representative of results from at least two independent donors. ICAM-1 was obtained from mouse spleens by the tissue lysis procedure as previously described (*18*) and was affinity-purified on a monoclonal antibody BE-29G1–Sepharose (Pharmacia) column. Human peripheral blood mononuclear cells were isolated from healthy donors as described [C. P. Nielson, R. E. Vestal, R. J. Sturm, R. Haeslip, *J. Allergy Clin. Immunol.* **88**, 801 (1990)]. Monocytes were removed from the cells by adherence to a T-175 culture flask (Nunc, Demark) at 37°C and 8% CO_2 for 30 min in RPMI 1640 medium supplemented with 10% calf serum. Recombinant human eotaxin, MIP-1α, MIP-1β, RANTES, MCP-1, MCP-3, IP-10, IL-8, GRO-α, C-X3-C-kine, DC-CK-1, MIP-3α, MIP-3β, 6-C-kine, I-309, SDF-1β, lymphotactin, and TARC were obtained from R&D Systems (Minneapolis, MN) or PeproTech (Rocky Hill, NJ). Recombinant IL-8 was also provided by A. Rott (Sandoz, Vienna). Synthetic human SDF-1α, MIP-3α, MIP-3β, TARC, and C-X3-C-kine were also prepared by two of the authors (M.A.S. and D.A.T.) by chemical ligation [P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. H. B. Kent, *Science* **266**, 776 (1994). Synthetic SDF-1α was also provided by I. Clark-Lewis. Chemokines were shown to be active by chemotaxis of human peripheral blood lymphocytes, eosinophils, or neutrophils, or of chemokine receptor–overexpressing lymphoma cell lines. There were no detectable differences among chemokines obtained from different sources in these assays.

35. $CD4^+$ cells were purified with the use of anti-CD4 Dynabeads and the DETACH-a-BEAD system (Dynal, Lake Success, NY). Mouse monoclonal antibodies to human CD45RO (clone UCHL1) and to human CD45RA (clone HI100) were obtained from PharMingen (San Diego, CA). Cells coated with these monoclonal antibodies were removed by incubation with microbeads conjugated with antibodies to mouse immunoglobulin, followed by magnetic depletion (Miltenyi Biotec, Auburn, CA). A portion of each processed cell subpopulation was stained with directly conjugated monoclonal antibodies and analyzed by flow cytometry to ascertain purity, which was ~99% for $CD4^+$ cells and ~97% for both $CD45RA^-$ $CD45RO^+$ and $CD45RA^+$ $CD45RO^-$ subpopulations.
36. C. R. Mackay, *Adv. Immunol.* **53**, 217 (1993).
37. M. E. Sanders *et al.*, *J. Immunol.* **140**, 1401 (1988).
38. E. L. Berg, L. M. McEvoy, C. Berlin, E. C. Butcher, *Nature* **366**, 695 (1993); U. H. von Andrian, S. R. Hasslen, R. D. Nelson, S. L. Erlandsen, E. C. Butcher, *Cell* **82**, 989 (1995).
39. M. B. Lawrence, E. L. Berg, E. C. Butcher, T. A. Springer, *Eur. J. Immunol.* **25**, 1025 (1995).
40. PNAd was obtained from human tonsils with the tissue lysis procedure previously described (*18*) and was affinity-purified on a monoclonal antibody MECA-79–Sepharose column. The inside walls of 100-μl microcap capillary tubes (Drummond, Broomall, PA) were coated with PNAd or a combination of PNAd plus ICAM-1. For double coatings, the tubes were coated first with 10 μl of diluted ICAM-1 for 8 hours at 4°C, the unbound ICAM-1 was removed, and the tubes were then coated with diluted PNAd by incubation overnight at 4°C. For single PNAd coating, only the second half of this procedure was performed. Before flow experiments, unbound PNAd was removed and the entire inside of the capillary was exposed to 100% calf serum at room temperature for 5 min. To cocoat the PNAd–ICAM-1 areas with immobilized chemokine (or medium alone as a control), we added 10 μl of 2 μM chemokine (in equilibrated RPMI 1640 with 10% calf serum) through the downstream end of the capillary and coaxed it into the same position as the area coated with PNAd and ICAM-1. Care was taken that no chemokine touched the capillary upstream of the PNAd and ICAM-1. The tube was then incubated for 5 min at ambient temperature. Immediately before the experiment, the unbound chemokine was washed out through the downstream end of the tube by infusing 5 ml of complete medium into the upstream end. Washing in this direction prevented chemokine from contacting the upstream areas. Cells were passed through the capillary in complete medium at a density of 1.5×10^6 per milliliter. The rate of flow was controlled by a Harvard 33 syringe pump (Harvard Apparatus, South Natick, MA). Experiments were performed at a flow rate of 1250 μl/min, which creates a wall shear stress of ~2.0 dynes/cm² for a capillary with an inner diameter of 1.025 mm, as calculated from Poiseuille's law for newtonian fluids with a viscosity of 0.01 poise. {Wall shear stress (dyne cm²) = mean flow velocity (mm/s) × [8/tube diameter (mm)] × viscosity (poise).} The interactions of cells with the coated areas were recorded on videotape, and the behavior of individual cells was analyzed frame by frame.
41. C. Berlin *et al.*, *Cell* **80**, 413 (1995).
42. We thank I. Clark-Lewis, U. H. von Andrian, and D. P. Andrew for synthetic SDF-1α, PNAd, and ICAM-1, respectively, used in preliminary experiments; E. P. Bowman, K. Youngman, and M. Hubbe for help with preliminary experiments and discussions; L. Rott and G. Haraldson for advice on cell separation; and S. Haugejorden-Brown, E. P. Bowman, E. F. Foxman, and R. A. Warnock for critical reading of the manuscript. Supported by NIH grant GM37734 and an award from the Department of Veterans Affairs (E.C.B.); NIH grant DK38707 to the FACS Core Facility of the Stanford Digestive Disease Center, and NIH Cancer Etiology, Prevention, Detection, and Diagnosis grant 5T32CA090302, NIH grant 1F32AI08930, and the Arthritis Foundation (J.J.C.).

