

The NMR Structure of the Nucleocapsid Protein from the Mouse Mammary Tumor Virus Reveals Unusual Folding of the C-Terminal Zinc Knuckle^{†,‡}

Daniel J. Klein, Philip E. Johnson, Eric S. Zollars, Roberto N. De Guzman, and Michael F. Summers*

Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, Maryland 21250

Received September 27, 1999; Revised Manuscript Received December 1, 1999

ABSTRACT: The nucleocapsid protein (NC) from the mouse mammary tumor virus (MMTV) has been overexpressed in *Escherichia coli* and purified to homogeneity for structural studies by nuclear magnetic resonance (NMR) spectroscopy. The protein contains two copies of a conserved zinc-coordinating “CCHC array” or “zinc knuckle” motif common to the nucleocapsid proteins of nearly all known retroviruses. The residues comprising and adjacent to the zinc knuckles were assigned by standard two-dimensional ¹H and three-dimensional ¹H–¹⁵N NMR methods; the rotational dynamic properties of the protein were determined from ¹⁵N relaxation experiments, and distance restraints derived from the nuclear Overhauser effect (NOE) data were used to calculate the three-dimensional structure. The ¹H–¹H NOE and ¹⁵N relaxation data indicate that the two zinc knuckles do not interact with each other, but instead behave as independently folded domains connected by a flexible 13-residue linker segment. The proximal zinc knuckle folds in a manner that is essentially identical to that observed previously for the two zinc knuckles of the human immunodeficiency virus type 1 nucleocapsid protein and for the moloney murine leukemia virus nucleocapsid zinc knuckle domain. However, the distal zinc knuckle of MMTV NC exhibits a rare three-dimensional fold that includes an additional C-terminal β -hairpin. A similar C-terminal reverse turn-like structure was observed recently in the distal zinc knuckle of the Mason-Pfizer monkey virus nucleocapsid protein [Gao, Y., et al. (1998) *Protein Sci.* 7, 2265–2280]. However, despite a high degree of sequence homology, the conformation and orientation of the β -hairpin in MMTV NC is significantly different from that of the reverse turn in MPMV NC. The results support the conclusion that structural features of NC zinc knuckle domains can vary significantly among the different genera of retroviridae, and are discussed in terms of the recent and surprising discovery that MMTV NC can facilitate packaging of the HIV-1 genome in chimeric MMTV mutants.

The mouse mammary tumor virus (MMTV)¹ was the first retrovirus isolated from a mammalian source and, except for the human immunodeficiency virus (HIV), has been the most-studied of the mammalian retroviruses (1). MMTV is a type B retrovirus that is passed from mother to offspring either through milk during nursing or genetically as a provirus. Infection by MMTV leads to carcinomas in the alveolar epithelial cells of the mammary gland. Only the female offspring develop carcinomas, which are stimulated by the female estrogen hormones via a poorly understood mechanism (2). Very recently, PCR methods were employed as a probe for genes closely related to those of MMTV in

healthy and cancerous human breast tissue (3). These studies probed for a 660 bp sequence of the MMTV *env* gene that exhibits a very low degree of homology to human endogenous retroviruses. The sequence was found in approximately 40% of the human breast cancer tissues that were examined, but was not found in normal human tissues, including breast tissue, or in other human cancers or cell lines. These findings led to the suggestion that a virus similar to MMTV may play a role in the etiology of a large proportion of human breast cancers (3).

As for all other known retroviruses, the MMTV genome encodes a Gag polyprotein that is produced in the host cell during the late stage of the infectious cycle. Approximately 1500 Gag proteins self-associate at the cell membrane, leading to budding and the formation of immature virus particles (4). Concomitant with budding, the Gag proteins are processed by the retroviral protease into the matrix (MA), capsid (CA), and nucleocapsid (NC) proteins, which undergo rearrangements to form the mature, infectious virion. Except for the spumaviruses, all retroviral NC proteins contain one or two copies of a conserved CCHC array (C-X₂-C-X₄-H-X₄-C, with C being cysteine, H histidine, and X a variable amino acid), sometimes termed a zinc knuckle (5, 6), that appears to function primarily in nucleic acid interactions.

[†] This research was supported by NIH Grant GM42561. D.J.K. was supported by a MARC U*STAR Affiliate fellowship, and P.E.J. was supported by a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship.

[‡] The coordinates for the N- and C-terminal domain have been deposited in the Protein Data Bank (1DSQ and 1DSV, respectively), and the chemical shifts have been deposited in the BMRB (4559).

* To whom correspondence should be addressed. Phone: (410) 455-2527. Fax: (410) 455-1174. E-mail: summers@hhmi.umbc.edu.

¹ Abbreviations: BME, β -mercaptoethanol; HIV-1, human immunodeficiency virus type 1; MMTV, mouse mammary tumor virus; MoMuLV, moloney murine leukemia virus; MPMV, Mason-Pfizer monkey virus; NC, nucleocapsid; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; τ_m , overall rotational correlation time.

Table 1: Electrospray Mass Spectrometry Results for the Recombinant MMTV NC Protein

Zn ₂ -NC protein ^a		apo-NC protein ^b		N-terminal cleavage site (Λ)	protein length (residues)
expt mass (Da)	calcd mass (Da)	expt mass (Da)	calcd mass (Da)		
10263 ± 2	10257	10139 ± 5	10137	M-A-A-Λ-A-M-R...	93
10333 ± 2	10328	10206 ± 3	10208	M-A-Λ-A-A-M-R...	94
10403 ± 3	10397	10277 ± 1	10279	M-Λ-A-A-A-M-R...	95

^a Protein in 1 mM ammonium acetate (pH 7.0). ^b Protein in 10% acetic acid/90% water.

NC proteins stabilize the RNA genome in mature virions (7–14), facilitate elongation of proviral DNA by reducing the level of reverse transcriptase pausing at stable stem–loop sites (15), and stabilize the proviral DNA after reverse transcription (16). In addition, the NC domains of the Gag precursor proteins are critical for viral genome recognition and encapsidation. Mutations that abolish zinc binding lead to noninfectious virions that lack their genomes (7–14), and mutations of conservatively substituted hydrophobic residues within the CCHC arrays can alter packaging specificity (13). Also, all the NC domains of the human immunodeficiency virus type 1 (HIV-1) and moloney murine leukemia viruses (MoMuLV) have been swapped, resulting in the specific packaging of the non-native genomes (14). These results indicate that the NC domains of Gag are capable of discriminating not only between viral and cellular RNAs but also between RNAs of different retroviruses.

In contrast to the studies described above, Aldovini and co-workers recently demonstrated that replacement of the HIV-1 NC coding sequence with that of the MMTV NC sequence results in HIV-1 genome incorporation at 50% of the wild-type level (17). Similarly, viruses produced from chimeric MMTV genomes, in which the native MMTV NC sequence is replaced with the corresponding HIV-1 NC sequence, preferentially package the MMTV genome. The MoMuLV NC protein contains only one CCHC motif, whereas the NC proteins from both HIV-1 and MMTV contain two CCHC motifs; it has been suggested that the packaging of HIV-1 RNA is not critically dependent on the amino acid context of the CCHC arrays, but instead depends on the availability of an NC protein with two CCHC zinc knuckles (17). To develop a better understanding of retroviral genome recognition and discrimination at the structural level, and as a first step toward understanding the structural biology of a virus that may be relevant to human breast cancer, we have overexpressed and purified the MMTV NC protein for NMR-based structural and dynamical studies.

