thus lowering the definition of the C^4 position. No intermolecular NOE was found for C^5, and the absence of chemical shift changes greater than 0.02 ppm for this base suggests that it is not involved in the binding of (14−53)NCp10. Regarding the latter, the presence of the pentanucleotide gives rise to new intramolecular NOE contacts not observed in free protein (*19*). This is the case for the methyl groups of Leu^{21} with the side chain of Ala^{27} and Ala^{36} and with the amide groups of Ala^{27}, Tyr^{28}, and Ala^{36}. These data and the intermolecular NOE cross signals between Leu^{21} side chain protons and C^2-$H_{1'}$ and C^2-H_5 show that d(ACGCC) complexation stabilizes a conformation in which a proximity between Leu^{21} and hydrophobic residues of both the zinc finger moiety and the nucleotide occurs (Figure 5).

The temperature variation of amide proton signals in the complex and free (14−53)NCp10 was measured over the 293−313 K range (Table 3 in Supporting Information), and the plots $\Delta\delta/\Delta T$ were found to be almost linear with slopes of −6.5 to −9.5 ppb/K for residues 14−21 and −5.0 to −9.5 ppb/K for residues 44−53. This supports the high solvent accessibility for these amide protons located in flexible regions devoid of well-structured motifs. In the central part, residues 22−44, most amide protons were protected from exchange with the solvent by numerous hydrogen bonds present in the folded structure formed by a loop spanning

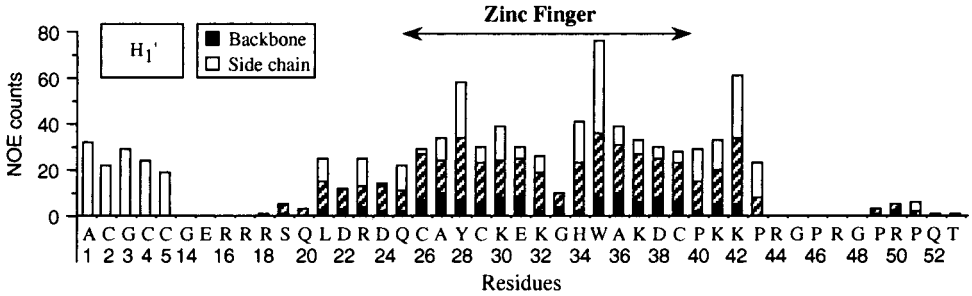

FIGURE 3: Number of restraints per residue corresponding to the 577 NOEs observed in the (14−53)NCp10−d(ACGCC) complex: 458 intramolecular distance restraints for the peptide, 79 for the nucleotide, and 40 intermolecular contacts were used for structure determination. Key: black, backbone−backbone; hashed, backbone−side chain; white, side chain−side chain interactions.

residues 22−25 and four loops encompassing residues 26−28, 31−33, 35−38, and 39−42, respectively. Upon complexation, several of these amide protons became even less accessible to the solvent as illustrated by positive differences in temperature variations between complexed and free forms. The most noticeable relative changes were observed for Asp^{24} (+2.8 ppb/K), Gln^{25} (+3.0 ppb/K), Ala^{27} (+4.5 ppb/K), Ala^{36} (+5.5 ppb/K), Asp^{38} (+2.5 ppb/K), Cys^{39} (+1.8 ppb/K), and Lys^{41} (+1.0 ppb/K).

In the complex, d(ACGCC) showed 79 intramolecular NOEs of which only 7 were interresidual. The structure of the bound single-stranded pentanucleotide differs from the standard nucleic acid conformations A and B as illustrated by the interresidue cross-peaks A^1-$H_{1'}$/C^2-H_6, A^1-$H_{2',2''}$/C^2-H_5, and A^1-$H_{2,\ 2''}$/C^2-H_6 and the absence of sequential $H_{3'}$/H_6, H_8 NOE cross signals expected for the B-type conformation and the absence of sequential H_6, H_8/$H_{1'}$ cross signals characteristic of the A-conformation.

Structure Calculations. The distribution of the total of 577 NOE distance restraints used in the structure calculations of the complex is depicted in Figure 3. This includes 129 sequential, 68 medium-range, and 124 long-range intramolecular NOEs. The distribution of the NOEs for each residue shows a large number of distance restraints in the zinc finger domain especially for Trp^{35} and Tyr^{28}. Most of the long-range NOEs were found in this region.

Structure calculations of complexes between a protein and a single-stranded nucleic acid are complicated by the difficulty of finding a population of reasonable low-energy structures for the oligonucleotide. Thus in de novo structure calculations using a simulated annealing protocol (*42*), the protein moiety is well defined but not the nucleotide moiety. The intramolecular NOE restraints in the nucleotide indicate that the conformation of d(ACGCC) in the complex differs from the idealized nucleotide conformations (A or B). Therefore, as already discussed (*26*) for the NCp7−d(ACGCC) complex, the most successful approach was to separately calculate the (14−53)NCp10 structure in the complex and to construct the nucleotide conformation by successive restrained energy minimizations from idealized conformations, applying first only intramolecular constraints and then adding the intermolecular NOE constraints in a docking step.

By this way two sets of structures of the (14−53)NCp10−d(ACGCC) complex denoted "A" and "B" according to the type of the idealized starting conformation of the nucleotide, were calculated.