6 October 1997; accepted 2 December 1997

Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3 Ψ-RNA Recognition Element

Roberto N. De Guzman, Zheng Rong Wu, Chelsea C. Stalling, Lucia Pappalardo, Philip N. Borer,* Michael F. Summers*

The three-dimensional structure of the human immunodeficiency virus–type 1 (HIV-1) nucleocapsid protein (NC) bound to the SL3 stem-loop recognition element of the genomic Ψ RNA packaging signal has been determined by heteronuclear magnetic resonance spectroscopy. Tight binding (dissociation constant, ~100 nM) is mediated by specific interactions between the amino- and carboxyl-terminal CCHC-type zinc knuckles of the NC protein and the G^7 and G^9 nucleotide bases, respectively, of the G^6-G^7-A^8-G^9 RNA tetraloop. A^8 packs against the amino-terminal knuckle and forms a hydrogen bond with conserved Arg^{32}, and residues Lys^3 to Arg^{10} of NC form a 3_{10} helix that binds to the major groove of the RNA stem and also packs against the amino-terminal zinc knuckle. The structure provides insights into the mechanism of viral genome recognition, explains extensive amino acid conservation within NC, and serves as a basis for the development of inhibitors designed to interfere with genome encapsidation.

All retroviruses encode a gag polyprotein that is produced in the host cell during the late stages of the infectious cycle and directs the encapsidation of two copies of the unspliced viral genome during virus assembly and budding (*1*). Concomitant with budding, the gag polyproteins are cleaved by the viral protease into the matrix (MA), capsid (CA), and nucleocapsid (NC) proteins, which rearrange during maturation to form infectious particles (*2*). Except for the spumaviruses, all retroviral NC proteins contain one or two CCHC-type zinc knuckle domains (Cys-X_2-Cys-X_4-His-X_4-Cys, where X = variable amino acid) (*3*) (Fig. 1A). These domains are critical for viral replication and participate directly in genome recognition and encapsidation (*4*, *5*). Mutations that abolish zinc binding lead to noninfectious virions that lack their genomes (*4*, *6*), and mutations of conservatively substituted hydrophobic residues within the CCHC arrays can alter RNA packaging specificity (*5*). In addition, entire NC domains of HIV-1 and Moloney murine leukemia virus (MoMuLV) have been swapped, resulting in the specific packaging of the non-native genomes (*6*).

Recognition of the HIV-1 genome occurs by means of interactions between NC and a ~120-nucleotide region of the unspliced viral RNA known as the Ψ-site, which is located between the 5′ long terminal repeat and the *gag* initiation codon (*7*). Extensive site-directed mutagenesis, chemical modification, nuclease accessibility mapping, and free energy computational studies indicate that the HIV-1 Ψ-site contains four stem-loop structures, denoted SL1 through SL4 (Fig. 1B) (*8*–*13*). Although mutagenesis experiments indicate that all four of these structures are important for efficient encapsidation (*13*, *14*), SL3 is of particular interest because its sequence is highly conserved among different strains of HIV-1 (*10*) despite heterogeneity at adjacent positions, and because linkage of SL3 to heterologous RNAs is sufficient to direct their recogni-

R. N. De Guzman, Z. R. Wu, C. C. Stalling, M. F. Summers, Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland–Baltimore County (UMBC), 1000 Hilltop Circle, Baltimore, MD 21250, USA.
L. Pappalardo and P. N. Borer, Department of Chemistry, Syracuse University, Syracuse, NY 13244, USA.
*To whom correspondence should be addressed.

tion and packaging into virus-like particles (9).