MATERIALS AND METHODS

Subcloning of MMTV NC. The coding region for the MMTV NC protein was subcloned into the plasmid pET-3d under the control of the T7 promoter. The tetracycline-resistant plasmid p202 (18) containing the partial sequence of the gag-pol region of MMTV strain C3H was obtained from the American Type Culture Collection (#45005). The NC coding region in p202 was PCR amplified using the 5′-primer CAG GGT ATG GCC ATG GCA GCA GCC ATG A (*Nco*I site underlined) and the 3′-primer GGA CGG ATC CTC TAC AAG TTT TTTG (*Bam*HI site underlined, stop codon italicized) and subcloned into the *Nco*I and *Bam*HI sites of pET-3d (Novagen, Madison, WI). PCR tests of several clones using primers for the T7 promoter and T7

terminator and subsequent DNA sequencing confirmed the correct sequence of MMTV NC. One clone, designated pRD43, was transformed into *Escherichia* strain BL21(DE3)-pLysE (Novagen) and used for protein expression.

Expression and Purification. A 25 mL ZB starter culture (19) supplemented with chloramphenicol (34 mg/L) and ampicillin (100 mg/L) was inoculated with 100 μL of a glycerol stock of pRD43 in BL21(DE3)pLysE and grown at 37 °C and 200 rpm for 16 h in a shaker incubator. The starter culture was added to a 2 L preparation containing 1.8 L of ZB, 200 mL of 10× M9 salt, 20 mL of 40% glucose, 2 mL of 1 M MgSO₄, 1 mL of 200 g/L ampicillin, 2 mL of 34 g/L chloramphenicol, and 1 mL of 0.2 M ZnCl₂, and the cells were grown at 37 °C and 270 rpm. After 4 h, the cells reached an A₆₀₀ of 0.6 and were induced by adding 2 mL of 1 M IPTG. Growth was continued for another 5 h before harvesting the cells by centrifugation at 6000 rpm for 10 min at 4 °C. The cells were resuspended in lysis buffer [50 mM Tris-HCl (pH 8.0), 10% glycerol, 0.1 M NaCl, 0.1 mM ZnCl₂, 5 mM dithiothreitol, and 2 mM EDTA]. Cell lysis was carried out at 4 °C with the addition of 10 mM PMSF, pepstatin A, and 1% deoxycholate in lysis buffer. The lysate was sonicated four or five times with 20 s bursts to reduce the viscosity. Nucleic acids were precipitated by adding 4% polyethyleneimine dropwise to a final concentration of 0.4%, and the mixtures were stirred for 15 min before centrifugation at 16 000 rpm for 20 min at 4 °C. The supernatant was collected, passed through a 0.45 μm filter, and loaded at a rate of 1 mL/min onto 20 mL Q-Sepharose and a 20 mL SP-Sepharose columns (Pharmacia) connected in series. The columns were then washed with 2 volumes of buffer A [50 mM Tris-HCl (pH 8.0), 10% glycerol, 0.1 M NaCl, 0.1 mM ZnCl₂, and 10 mM BME (β-mercaptoethanol)], and the NC protein was eluted from the SP-Sepharose column with a 200 mL linear gradient of 30 to 40% buffer B [50 mM Tris-HCl (pH 8.0), 10% glycerol, 1.0 M NaCl, 0.1 mM ZnCl₂, and 10 mM BME]. Fractions containing the NC protein were pooled (18 mL) and loaded at a rate of 0.5 mL/min onto a 300 mL Sephadex G-50 column (Pharmacia) using buffer A. The eluted protein was concentrated and dialyzed into NMR buffer [25 mM [²H₃]acetate (pH 7), 25 mM NaCl, 0.1 M ZnCl₂, and 0.1 mM BME] using an Amicon Centriprep-3.0 device.

Mass Spectrometry. Aliquots (500 μL) of 100 μM MMTV NC protein were exchanged into 10% acetic acid/90% water so mass spectra for the denatured apo-NC protein could be obtained, and into 1 mM ammonium acetate (pH 7.0) so the spectrum of the native Zn₂-NC protein could be obtained. Measured and calculated molecular masses are given in Table 1. The mass spectra obtained for these samples reveal the presence of two minor forms of the protein that result from different N-terminal truncations, with the predominant form

corresponding to the 94-residue native protein, numbered as follows: (NH₃⁺)-A¹AAMRGQKY S¹⁰TFVKQTYGG G²⁰KGGQGAEGP V³⁰CFSCGKTGH I⁴⁰RKDCCKDEKG S⁵⁰KRAPPGLCP R⁶⁰CKKGYHWKS E⁷⁰CKSKFDKDG N⁸⁰PLPPLTNA E⁹⁰NSKNL.

NMR Spectroscopy. NMR data were collected using Bruker DMX 600 MHz and GE Omega PSG 600 MHz spectrometers equipped with triple-resonance probes and with samples at 25 °C and pH 7.0. Quadrature detection in the indirect dimensions was achieved with States–TPPI phase cycling (20), and water signals were suppressed with water flip-back pulses (21), pulsed field gradients (22), or presaturation during the relaxation delay. Decoupling in ¹⁵N-labeled samples was achieved using WALTZ16 modulation (23). NOE data were obtained from two-dimensional homonuclear NOESY (mixing times τ_m of 50 and 200 ms) (24, 25), three-dimensional ¹⁵N-edited NOESY-HSQC ($\tau_m = 200$ ms) (26), and three-dimensional ¹⁵N,¹⁵N-edited HMQC-NOESY-HSQC ($\tau_m = 120$ ms) (27) NMR spectra. Side chain signals were assigned using three-dimensional ¹⁵N-edited TOCSY data (28) obtained with a 70 ms clean-MLEV-17 mixing period (29) and sensitivity-improved gradient coherence selection (30). Three-dimensional HNHA data (31) were also obtained to allow discrimination of α -protons, and β -protons were assigned using three-dimensional HNHB data (32). NMR data were processed with NMRPipe (33) and analyzed with NMRView (34).

