High-Resolution Solution NMR Structure of the Minimal Active Domain of the Human Immunodeficiency Virus Type-2 Nucleocapsid Protein[†,‡]

Yoshio Kodera,[§,||] Kazuki Sato,[§] Tomonori Tsukahara,[⊥] Hiroyoshi Komatsu,[⊥] Tadakazu Maeda,[||] and Toshiyuki Kohno*[,§]

Mitsubishi Kasei Institute of Life Sciences, Machida, Tokyo 194-8511, Japan, and Department of Physics, School of Science, and Department of Immunology, School of Allied Health Sciences, Kitasato University, Sagamihara, Kanagawa 228-8555, Japan

Received July 29, 1998; Revised Manuscript Received October 8, 1998

ABSTRACT: The retroviral nucleocapsid (NC) protein is a multifunctional protein essential for RNA genome packaging and viral infectivity. The NC protein, NCp8, of the human immunodeficiency virus type-II (HIV-2) is a 49 amino acid peptide containing two zinc fingers, of the type C-X_2-C-X_4-H-X_4-C, connected by seven amino acid residues, called the "basic amino acid cluster." It has been shown that the N-terminal zinc finger flanked by the basic amino acid cluster is the minimal active domain for the specific binding to viral RNA and other functions. However, the structure−activity relationships of NCp8 have not been investigated in detail. In the present study, the three-dimensional structure of a 29 amino acid peptide, including the minimal active domain (NCp8-f1), was determined by two-dimensional ^{1}H NMR spectroscopy with simulated annealing calculations. A total of 15 converged structures of NCp8-f1 were obtained on the basis of 355 experimental constraints, including 343 distance constraints obtained from nuclear Overhauser effect connectivities, 12 torsion angle (ϕ, χ^1) constraints, and four constraints for zinc binding. The root-mean-square deviation of the 15 converged structures was 0.29 ± 0.04 Å for the backbone atoms (N, C^α, C) and 1.27 ± 0.13 Å for all heavy atoms. Interestingly, the basic amino acid cluster itself was defined well, with a loop-like conformation in which three arginine residues in the cluster and one arginine residue in the zinc finger are located approximately in the same plane of the molecule and are exposed to the solvent. The structure−activity relationships are discussed on the basis of the comparison of this well-defined structure with those of other NC proteins.

Nucleocapsid proteins (NC[1] proteins), structural proteins of retroviruses, are generated upon processing the gag polyprotein by the viral protease and are essential for various steps during retroviral replication. The NC proteins specifically discriminate the viral RNA from the cellular RNAs and facilitate dimerization and packaging of two RNA genomes into maturing viruses. In addition, the proteins possess many other functions: they enhance binding of the tRNA primer to the viral RNA, stimulate the reverse transcriptase, improve the efficiency of the integrase, and protect the RNA against nucleases. Each retroviral NC protein has either one or two conserved zinc finger domain(s), with the amino acid sequence Cys-X_2-Cys-X_4-His-X_4-Cys, which coordinate one zinc ion with very high affinity (X represents amino acids that vary among the viruses) (*1*, *2*).

Two types of human immunodeficiency viruses have been reported, HIV-1 and HIV-2, which contain NC proteins called NCp7 and NCp8, respectively (Figure 1). Both NCp7 and NCp8, which are 67% identity, contain two zinc fingers connected by seven amino acid residues, called the basic amino acid cluster. In the case of NCp7, several genetic studies have shown that mutations either of the basic amino acid residues in the cluster or of the first zinc finger result in the production of noninfectious viral particles (*3*, *4*). Moreover, site-directed mutagenesis that substituted a leucine residue for Pro^{31} resulted in the formation of noninfectious and immature viral particles (*5*). In vitro investigations demonstrated that the binding of the second zinc finger flanked by the basic amino acid cluster with viral RNA is very weak and nonspecific, whereas that of the first zinc finger flanked by the basic amino acid cluster is sufficient for specific RNA binding.

[†] This work was supported in part by the Ministry of Education, Science, Sports, and Culture of Japan, Kanagawa Academy of Science and Technology (KAST), and Kitasato University Research Grant for Young Researchers.

[‡] Atomic coordinates for the 15 converged structures of NCp8-f1 are deposited with Protein Data Bank, Brookhaven National Laboratories, Long Island, NY 11973, under the accession code 1nc8.

* To whom correspondence should be addressed. Fax: +81-427-24-6317. E-mail: tkohno@libra.ls.m-kagaku.co.jp.

[§] Mitsubishi Kasei Institute of Life Sciences.

[||] Department of Physics, Kitasato University.

[⊥] Department of Immunology, Kitasato University.

[1] Abbreviations: CD, circular dichroism; COSY, correlation spectroscopy; DQF−COSY, double quantum-filtered COSY; DSS, 4,4-dimethyl-4-silapentane-1-sulfonic acid; HIV, human immunodeficiency virus; HIV-1, human immunodeficiency virus type-I; HIV-2, human immunodeficiency virus type-II; HPLC, high-performance liquid chromatography; NC, nucleocapsid; NCp7, nucleocapsid protein of HIV-1; NCp8, nucleocapsid protein of HIV-2; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PE-COSY, primitive exclusive COSY; TOCSY, total correlated spectroscopy; rmsd, root-mean-squared deviation; WATERGATE, water suppression by gradient-tailored excitation; Standard abbreviations are used for usual amino acids.

10.1021/bi981818o CCC: $15.00
Published on Web 11/26/1998

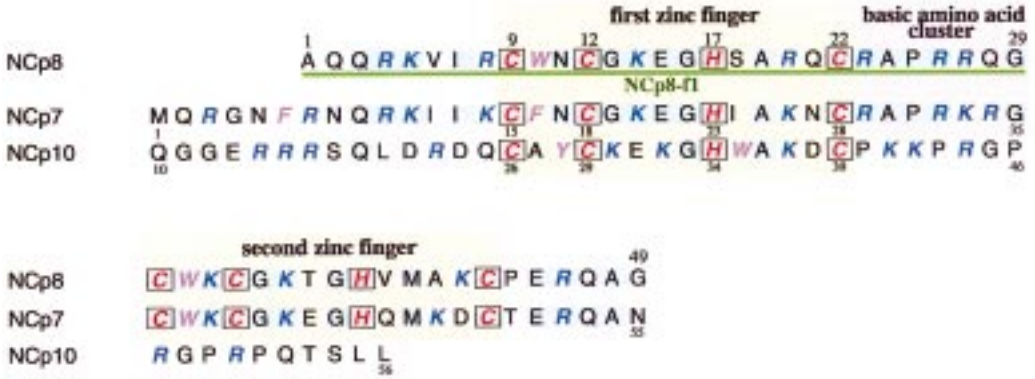

FIGURE 1: Amino acid sequences of the nucleocapsid protein of HIV-2 (NCp8), HIV-1 (NCp7), and Molony murine leukemia virus (NCp10). The program ALSCRIPT (*42*) was used to display the sequence alignment. The amino acid residues of NCp7 and NCp10 indicated by the light blue boxes are identical with those of NCp8, and the residues of NCp8 done by the light blue boxes are identical with those of NCp7.

The three-dimensional structure of the NCp7 and its fragment have been determined by ^{1}H NMR spectroscopy (*6*−*9*). These studies have shown that similar folding occurs for the two zinc fingers in the presence of zinc, which is consistent with the structures determined by previous studies of the zinc fingers of NCp7 (*10*, *11*). Recently, the dynamical behavior of NCp7 (*12*) and the three-dimensional structure of the NCp7−RNA complex (*13*) have been studied. These results indicate that the basic amino acid cluster adopts a stable conformation in its complex form, though that in its free state does not.