We prepared a 20-nucleotide RNA containing the sequence of SL3 (Fig. 1C) for NC-binding and structural studies by nuclear magnetic resonance (NMR) (*15*). RNA samples at natural isotopic abundance and enriched in ^{15}N and ^{13}C were prepared with T7 RNA polymerase (*15*), and recombinant HIV-1 NC protein (strain NL4-3; unlabeled, ^{15}N-labeled, and $^{15}N,^{13}C$-labeled) was expressed in *Escherichia coli* and purified under nondenaturing conditions (*16*). Samples of RNA were titrated with protein to equimolar concentrations, affording a 1:1 NC-SL3 complex with an apparent molecular weight of 13.7 kD (*17*). Tight binding (dissociation constant K_d, ~100 nM) (*18*) occurs in the slow exchange regime of the NMR chemical shift time scale. Analysis of the homo- and heteronuclear correlated NMR data enabled complete assignment of the protein and RNA signals and the identification of 59 direct intermolecular nuclear Overhauser effects (NOEs) (*19*). A portion of the two-dimensional (2D) NOE spectrum (Fig. 2A) shows intermolecular NOEs from the aromatic protons of Phe^{16} and Trp^{37} to A^8 and G^7, respectively, as well as strong intramolecular NOEs. All intermolecular NOEs were assigned unambiguously in 3D pulsed-field gradient–edited ^{13}C-filtered,^{12}C-detected NOE data. For example, Ala^{25}-CH_3 exhibits NOEs with A^8-H1′, A^8-H2, A^8-H8 (spin diffusion), and G^9-H8; Ile^{24}-δCH_3 interacts with G^9-H1′ and G^9-H8; and Lys^{26}-Hα interacts with A^8-H2 (Fig. 2B).

A total of 719 experimental distance restraints (average of 19.4 restraints per refined residue) identified from the NOE data were used to generate an ensemble of 25 distance geometry structures with the program DYANA (*20*). Stereo views of the best fit superpositions and statistical information for the structure calculations (Fig. 3) demonstrate that the calculations led to good convergence. The stem nucleotides (through the C^5-G^{10} base pair) form an A helix (Fig. 4A). The G^6 nucleobase of the G^6-G^7-A^8-G^9 RNA tetraloop stacks on the C^5 base and forms a G^6-O6–G^9-NH_2 hydrogen bond. The remaining tetraloop bases project away from the stem and interact directly with the NC protein (Fig. 4B).

The NC protein consists of two zinc knuckle domains (F1 = Val^{13} to Ala^{30}; F2 = Gly^{35} to Glu^{51}) separated by a basic linker segment (Pro^{31}-Arg^{32}-Lys^{33}-Lys^{34}) and flanked by NH_2- and COOH-terminal tails (Met^1 to Thr^{12} and Arg^{52} to Asn^{55}, respectively). Residues Lys^3 to Arg^{10} of the NH_2-terminal tail form a 3_{10} helix that binds within the RNA major groove, and the zinc knuckles interact with the exposed bases of the RNA tetraloop (Fig. 4A). G^9 interacts specifically with the F1 knuckle by binding to a hydrophobic cleft formed by the side chains of conservatively substituted residues Val^{13}, Phe^{16}, Ile^{24}, and Ala^{25}. The Phe^{16} and Ala^{25} backbone NH groups located at the bottom of the cleft form hydrogen bonds with G^9-O6, and the Lys^{14}-CO backbone oxygen forms a hydrogen bond with G^9-H1 (Fig. 4C). G^7 interacts in a very similar manner with the F2 knuckle, with the nucleotide base packing in a hydrophobic cleft formed by conservatively substituted Trp^{37}, Gln^{45}, and Met^{46} side chains; the exocyclic G^7-O6 oxygen forms a hydrogen bond with the backbone NH atoms of Trp^{37} and Met^{46}, and the G^7-H1 proton forms a hydrogen bond with Gly^{35}-CO. Thus, both zinc knuckles bind specifically to exposed guanosines and form hydrogen bonds to groups that normally engage in Watson-Crick hydrogen bonding in A helices, and this may serve as the primary mode by which CCHC zinc knuckles contribute to sequence-specific RNA binding (*21*). This binding mechanism is substantially different from that of CCHH-type zinc fingers from eukaryotic transcription factors; the latter mechanism mainly involves interactions between side chains of α-helical residues and base pairs in the

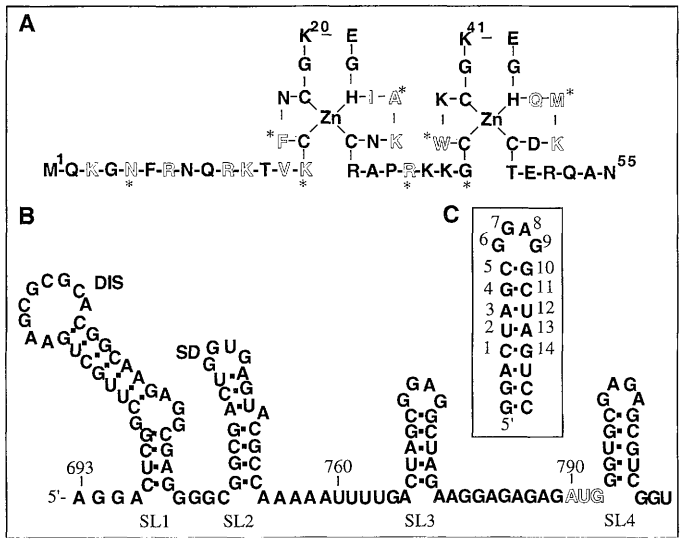

Fig. 1. (**A**) Amino acid sequence of the HIV-1_{NL4-3} NC protein showing the zinc-binding modes of the two CCHC-type zinc knuckles. Residues that contact the RNA in the NC-SL3 complex are denoted by open letters; asterisks denote residues involved in intermolecular hydrogen bonding. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; T, Thr; V, Val; and W, Trp. (**B**) Nucleotide sequence and secondary structure of the HIV-1_{NL4-3} Ψ-sequence (*12*). The dimer initiation and major splice donor sites are labeled DIS and SD, respectively, and the gag initiation codon (AUG) is given in open letters. (**C**) Sequence of the RNA construct used in our studies (*15*).