NMR Relaxation Experiments. ¹⁵N T_1 and T_2 relaxation measurements and ¹⁵N–{¹H} NOE measurements were conducted using conventional pulse sequences (35–37) that included water flip-back (21) and WATERGATE read pulses for water saturation prior to signal acquisition (22). To minimize systematic heating problems, data were collected in an interleaved manner with alternating relaxation times. The T_1 data were collected using relaxation recovery delays of 20.0, 100.3, 200.6, 351.1, 501.6, 702.2, 902.8, and 1203.7 ms, and the T_2 data were collected with 7.9, 23.8, 39.7, 47.7, 79.5, 95.4, 135.2, and 198.8 ms relaxation recovery delays. An additional T_2 experiment was conducted using a 3 s predelay. The ¹⁵N–{¹H} NOE spectrum was collected with a 3 s predelay and 3 s proton saturation period interleaved with a reference spectrum that included a 6 s predelay and no saturation period. Relaxation rates were determined by fitting the cross-peak intensities to an exponential decay. The ¹⁵N–{¹H} NOE values were obtained from the ratio (I/I_0) of the intensity of a cross-peak in the presence (I) and absence (I_0) of proton saturation, and errors were estimated from the baseline noise in the two spectra. The relaxation data were analyzed using the program Quadric Diffusion, version 1.11 (38), provided by A. G. Palmer (Columbia University, New York, NY).

Structure Calculations. Structure calculations were performed using DYANA (39). Distance restraints of 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å were employed for strong, medium, and weak cross-peaks, respectively, observed in the two- and three-dimensional NOESY spectra. An additional 0.5 Å was added for NOEs involving methyl protons. Eighteen restraints per zinc knuckle were included to enforce Zn–S and Zn–N bond distances of 2.35 and 2.05 Å, respectively, and to ensure proper hybridization of the His N^ε and Cys S^γ atoms. Hydrogen bond restraints for elements of regular secondary structure (including helical and β -sheet segments and the

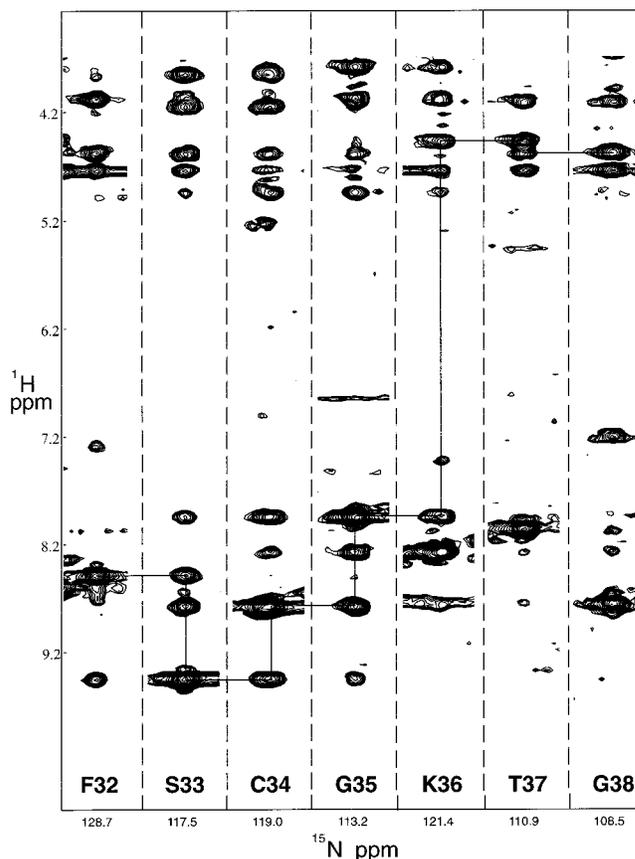


FIGURE 1: Selected portions of the three-dimensional ¹⁵N-edited NOESY-HSQC spectrum of the MMTV nucleocapsid protein showing sequential H^N–H^N and H^α–H^N NOE connectivities observed for residues Phe32–Gly38 of the N-terminal zinc knuckle.

rubredoxin knuckles) were identified on the basis of redundant and internally self-consistent NOE cross-peak patterns (40). Deuterium exchange experiments were not employed since we have previously found that the NH protons of these small zinc knuckle domains are not well protected, and substantial NH exchange generally occurs on the time scale of these experiments (tens of minutes). Numerous cross-peaks observed in the NOESY spectra were not used as distance restraints. For example, protons with NOE signals to a specific prochiral proton often gave rise to a signal with a weaker intensity to the other prochiral proton, and in these cases, restraints were not included for the weaker of the two cross-peaks. In addition, most aliphatic protons of residues with exposed, charged side chains were not restrained as they appear to be disordered in solution.

RESULTS AND DISCUSSION

Resonance Assignments. Signal assignments were made for the native protein by standard analysis of two-dimensional homonuclear correlated spectra and three-dimensional data obtained for the ¹⁵N-labeled protein (41). Rapid H^N proton exchange precluded unambiguous assignment of residues Ala1–Gly28, Lys48–Pro55, and Asn91–Leu95. Selected portions of the ¹⁵N-edited NOESY-HSQC spectra showing sequential NOE connectivities for residues comprising part of the N-terminal zinc knuckle domain are shown in Figure 1. The proton signals of the structured regions of the MMTV NC could be almost completely assigned using the phase-sensitive double-quantum-filtered COSY, HOHAHA, NOE-

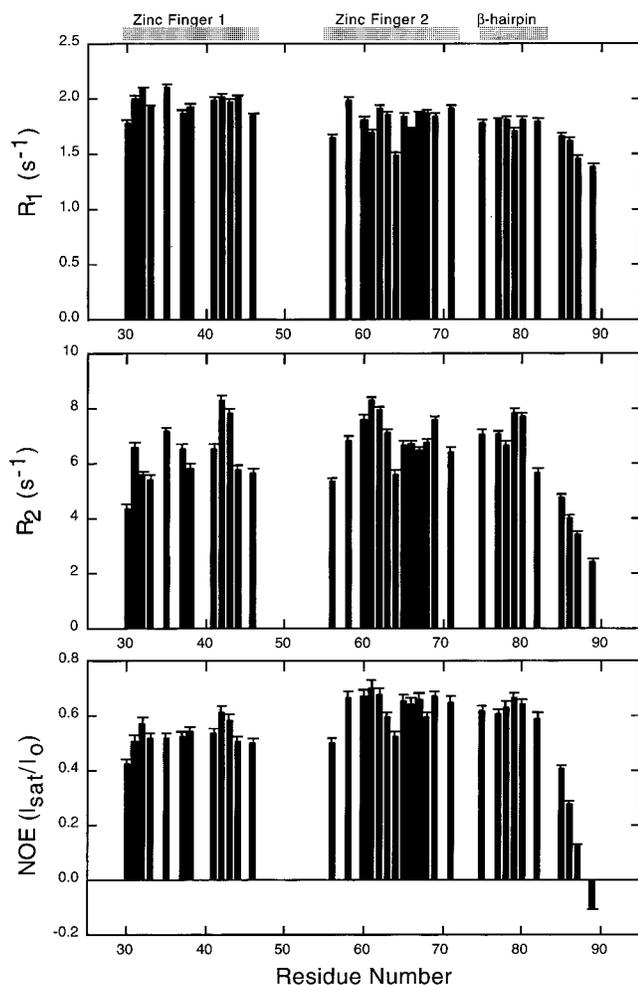


FIGURE 2: ^{15}N backbone amide R_1 (top) and R_2 (middle) relaxation rates and $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE values (bottom) obtained for the MMTV NC protein. Residues that comprise the N- and C-terminal zinc knuckles are given at the top of the figure.