The 20 best structures of each set were superimposed with respect to the backbone (N, Cα, C′,O) of residues 21−43,

Table 2: Energetic and Structural Statistics for the (14−53)NCp10−d(ACGCC) Complex (20 Structures)

av energy analysis (kcal/mol)	
E (total)	−103 ± 18.9
E (restraints)	9 ± 1.7
E (nonbonded)	−161 ± 8.2
rmsd from exptl distance restraints (577)	0.02 ± 0.002
rmsd from idealized covalent geometry	
bonds (Å) (835)	0.006 ± 0.0003
angles (deg) (1517)	3.00 ± 0.062
dihedral ω (deg) (39)	7.08 ± 0.552
rmsd of Cartesian coordinates (20 structures with respect to lowest energy structure) (Å)	
backbone (N, Cα, C′,O)	
14−53	5.58 ± 1.53
21−43	0.96 ± 0.29
21−39	0.78 ± 0.27
26−39	0.65 ± 0.27
all atoms	
14−53	6.74 ± 1.76
21−43	1.88 ± 0.50
21−39	1.82 ± 0.51
26−39	1.71 ± 0.50

giving an average rmsd of 0.68 ± 0.10 Å for A and 0.67 ± 0.13 Å for B. The structural statistics are depicted in Table 2. For the pentanucleotide, the rmsd were A^1 0.44 ± 0.13 Å, C^2 0.55 ± 0.12 Å, G^3 0.47 ± 0.20 Å, C^4 0.46 ± 0.27 Å, and C^5 0.36 ± 0.19 Å in A and A^1 0.24 ± 0.20 Å, C^2 0.21 ± 0.07 Å, G^3 0.26 ± 0.14 Å, C^4 0.27 ± 0.11 Å, and C^5 0.44 ± 0.20 Å in B. The rmsd values are relatively weak and are in favor of a complex belonging to series B. The position of the nucleotide bases in the complex with NCp10 is similar in both series A and B (rmsd A^1−C^4 heavy atoms 2.54 Å and excluding G^3 1.81 Å). However, the B series shows lower total and distance restraint energies and reproduces better the stacking of Trp^{35} to G^3.

To verify that the position of the nucleotide in the complex is not biased by the docking protocol, the nucleotide was removed from the complex, placed 20 Å aside the protein, and rotated by 180° about the binding axis. A restrained dynamical simulated annealing protocol was used, showing that the same complex with Trp insertion between G^3 and C^4 is obtained.

Structure of the (14−53)NCp10−d(ACGCC) Complex. The interaction of (14−53)NCp10 with the pentanucleotide d(ACGCC) leads to important changes in the conformation of both ligands.

Thus, a stable conformation of the sequence outside the zinc finger comprising residues Leu^{21} to Cys^{26} is induced by the pentanucleotide. Residue Leu^{21} comes into close contact with Ala^{27} and Ala^{36}, as shown by the NOE effects

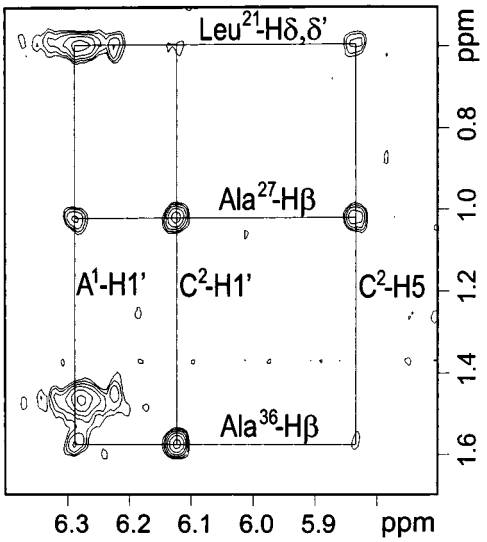

FIGURE 4: NOESY region of (14−53)NCp10−d(ACGCC) (τ_m 100 ms, ^2H_2O, pH 6.5, 2 mM) showing intermolecular contacts. The cross-signals between A^1-$H_{1'}$, C^2-$H_{1'}$, -H_5, and the methyl groups of Leu^{21}, Ala^{27}, and Ala^{36} (and between the methyl groups, not shown) illustrate the presence of a hydrophobic cluster at the interface region.

between the methyl groups of Leu^{21} and Ala^{27} present only in the complex and the NOEs between the amide groups of Ala^{27}, Ala^{36}, Tyr^{28}, Cys^{29}, and the δ,δ'-CH_3 of Leu^{21}, which is also in contact with C^2-$H_{1'}$, H_5, in the complex. A NOESY section showing the methyl groups of Ala^{27}, Ala^{36}, and Leu^{21} close to A^1-$H_{1'}$, C^2-$H_{1'}$, H_5 is reported in Figure 4. The reorientation seems to be caused by the presence of nucleotides A^1, C^2, and G^3 which induce the formation of a hydrophobic binding pocket including Leu^{21} (see Figure 5). The side chain of Trp^{35} is found inserted between G^3 and C^4 (Figure 6), in agreement with the experimental findings described before, and leads to a chain lengthening of the nucleotide up to 6.2 Å for $dC_{1'}(G^3, C^4)$.

The 20 energy-minimized structures deduced from the simulated annealing calculation are shown in Figure 6. The rmsd residue by residue is illustrated by the thickness of the ribbon (range 0.4−15.2 Å). The structural change of NCp10 upon complexation is reflected by the changes of the backbone rmsd values (N, Cα, C′, O) for the best structure as compared to that of free NCp10 (*14*), which are 1.27 Å for the residues of the zinc box, 26−39, 1.57 Å for the 26−43 sequence, 3.27 Å for the 21−39 sequence, and 3.42 Å for the 21−43 sequence. Among the five turn elements, 22−25, 26−28, 31−33, 35−38, and 39−42, found in the central structured domain of NCp10, only the first one was slightly modified in the complex. The ϕ, ψ angles (given in degrees) of the best structure for the $i+1$ and $i+2$ residues in the β-turns and $i+1$ in the γ turn are as follows: 22−25, β_I-turn type (−54, −52, −63, −57); 26−28, γ_i-turn type (−84, 46); 31−33, 3_{10}-turn (−35, −57); 35−38, β_{III}-turn (−60, −16, −65, −59); 39−42, β_{III}-turn (−79, −32, −74, −14).

The β_I-turn formed by residues 22−25 positions the side chain of Arg^{23} in a manner that it could interact with the phosphate group P-III or P-IV in agreement with the NMR data. The amide protons of residues 23 and 24 are hydrogen bonded to Asp^{22} −COO−. In the zinc finger domain, residues 26−28 form an inverse γ-turn [hydrogen bond between Tyr^{28} (NH) and Cys^{26} (CO) already observed in free NCp10 (*14*)]. The region 31−35 differs in the free and the complexed form.