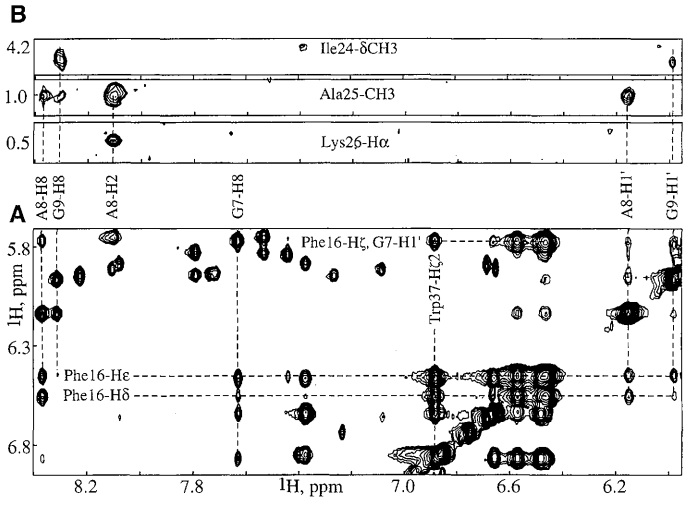

Fig. 2. (**A**) Portion of the 800-MHz 2D NOE spectrum obtained for the NC-SL3 complex in D_2O solution. Intermolecular cross-peaks involving the aromatic protons of Phe^{16} (F1 knuckle) and Trp^{37} (F2 knuckle) are labeled. Strong intramolecular Trp^{37}-Phe^{16} cross-peaks (labeled) are indicative of interknuckle packing in the complex. (**B**) Selected strips from the 800-MHz 3D ^{13}C-filtered,^{12}C-detected HMQC-NOESY data obtained for the NC-SL3 complex in D_2O showing unambiguously assigned intermolecular NOE cross-peaks associated with the Ile^{24}-δCH_3, Ala^{25}-CH_3, and Lys^{26}-Hα protons.

DNA major groove (*22*, *23*).

The base of the remaining tetraloop nucleotide, A^8, makes hydrophobic contacts with the Ala^{25}-CH_3, Phe^{16}-$C\beta H_2$, and Asn^{17}-$C\beta H_2$ groups of the F1 knuckle (Fig. 4B) and forms a hydrogen bond with the side-chain Nε-H proton of Arg^{32} (Fig. 4D). This arginine is highly conserved among the known strains of HIV-1 (substituted by Lys in only three of 94 published sequences) (*24*), whereas most of the other basic sites in NC do not substantially discriminate between Arg and Lys residues. Mutation of Arg^{32} to Gly results in a reduction in genome packaging (to 10% of that found in the wild type) and abolishes infectivity (*25*). Thus, the Arg^{32}-A^8 hydrogen bond provides a rationale for the high conservation of Arg^{32} and its extreme sensitivity to site-directed mutagenesis.

The HIV-1 NC protein contains 14 additional basic sites with conserved Arg or Lys residues (*24*), 10 of which participate in intra- or intermolecular interactions in the NC-SL3 complex. Salt bridges involving the Lys^{38}-Glu^{51} and Lys^{14}-Glu^{21} pairs appear to stabilize the folding of the F2 and F1 domains, respectively, and a salt bridge between Lys^{33} and Glu^{42} appears to stabilize F2 knuckle-linker interactions. The side chain NH_3^+ of Lys^{47} is located between the 3′- and 5′-phosphodiesters of G^7 in a manner that neutralizes repulsions and anchors the F2 knuckle to RNA. Similarly, Lys^{26} anchors the F1 knuckle to RNA through electrostatic interactions with the 3′-phosphodiester of G^9. The side chains of Lys^{20}, Arg^{29}, Lys^{34}, and Lys^{41} project into solution and do not form salt bridges in the complex.

Conserved basic sites also exist at positions 3, 7, 10, and 11 of the 3_{10} helix (*24*), and in the NC-SL3 complex these residues make the following electrostatic contacts with phosphodiester groups of the RNA stem: $Lys^3 \cdots G^{10}$ (3′-P), $Arg^7 \ldots C^1$ (3′- and 5′-P), $Arg^{10} \cdots U^2$ (3′-P, 5′-P, or both), and $Lys^{11} \cdots G^4$ (3′-P) (Fig. 4D). In addition to these nonspecific interactions, the side-chain carbonyl of Asn^5 forms a hydrogen bond with the exposed exocyclic NH_2 proton of C^{11}, and the Asn^5-NH_2 group is poised to interact with the 2′-hydroxyl of G^9 and the N7 atom of G^{10}. Interestingly, Asn^5 is also highly conserved among the known strains of HIV-1 (94% Asn and 3% Gln, compared with 33% Asn and 62% Gly for Asn^8), and Asn^5 is the only NC residue that makes specific hydrogen-bonding contacts with the RNA stem. Although site-directed mutagenesis of Asn^5 has not been performed, its high conservation is consistent with the structural implication that this residue is important for RNA recognition.