SY, HNHA, and HNHB spectra. No ^1H signals were assigned for the backbone amide protons of the unstructured regions (Ala1–Gly28, Lys48–Pro55, and Asn91–Leu95) and Lys74. These regions are characterized by a lack of $^1\text{H}\text{-}^1\text{H}$ NOE cross-peaks and severe chemical shift degeneracy. Residues comprising the CCHC arrays exhibited sequential $\text{H}^{\text{N}}_i\text{-}\text{H}^{\text{N}}_{i+1}$, $\text{H}^{\alpha}_i\text{-}\text{H}^{\text{N}}_{i+1}$, and $\text{H}^{\text{N}}_i\text{-}\text{H}^{\delta}_{i+1}$ (for prolines) NOEs, enabling spectral assignments to be made by sequential assignment strategies (40). In a few cases, aliphatic protons of residues with longer side chains could not be assigned unambiguously due to signal overlap in the ^1H dimension. Proton and nitrogen NMR chemical shift assignments are given in Table S1 of the Supporting Information.

^{15}N Backbone Relaxation Data. ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE values measured for 12 of the 16 non-proline residues of the N-terminal CCHC array and 19 of the 24 non-proline residues of the C-terminal array are shown in Figure 2. Values for residues not included in Figure 2 are missing due to spectral overlap, which precluded the accurate measurement of peak intensities. Since it was previously reported that, in certain circumstances, a longer relaxation delay might be necessary to reduce sample heating which in turn would effect the T_2 values (42), a second T_2 experiment was carried out with a 3 s (as opposed to 1 s) relaxation delay. However,

in this study there were no significant differences between the T_2 values collected with a 1 or 3 s relaxation delay.

Inspection of the $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE spectra revealed that the only regions of the MMTV NC protein that are structured are the two zinc knuckles. Thus, even though we were unable to assign all of the residues of the N- and C-terminal tails and the linker segment, *all of the unassigned backbone amide groups exhibited negative $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE values*, consistent with a random coil structure. Qualitatively, we also observe that the $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE values for assigned residues (Figure 2) are uniformly larger in the C-terminal zinc knuckle and β -hairpin than in the N-terminal zinc knuckle. The average $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE value in the N-terminal zinc knuckle is 0.53 ± 0.04 , while the average value for the C-terminal array is 0.63 ± 0.07 . This indicates that the C-terminal zinc knuckle tumbles more slowly in solution than the N-terminal zinc knuckle, as is expected since the C-terminal zinc knuckle contains the additional β -hairpin structure and has a larger mass (see below).

The rotational diffusion properties of the two zinc knuckles were analyzed independently using the program quadric_diffusion (38). Only the amides with relatively little internal motion, as judged by $^{15}\text{N}\text{-}\{^1\text{H}\}$ NOE values of >0.5 , were used for this analysis. Rotational diffusion tensors and the rotational correlation times were calculated for isotropic, axial symmetric, and anisotropic diffusion models, in a manner similar to that previously described (43, 44), and the results are shown in Table 2. We find that the data for the N-terminal zinc knuckle best fit an isotropic rotational diffusion model, with an effective rotational diffusion constant [$D_{\text{iso}} = (D_{xx} + D_{yy} + D_{zz})/3 = 1/(6\tau_m)$] of $(3.38 \pm 0.02) \times 10^{-7} \text{ s}^{-1}$ (corresponding to a τ_m of 4.93 ns). The C-terminal zinc knuckle fits better to an axially symmetric rotational diffusion model, with a D_{iso} of $(3.13 \pm 0.04) \times 10^{-7} \text{ s}^{-1}$ ($\tau_m = 5.32$ ns). The fact that these values of D_{iso} are significantly different is a strong indication that the two zinc knuckles of MMTV NC do not undergo rotationally correlated motion and do not form a single globular domain.

Structure Determination. Elements of regular secondary structure were identified on the basis of observed cross-peak patterns and intensities observed in the two- and three-dimensional NOESY spectra. Sequential, midrange, and long-range cross-peaks for the backbone H^{N} and H^{α} protons, and their corresponding secondary structural elements, are summarized in Figure 3. The pattern of sequential and long-range NOEs observed for the residues comprising the N-terminal zinc knuckle was similar to that observed for the HIV-1 NC zinc knuckles. For example, intense sequential $\text{H}^{\text{N}}\text{-}\text{H}^{\text{N}}$ cross-peaks characteristic of a 3_{10} -helix were observed for residues Lys41–Cys44, and cross-peak patterns characteristic of metal-coordinating rubredoxin turns were observed for residues Cys31–Lys36 and Cys58–Cys63 (5).

NOEs characteristic of a β -hairpin were observed for residues Phe75–Leu82 that immediately follow the C-terminal CCHC array (Figure 4). The presence of interstrand Phe75 $\text{H}^{\alpha}\text{-}\text{Pro81}$ H^{α} (strong) and Asp76 $\text{H}^{\text{N}}\text{-}\text{Asn80}$ H^{α} (weak) NOE cross-peaks, as well as other cross-strand NOEs involving side chain protons, is consistent with the presence of hydrogen bonds from the backbone H^{N} protons of Asp76, Asn80, and Leu82 to the backbone carbonyl oxygen atoms of Asn80, Asp76, and Lys74, respectively (Figure 4). Numerous NOEs are observed between protons of the

Table 2: Rotational Diffusion Parameters for the Two Zinc Knuckles of the MMTV NC Protein

model	$D_{\text{iso}} (\times 10^{-7} \text{ s}^{-1})^a$	$2D_{zz}/(D_{xx} + D_{yy})$	D_{xx}/D_{yy}	θ (rad) ^b	ϕ (rad) ^b	Ψ (rad) ^b	χ^2	F
N-terminal zinc knuckle ^c								
isotropic	3.38 ± 0.02	—	—	—	—	—	265	—
axially symmetric	3.07 ± 0.15	2.0 ± 0.5	—	1.6 ± 0.3	3.2 ± 0.5	—	138	1.8
anisotropic	3.43 ± 0.04	2.0 ± 0.1	2.2 ± 0.3	1.6 ± 0.1	-0.5 ± 0.8	-0.01 ± 0.13	67	1.03
C-terminal zinc knuckle and β -hairpin ^d								
isotropic	3.13 ± 0.01	—	—	—	—	—	182	—
axially symmetric	3.13 ± 0.04	1.1 ± 0.2	—	1.2 ± 0.2	3.5 ± 0.2	—	106	3.2
anisotropic	3.13 ± 0.01	1.28 ± 0.04	1.15 ± 0.05	1.6 ± 0.1	0.08 ± 1.33	-0.05 ± 0.24	100	0.28

^a $\tau_m = (1/6D_{\text{iso}})$. ^b The euler angles describe the orientation of the diffusion tensor in the frame of reference of the inertial tensor. ^c Residues V30–K45. ^d Residues L57–L82.