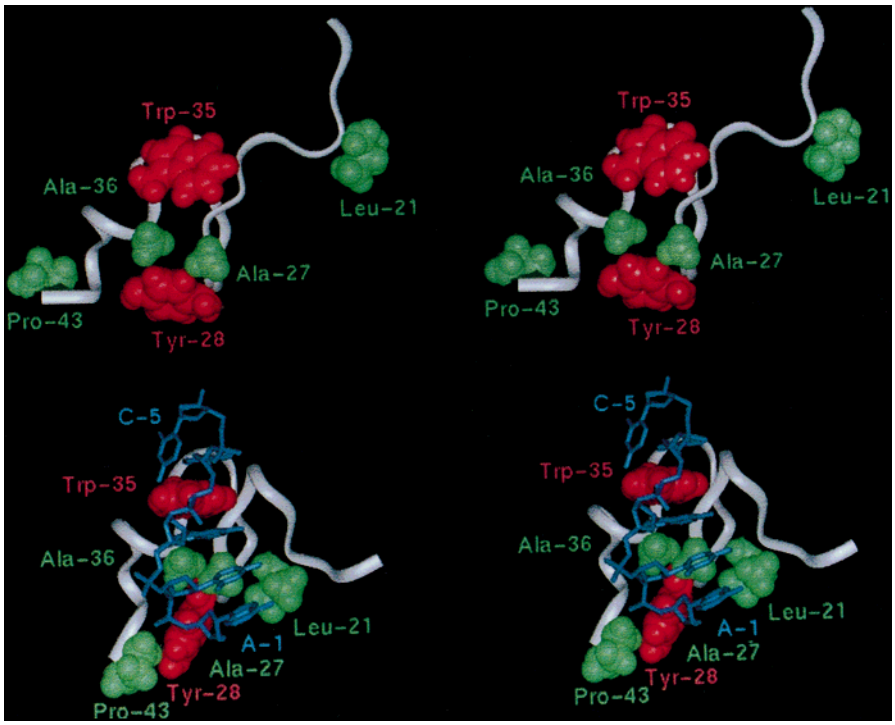

FIGURE 5: Stereoviews (ribbon representation) of the residues 18−43 in free NCp10 (top) and in the complex (14−53)NCp10−d(ACGCC) (bottom) with superimposition of the backbone atoms (N, Cα, C′,O) of residues 26−39. The side chains of Leu^{21}, Ala^{27}, Ala^{36}, and Pro^{43} (green) and Tyr^{28} and Trp^{35} (red) are shown with van der Waals surfaces. The rigid zinc finger domain does not alter significantly its overall structure upon complexation. Major conformational changes occur close to the zinc array due to the participation of several residues in the binding process. A fourth loop is stabilized, and the methyl groups of Leu^{21} approach Ala^{27} to form a hydrophobic cluster which participates in the binding process of d(ACGCC) (heavy atoms given in blue). The positions of both aromatic residues remain unchanged with only a variation of their side chain conformation.

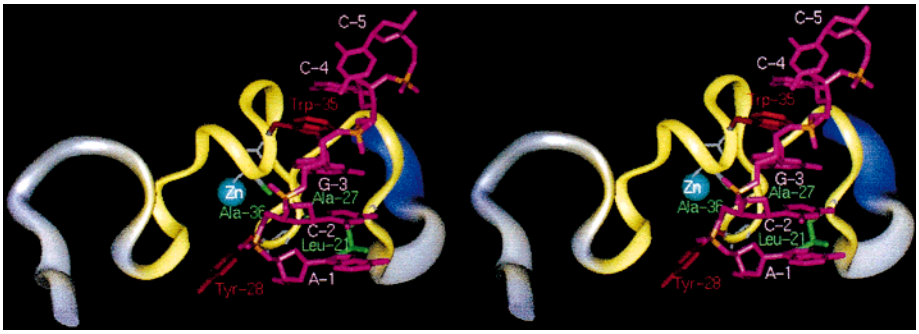

FIGURE 6: Stereo representation of the complex (14−53)NCp10−d(ACGCC). The width and color of the ribbon reflect the backbone rmsd by monomer calculated from 20 energy-minimized final structures superimposed with respect to the backbone of residues 21−43. This shows that he N-terminal (14−20, to the right) and C-terminal sequences (44−53, to the left) are more flexible. The lengthening of the ribose phosphate backbone is caused by Trp^{35} intercalation between the successive G^3-C^4 bases. The G^3 base has turned relative to C^2 and A^1 and is located in a hydrophobic cleft allowing the stacking of Trp^{35} on G^3 and the formation of hydrogen bonds.

The γ-turn (*30*−*32*) stabilized in free NCp10 by a tight hydrogen bond between Trp^{35} (NH) and Gly^{33} (CO) was not observed in the complex. The highly conserved residue Gly^{33} is in the 3_{10}-type turn of residues 31−33 forming a hydrogen bond between Gly^{33} (NH) and Glu^{31} (CO) in some structures. It can be observed that there are similar structural elements in NCp7 with Gly^{22} and Gly^{43} forming almost the same 3_{10} bend (*17*, *45*, *46*). This region is further stabilized by zinc coordination of the His^{34} side chain and salt bridges between Lys^{32} and Lys^{41} side chains and Glu^{31}. Residues Lys^{30}, Lys^{32}, and Lys^{41}, located on the same side of the complex structure, are easily accessible and could participate in additional binding. Residues 35−38 form a β_{III}-turn stabilized by a hydrogen bond between Asp^{38} (HN) and Trp^{35} (CO) which forms a second hydrogen bond to Cys^{39} (NH). A β_I-turn [Lys^{42} (NH), Cys^{39} (CO)] is found for residues 39−42. The last two turns were also found in free NCp10 (*14*).

A hydrophobic cleft formed by Trp^{35} and by nearly perfectly aligned residues Leu^{21}, Ala^{27}, and Ala^{36} (CC distance of the methyl groups of Leu^{21}/Ala^{27}, 4.40 Å, and Ala^{27}/Ala^{36}, 3.60 Å) fits to complementary aromatic rings of the nucleic acid bases. Most of the observable intermolecular NOEs correspond to amino acids belonging to this hydrophobic cleft (see Table 1). Tyr^{28} was found to be too far apart to interact with the nucleotide bases (Tyr^{28}-Cγ/A^1-N_9 8.60 Å, Tyr^{28}-Cγ/C^2−N_1 10.18 Å).