Strong-intensity interfinger NOE cross-peaks observed in the complex, but not for the free protein (*26*), indicate that binding is also promoted by the formation of extensive intraprotein interactions (Fig. 4B). The 3_{10} helix packs tightly against the F1 knuckle by hydrophobic interactions involving conservatively substituted Phe^4 of the 3_{10} helix and Val^{13} and Ile^{24} of the F1 knuckle. In addition, the zinc knuckles pack tightly together as a result of hydrophobic interactions between Trp^{37} of F2 and Phe^{16}, Asn^{17}, and Gly^{19} of F1, and by a hydrogen bond from Trp^{37}-Hε1 to the backbone carbonyl of Phe^{16}. Finally, residues that link the two zinc knuckles (Pro^{31}-Lys^{34}) adopt a single conformation that is stabilized by extensive hydrogen bonding (Asn^{17}-NH^E to Cys^{28}-Sγ, Asn^{17}-NH^Z to Pro^{31}-CO, and Asn^{17}-Oδ1 to Lys^{33}-NH). Mutations likely to destabilize the linker structure, such as $Ala^{30} \rightarrow$ Pro and $Pro^{31} \rightarrow$ Leu, lead to poorly infectious and noninfectious particles, respectively (*25*), and thus it appears that high-affinity binding to SL3 is mediated by the formation of extensive intramolecular interactions, in addition to the specific protein–nucleic acid interactions described above. These findings are analogous to those observed recently in a DNA complex with a three–zinc finger domain of transcription factor IIIA, where flexible linker segments become structured and numerous finger-finger contacts are made upon binding to DNA (*23*).

The SL3 RNA-NC complex differs from other structurally characterized protein-RNA complexes (*27*), most of which are characterized by purine-purine base pairs that widen the major groove and allow penetration of α-helical (*28*) or

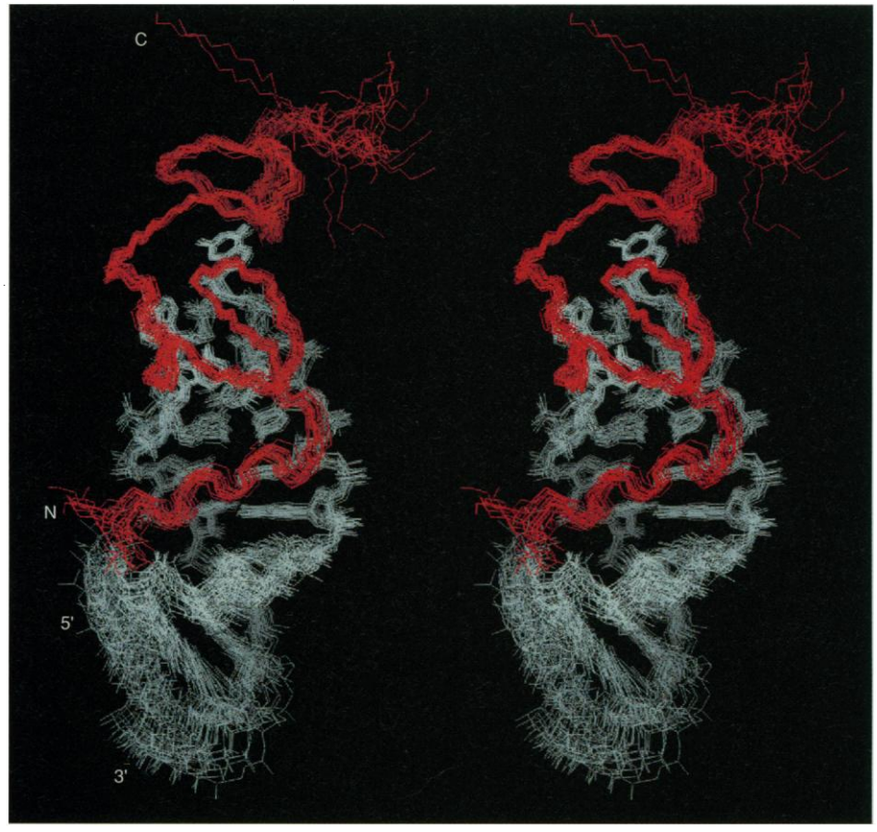

Fig. 3. Stereo view of the best fit superposition of the HIV-1 NC-SL3 RNA complex (backbone C, Cα, and N atoms of residues Lys^3 to Glu^{51} of NC, and all heavy atoms of SL3 RNA nucleotides C^1 to G^{14}; NC in red and RNA in white). Distance restraints: total, 719; average number of restraints per refined residue, 19.4; intraresidue, 100; sequential, 162; medium and long range, 158; intermolecular, 59; hydrogen bond (four per hydrogen bond), 240. Target function: mean, 0.37 ± 0.05 Å^2; maximum, 0.42 Å^2; minimum, 0.25 Å^2. Individual violations: average maximum, 0.11 ± 0.02 Å; maximum, 0.17; average number of violations (>0.1 Å) per structure, 2.8 ± 0.3. Pairwise root-mean-square deviations relative to mean atom positions: RNA residues C^1 to G^{14} (all heavy atoms), 0.59 ± 0.10 Å; NC residues Lys^3 to Glu^{51} (backbone heavy atoms), 0.36 ± 0.11 Å; backbone atoms of NC (Lys^3 to Glu^{51}) plus all heavy atoms of RNA (C^1 to G^{14}), 0.63 ± 0.11 Å; all heavy atoms of residues Lys^3 to Glu^{51} plus C^1 to G^{14}, 0.93 ± 0.12 Å.