Despite a number of studies of the biological functions and the three-dimensional structure of NCp7, only a few studies have been done for NCp8 of HIV-2. A competitive UV cross-linking assay of NCp8 and NCp8-derived synthetic peptides (*14*) showed that the peptide corresponding to either the first or second zinc finger flanked by the basic amino acid cluster interacts specifically with viral RNA. Although the secondary structure of the first zinc finger of NCp8 has been reported (*15*), its three-dimensional structure has not yet been determined. Thus, the determination of the three-dimensional structure of NCp8 and the comparison of the NCp8 and NCp7 structures would yield important information to clarify the recognition mechanism of the viral RNA by NC proteins.

In the present study, the three-dimensional structure of the minimal active domain of NCp8, NCp8-f1, has been determined. In particular, the structure of the basic amino acid cluster was determined in detail for the first time. The structure−activity relationships will be discussed, on the basis of the comparison of the structure of NCp8 with those of other NC proteins.

MATERIALS AND METHODS

Peptide Synthesis. Fmoc amino acids and other reagents used with the synthesizer were purchased from Applied Biosystems Japan (Chiba, Japan). Other reagents for peptide synthesis were purchased from Peptide Institute (Osaka, Japan) or Kokusan Chemical Works Ltd. (Tokyo, Japan). The standard solid-phase methodology was employed to synthesize NCp8-f1, using a Perkin-Elmer Applied Biosystems Japan model 431A peptide synthesizer. The peptide was purified by a Shimadzu LC-8A system with a Cosmosil 5C18-AR reversed-phase C_{18} column (20 × 250 mm), and its purity was confirmed by a Shimadzu LC-6A system with a C_{18} column (4.6 × 250 mm). The peptide's identity was confirmed by analytical HPLC and TOF-MS spectra recorded on a Voyager-DE mass spectrometer (PerSeptive Biosystems).

Circular Dichroism Measurements. Circular dichroism (CD) spectra of NCp8-f1 in the presence and absence of zinc were recorded on a Jasco J-600 spectropolarimeter. The spectra were acquired at 25 °C over 185−260 nm, using a cuvette with a path length of 1 mm, a sensitivity of 50 mdeg/cm, and a scan speed of 100 nm/min. The concentrations of NCp8-f1 and $ZnCl_2$ dissolved in H_2O were approximately 15 and 17 μM, respectively. The spectra were recorded digitally and each signal was averaged four times, followed by subtraction of an equally signal-averaged solvent baseline by feeding through a data processor.

Analytical Ultracentrifugation. Analytical ultracentrifugal analyses, using the method of sedimentation equilibrium, were performed with a Beckman XL-A analytical ultracentrifuge and an An-60Ti rotor. The Zn^{2+}-bound NCp8-f1 (1 mM, pH 5.8) was prepared by diluting a portion of the NMR sample (5 mM NCp8-f1). This diluted sample showed the same 1D ^{1}H NMR spectrum as that of the 5 mM sample (data not shown). NCp8-f1 (1 mM), in the absence of zinc at pH 3.0, was also prepared for comparison. For each sample, centrifugation was carried out at 15 °C. The initial rotor speed was set at 3000 rpm. The speed was then increased to 48 000 rpm, that is, the maximum speed for the cell. Scans were collected at 314 nm for both samples after 12 and 24 h. Both scans were compared to ensure that equilibrium had been reached. The partial specific volumes calculated for Zn^{2+}-bound and Zn^{2+}-free NCp8-f1 were 0.6940 and 0.7057 mL/g, respectively.

NMR Spectroscopy. The samples for NMR experiments contained 5.0 mM NCp8-f1 and 5.5 mM $ZnCl_2$ in either 90% H_2O/10% ^2H_2O or 99.96% ^2H_2O at pH 5.8. Under these conditions, the NMR spectra were well-resolved, and no aggregation was observed over a period of several months, during which the oligomerization state of Zn^{2+}-bound NCp8-f1 was examined using analytical ultracentrifugation, as described above. NMR measurements were performed using standard pulse sequences and phase cycling on a Bruker DMX-500 or AMX-500 spectrometer operating at 500 MHz for the proton frequency. All two-dimensional NMR spectra were acquired in a phase-sensitive mode, using the time-proportional phase incrementation (*16*) for quadrature detec-

tion in the t_1 dimension. NOESY spectra (*17, 18*) were recorded with mixing times of 80, 100, 200, and 300 ms at temperatures of 5, 15, and 25 °C. TOCSY spectra were recorded using a MLEV-17 pulse scheme (*19*) with isotropic mixing times of 50 and 80 ms. The suppression of the solvent resonance was achieved using the WATERGATE scheme in both the NOESY and TOCSY measurements (*20*). DQF-COSY (*21*) and PE-COSY (*22*) spectra were recorded to obtain the constraints for the torsion angles and the stereospecific assignments. Selective irradiation during the relaxation delay period was used to suppress the solvent resonance. The data sizes used for acquisition were 512 (t_1) × 8192 (t_2) for DQF-COSY and PE-COSY and 512 × 2048 otherwise, and the spectral width was 6250 Hz.

Data processing was performed on either a Bruker X-32 UNIX workstation with UXNMR software or a Silicon Graphics INDIGO2 workstation using the program package NMRDraw and NMRPipe (*23*). Phase-shifted sine-squared window functions were applied prior to Fourier transformation, with shifts from $\pi/3$ to $\pi/2$ in both dimensions. Final matrix sizes were 1028 × 8192 real points for COSY and PE-COSY and 2048 × 2048 otherwise.

To detect the slowly exchanging backbone amide protons, we recorded a NOESY spectrum with a mixing time of 200 ms and at pH 8.7. The intensity of each intraresidue NH−C^{α}H cross-peak was then compared with that obtained at pH 5.8. The chemical shifts of C^{α}H were not significantly changed when the pH was changed from pH 5.8 to 8.7.

Chemical shifts were referenced to the methyl resonance of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) used as an internal standard. Complete sets of two-dimensional spectra were recorded at 15 °C and pH 5.8 (uncorrected meter readings).

Distance Constraints and Structure Calculations. Interproton distance constraints were obtained from the NOESY spectra with a mixing time of 200 ms. Quantitative determination of the cross-peak intensities was based on the counting of the contour levels. All NOE data were divided into four classes, strong, medium, weak, and very weak, corresponding to distance upper limits of 2.5, 3.5, 5.0, and 6.0 Å, respectively, in the interproton distance constraints. Pseudoatoms were used for nonstereospecifically assigned protons, and intraresidue and long-range correcting factors were added to the distance restraints, respectively (*24*). In addition, 0.5 Å was added to the upper limits for distance restraints involving methyl protons.

The backbone NH−C^{α}H coupling constants were estimated from the DQF-COSY spectrum, and were converted to the backbone torsion angle ϕ constraints according to the following rules (*25*): for $^3J_{NH-C^{\alpha}H}$ less than 5.5 Hz and greater than 8.0 Hz, the ϕ angles were constrained in the ranges of $-65° \pm 25°$ and $-120° \pm 40°$, respectively. The range of χ^1 side chain torsion angle constraints and the stereospecific assignments of the prochiral β-methylene protons were obtained by using the $^3J_{\alpha\beta}$ coupling constants combined with the intraresidue NH−C^{β}H NOEs observed with a mixing time of 200 ms (*26*). The $^3J_{\alpha\beta}$ coupling constants were determined by inspecting the PE-COSY spectrum in ^2H_2O.

All calculations were carried out on an HP 9000/720 workstation with the X-PLOR 3.1 program (*27*). The three-dimensional structures were calculated on the basis of the experimentally derived distance and torsion angle constraints, using a dynamically simulated annealing protocol starting from a template structure with randomized backbone ϕ and ψ torsion angles.