In the refined structure, the intermolecular hydrogen bonds are G^3 (NH) with Asp^{22} (CO), Gln^{25} (CO); C^4 (NH_2) with Asp^{24} (CO); Ala^{27} (HN) and Ala^{36} (HN) with G^3 (O_6); the guanidinium group of Arg^{23} with P-III, P-IV. Lys^{42} (NH_3^+), which was shown by point mutation experiments to be essential for NCp10 functioning (*5*), interacts with the phosphate group P-I of d(ACGCC), whereas Lys^{41} was not found in the proximity of the oligonucleotide but forms a salt bridge with Glu^{31}. Lys^{37} establishes contacts to P-II, and finally all phosphate groups of the nucleotide have potential partners in NCp10 to form salt bridges.

DISCUSSION

The aim of this work was to investigate the mechanism of single-stranded nucleic acid recognition by NC proteins through the structural determination of the complex formed between (14−53)NCp10 from MoMuLV and d(ACGCC). It could be argued that NCps interact mainly with RNA and not DNA sequences. However, both types of single-stranded nucleotides show similar binding properties with NCps as shown by various experimental methods (*26*, *27*, *30*), inasmuch as a guanosine residue is present in the nucleic acid. The pentanucleotide d(ACGCC) was therefore selected for this study because it allows a comparison with its complex with NCp7 studied in our laboratory (*21*, *26*) and with the N-terminal zinc finger of NCp7 of HIV-1 MN (*16*).

Interestingly, the peptide (24−42)NCp10 restricted strictly to the zinc finger domain did not bind d(ACGCC), as shown by both NMR and fluorescence. This is in accordance with biochemical experiments performed with (24−41)NCp10 and a variety of single-stranded DNA and RNA sequences (*29*). Under the same experimental conditions the 13−30 sequence corresponding to the N-terminal zinc finger of NCp7 of MN (*17*) or Mal strains (*29*), which have only one more positive charge than the (24−42)NCp10, was shown to interact with d(ACGCC). The smallest fragment of NCp10 able to bind d(ACGCC) was the peptide (14−53)NCp10, which exhibited an apparent affinity constant for d(ACGCC), $K_{aff} = 3 \times 10^7$ M^{-1}, not significantly different from that of NCp10. The differences in d(ACGCC) recognition observed between (24−42)NCp10 and the extended (14−53)NCp10 finger domain can be explained by the nature of the intermolecular interactions observed in the (14−53)NCp10−d(ACGCC) complex. Thus, in solution, the free (14−53)NCp10 has the same structural characteristics as the entire (1−56)NCp10 (*14*) with a finger domain highly constrained by various loops around the zinc ion and flexible N- and C-terminal sequences. The folding of the NCp10 zinc finger was roughly maintained in the complex with d(ACGCC), but two important modifications were observed. The first one was the stabilization of a preferential conformation of the peptide leading to a spatial proximity of the hydrophobic residues Leu^{21}, Ala^{27}, and Ala^{36} (Figure 5). This modification leads to a folding of residues 21−27 and to the formation of a well-defined binding site in which the guanosine residue fits. The second modification was a conformational change associated with a slight modification in the structure of the finger around the highly conserved Gly^{33} residue. This structural adaptation of the CCHC box is very likely required for optimal recognition of the oligonucleotide target and could explain why this glycine, devoid of steric hindrance, cannot be replaced by other amino acids (*47*).

The coordinates of the (14−53)NCp10−d(ACGCC) complex show that several hydrophobic and some ionic interac-