β-sheet (*29*) segments, or by combinations of interactions (*30*). In the NC-SL3 complex, non–A-helical torsion angles associated with the G^9 phosphodiesters lead to a kink in the RNA backbone and a widening of the major groove, allowing penetration of the smaller 3_{10} helix. The structure of the tetraloop differs markedly from those of the GNRA class (*31*), in which three of the bases are stacked and involved in intramolecular hydrogen bonding. In this respect, the NC-SL3 complex is similar to that of the bacteriophage MS2 coat protein–operator stem-loop structure in which exposed tetraloop bases participate in specific intermolecular hydrogen bonding (*32*). In general, protein-RNA interactions occur by means of an adaptive binding mechanism in which flexible RNA nucleotides become ordered upon binding (*27*), and this also appears to be the case for the SL3 RNA (*33*).

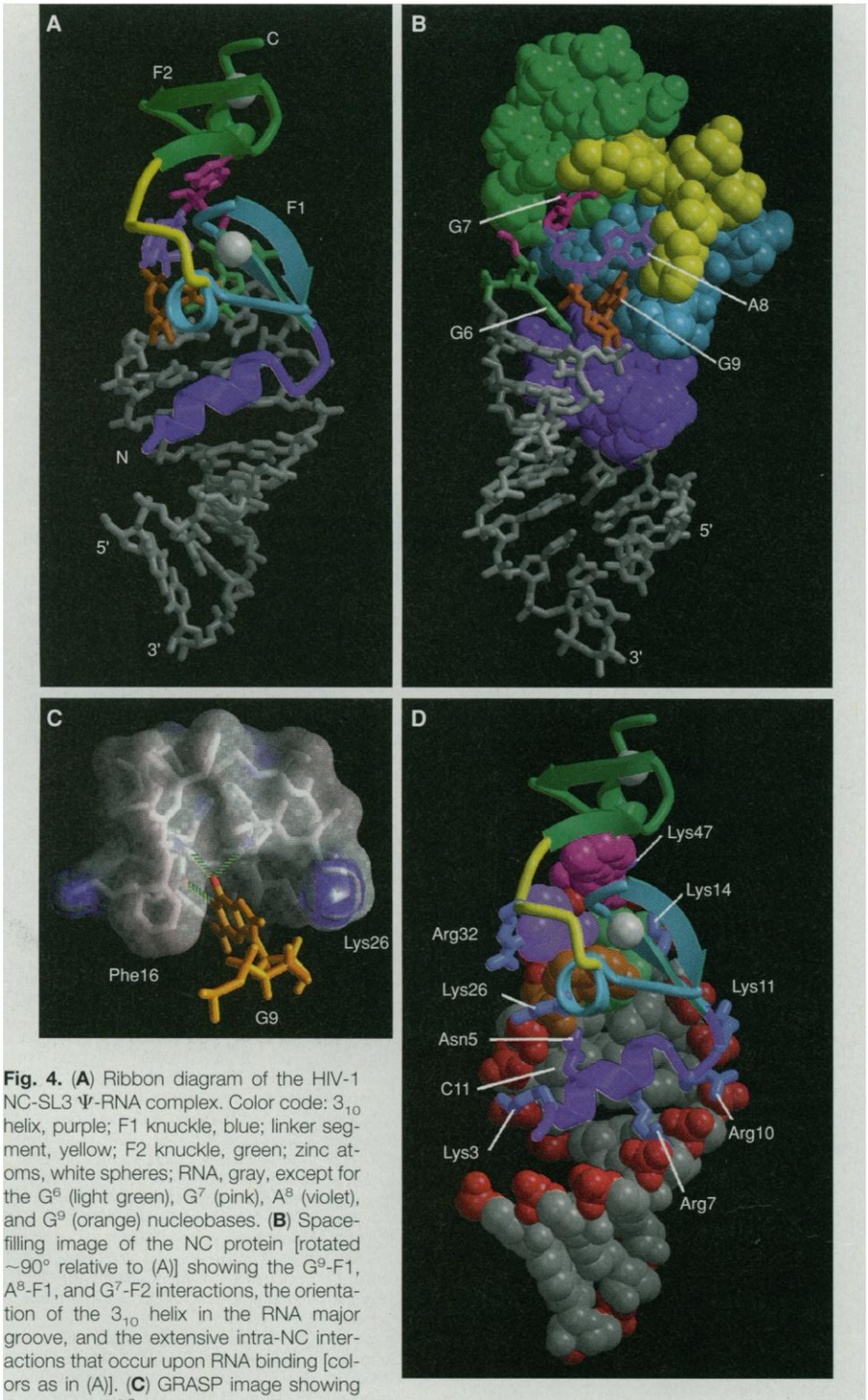

Fig. 4. (**A**) Ribbon diagram of the HIV-1 NC-SL3 Ψ-RNA complex. Color code: 3_{10} helix, purple; F1 knuckle, blue; linker segment, yellow; F2 knuckle, green; zinc atoms, white spheres; RNA, gray, except for the G^6 (light green), G^7 (pink), A^8 (violet), and G^9 (orange) nucleobases. (**B**) Space-filling image of the NC protein [rotated ~90° relative to (A)] showing the G^9-F1, A^8-F1, and G^7-F2 interactions, the orientation of the 3_{10} helix in the RNA major groove, and the extensive intra-NC interactions that occur upon RNA binding [colors as in (A)]. (**C**) GRASP image showing the nature of G^9 nucleobase binding to the hydrophobic cleft of the F1 knuckle (intermolecular hydrogen bonds are shown in green). G^7 binds the F2 knuckle in a similar manner. (**D**) Space-filling representation of the SL3 RNA in the NC-SL3 complex [same orientation as in (A)] showing relative proximities of conserved NC basic residues (blue) to the RNA phosphodiesters (red). Conserved residue Asn^5, which forms hydrogen bonds with C^{11}-NH_2, G^{10}-N7, and G^9-O2′, is shown in purple, and the hydrogen bond between conserved Arg^{32} and A^8 is also shown.