Evaluation Methods. The convergence of the calculated structures was evaluated in terms of the structural parameters, that is, the rmsd from the experimental distance and dihedral constraints, the values of the energy statistics (F_{NOE}, F_{tor}, F_{repel}, and E_{L-J}), and the rmsd from idealized geometry. The distributions of the backbone dihedral angles of the final converged structures were evaluated by the representation of the Ramachandran dihedral pattern, indicating the deviations from the sterically allowed (ϕ, ψ) angle limits (*28*). The degrees of angular variation among the converged structures were further assessed by using an angular order parameter *S* (*29*). The order parameter *S* was calculated using following equation:

$$S = \frac{1}{N}\left(\left(\sum_{j=1}^{N} \sin\theta_j\right)^2 + \left(\sum_{j=1}^{N} \cos\theta_j\right)^2\right)^{1/2}$$

in which *N* is the total number of converged structures and θ_j is a particular dihedral angle of the *j*th structure of the total number of structures.

RESULTS

Confirmation of the Zinc Binding. The CD and ^{1}H NMR 1D spectra (data not shown) were measured to confirm the zinc binding to NCp8-f1. The CD spectrum of NCp8-f1 in the absence of zinc at pH 3.0 is characteristic of a random coil conformation. A ^{1}H NMR 1D spectrum obtained under the same conditions shows poor dispersion of the amide proton resonances, indicating that this peptide does not adopt an ordered conformation under these conditions. Upon the addition of 1.1 mol of $ZnCl_2$ per mole of the peptide and an increase of the pH to 5.8, remarkable changes were observed in the CD spectrum: the minimum at 180 nm was decreased in intensity and shifted to 200 nm, and a new maximum appeared at 220 nm. Such spectral changes were also observed in NCp7 (*8*) and the first zinc finger fragment of NCp8 (*15*). Simultaneously, the amide and methyl proton resonances were better resolved in the presence of Zn^{2+} ion than in its absence. These results demonstrate that an ordered conformation is induced in NCp8 upon zinc binding.

Analyze the Oligomerization State of NCp8-f1. The oligomerization state of NCp8-f1 in the presence of zinc, under the conditions of the NMR measurements, was analyzed by equilibrium sedimentation. Analytical runs in the presence (pH 5.8) and the absence (pH 3.0) of zinc gave similar results at the rotor speed of 48 000 rpm, and the calculated mass was about 1500 for both samples, which indicates that NCp8-f1 exists in a monomer state under both conditions.

Sequential Resonance Assignments. Sequence-specific resonance assignments were achieved according to the standard method established by Wüthrich and co-workers (*30*). Since valine, isoleucine, and histidine residues each appear only once in the primary sequence of NCp8-f1, these residues were used as starting points in the sequential assignment process. In particular, the spin systems for Val^6 and Ile^7 were assigned by the observation of the magnetiza-

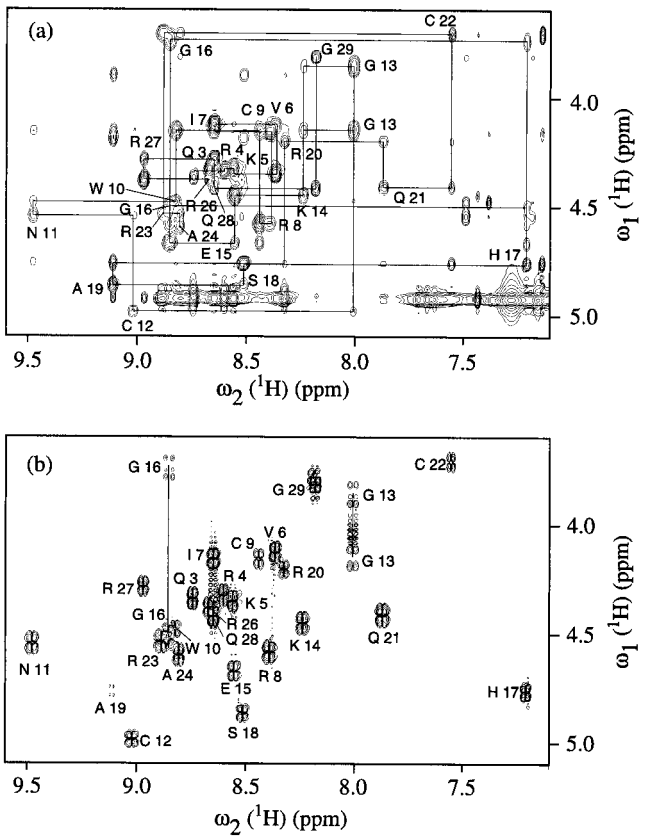

FIGURE 2: Portions of the 500 MHz two-dimensional NMR spectra of 5.0 mM NCp8-F1 in 90% H_2O/10% ^2H_2O at pH 5.8 and 15 °C. (a) Sequential $d_{\alpha N}(i,i+1)$ NOE connectivities for residues 1−29 in the NOESY spectrum observed with a mixing time of 200 ms. (b) The fingerprint region of the DQF-COSY spectrum corresponding to the same region as in spectrum (a). Intraresidue NH−C$^\alpha$H cross-peaks are labeled with the residue number by standard single-letter amino acid abbreviations.

tion transfer from C$^\alpha$H to C$^\gamma$H$_3$ of Val and that from C$^\alpha$H to C$^\gamma$H$_3$ and C$^\delta$H$_3$ for Ile in the TOCSY spectrum. The six arginine residues were easily distinguished from the other residues, on the basis of the observed cross-peaks between N$^\epsilon$H and C$^\alpha$H, C$^\beta$H$_2$, C$^\gamma$H$_2$, and C$^\delta$H$_2$ in the TOCSY spectrum, and were used to confirm the sequence-specific resonance assignments. Figure 2a shows the C$^\alpha$H−NH fingerprint region of the NOESY spectrum containing sequential $d_{\alpha N}(i, i+1)$ connectivities. The interresidue NOE cross-peaks were distinguished from the intraresidue ones by comparing the NOESY spectrum with the TOCSY and DQF-COSY spectra (Figure 2b). The Pro^{25} residue was assigned by the observation of a strong sequential NOE between the Ala^{24} C$^\alpha$H and the Pro^{25} C$^\delta$H$_2$. The presence of this NOE cross-peak indicates that the Pro^{25} in NCp8-f1 adopts the trans configuration. Figure 3 summarizes the sequential and medium-range NOE connectivities observed in the 200 ms NOESY spectrum, together with the $^3J_{NH-C^\alpha H}$ coupling constants as estimated below. The complete proton resonance assignments of NCp8-f1 are available as Supporting Information.

Dihedral Angles and Stereospecific Assignments. A total of 10 backbone torsion ϕ angles for NCp8-f1 were constrained: four residues (Trp^{10}, Ala^{19}, Cys^{22}, and Arg^{26}) with $^3J_{NH-C^\alpha H}$ less than 5.5 Hz and six residues (Ile^7, Arg^8, Cys^{12}, Ser^{18}, Gln^{21}, and Arg^{23}) with $^3J_{NH-C^\alpha H}$ greater than 8.0 Hz. Backbone dihedral constraints were not applied for residues with $^3J_{NH-C^\alpha H}$ values between 5.5 and 8.0 Hz.

The stereospecific assignments together with the conformations around the C$^\alpha$−C$^\beta$ bonds for two of the 20 nondegenerate β-methylene protons in NCp8-f1, that is, Cys^{12} and Ser^{18}, were established. Both have g^2g^3 conformations about the C$^\alpha$−C$^\beta$ bonds, and the χ^1 side chain torsion angle was constrained in the range 60° ± 40° (*31*).