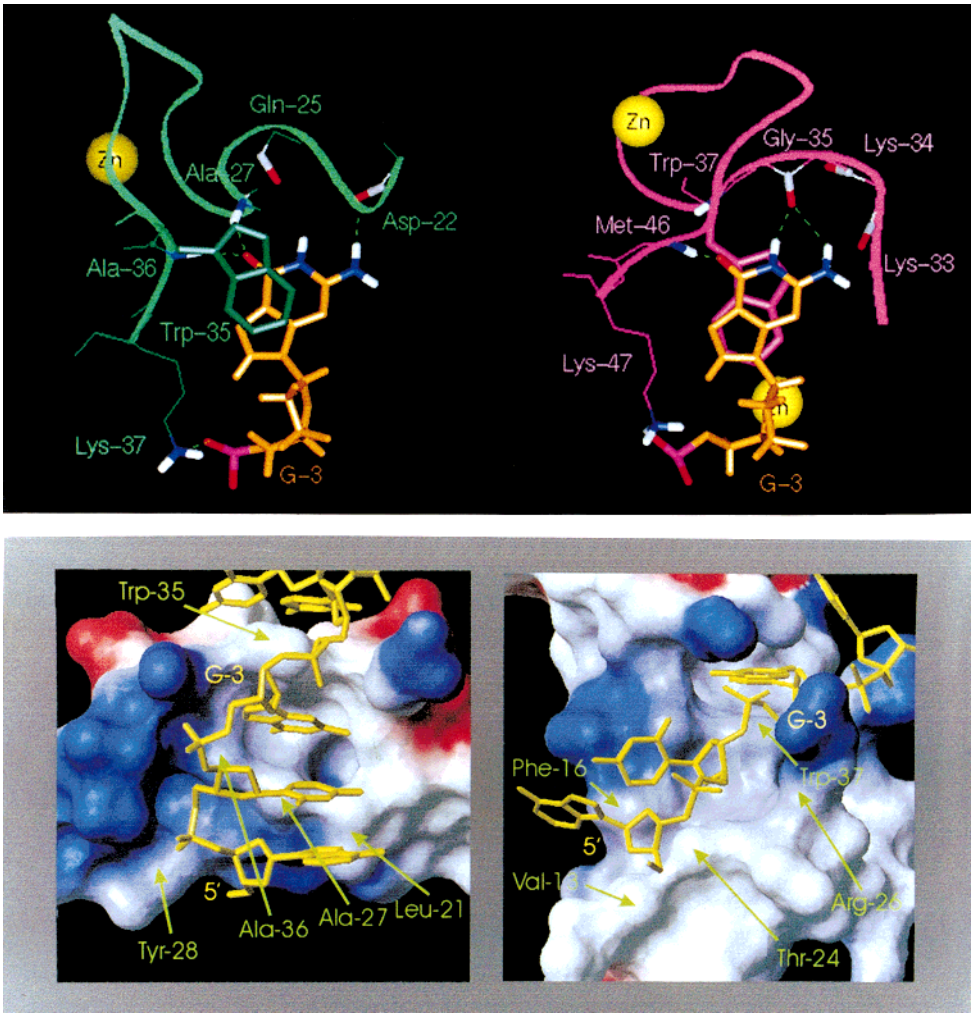

FIGURE 7: (Top) Analogies in the complexes (14−53)NCp10−d(ACGCC) (left) and (12−53)NCp7−d(ACGCC) (right) regarding the position of the dG^3 base (intermolecular hydrogen bonds are given as stippled lines). Both NCs bind G^3 in a similar manner using the same hydrogen bond network and a stacking of Trp on dG^3. The NCp10 zinc finger exhibits a folding similar to that of the C-terminal zinc finger of NCp7. However, the Trp residue, located after His^{35} in NCp10 and after Cys^{36} in NCp7, is found intercalated between dG^3 and dC^4 in the NCp10 complex and between dC^2 and dG^3 in the NCp7 complex. The hydrogen bond donors and acceptors involved in Gua binding in complexed NCp10 (NH_3, Lys^{37}; NH, Ala^{36}, Ala^{27}; CO, Gln^{25}, Asp^{22}) have equivalents in complexed NCp7 (NH_3, Lys^{47}; NH, Met^{46}, Trp^{37}; CO, Gly^{35}, Lys$^{34/33}$). (Bottom) Surface model of the electrostatic potential of the NCs (heavy atoms) in (14−53)NCp10−d(ACGCC) (left) and (12−53)NCp7−ACGCC (right). Positive potentials are given in blue and negative in red. The Gua residue is bound into a lipophilic pocket flanked by Trp which is on different sides of the dG^3 base in the two complexes. Successions of aligned hydrophobic residues lead to similar intermolecular hydrophobic surfaces exposed to the nucleotide in both NCs. They are composed of the side chains of Leu^{21}, Ala^{27,36}, and Tyr^{28} in NCp10 and Val^{13}, Thr^{24}, Ala^{25}, the β,γ-CH_2 of Arg^{26}, and Phe^{16} in NCp7. The hydrophobic groups contact A^1 and C^2 in both NCs (left and right), but the positions of the Trp residues (stacked on G^3) illustrate that the oligonucleotide orientation in both NC complexes is different. This figure was prepared with the program MOLMOL (*63*).

tions contribute to its stabilization. The hydrophobic contacts involve residues Ala^{27}, Tyr^{28}, Trp^{35}, and Ala^{36}, which belong to the finger domain, and also Leu^{21} and possibly Pro^{43}, which are located on both sides of the CCHC box (Figures 6 and 7).

These amino acids probably play a crucial role in complex formation since the peptide (24−42)NCp10, which contains only two alanines as lipophilic residues, is unable to bind d(ACGCC), and similar observations with (24−41)NCp10 and other nucleotides have been recently reported (*29*). The role of Leu^{21} in the formation of a binding site for (14−53)NCp10 may explain the in vivo RNA packaging defect of a NCp10 mutant characterized by a deletion of residues 16−23 and of a chimeric mutant in which the NCp10 zinc finger domain was provided with the flanking residues present on both sides of the two CCHC boxes of HIV-1 NCp7 (*48*). In the case of NCp7, interaction of d(ACGCC) with the hydrophobic cluster formed by Val^{13}, Ile^{24} (Thr^{24}), and Ala^{25} in the proximal zinc finger (13−30)NCp7 is sufficient for d(ACGCC) interaction (*16*, *49*), but the presence of the C-terminal domain containing the Trp^{37} residue in spatial proximity to the N-terminal CCHC box increased the affinity for d(ACGCC) and was shown to be necessary for the formation of a biologically relevant complex (*21*, *26*, *31*). Accordingly, replacement of Trp^{37} by Ala or Leu led to a considerable loss of virus infectivity (*20*, *50*) probably related to the drastic reduction in NCp7-induced formation of the tRNA−RNA duplex necessary for retrotranscription (*31*, *51*).

The basic residues in NCp10, which have been shown to be essential for in vitro RNA dimerization and tRNA^{Pro} annealing (*5*, *6*), have numerous contacts with the nucleotide in the complex (data not shown). Some of them, such as Lys^{37} and Lys^{42}, are hydrogen bonded to P-III and P-I,

respectively. The highly conserved residue Lys[42], which seems to be placed in a key position for ionic interaction, is probably equivalent to Lys[36] in the avian retrovirus NCp12. Accordingly, the mutation of Lys[36] by Ile was reported to completely block RNA packaging (*52*). Likewise in NCp7, the replacement by Gly of Arg[32] located in the linking sequence at a position similar to Lys[42] in NCp10 and Lys[36] in NCp12 led to a complete loss of virus infectivity (*53*). In addition to Leu[21], the three basic residues Arg[16,17,18], shown to be essential for RNA dimerization in vitro (*54*) and for virus infectivity of NCp10 (*55*), are lacking in (14−42) NCp10, which was found unable to bind d(ACGCC). This suggests that electrostatic attractions between positively charged amino acids of the protein and the ribose phosphate backbone of the pentanucleotide play an important role in the kinetic component of the binding process as already shown in the case of DNA intercalators (*56*).

As illustrated in the case of NCp7 and NCp10 (this study), the main components governing the recognition of nucleic acids by NCps are the aromatic residues present in one or two fingers. The side chains of these amino acids are involved in stacking interactions with the bases (*26*) and are embedded in a cluster of hydrophobic amino acids constituting a platform for van der Waals interactions (*27*). Accordingly, in RSV NCp12, which contains two zinc fingers, deletion of the proximal finger, which shows similarities to that of NCp10, produces a complete loss of infectivity while deletion of the distal finger devoid of aromatic residues has no significant effect (*57*).