In general, retroviral genome recognition occurs in the cytosol before budding, and its mechanism is difficult to study and appears complex. Although SL3 alone is sufficient to direct packaging of heterologous RNAs (*9*), its deletion from the native genome does not fully abrogate packaging (*13*). Deletion of stem loops SL1, SL3, and SL1+SL3 reduces packaging to 19%, 12%, and 5%, respectively, of that found in the wild type (*13*). Also, isolated SL1, SL3, and SL4 RNAs have affinities of ~100 to 200 nM for the NC protein (compared with ~50 nM for the intact Ψ-site) (*12*). Thus, it is likely that in vivo packaging involves more than one gag polyprotein, and in this regard, the inherent flexibility of NC may permit binding of different gag proteins to the other stem loops through different subsets of inter- and intramolecular interactions.

In summary, the NC-SL3 structure provides a rationale for the high conservation of more than 50% of the amino acids in NC, explains the available mutagenesis data, and reveals molecular-level details associated with HIV-1 genome recognition. The NC protein plays essential roles in both early and late stages of the viral replication cycle and is thus an attractive antiviral target. The mutationally intolerant CCHC domains of the HIV-1 NC protein are susceptible to attack by antiviral agents that eject zinc from the zinc knuckles (*34*), at least two of which are undergoing clinical trials for the treatment of acquired immunodeficiency syndrome (*35*). Our studies provide the basis for an alternative rational drug design strategy that involves the development of inhibitors that interfere with genome recognition and packaging by competing with the NC or RNA binding sites.

REFERENCES AND NOTES

1. C. Dickson, R. Eisenman, R. Fan, E. Hunter, N. Teich, in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1984), vol. 1, pp. 513–648.
2. R. J. Mervis *et al.*, *J. Virol.* **62**, 3993 (1988); L. E. Henderson *et al.*, *ibid.* **66**, 1856 (1992).