Secondary Structure. The regular secondary structure elements in the NCp8-f1 molecule were identified according to standard criteria, using the sequential and medium-range NOEs summarized in Figure 3. In the N-terminal region of NCp8-f1 (residues 1−8), rather intense sequential $d_{\alpha N}$ and

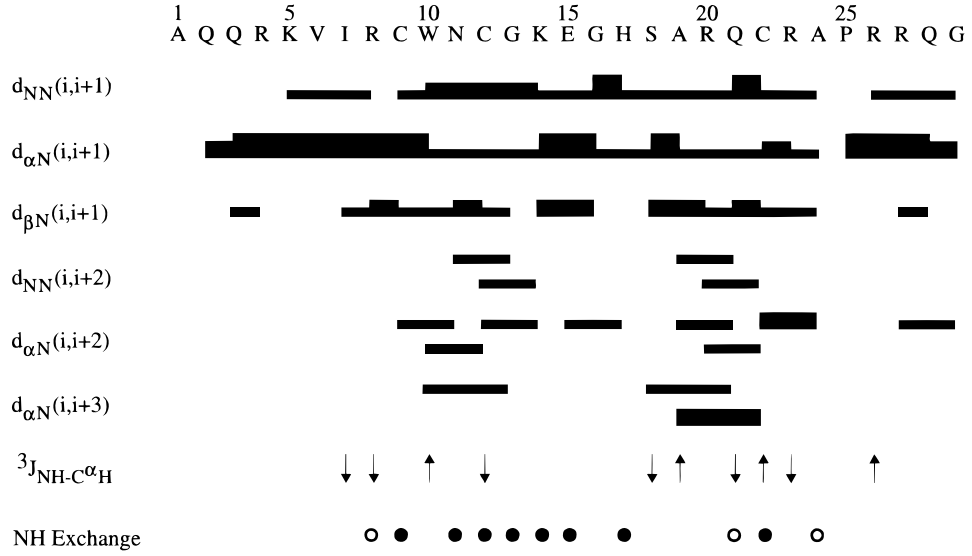

FIGURE 3: Summary of the sequential NOE connectivities and the $^3J_{NH-C^\alpha H}$ coupling constants observed in NCp8-f1. These structural parameters were used for the sequence-specific assignments and the identification of secondary structure elements in NCp8-f1. The sequential NOEs, d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$, and the medium-range NOEs, $d_{NN}(i,i+2)$, $d_{\alpha N}(i,i+2)$, and $d_{\alpha N}(i,i+3)$, are indicated by bars between two residues. The NOEs are classified as strong, medium, and weak according to the height of the filled bars. Values of $^3J_{NH-C^\alpha H}$ coupling constants are indicated by ↑ (>8.0 Hz) and ↓ (<5.5 Hz) symbols. Slowly exchanging backbone amide protons that are still observed in a NOESY spectrum at pH 8.7 with intensities of more than half and of a quarter, relative to the corresponding resonances in a NOESY spectrum at pH 5.8, are indicated by filled and open circles, respectively.

Table 1: Structural Statistics for NCp8-fl[a]

structural parameter	15 converged structures	mean structure
RMS deviations from experimental distance constraints (Å)		
all (343)	0.014 ± 0.003	0.008
intraresidue (94)	0.004 ± 0.005	0.005
sequential (101)	0.008 ± 0.002	0.010
medium range ($\|i - j\| < 5$) (51)	0.019 ± 0.005	0.009
long range ($\|i - j\| \geq 5$) (93)	0.020 ± 0.005	0.015
RMS deviation from experimental dihedral constraints (deg) (10)	0.357 ± 0.096	0.294
energetic statistics (kcal mol^{-1})[b]		
F_{NOE}	3.61 ± 1.45	1.88
F_{tor}	40.53 ± 0.32	40.2
F_{repel}	4.38 ± 1.13	4.52
E_{L-J}	−66.6 ± 11.4	−67.6
RMS deviations from idealized geometry		
bonds (Å)	0.002 ± 0.0002	0.002
angles (deg)	0.561 ± 0.014	0.559
impropers (deg)	0.374 ± 0.020	0.354

[a] The 15 converged structures refer to the final set of dynamical simulated annealing structures starting from 100 initial random structures; the mean structure was obtained by restrained minimization of the averaged coordinates of the 15 individual structures. The number of each experimental constraint used in the calculations is given in parentheses. [b] F_{NOE}, F_{tor}, and F_{repel} are the energies related to the NOE violations, the torsion angle violations, and the van der Waals repulsion term, respectively. The values of the force constants used for these terms are the standard values, as depicted in the X-PLOR 3.1. manual. E_{L-J} is the Lennard−Jones van der Waals energy calculated with the CHARMm (*41*) empirical energy functions.

weak d_{NN} NOEs were observed. However, no other $d_{\beta N}$ or medium-range NOEs were observed, except for the $d_{\beta N}$ NOEs between Gln[3] and Arg[4]. These observations indicate that no regular secondary structure is present in the N-terminal region of NCp8-f1.

In the basic amino acid cluster, ^{23}RAPRRQG^{29}, two $d_{\alpha N}$-(i,i+2) NOEs were observed, suggesting the presence of two turn-like conformations. The Ala[24]-Pro[25] bond is considered to be fixed in the trans conformation, and not in equilibrium between the cis and trans conformations, because no significant changes were observed in the chemical shifts of the amide protons in this cluster when the temperature was increased from 278 to 313 K (data not shown). Interestingly, several NOEs were observed between the backbone protons in the basic amino acid cluster and the side chain protons of the zinc finger (Trp[10], Asn[11], and Cys[12]), as shown in Figure 4.

Structure Calculations and Evaluation. The three-dimensional structure of NCp8-f1 in solution was calculated on the basis of 12 dihedral and 343 distance constraints. The 343 distance constraints include 94 intraresidue and 245 interresidue NOE distance constraints, and an additional four constraints between the zinc ion and the sulfur atoms of Cys[9], Cys[12], and Cys[22] and between the zinc ion and the N$^{\epsilon}$ of His[17] (*9*).

Simulated annealing calculations were started from 100 initial random structures. We selected 15 final structures with total energies of less than 65 kcal·mol^{-1} and Lennard-Jones van der Waals energies of less than −55 kcal·mol^{-1}. These structures showed good agreement with the NMR input constraints, with no NOE distance and torsion angle violations larger than 0.5 Å and 5°, respectively. The structural statistics for the mean and the 15 converged structures were evaluated in terms of the structural parameters, as shown in Table 1. The deviations from idealized covalent geometry were very small, and the Lennard-Jones van der Waals energies were large and negative, indicating that no distortions and no nonbonded bad contacts existed in the 15 converged structures (Table 1). In a Ramachandran-type plot showing the backbone dihedral angles of residues 7−29 of the final 15 converged structures (see Supporting Information), the (ϕ, ψ) values fall in the allowed regions.

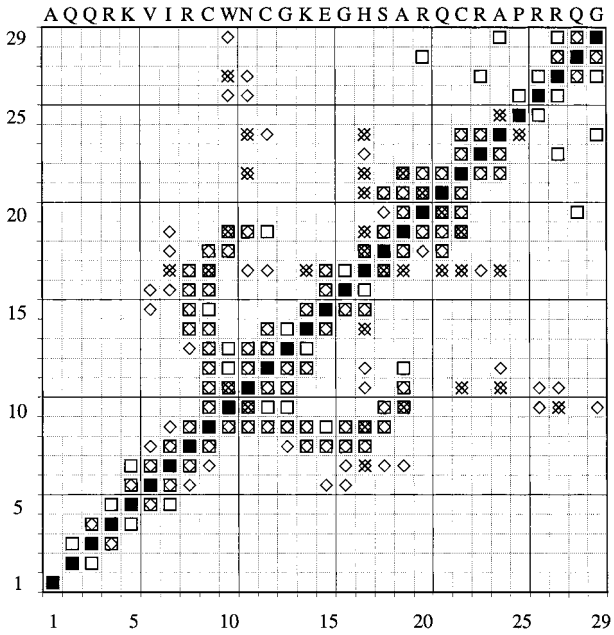

FIGURE 4: Diagonal plot of NOEs for the entire peptide: black boxes, intraresidue NOEs; open boxes, backbone/backbone NOEs; diamonds, backbone/side chain NOEs; crosses, side chain/side chain NOEs.