FIGURE 3: Summary of backbone NOE connectivities and the deduced secondary structural elements of the MMTV nucleocapsid protein. Thick, medium, and thin bands correspond to strong, medium, and weak cross-peaks, respectively, observed in the NOESY data.

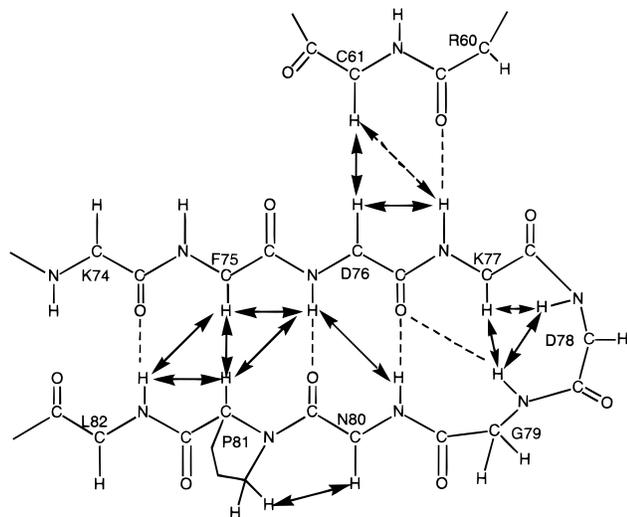


FIGURE 4: Alignment of the β -strands that make up the C-terminal β -hairpin in the distal zinc knuckle of MMTV NC. The solid arrows represent unambiguously assigned NOEs; the dashed arrow represents an NOE that could not be unambiguously assigned due to significant signal overlap, and the dotted lines represent hydrogen bonds.

β -hairpin and the preceding residues of the CCHC array, including Asp76 H^α –Cys61 H^α and H^β , Cys61 H^α –Asp76 H^β , Cys61 H^β –Asp76 H^β , Leu82 CH_3 –Cys61 H^β , and Ser73 H^β –Cys61 H^β protons. These NOEs position the β -hairpin adjacent to the rubredoxin-like turn of the zinc knuckle (Figure 4).

An ensemble of 20 structures were calculated using 365 NOE-derived distance restraints, including 143 restraints for

the N-terminal domain and 222 restraints for the C-terminal domain. This corresponds to an average of 12.4 and 13.6 restraints per refined residue for the N- and C-terminal knuckles, respectively. The quality of the structures of the N- and C-terminal domains is good, as evidenced by the low overall target functions (0.025 and 0.18 \AA^2 , respectively) and high internal convergence (rmsd for the best fit superpositions of backbone heavy atoms of 0.32 and 0.54 \AA , respectively; see Figure 5 and Table 3).

Description of the Structure and Comparison with Other Nucleocapsid Proteins. As indicated above, no interproton NOEs were observed between the N- and C-terminal zinc knuckles, and the heteronuclear NOE data indicate that the two domains tumble with different rotational correlation times. These data indicate that the MMTV NC protein may be best described as a flexible polypeptide chain with independently folded and noninteracting N- and C-terminal zinc knuckle domains. Ribbon representations of the N- and C-terminal zinc knuckle structures with the lowest residual restraint violations are shown in Figure 6.

Residues Val30–Gly38 and Leu57–Tyr65 of the N- and C-terminal zinc knuckles, respectively, form antiparallel β -hairpins that contain rubredoxin-type turns (5, 45). Within the N-terminal zinc knuckle, the backbone amide protons of residues Ser33 and Cys34 form hydrogen bonds to the sulfur atom of Cys31, and the amide proton of Lys36 forms a hydrogen bond with the sulfur atom of Cys34. Analogous $\text{NH}\cdots\text{S}$ hydrogen bonds exist in the C-terminal zinc knuckle, with the amide protons of Arg60 and Cys61 forming hydrogen bonds with S' of Cys58, and the amide proton of Lys63 forms a hydrogen bond with S' of Cys61. These $\text{NH}\cdots$

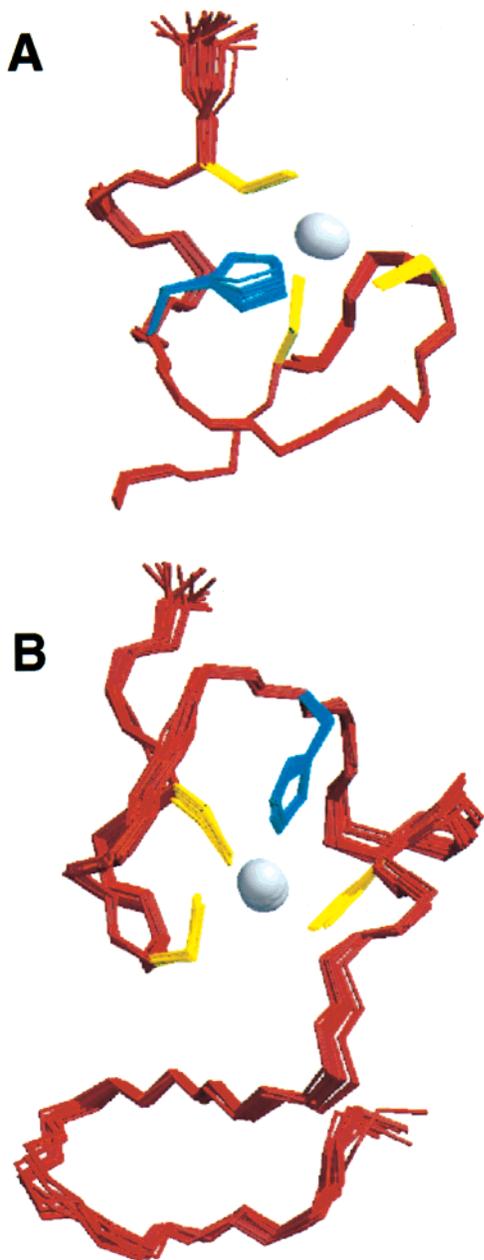


FIGURE 5: Superposition of the ensemble of the 20 refined structures of the N-terminal (A) and C-terminal (B) zinc knuckles of the MMTV NC protein. Backbone C', C α , and N atoms are displayed in red, zinc atoms in gray, and the side chains of the zinc-coordinating cysteines and histidine in yellow and blue, respectively. This figure was generated with the program Molscript (59).

S hydrogen bonding patterns have been observed in every other structurally characterized CCHC zinc knuckle, as well as in numerous other zinc knuckle families (46). In several cases, the existence of NH \cdots S hydrogen bonds has been observed directly by the detection of H-bond-mediated scalar coupling (47–51). Residues His39 and His66 of the N- and C-terminal zinc knuckles, respectively, coordinate zinc atoms through their N ϵ^2 atoms, and residues Lys41–Cys44 and Lys68–Cys71 form short 3_{10} -helices. These structural elements are identical to those observed previously for the HIV-1 and MoMuLV NC zinc knuckles (5, 52–56).