It is important to observe that, in the (14−53)NCp10−d(ACGCC) complex, Tyr[28] (Figure 6) does not directly participate in the oligonucleotide recognition, suggesting that it could be involved in the binding of important functional viral proteins such as RT or integrase. Accordingly, the replacement of Tyr[28] by Ser in MoMuLV affects genomic RNA packaging only moderately but results in noninfectious virions (*57*). Similar results were observed in Friend MuLV mutants in which the replacement of Tyr[28] by Ser or Gly led to a complete loss of virus infectivity although only a 50% reduction in genomic RNA content was measured in the virus particles.

An important result of this NMR study is the clear demonstration of the insertion of Trp between two successive bases and its stacking to dG. Trp[35] is inserted between the G^3 and C^4 bases of d(ACGCC) whereas in the case of NCp7, Trp[37] is inserted in the C^2-G^3 sequence (*26*).

The dG residue involved in NCps recognition directs its base moiety to a cleft (*27*) allowing Trp stacking and specific hydrogen bonds to be formed (Figure 7). Thus the guanine is not available for base pairing, accounting for the preference of NCps for unpaired nucleic acid binding. This is illustrated by the complex formed by the HIV-1 NCp7 and the stem−loop sequence SL3 (*27*) in which the protein interacts by its zinc finger domain with nonpaired nucleotides of the loop. In the complex, the loop residue G^7, which has its base moiety flipped out of the loop, was found to be oriented toward the C-terminal zinc finger of NCp7 and stacked to Trp[37] as in the case of the (12−53)NCp7−d(ACGCC) complex (*26*).

The affinity of the pentanucleotide d(ACGCC) to the isolated N-terminal CCHC zinc finger of $NCp7(F1)^{MN}$ (*16*) is weaker than for (12−53)NCp7 or NCp7 (*26*). This is essentially caused by the absence of the binding pocket formed by the two successive zinc fingers which allows guanine/tryptophan interaction (*26, 27*). This accounts for the specificity of NCps for guanine in the nucleotide sequences of oligonucleotides.

The zinc finger of NCp10 (*14*) has a folding almost identical to that of the two zinc fingers of NCp7 (*15, 45*). The comparison of the NCp10−d(ACGCC) complex with NCp7−d(ACGCC) (*27*) (Figure 7, top) following superposition of G^3 in both structures shows the similarity of NCp10 zinc finger spatial disposition with that of the NCp7 C-terminal zinc finger which carries the Trp residue. Thus, similar side chain positions are observed for Arg[23], Lys[32], Ala[36], Lys[37], and Lys[42] in NCp10 and Arg[32], Lys[41], Met[46], Lys[47], and Lys[38] in NCp7, respectively. In the complex with d(ACGCC), the tryptophan side chain of NCp7 is placed between C^2 and G^3 in NCp7 (*26*) and between G^3 and C^4 in NCp10 (this study). Nevertheless, the hydrogen bonds found for G^3 in the NCp10−d(ACGCC) complex [G^3 (NH_2)/Asp[22] (CO); G^3 (NH)/Asp[22] (CO), Gln[25] (CO); G^3 (O_6)/Ala[27] (NH), Ala[36] (HN); G^3 (PO_4)/Lys[37] (NH_3^+)] correspond to those found for G^3 in the NCp7−d(ACGCC) complex [G^3 (NH_2)/Lys[33] (CO), Lys[34] (CO); G^3 (NH, NH_2)/Gly[35] (CO); G^3 (O_6)/Met[46] (NH); G^3 (PO_4)/Lys[47] (NH_3^+)] (*26*). In addition, an almost identical orientation of the loop G^7 residue in the NCp7−SL3 complex leads to a similar hydrogen bond pattern [G^7 (NH)/Gly[35] (CO); G^7 (O_6)/Trp[37] (NH), Met[46] (NH); G^7 (PO_4)/Lys[47] (NH_3^+)] (*27*). The succession of hydrophobic groups Tyr[28], Leu[21], Ala[27], and Ala[36] in the zinc finger of NCp10, which are in contact with the hydrophobic protons of A^1 and C^2, was also observed in the NCp7−d(ACGCC) complex. This includes Phe[16], Val[13], Thr[24], Ala[25], and the β,γ (CH_2) part of Arg[26]. In NCp7, these residues are located on the N-terminal finger and contact the nucleotide to form the wall of a deep cleft enclosing the ligand (Figure 7, bottom).

The recently published solution structure of another nucleocapsid protein of the Mason−Pfitzer monkey virus (MPMV) (*58*) shows that this NCp could exhibit binding properties with single-stranded nucleic acids similar to those delineated here for NCp10 or NCp7 binding. Its CCHC zinc fingers fold similarly around the metal as in NCp7 and NCp10. The proximal zinc finger of MPMV contains two Phe residues, and a Trp residue is present in the second finger, which is located in the same key position as Trp[37] in NCp7 and Trp[35] in NCp10, respectively. The fluorescence of this amino acid Trp[62] is almost completely quenched during the interaction with nucleic acids (*58*) exactly as this occurs with NCp7 and NCp10. It is therefore tempting to hypothesize that the mechanism of nucleic acid recognition of these three NCps is similar. This could account for the fact that NCp10 is able to substitute for NCp7, although with a weaker efficiency, in potentiating dimerization of genomic RNA of HIV-1 (1−415) in vitro (*59*) and that NCp7 is active in a primer extension assay with a MoMuLV RNA template (*29*).