The convergence for the final set of 15 structures was evaluated by the atomic rmsd values (Table 2). The rmsds from the averaged coordinate positions in the region of residues 7−29 were 0.29 ± 0.04 Å for the backbone atoms (N, C$^{\alpha}$, C) and 1.27 ± 0.13 Å for all heavy atoms. For the same atom selection, the average pairwise rmsd values of the 15 individual structures were 0.42 ± 0.13 and 1.85 ± 0.28 Å, respectively. However, for the N-terminal region of NCp8-f1 (residues 1−5), the rmsds were greater than 3 Å for the backbone atoms (N, C$^{\alpha}$, C) and all heavy atoms

Table 2: Root-Mean-Squared Differences for the 15 Converged Structures of NCp8-fl[a]

	atomic rmsd of 15 converged structures versus mean structure (Å)		average pairwise rmsd for 15 converged structures (Å)	
	residues 7−29	residues 9−22	residues 7−29	residues 9−22
backbone				
(N, C^{α}, C)	0.29 ± 0.04	0.25 ± 0.04	0.42 ± 0.13	0.36 ± 0.10
(N, C^{α}, C, O)	0.33 ± 0.04	0.31 ± 0.05	0.47 ± 0.11	0.44 ± 0.15
all heavy atoms	1.27 ± 0.13	0.88 ± 0.16	1.85 ± 0.28	1.28 ± 0.24

[a] The rmsd values were obtained by best-fitting the backbone atom (N, C^{α}, C, O) coordinates for all of the residues of the 15 converged structures. The numbers given for the backbone and all heavy atoms represent the mean values ± standard deviations.

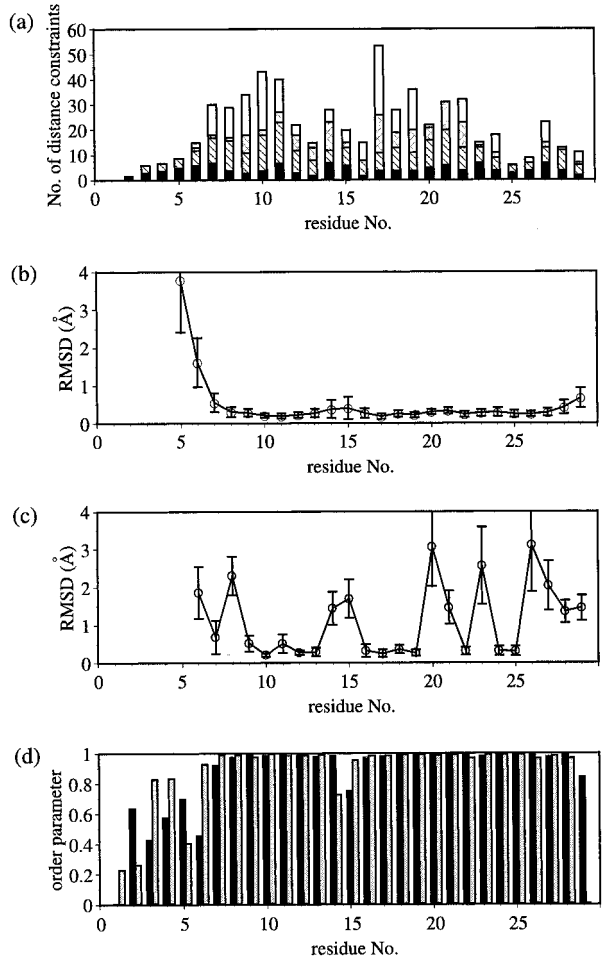

FIGURE 5: Backbone variability by residue. (a) Distribution of the number of experimental distance constraints as a function of the sequence position of NCp8-f1: filled bars, intraresidue NOEs; hatched bars, sequential NOEs; stippled bars, medium-range NOEs; open bars, all long-range constraints. (b and c) Distributions of the rmsd of the backbone (b) and all heavy atoms (c) from the mean structure as a function of residue number are shown together with their standard deviations. (d) Filled bars, the order parameter S of the ϕ; shaded bars, the order parameter S of the ψ, 0 = randomly distributed, and 1 = perfectly aligned.

(not shown). These large values confirm the absence of an ordered conformation in this region. The distribution of the atomic rmsds for the 15 converged structures about the mean structure, as a function of the residue number, is shown for the backbone atoms (Figure 5b) and all heavy atoms (Figure 5c). The backbone structure is well-defined throughout all of the residues of the zinc finger and the basic amino acid cluster, that is, Ile^{7}-Gly^{29}. The convergence was further assessed by the (ϕ, ψ) spacing for all selected structures. Figure 5d shows the distribution per residue of the order parameter S, indicating the homogeneity of the backbone dihedral angles. The majority of the backbone dihedral angles are defined well ($S > 0.9$), except for those of residues 1−6, Lys^{14}, Glu^{15}, and Gly^{29}, which have lower angular order parameters ($S < 0.8$). These data again confirm that the N-terminal part of NCp8-f1 does not adopt an ordered conformation. The relatively low S values of Lys^{14} and Glu^{15} (Figure 5d) reflect the limited number of NOE distance constraints for Gly^{13} and Gly^{16} (Figure 5a).

Description of the Three-Dimensional Structure. Figure 6a shows a stereopair representation of the best-fit superposition of the backbone atoms (N, C^{α}, C) of all of the residues for the 15 converged structures of NCp8-f1, which was obtained by best-fitting the backbone atom (N, C^{α}, C, O) coordinates for residues 7−28. Residues His^{17}-Cys^{22} adopt one turn of a helical conformation. In Figure 6, parts b and c, the side chains of the residues coordinating to the zinc are presented along with the backbone atoms. The side chains of Cys^{12}, His^{17}, and Cys^{22} converge very well, indicating that the tight binding of the zinc ion with these residues stabilizes the compact, knuckle-like conformation of the zinc finger.

To determine which nitrogen atom of His^{17}, N^{ϵ} or N^{δ} is bound to the zinc ion, we performed two types of distance geometry calculations. The 10 converged structures obtained by the first set of calculations without constraints on the zinc ion were very close to those shown in Figure 6. The second set of calculations was performed with the same distance constraints as those used to obtain the structures in Figure 6, except for the connections between His and the zinc ion, where the distance constraints of 2.0 Å were used between the His N^{δ} and the zinc, rather than between the His N^{ϵ} and the zinc. The structures calculated with these modified constraints had total energies of more than 120 kcal·mol^{-1} and were not converged. These results clearly indicate that the zinc ion is coordinated to the His via N^{ϵ}, as found in the zinc finger of NCp7 (*9*).

DISCUSSION

For the NC proteins of HIV-1 and HIV-2, the first zinc finger and the basic amino acid cluster are essential for their biological activities and are referred to as the minimal active domain. However, the precise three-dimensional structure of this basic amino acid cluster in the domain has not yet been determined, even though many structural analyses have been performed on the domain, and the structures of the zinc finger regions have been precisely determined for the NCp7 of HIV-1 (*7*−*9*). In the present study, we have determined the precise three-dimensional structure of this domain, including the basic amino acid cluster. Below, we will discuss the structure−activity relationships of the minimal active domain of NCp8 in comparison with those of NCp7 and other NC proteins.