The C-terminal zinc knuckle contains an additional β -hairpin (residues Lys74–Leu82) that is not present in the

HIV-1 and MoMuLV zinc knuckles. Residues Asp76–Asn80 of the β -hairpin form a five-residue turn that is best described as a type II turn followed by a Gly79 β -bulge. Within this turn, the backbone amide protons of Gly79 and Asn80 are oriented to form hydrogen bonds to the backbone carbonyl of Asp76. In addition, the side chain NH δ^2 proton of Asn80 is poised to form a hydrogen bond with the backbone carbonyl of Lys77 (Figure 7). This hydrogen bonding pattern is consistent with unusual downfield ^{15}N and ^1H chemical shifts observed for the Asn80 NH $_2$ group. The β -hairpin packs tightly against the rubredoxin-like turn, and is stabilized by both hydrophobic and hydrogen bonding interactions. In particular, the backbone amide of Lys77 forms a hydrogen bond with the backbone carbonyl of Arg60, and the side chain oxygen of Ser73 is poised to allow hydrogen bonding between the OH hydrogen and the sulfur atom of either Cys61 or Cys71. In addition, the side chain of Leu82 makes hydrophobic contacts with the H $^\beta$ protons of Cys61, the sulfur atom of Cys71, and the β -protons of Ser73 and Asp76 (Figure 8). The C-terminal zinc knuckle is highly basic, and exposes a total of seven Arg and Lys residues in a nearly circular pattern about the circumference of the domain (Figure 8).

Very recently, the NMR structure of the NC protein from the Mason-Pfizer monkey virus (MPMV), a prototypical D-type retrovirus, was determined by NMR methods (57). This protein also contains two independently folded CCHC-type zinc knuckles. The structure of the N-terminal zinc knuckle of MPMV NC could not be determined due to internal conformational averaging on the microsecond to millisecond time scale; however, analysis of chemical shifts and NOE cross-peak patterns indicates that its fold is similar to that observed previously for the HIV-1 and MoMuLV NC zinc knuckles (57). High-quality NMR data were obtained for the C-terminal array of MPMV NC, which was shown to contain an additional reverse turn-like structure that packs against the rest of the CCHC array. Indeed, Giedroc and co-workers predicted that the C-terminal zinc knuckle of the MMTV NC protein, as well as the NC proteins of other D-type retroviruses, would contain this additional structure, on the basis of sequence comparisons (57). Our finding of an additional β -hairpin in the C-terminal CCHC array of the MMTV NC protein is consistent with this prediction. However, the β -hairpin observed in our structure is substantially different from that observed in the MPMV NC structure and predicted for MMTV NC. In particular, in the MPMV NC protein, residues Thr70–Gln73 form a “turn-like structure” and residues Asn75–Ile77 form a “tight γ -turn” that includes a hydrogen bond from the backbone amide proton of Ile77 to the backbone carbonyl oxygen of Asn75 (57). In contrast, in the MMTV NC structure, corresponding residues Phe75–Lys77 and Asn80–Leu82 do not form turns, but instead form antiparallel strands of the β -hairpin, with the intervening residues forming a turn as described above (see Figures 4 and 7). These differences are surprising in view of the high degree of sequence homology observed for the corresponding residues of the MPMV and MMTV NC proteins (57).

On the other hand, several long-range NOEs involving protons of the β -hairpin are observed in both the MPMV and MMTV NC proteins. For example, in MPMV, the Ile77 methyl protons exhibit NOEs to the H $^\beta$ protons of Cys56

Table 3: Distance Geometry Restraints and Structural Statistics for the 20 Final Structures

	N-terminal knuckle ^a	C-terminal knuckle ^b	total
no. of upper limit distance restraints			
intraresidue	22	30	52
sequential	28	47	75
medium- and long-range	43	73	116
hydrogen bond	36	58	94
Zn–ligand	14	14	28
total NMR-derived restraints	143	222	365
average no. of restraints per refined residue	12.4	13.6	13.1
distance violations			
mean total penalty (Å ²)	0.025	0.18	
maximum total penalty (Å ²)	0.031	0.22	
minimum total penalty (Å ²)	0.021	0.14	
pairwise rmsd relative to mean atom positions			
knuckle 1, Pro29–Glu46 (backbone heavy atoms) (Å)	0.32 ± 0.11		
knuckle 2, Gly56–Pro83 (backbone heavy atoms) (Å)	0.54 ± 0.18		

^a Pro29–Glu46. ^b Gly56–Pro83.

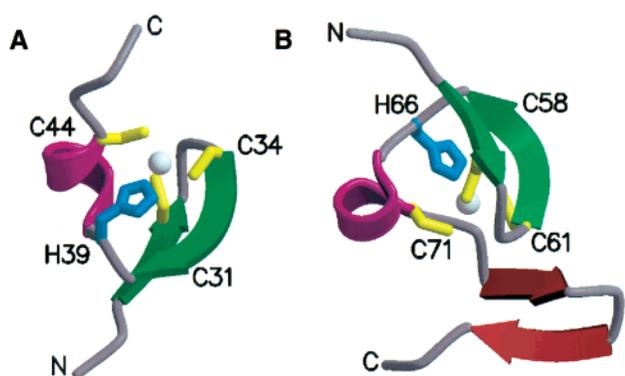


FIGURE 6: Ribbon representations of the N-terminal (A) and C-terminal (B) zinc knuckle domains of the MMTV NC protein. The unusual β -hairpin at the C-terminus of the distal zinc knuckle is colored red in panel B. This figure was generated with Molscript and rendered using Raster3D (60).

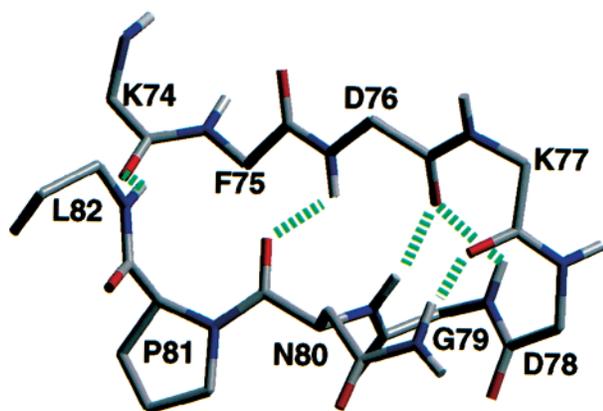


FIGURE 7: Main chain representation of the C-terminal β -hairpin of the distal MMTV NC zinc knuckle. The side chains of Asn80 and Pro81 are also shown. Green dashed lines represent hydrogen bonds deduced from the NOE data that contribute to the stability of the turn. This figure was generated using SETOR (61).

and Ser68 (57), and analogous NOEs in the MMTV NC are observed from the corresponding methyl protons of Leu82 to the H^β protons of Cys61 and Ser73. As a result, the conserved hydrophobic side chains of Leu82 in MMTV NC and Ile77 in MPMV both pack against similar portions of the rubredoxin turns, and the side chain OH protons of conserved serines at positions 73 (MMTV) and 68 (MPMV)