The biological relevance of these structural studies is supported by the results of numerous experiments. RNA dimerization was shown to occur by specific recognition of well-defined stem−loop structures present in the Ψ domain (*60*). In the case of MoMuLV, the most important stem−loop structure encompasses the 283−298 nucleotide sequence

and contains a single-stranded AGCU loop (*61*). NCp10 potentiates the dimerization of two 283−298 sequences, a process which could be due to the intercalation of NCp10 between the GC sequence of AGCU followed by the fraying of this stem−loop structure and formation of a thermodynamically more stable dimeric structure (*61*). Mutations affecting the structure of the zinc finger domain and giving rise to viruses devoid of infectivity severely decrease the packaging of RNA and increase its sensitivity to nucleases (*8, 15, 20*). Nevertheless, significant amounts of genomic RNA were found in all these noninfectious virus particles, indicating that the bona fide structure of the zinc finger(s) of NCps is crucially required at another essential step of the retrovirus life cycle (review in ref *2*). This could be the reverse transcription, which requires an initial annealing of the tRNA primer and its viral RNA template at the level of the primer binding site (PBS). This process could be facilitated by a Trp-dependent partial unfolding of the highly structured RNA template and the tRNA, thus facilitating their annealing (*62*). Accordingly, the replacement of Trp^{37} by Leu in NCp7 inhibits the tRNA−RNA annealing process (*31*). On the other hand, several amino acids of the zinc finger domain of NCp10 are not directly involved in the complex with d(ACGCC), and their solvent-accessible side chains are therefore free for possible additional specific interactions with, for instance, RT, as recently shown in the case of NCp7 (*12*). Modifications of the structure of NCps could hinder these interactions, accounting for the loss of virus infectivity associated with mutations in the finger domains (*17, 19, 20*).

In conclusion, the solution structure of the 1:1 complex of NCp10 with d(ACGCC) shows that it interacts with single-stranded nucleic acids through the same binding mode that other nucleocapsid proteins such as NCp7 use. The main characteristics are binding to a hydrophobic groove (*27*), Trp stacking (*26*), and stabilization of the nucleotide backbone by basic residues (*26, 27*). Moreover, this structural study accounts for most of the results of mutations in NCp10 and in NCp7 (*3, 4, 14, 20, 47, 50, 53*). Finally, it provides topological elements for the rational design of compounds able to selectively inhibit critical functions of NCps. Recently, polytherapies with RT and protease inhibitors have been shown to significantly improve the treatment of AIDS. Nevertheless, these drugs led to resistance arising from virus adaptional mutations (*32*). This problem could be overcome with compounds targetted to NCps zinc finger(s) since mutations in this (these) highly structured domain(s) appear to result in loss of viral infectivity.

ACKNOWLEDGMENT

We thank N. Morellet, H. de Rocquigny, and J.-L. Darlix for helpful discussions during this work and V. Theilleux and T. Huynh-Dinh for the synthesis of d(ACGCC). We gratefully acknowledge C. Dupuis for assistance in drafting the manuscript and M. Beinfeld for stylistic revision.

SUPPORTING INFORMATION AVAILABLE

Tabulated data of proton chemical shifts, NH temperature gradients, and distance restraints. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Coffin, J. N. (1985) in *RNA Tumor Viruses* (Weiss, R., Teich, N., Varmus, H., and Coffin, J., Ed.) Vol. I, pp 261−368, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Darlix, J.-L., Lapadat-Tapolsky, M., De Rocquigny, H., and Roques, B. P. (1995) *J. Mol. Biol. 254*, 523−537.
3. Gorelick, R. J., Henderson, L. E., Hanser, J. P., and Rein, A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8420−8424.
4. Housset, V., De Rocquigny, H., Roques, B. P., and Darlix, J.-L. (1993) *J. Virol. 67*, 2537−2545.
5. Prats, A.-C., Sarih, L., Gabus, C., Litvak, S., Keith, G., and Darlix, J.-L. (1988) *EMBO J. 7*, 1777−1783.
6. Darlix, J.-L., Vincent, A., Gabus, C., de Rocquigny, H., and Roques, B. P. (1993) *C. R. Acad. Sci., Ser. III 316*, 763−771.
7. You, J.-C., and McHenry, C. S. (1994) *J. Biol. Chem. 269*, 31491−31495.
8. Dorfmann, T., Luban, J., Goff, S. P., Haseltine, W. A., and Göttlinger H. G. (1993) *J. Virol. 67*, 6159−6169.
9. Lapadat-Tapolsky, M., De Rocquigny, H., Van Gent, D., Roques, B. P., Plasterk, R., and Darlix, J.-L. (1993) *Nucleic Acids Res. 21*, 831−839.
10. Risco, C., Menéndez-Arias, L., Copeland, T. D., Pinto da Silva, P., and Oroszlan, S. (1995) *J. Cell Sci. 108*, 3039−3050.
11. De Rocquigny, H., Petitjean, P., Tanchou, V., Décimo, D., Drouot, L., Delaunay, J. L., Darlix, J. L., and Roques, B. P. (1997) *J. Biol. Chem. 272*, 30753−30759.
12. Druillennec, S., Caneparo, A., de Rocquigny, H., and Roques, B. P. (1999) *J. Biol. Chem. 274*, 11283−11288.
13. Mély, Y., Cornille, F., Fournié-Zaluski, M.-C., Darlix, J.-L., Roques, B. P., and Gérard, D. (1991) *Biopolymers 31*, 899−906.
14. Déméné, H., Jullian, N., Morellet, N., De Rocquigny, H., Cornille, F., Maigret, B., and Roques, B. P. (1994) *J. Biomol. NMR 4*, 153−170.
15. Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J.-L., and Roques, B. P. (1992) *EMBO J. 11*, 3059−3065.
16. South, T. L., and Summers, M. F. (1993) *Protein Sci. 2*, 3−19.
17. Morellet, N., De Rocquigny, H., Mély, Y., Jullian, N., Déméné, H., Ottmann, M., Gérard, D., Darlix, J.-L., Fournié-Zaluski, M.-C., and Roques, B. P. (1994) *J. Mol. Biol. 235*, 287−301.
18. Lee, B. M., de Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) *J. Mol. Biol. 279*, 633−649.
19. Déméné, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N., Mély, Y., Darlix, J.-L., Fournié-Zaluski, M.-C., Saragosti, S., and Roques, B. P. (1994) *Biochemistry 33*, 11707−11716.
20. Gorelick, R. J., Chabot, D. J., Ott, D. E., Gagliardi, T. D., Rein, A., Henderson, L. E., and Arthur, L. O. (1996) *J. Virol. 70*, 2593−2597.
21. Roques, B. P., Morellet, N., de Rocquigny, H., Déméné, H., Schüler, W., and Jullian, N. (1997) *Biochimie 79*, 673−680.
22. Rice, W. G., Supko, J. G., Malspeis, L., Buckheit, R. W., Jr., Clanton, D., Bu, M., Graham, L., Schaeffer, C. A., Turpin, J. A., Domalga, J., Gliotti, R., Bader, J. P., Halliday, S. M., Coren, L., Sowder, R. C., II, Arthur, L. O., and Henderson, L. E. (1995) *Science 270*, 1194−1197.
23. Druillennec, S., Dong, C. Z., Escaich, S., Gresh, N., Bousseau, A., Roques, B. P., and Fournié-Zaluski, M. C. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 4886−4891.
24. Karpel, R. L., Henderson, L. E., and Oroszlan, S. (1987) *J. Biol. Chem. 262*, 4961−4967.
25. Tsuchihashi, Z., and Brown, P. O. (1994) *J. Virol. 68*, 5863−5870.
26. Morellet, N., Déméné, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournié-Zaluski, M. C., and Roques, B. P. (1998) *J. Mol. Biol. 283*, 419−434.
27. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−388.
28. Delahunty, M. D., South, T. L., Summers, M. F., and Karpel, R. L. (1992) *Biochemistry 31*, 6461−6469.
29. Wu, W., Henderson, J. E., Copeland, T. D., Gorelick, R. J., Bosche, W. J., Rein, A., and Lewin, J. G. (1996) *J. Virol. 70*, 7132−7142.