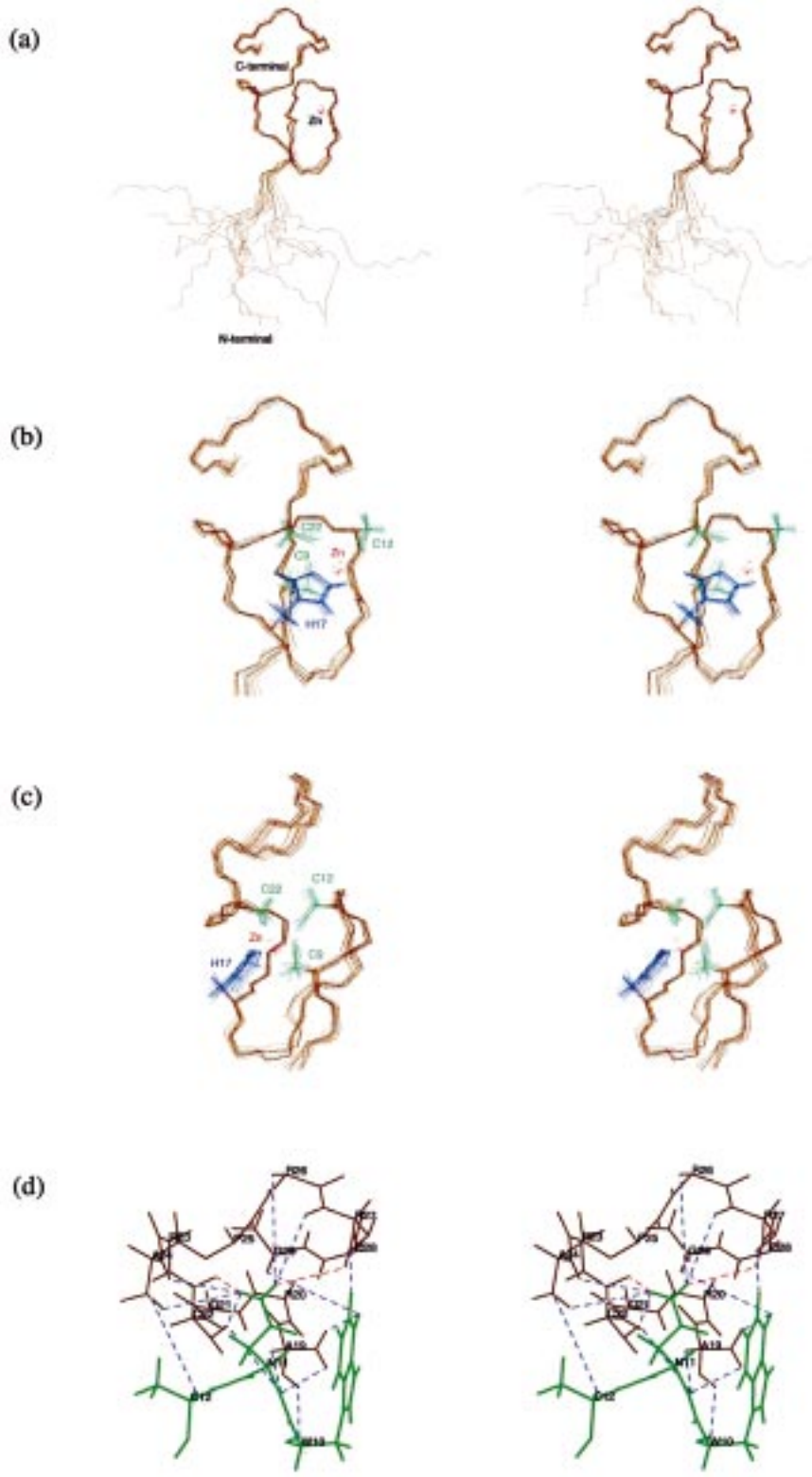

FIGURE 6: Stereopairs for the 15 converged structures of NCp8-f1. These are the results of the best-fit superposition of the backbone atoms (N, C^{α}, C, and O) for all of the residues of the molecule, drawn by using MIDAS Plus (*43*). (a) Stereopair of backbone heavy atoms (N, C^{α}, and C) for the complete sequence. (b) Stereopair of backbone heavy atoms (N, C^{α}, and C) and all side chain atoms for the residues Ile^{7} to Gly^{29} at the same orientation as in (a). (c) Stereopair for the same sequence as in (b) obtained by a 90° rotation about the vertical axis of stereopair (b). (d) Representation of strong observed NOEs (blue dashed lines) and the putative hydrogen bonds (red dashed lines) between the zinc finger (green) and the basic amino acid cluster (dark brown) in NCp8-f1.

The Structure of NCp8-f1. The conformation of the N-terminal region of the zinc finger seems to be stable, as suggested by the observation of HN−H^{α} NOESY cross-peaks for Cys^{9} and Asn^{11}-Lys^{14} even at pH 8.7 (Figure 3). It is possible that this region is stabilized by the same hydrogen bond network as that in NCp7 (*10*). The His^{17}-Glu^{21} region adopts a helical conformation. This conformation is also found in the corresponding region of NCp7 (*11*) and NCp10 from Moloney murine leukemia virus (*32*) (Figure 1). Interestingly, for most of the NC proteins, the position

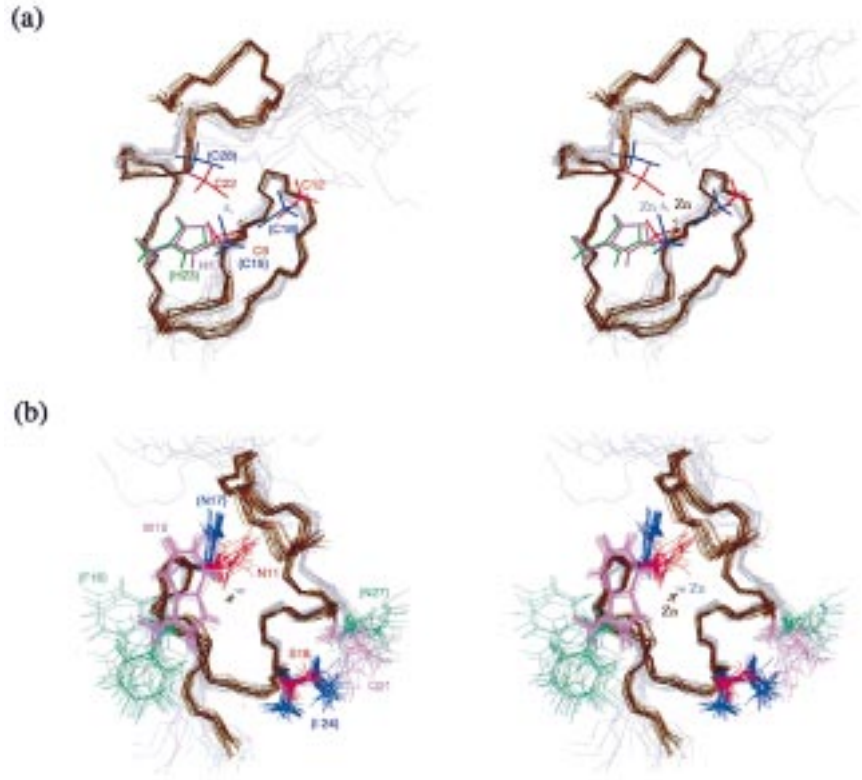

FIGURE 7: (a) Stereopairs of backbone heavy atoms (N, C^{α}, and C) and the side chains bound with zinc for the 15 and 20 converged structures of NCp8-f1 (dark brown) and NCp7 (gray), respectively. The zinc-bound side chains of NCp8-f1 and NCp7 are colored red/purple and blue/green, respectively. These are the results of the best-fit superposition of the backbone atoms (N, C^{α}, C) for the residues of the zinc finger region of the molecules. The NCp7 structure coordinates were extracted from the Brookhaven Protein Data Bank entry 1AAF. (b) is obtained by a 90° rotation about the vertical axis of the stereopairs (a). The side chains of Trp^{10}, Asn^{11}, Ser^{18}, and Gln^{21} (NCp8-f1), and Phe^{16}, Asn^{17}, Ile^{24}, and Asn^{27} (NCp7) are presented instead of the side chains bound with zinc.

corresponding to Arg^{20} of NCp8 is occupied by a basic amino acid residue (*33*). Taken together, it is likely that the helical conformation enables the basic amino acid residue to protrude from the knuckle of the zinc finger and makes it easy for this residue to interact with RNA or other proteins.