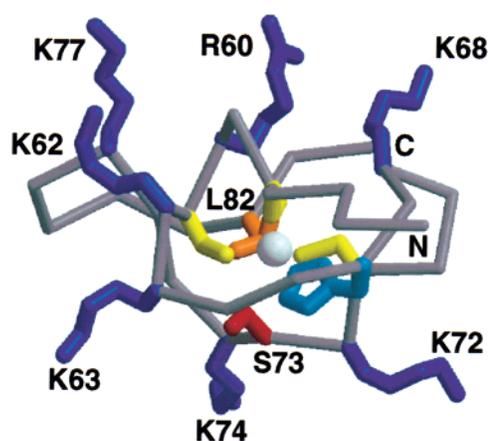


FIGURE 8: Structure of the MMTV NC distal zinc knuckle showing buried hydrophobic residues Leu82 (orange) and Ser73 (red), the zinc-coordinating cysteines (yellow) and histidine (light blue), and the exposed basic residues (dark blue).

both appear to hydrogen bond with a cysteine sulfur atom. Also, as indicated above, the side chain NH_2 group of Asn80 in MMTV NC is poised to form a hydrogen bond with the backbone carbonyl of Lys77, resulting in downfield 1H and ^{15}N chemical shifts, and the analogous Asn75 side chain NH_2 group of MPMV NC also exhibits downfield-shifted ^{15}N and 1H NMR signals. However, we do not observe NOEs between the methyl protons of Leu82 and the H^β protons of Lys62, as was observed in MPMV for the corresponding residues Ile77 and Lys57. Thus, the differences between the MMTV and MPMV NC zinc knuckle structures appear to be due to differences in the experimental NMR data.

CONCLUSIONS

The studies presented here provide the first structure determination of an NC protein from a mammalian B-type retrovirus. Whereas the N-terminal CCHC array adopts the generic three-dimensional fold exhibited by the HIV-1 NC and MoMuLV zinc knuckles, the C-terminal array has an unusual structure that includes an additional, well-defined β -hairpin. Backbone ^{15}N relaxation data indicate that residues comprising the N- and C-terminal tails and the central residues that link the domains are unstructured. The linker residues of the HIV-1 NC protein are also flexible, and this

allows the two zinc knuckles to interact with each other in a weak manner when the protein is free in solution (43, 55, 56, 58) and in a much stronger manner when the protein binds to nucleic acids (6). However, the two zinc knuckles of MMTV NC do not appear to interact with each other in the free protein, but instead behave as independently folded and stable domains on a flexible polypeptide chain. This appears to be a common feature of retroviral NC proteins, which may be important for adaptive binding to different nucleic acid targets (43). For example, as part of the Gag precursor, NC must bind specifically to one or more RNA target sites within the retroviral Ψ -RNA packaging signal during the assembly stage of the viral replication cycle, and during the infectivity stage, the protein performs a chaperone function by binding to both viral RNA and proviral DNA.

We are unable to explain the differences observed between the structures of the closely related, distal zinc knuckles of the MMTV and MPMV NC proteins, or the differences in the relative lability of the proximal zinc knuckles. Although the MPMV data were collected with a sample pH of 6.3, compared to a pH of 7.0 for MMTV, the ^1H - ^{15}N HSQC spectrum of MMTV NC was unaffected upon adjusting the pH to 6.3, and it is therefore unlikely that differences in pH were responsible for the different conformation of the β -hairpin, or for the high conformational dynamics reported for the N-terminal zinc knuckle of the MPMV NC protein (57). At this point, it appears that subtle differences in amino acid composition may be responsible for the observed differences in protein structure and lability.

The biological implications for the presence of the unusual C-terminal zinc knuckle in MMTV NC are not clear. Our findings are in many ways surprising, given the recent observation by Aldovini and co-workers that the HIV-1 and MMTV NC arrays can be swapped without producing a major adverse impact on RNA packaging. The studies presented here reveal that the HIV-1 and MMTV NC proteins are actually quite different. First, the linker that connects the CCHC arrays in HIV-1 NC is very short, and as a consequence, the two zinc knuckles exhibit transient interknuckle interactions in the absence of RNA (43, 55). Very strong interknuckle interactions with specific hydrogen bonding and hydrophobic packing are observed when HIV-1 NC is bound to fragments of the HIV-1 Ψ -RNA recognition site (6). On the other hand, the linker in MMTV NC is much longer and the protein does not exhibit transient interknuckle interactions. In addition, as discussed above, the C-terminal zinc knuckle of the MMTV NC protein contains a C-terminal β -hairpin that is not present in the HIV-1 NC protein.

Preliminary modeling experiments indicate that the additional β -hairpin would not interfere with interknuckle interactions observed in the structure of the HIV-1 NC protein when it is bound to the SL3 Ψ -RNA stem-loop recognition element (6), although it could interfere with other potential interknuckle interactions. On the other hand, the C-terminal zinc knuckle of MMTV NC contains a proline at the position immediately following the first cysteine (Pro59). This proline is conserved among the NC proteins of the B- and D-type retroviruses, but is never a proline in the lentiviruses such as HIV-1. This is a significant substitution, as the backbone amide proton at this position in both of the HIV-1 NC zinc knuckles participates in intermolecular hydrogen bonding to exposed guanosine nucleobases in the

HIV-1 NC-SL3 complex (6). Thus, the placement of a proline at this position is likely to preclude the specific binding of the C-terminal MMTV zinc knuckle to guanosine bases, and this suggests that this class of retroviral-type CCHC zinc knuckles recognizes either a different nucleobase or possibly a completely different RNA structural element.

The identity of the MMTV RNA packaging sequence has yet to be experimentally determined. However, as indicated above, MMTV NC appears to be capable of packaging the HIV-1 genome in chimeric HIV-1 virions that contain the MMTV NC sequence. Efforts are currently underway in our laboratory to study the binding of the MMTV NC protein to the HIV-1 Ψ -RNA to identify the potential nucleic acid interactive properties of its unusual C-terminal zinc knuckle and to determine how proteins as different as MMTV and HIV-1 NC are capable of recognizing and binding to a common RNA target.

ACKNOWLEDGMENT

We thank Justin Wu and Ramon Campos-Olivas (Howard Hughes Medical Institute, University of Maryland Baltimore County) for helpful discussions, Arthur G. Palmer (Columbia University) and Neil Farrow (DuPont Pharmaceutical Co.) for programs used for the analysis of the relaxation data, Dan Fabris (University of Maryland Baltimore County) for performing the mass spectrometry measurements, and Michael E. Zawrotney and Robert Edwards (Howard Hughes Medical Institute, University of Maryland Baltimore County) for technical assistance.