30. Fisher, R. J., Rein, A., Fivash, M., Urbaneja, M. A., Casas-Finet, J. R., Medaglia, M., and Henderson, L. E. (1998) *J. Virol. 72*, 1902−1909.
31. Rémy, E., de Rocquigny, H., Petitjean, P., Muriaux, D., Theilleux, V., Paoletti, J., and Roques, B. P. (1998) *J. Biol. Chem. 273*, 4819−4822.
32. Wong, J. K., Hezareh, M., Günthard, H. F., Havlir, D. V., Ignacio, C. C., Spina, C. A., and Richman, D. D. (1997) *Science 278*, 1291−1295.
33. De Rocquigny, H., Ficheux, D., Gabus, C., Allain, B., Fournie-Zaluski, M.-C., Darlix, J.-L., and Roques, B. P. (1993) *Nucleic Acids Res. 21*, 823−829.
34. Adam, S., Taboury, J. A., Taillandier, E., Popinel, A., Huynh-Dinh, T., and Igolen, J. (1986) *J. Biomol. Struct. Dyn. 3*, 873−885.
35. Rance, M., Sœrensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
36. Griesinger, C., Sœrensen, O. W., and Ernst, R. R. (1985) *J. Am. Chem. Soc. 107*, 6394−6396.
37. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1982) *J. Chem. Phys. 71*, 4546−4553.
38. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
39. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
40. Plateau, P., and Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
41. Bax, A., and Morris, G. A. (1981) *J. Magn. Reson. 42*, 501−505.
42. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett. 239*, 129−136.
43. Gorenstein, D. G., Luxon, B. A., Goldfield, E., Lai, K., and Vegeais, D. (1982) *Biochemistry 21*, 580−589.
44. Delepierre, M., Van Heijenoort, C., Igolen, J., Pothier, J., Le Bret, M., and Roques, B. P. (1989) *J. Biomol. Struct. Dyn. 7*, 557−590.
45. South, T. L., Blake, P. R., Have, O. R., and Summers, M. F. (1991) *Biochemistry 30*, 6042−6349.
46. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry 29*, 329−340.
47. Méric, C., and Goff, S. P. (1989) *J. Virol. 63*, 1558−1568.
48. Berkowitz, R. D., Ohagen, A., Höglund, S., and Goff, S. P. (1995) *J. Virol. 69*, 6445−6456.
49. Dannull, J., Surovoy, A., Jung, G., and Moelling, K. (1994) *EMBO J. 13*, 1525−1533.
50. Gorelick, J. R., Nigida, S. M., Jr., Bess, J. W., Jr., Arthew, L. O., Henderson, L. E., and Rein, A. (1990) *J. Virol. 64*, 3207−3211.
51. De Rocquigny, H., Delaunay, T., Petitjean, P., Fournié-Zaluski, M. C., and Roques, B. P. (1998) *Regards Biochimie 2*, 44−50.
52. Fu, X., Katz, R. A., Skalka, A. M., and Leis, J. (1988) *J. Biol. Chem. 263*, 2140−2145.
53. Ottmann, M., Gabus, C., and Darlix, J. L. (1995) *J. Virol. 69*, 1778−1784.
54. De Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M. C., Roques, B. P., and Darlix, J. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6472−6476.
55. Housset, V., de Rocquigny, H., Roques, B. P., and Darlix, J. L. (1993) *J. Virol. 67*, 2537−2545.
56. Gaugain, B., Barbet, J., Capelle, N., Roques, B. P., and Le Pecq, J. B. (1978) *Biochemistry 17*, 5078.
57. Méric, C., Gouilloud, E., and Spahr, P. F. (1988) *J. Virol. 62*, 3328−3333.
58. Gao, Y., Kaluarachchi, K., and Giedroc, D. P. (1998) *Protein Sci. 7*, 2265−2280.
59. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F., and Barré-Sinoussi, F. (1990) *J. Mol. Biol. 216*, 689−699.
60. Lever, A. M. L., Göttlinger, H., Haseltine, W., and Sodroski, J. (1989) *J. Virol. 63*, 4085−4087.
61. Girard, P. M., de Rocquigny, H., Roques, B. P., and Paoletti, J. (1996) *Biochemistry 35*, 8705−8714.
62. Li, X., Quan, Y., Arts, E. J., Li, Z., Preston, B. D., de Rocquigny, H., Roques, B. P., Darlix, J.-L., Kleiman, L., Parniak, M. A., and Wainberg, M. A. (1996) *J. Virol. 70*, 4996−5004.
63. Konradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.

BI990378D