The basic amino acid cluster, Arg^{23}-Gly^{29}, also adopts a well-defined conformation (Figure 6). Figure 6d shows the mean structure of the cluster together with that of the N-terminal region of the zinc finger, Trp^{10}- Cys^{12}. The 16 NOESY cross-peaks are observed between the side chain atoms of the N-terminus of the zinc finger and the backbone atoms of the cluster. The strong observed NOEs are represented by blue dashed lines in Figure 6d. It is noteworthy that the side chain of Asn^{11} protrudes in the center of the cluster. In addition, the signals of Asn^{11} $N^{\delta}H$ were observed even at pH 8.7 (data not shown). These results indicate that the local structure of the cluster is stable. The conformation of the cluster is likely to be stabilized by the hydrogen bonds between the cluster and the zinc finger: between Arg^{27} O and Asn^{11} $N^{\delta}H$ and between Cys^{22} O and Asn^{11} $N^{\delta}H$ (the red dashed lines in Figure 6d), as suggested by the short distances between these atoms in the mean structure.

Structural Comparison with NCp7 from HIV-1. Figure 7, parts a and b, show the best-fit superposition of these structures in NCp8-f1 (dark brown) and in NCp7 (gray), which were fitted using the backbone atom (N, C^{α}, C) coordinates in the zinc finger regions. For the zinc finger, the conformations of the two N-terminal regions are similar to each other. However, the position of the zinc ions, the conformation of the C-terminal regions, and the direction of the side chains of Trp^{10} and Asn^{11} of NCp8 are rather different from those corresponding to NCp7. The conformational difference of the C-terminal region would result from the fact that Ile^{24} of NCp7 is sterically larger than Ser^{18} of NCp8, which induces the different position of the zinc ions.

The most significant difference between the two structures lies in the basic amino acid cluster region: the cluster of NCp8-f1 adopts a well-defined conformation, whereas that of NCp7 does not. Recently, a ^{15}N NMR relaxation study NCp7 (*12*) has been reported where the conformation of the basic amino acid cluster is discussed. This study concluded that the conformation of the cluster of NCp7 was not stable and the solution behavior of the cluster might be considered as a rapid equilibrium between several different conformations. If the basic amino acid cluster of the full-length NCp8 behaves like that of NCp7, then the reasons why the cluster of NCp8-f1 adopts a well-defined conformation would be explained as follows. Because the conformation of the C-terminal region of the zinc finger of NCp8-f1 is different from that of NCp7 as mentioned above, and/or NCp8-f1 does not contain the second zinc finger, it becomes easy for the cluster to locate in the vicinity of the N-terminal region of the zinc finger. This conformation of the cluster would correspond to one of the several equilibrium conformations which has been observed in NCp7 (*12*). Furthermore, the hydrogen bond between Asn^{11} $N^{\delta}H$ and the backbone oxygen in the cluster may stabilize the conformation.

It is not certain whether the conformation of the basic amino acid cluster of the full-length NCp8 is stable unlike that of NCp7. However, the three-dimensional structure of NCp8-f1 determined in this study is significant for the biological activity, because NCp8-f1 itself interacts specifically with viral RNA (*34*). Below, we will discuss the structure−activity relationships of NC proteins in more detail.

Structure−Activity Relationship of the Minimal Active Domain. Both NCp8 and NCp7 have two successive zinc fingers linked by the basic amino acid cluster. Many investigations of the biological functions of the NCp7-derived peptide and the mutant peptide have indicated that both the first zinc finger and the basic amino acid cluster are essential for specific RNA binding (*5*, *35*, *36*). Deletions of short sequences containing these basic residues lead to a complete loss of nucleic acid binding activity in vitro (*35*, *36*). Point mutations in the basic amino acid cluster, ^{29}RAPRKKG^{35}, lead to a very poorly infectious virus (A30P) or a noninfectious virus (P31L and R32G), and the replacement of ^{32}RKK^{34} by SSS results in a decrease of the virus titer by 100-fold (*5*). On the other hand, the NC protein of Moloney murine leukemia virus, NCp10, plays same role in the viral replication as NCp7 and NCp8 though it has only one zinc finger (Figure 1). The zinc finger and the flanking basic residues of NCp10 were shown to be necessary for the production of infectious virions. Among others, the Lys^{41} and Lys^{42} residues, which follow the zinc finger, are critical for virus infectivity (*37*). Furthermore, the substitution of a glycine residue for Tyr^{28}, which corresponds to Trp^{10} in NCp8-f1, leads to an inhibition of virus replication through an alteration in viral genomic RNA binding (*38*). Therefore, comparison of the structures of NCp8-f1 and NCp10 would yield important clues for elucidating the structural role of the minimal active domains of NCp8 and NCp7.

Figure 8a shows the locations of Arg^{8}, Arg^{20}, Arg^{26}, Arg^{27}, and Trp^{10} in the mean structure of NCp8-f1, and Figure 8b is the surface charge representation of NCp8-f1, with the same orientation as that in Figure 8a. These figures show that all five of these residues are located approximately in the same plane of the molecule, and that the four arginine residues are exposed to the solvent. The three-dimensional structure of this region is in accord with that of NCp10 (*32*) in the following two respects. (1) The four arginine residues, Arg^{8}, Arg^{20}, Arg^{26}, and Arg^{27}, of NCp8-f1 are almost at the same positions as those of the four lysine residues, Lys^{30}, Lys^{37}, Lys^{41}, and Lys^{42}, of NCp10, respectively. Among others, Arg^{20} (and the corresponding Lys^{37} in NCp10) is located in the helix in the C-terminal region and protrudes from the zinc finger knuckle to the solvent. (2) In both structures, the aromatic residue in the N-terminal region of the zinc finger (Trp^{10} in NCp8-f1 and Tyr^{28} in NCp10) is in close proximity to two basic residues in the cluster (Arg^{26} and Arg^{27} in NCp8-f1, and Lys^{41} and Lys^{42} in NCp10), despite the fact that the positions of these residues in the primary sequence are not in agreement with each other as shown in Figure 1. On the basis of these facts, it may be concluded that the configuration of these basic amino acid residues and the aromatic residue is important for the specific recognition of the RNA.

A recent heteronuclear NMR study (*13*) determined the three-dimensional structure of NCp7 bound with the SL3 stem−loop recognition element. It was shown that the basic amino acid cluster adopts a stable conformation, and that Lys^{14}, Phe^{16}, Lys^{26}, and Arg^{32} form an active surface. Furthermore, the side chain of Asn^{17}, corresponding to Asn^{11} of NCp8, is connected, by hydrogen bonds, to the backbones of Pro^{31} and Lys^{33} in the basic amino acid cluster. These observations suggest that the structure of the active surface of NCp8-f1, determined in the present study, is similar to that of NCp7 in the RNA-bound form.