SUPPORTING INFORMATION AVAILABLE

^1H and ^{15}N NMR chemical shift assignments for MMTV NC. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

1. Teich, N. (1984) in *RNA Tumor Viruses* (Wiess, R., Teich, N., Varmus, H., and Coffin, J., Eds.) pp 25–208, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Dickson, C., and Peters, G. (1983) in *Current Topics in Microbiology and Immunology* (Vogt, P. K., and Koprowski, H., Eds.) pp 1–34, Springer-Verlag, New York.
3. Wang, Y., Holland, J. F., Bleiweiss, I. J., Melana, S., Liu, X., Pelisson, I., Cantarella, A., Strelrecht, K., Mani, S., and Pogo, B. G.-T. (1995) *Cancer Res.* 55, 5173–5179.
4. Vogt, V. M., and Simon, M. N. (1999) *J. Virol.* 73, 7050–7055.
5. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry* 29, 329–340.
6. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science* 279, 384–388.
7. Gorelick, R. J., Henderson, L. E., Hanser, J. P., and Rein, A. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 8420–8424.
8. Gorelick, R. J., Nigida, J., Stephen, M., Bess, J. W., Arthur, L. O., Henderson, L. E., and Alan, R. (1990) *J. Virol.* 64, 3207–3211.
9. Gorelick, R. J., Chabot, D. J., Alan, R., Henderson, L. E., and Arthur, L. O. (1993) *J. Virol.* 67, 4027–4036.
10. Jentoft, J. E., Smith, L. M., Fu, X., Johnson, M., and Leis, J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 7094–7098.
11. Dupraz, P., Oertle, S., Méric, C., Damay, P., and Spahr, P.-F. (1990) *J. Virol.* 64, 4978–4987.
12. Méric, C., Gouilloud, E., and Spahr, P.-F. (1988) *J. Virol.* 62, 3328–3333.
13. Méric, C., and Goff, S. P. (1989) *J. Virol.* 63, 1558–1568.

14. Berkowitz, R. D., Ohagen, A., Høglund, S., and Goff, S. P. (1995) *J. Virol.* 69, 6445–6456.
15. Wu, W., Henderson, L. E., Copeland, T. D., Gorelick, R. J., Bosche, W. J., Rein, A., and Levin, J. G. (1996) *J. Virol.* 70, 7132–7142.
16. Lapadat-Tapolsky, M., De Rocquigny, H., Van Gent, D., Roques, B., Plasterk, R., and Darlix, J.-L. (1993) *Nucleic Acids Res.* 21, 831–839.
17. Poon, D. T. K., Li, G., and Aldovini, A. (1998) *J. Virol.* 72, 1983–1993.
18. Majors, J. E., and Varmus, H. E. (1981) *Nature* 289, 253–258.
19. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* 185, 60–89.
20. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) *J. Magn. Reson.* 85, 393–399.
21. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 12593–12593.
22. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* 2, 661–665.
23. Shaka, A. J., Keler, J., and Freeman, R. (1983) *J. Magn. Reson.* 53, 313–340.
24. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546–4553.
25. Macura, S., and Ernst, R. R. (1980) *Mol. Phys.* 41, 95–117.
26. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) *Biochemistry* 28, 6150–6156.
27. Ikura, M., Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc.* 112, 9020–9022.
28. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc.* 111, 1515–1517.
29. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) *J. Am. Chem. Soc.* 110, 7870–7872.
30. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) *J. Biomol. NMR* 4, 845–858.
31. Vuister, G. W., and Bax, A. (1993) *J. Am. Chem. Soc.* 115, 7772–7777.
32. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) *J. Magn. Reson.* 95, 636–641.
33. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* 6, 277–293.
34. Johnson, B. A., and Blevins, R. A. (1994) *J. Biomol. NMR* 4, 603–614.
35. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* 28, 8972–8979.
36. Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A., and Torchia, D. A. (1992) *J. Magn. Reson.* 97, 359–375.
37. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* 33, 5984–6003.
38. Lee, L. K., Rance, M., Chazin, W. J., and Palmer, A. G. I. (1997) *J. Biomol. NMR* 9, 287–298.
39. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) *J. Mol. Biol.* 273, 283–298.
40. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
41. Clore, G. M., and Gronenborn, A. M. (1991) *Science* 252, 1390–1399.
42. Gagne, S. M., Tsuda, S., Spyropoulos, L., Kay, L. E., and Sykes, B. D. (1998) *J. Mol. Biol.* 278, 667–686.
43. Lee, B. M., De Guzman, R. N., Turner, B. G., Tjandra, N., and Summers, M. F. (1998) *J. Mol. Biol.* 279, 633–649.
44. Campos-Olivas, R., and Summers, M. F. (1999) *Biochemistry* 38, 10262–10271.
45. Adman, E., Watenpaugh, E. D., and Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 4854–4858.
46. Blake, P. R., and Summers, M. F. (1994) in *Advances in Inorganic Biochemistry* (Marzilli, L. G., and Eichhorn, G., Eds.) pp 201–228, Elsevier, New York.
47. Blake, P. R., Park, J. B., Adams, M. W. W., and Summers, M. F. (1992) *J. Am. Chem. Soc.* 114, 4931–4933.
48. Blake, P. B., Lee, B., Park, J.-B., Zhou, Z. H., Adams, M. W. W., and Summers, M. F. (1994) *New J. Chem.* 18, 387–395.
49. Blake, P. R., Lee, B., Summers, M. F., Adams, M. W. W., Park, J.-B., Zhou, Z. H., and Bax, A. (1992) *J. Biomol. NMR* 2, 527–533.
50. Pérez-Alvarado, G. C., Miles, C., Michelsen, J. W., Louis, H. A., Winge, D. R., Beckerle, M. C., and Summers, M. F. (1994) *Nat. Struct. Biol.* 1, 388–398.
51. Gardner, K. H., Adnerson, S. F., and Coleman, J. E. (1995) *Nat. Struct. Biol.* 2, 898–905.
52. South, T. L., Blake, P. R., Hare, D. R., and Summers, M. F. (1991) *Biochemistry* 30, 6342–6349.
53. Omichinski, J. G., Clore, G. M., Sakaguchi, K., Appella, E., and Gronenborn, A. M. (1991) *FEBS Lett.* 292, 25–30.
54. 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. C. I., Hare, D. R., and Arthur, L. O. (1992) *Protein Sci.* 1, 563–574.
55. Morellet, N., Jullian, N., De Rocquigny, H., Maignet, B., Darlix, J.-L., and Roques, B. P. (1992) *EMBO J.* 11, 3059–3065.
56. Déméné, H., Dong, C. Z., Ottmann, M., Rouyez, M. C., Jullian, N., Morellet, N., Mely, Y., Darlix, J. L., Fournié-Zaluski, M. C., Saragosti, S., and Roques, B. P. (1994) *Biochemistry* 33, 11707–11716.
57. Gao, Y., Kaluarachchi, K., and Giedroc, D. P. (1998) *Protein Sci.* 7, 2265–2280.
58. Mély, Y., De Rocquigny, H., Morellet, N., Roques, B. P., and Gérard, D. (1996) *Biochemistry* 35, 5175–5182.
59. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* 24, 946–950.
60. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* 277, 505–524.
61. Evans, S. V. (1993) *J. Mol. Graphics* 11, 134–138.

BI9922493