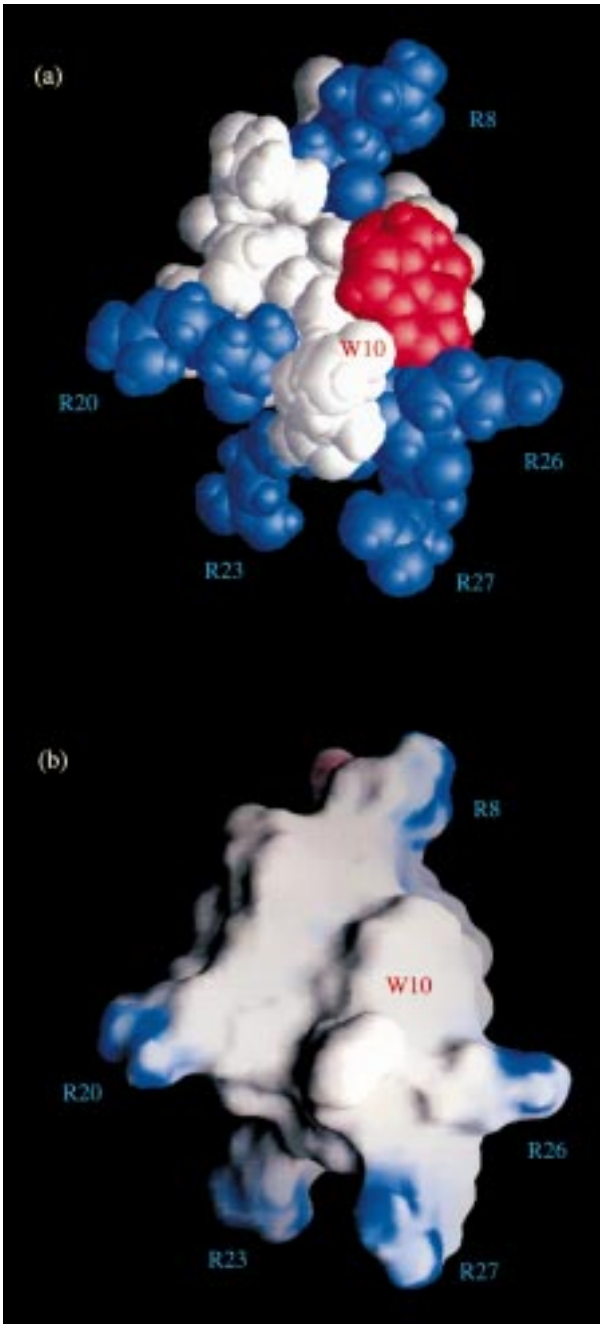

FIGURE 8: CPK (a) and electrostatic potential (b) representations of residues Ile^{7}-Gly^{29} of NCp8-f1. (a) Basic amino acid residues and Trp^{10} are drawn in blue and red, respectively. (b) Positive and neutral potentials are drawn in blue and white, respectively. These figures were prepared using the program GRASP (*44*).

In the structure of the complex of NCp7 with SL3 RNA, the second zinc finger of NCp7 does not contact a wide area of the SL3 RNA (*13*). This observation is in good agreement with the observation that the NCp7-derived peptide, consisting of the second zinc finger and the basic amino acid cluster,

does not exhibit specific RNA binding activity (*35*, *39*). On the other hand, the NCp8-derived peptide that includes the same region, the second zinc finger and the basic amino acid cluster, has the ability to recognize viral RNA specifically (*40*). In NCp7, residues Lys^3-Arg^{10} form a 3^{10} helix upon complex formation and are inserted into the major groove of the RNA stem. However, the N-terminal region of NCp8, consisting of six amino acid residues, would be too short to form a similar 3^{10} helix. These observations suggest that the RNA recognition mechanisms may be slightly different between NCp8 and NCp7. Therefore, it would be important to study the structure of NCp8 and its complex with RNA, and to compare them with the other retroviral NC proteins to clarify the mechanisms underlying specific RNA recognition and other biological functions.

ACKNOWLEDGMENT

We thank Dr. Frank Delaglio at NIH for providing us with the NMRPipe/NMRDraw software, and Dr. Kaori Wakamatsu at Gunma University, Drs. Jae Il Kim and Takao Matsuzaki, and Ms. Kuniko Kobayashi at Mitsubishi Kasei Institute of Life Sciences for useful discussions.

SUPPORTING INFORMATION AVAILABLE

One table of NMR chemical shifts of ^1H resonances of NCp8-f1 and one figure, a Ramachandran plot of the backbone conformational (ϕ, ψ) angles for all residues of the 15 converged structures of NCp8-f1; the circles (○) indicate glycine residues and the crosses (×) indicate other residues (3 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Henderson, L. E., Copeland, T. D., Sowder, R. C., Smythers, G. W., and Oroszlan, S. (1981) *J. Biol. Chem. 256*, 8400−8406.
2. Covey, S. N. (1986) *Nucleic Acids Res. 14*, 623−633.
3. Dupraz, P., Oertle, S., Meric, C., Damay, P., and Spahr, P. F. (1990) *J. Virol. 64*, 4978−4987.
4. Fu, X. D., Katz, R. A., Skalka, A. M., and Leis, J. (1988) *J. Biol. Chem. 263*, 2140−2145.
5. Ottmann, M., Gabus, C., and Darlix, J. L. (1995) *J. Virol. 69*, 1778−1784.
6. Morellet, N., De Rocquigny, H., Mély, Y., Jullian, N. H. D., Ottmann, M., Gérard, D., Darlix, J. L., Fournie-Zaluski, M. C., and Roques, B. P. (1994) *J. Mol. Biol. 235*, 287−301.
7. Morellet, N., Jullian, N., De Rocquigny, H., Maigret, B., Darlix, J. L., and Roques, B. P. (1992) *EMBO J. 11*, 3059−3065.
8. Omichinski, J. G., Clore, G. M., Sakaguchi, K., Appella, E., and Gronenborn, A. M. (1991) *FEBS Lett. 292*, 25−30.
9. Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W. J., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R., III, Hare, D. R., and Arthur, L. O. (1992) *Protein Sci. 1*, 563−574.
10. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry 29*, 329−340.
11. South, T. L., Blake, P. R., Hare, D. R., and Summers, M. F. (1991) *Biochemistry 30*, 6342−6349.
12. Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) *J. Mol. Biol. 279*, 633−649.
13. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−388.
14. Komatsu, H., Tsukahara, T., and Tozawa, H. (1996) *Biochem. Mol. Biol. Int. 38*, 1143−1154.
15. Laussac, J. P., Peyrou, G., Mazargui, H., Erard, M., Bourdonneau, M., and Chan, M. T. (1993) *New J. Chem. 17*, 607−612.
16. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
17. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
18. Macura, S., Huang, Y., Suter, D., and Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259−281.
19. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−359.
20. Piotto, M., Saudek, V., and Sklenár, V. (1992) *J. Biomol. NMR 2*, 661−665.
21. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479.
22. Mueller, L. (1987) *J. Magn. Reson. 72*, 191−196.
23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277−293.
24. Wüthrich, K., Billeter, M., and Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
25. Pardi, A., Billerter, M., and Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741−751.
26. Hyberts, S. G., Marki, W., and Wagner, G. (1987) *Eur. J. Biochem. 164*, 625−635.
27. Brünger, A. T. (1993) *X-PLOR Manual, Version 3.1*, Yale University, New Haven, CT.
28. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol. 7*, 95−99.
29. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736−751.
30. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley, New York.
31. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., and Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611−639.
32. Déméné, H., Jullian, N., Morellet, N., De Rocquigny, H., Cornille, F., Maigret, B., and Roques, B. P. (1994) *J. Biomol. NMR 4*, 153−170.
33. Darlix, J. L., Lapadat-Tapolsky, M., De Rocquigny, H., and Roques, B. P. (1995) *J. Mol. Biol. 254*, 523−537.
34. Komatsu, H., and Tozawa, H. (1994) *AIDS 8*, 1227−1235.
35. Dannull, J., Surovoy, A., Jung, G., and Moelling, K. (1994) *EMBO J. 13*, 1525−1533.
36. Schmalzbauer, E., Strack, B., Dannull, J., Guehmann, S., and Moelling, K. (1996) *J. Virol. 70*, 771−777.
37. Housset, V., De Rocquigny, H., Roques, B. P., and Darlix, J. L. (1993) *J. Virol. 67*, 2537−2545.
38. Meric, C., and Goff, S. P. (1989) *J. Virol. 63*, 1558−1568.
39. De Rocquigny, H., Gabus, C., Vincent, A., Fournié-Zaluski, M. C., Roques, B., and Darlix, J. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6472−6476.
40. Tsukahara, T., Komatsu, H., Kubo, M., Obata, F., and Tozawa, H. (1996) *Biochem. Mol. Biol. Int. 40*, 33−42.
41. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
42. Barton, G. J. (1993) *Protein Eng. 6*, 37−40.
43. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics 6*, 13−27.
44. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.

BI981818O