3. Sometimes referred to as CCHC box or CCHC zinc finger–like arrays [L. E. Henderson, T. D. Copeland, R. C. Sowder, G. W. Smythers, S. Oroszlan, *J. Biol. Chem.* **256**, 8400 (1981); J. M. Berg, *Science* **232**, 485 (1986); M. F. Summers, T. L. South, B. Kim, D. R. Hare, *Biochemistry* **29**, 329 (1990)].
4. R. J. Gorelick, L. E. Henderson, J. P. Hanser, A. Rein, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8420 (1988); R. J. Gorelick *et al.*, *J. Virol.* **64**, 3207 (1990); R. J. Gorelick, D. J. Chabot, R. Alan, L. E. Henderson, L. O. Arthur, *ibid.* **67**, 4027 (1993); J. E. Jentoft, L. M. Smith, X. Fu, M. Johnson, J. Leis, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7094 (1988); P. Dupraz, S. Oertle, C. Méric, P. Damay, P.-F. Spahr, *J. Virol.* **64**, 4978 (1990); C. Méric, E. Gouilloud, P.-F. Spahr, *ibid.* **62**, 3328 (1988).
5. C. Méric and S. P. Goff, *J. Virol.* **63**, 1558 (1989).
6. R. D. Berkowitz, A. Ohagen, S. Hoglund, S. P. Goff, *ibid.* **69**, 6445 (1995).
7. A. M. L. Lever, H. G. Göttlinger, W. A. Haseltine, J. G. Sodroski, *ibid.* **63**, 4085 (1989); A. Aldovini and R. A. Young, *ibid.* **64**, 1920 (1990); F. Clavel and J. M. Orenstein, *ibid.*, p. 5230; M. Poznansky, A. M. L. Lever, L. Bergeron, W. Haseltine, J. Sodroski, *ibid.* **65**, 532 (1991).
8. G. P. Harrison and A. M. L. Lever, *ibid.* **66**, 4144 (1992).
9. T. Hayashi, T. Shioda, Y. Iwakura, H. Shibuta, *Virology* **188**, 590 (1992).
10. T. Hayashi, Y. Ueno, T. Okamoto, *FEBS Lett.* **327**, 213 (1993).
11. K. Sakaguchi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5219 (1993).
12. J. Clever, C. Sassetti, T. G. Parslow, *J. Virol.* **69**, 2101 (1995); M. S. McBride and A. T. Panganiban, *ibid.* **70**, 2963 (1996).
13. J. L. Clever and T. G. Parslow, *ibid.* **71**, 3407 (1997).
14. J. Luban and S. P. Goff, *ibid.* **68**, 3784 (1994).
15. The RNA used in our studies contains self-complementary (5′)-G^{-3}-G^{-2}-A^{-1} and U^{+1}-C^{+2}-C^{+3}-(3′) nucleotides appended at the 5′- and 3′-termini, respectively, of the 14-nucleotide SL3 stem loop (Fig. 1) for high-yield synthesis by in vitro transcription [J. F. Milligan and O. C. Uhlenbeck, *Methods Enzymol.* **180**, 51 (1989)] with T7 RNA polymerase [J. Grodberg and J. J. Dunn, *J. Bacteriol.* **170**, 1245 (1988)] and to stabilize the stem. RNA oligonucleotides were prepared and purified to single-nucleotide resolution by gel electrophoresis [J. R. Wyatt, M. Chastain, J. D. Puglisi, *Biotechniques* **11**, 764 (1991)]. Yields varied from 110 to 140 optical density units (ε_{260} = 218.1 mM^{-1}) per 30-ml transcription reaction. Uniformly ^{15}N,^{13}C-labeled nucleotide triphosphates used for the synthesis of isotopically labeled RNA were prepared from *Methylophilus methylotrophus* grown with [^{15}N]ammonium sulfate and [^{13}C]methanol and purified as described [R. T. Batey, J. L. Battiste, J. R. Williamson, *Methods Enzymol.* **261**, 300 (1995)].
16. The NC coding region of HIV-1_{NL4-3} [A. Adachi *et al.*, *J. Virol.* **59**, 284 (1986)] was polymerase chain reaction–amplified and subcloned into the Nde I–Bam HI sites of pET-3a (Novagen). The resulting construct, pRD2, was transformed into *E. coli* BL21(DE3)pLysE for overexpression of the NC protein. The protein was purified by SP-Sepharose and Sephadex G-50 column chromatographies [adapted from Z. Ji, G. J. Klarmann, B. D. Preston, *Biochemistry* **35**, 132 (1996); J. C. You and C. S. McHenry, *J. Biol. Chem.* **268**, 16519 (1993)]. Uniformly ^{15}N- and ^{15}N,^{13}C-labeled protein was obtained by growth in minimal media containing [^{13}C]glucose or $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively (yield = 10 to 13 mg per liter of rich media and 5 to 7 mg per liter in minimal media). Sample purity was confirmed by SDS–polyacrylamide gel electrophoresis and mass spectrometry (M_r of apoprotein, observed = 6369 ± 2, calculated = 6369; M_r of Zn-bound protein, observed = 6501 ± 2, calculated = 6500).
17. The NC:RNA stoichiometry was determined to be 1:1 (and not 2:2, for example, for two proteins bound to an RNA duplex) by gel filtration chromatography. The free protein (M_r = 6500), free acid RNA (M_r = 6678), and NC-RNA complex elute with apparent molecular weights of 8.0, 7.6, and 13.7 kD, respectively.
18. The equilibrium constant (K_d = [NC][RNA]/[NC–RNA] ~ 100 nM) was estimated by native gel electrophoresis (15% acrylamide), performed using dilutions of an NMR-characterized NC-SL3 sample. Equilibrium concentrations of free RNA were determined by comparison of band intensities with those of known concentrations of SL3 RNA.
19. NMR data were collected with Bruker AVANCE (800 MHz), DMX (600 MHz), and GE Omega PSG (600 MHz) NMR spectrometers for NC-RNA samples (0.8 to 1.7 mM) containing unlabeled NC and RNA, ^{13}C,^{15}N-labeled NC bound to unlabeled RNA, and unlabeled NC bound to ^{13}C,^{15}N-labeled RNA, in D_2O and 90% H_2O–10% D_2O containing 25 mM NaCl, 0.1 mM $ZnCl_2$, 0.1 mM β-mercaptoethanol, and 25 mM acetate-d_3 buffer (pH 6.5). Stoichiometric binding was monitored by ^1H 1D and 2D homonuclear correlated NMR. NMR data were processed with Felix (Molecular Simulations Inc.) or NMRPipe [F. Delaglio *et al.*, *J. Biomol. NMR* **6**, 277 (1995)] and analyzed with NMRVIEW [B. A. Johnson and R. A. Blevins, *ibid.* **4**, 603 (1994)]. Water signal suppression was achieved with flip-back pulses [S. Grzesiek and A. Bax, *J. Am. Chem. Soc.* **115**, 12593 (1993)], pulsed-field gradients [M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **2**, 661 (1992)], or presaturation pulses during the relaxation delay. Signal assignments were made with data derived from 3D HNCA and HN(CO)CA [S. Grzesiek and A. Bax, *J. Magn. Reson.* **96**, 432 (1992)], 3D CBCA(CO)NH [S. Grzesiek and A. Bax, *J. Am. Chem. Soc.* **114**, 6291 (1992)], 3D C(CO)NH [S. Grzesiek, J. Angliser, A. Bax, *J. Magn. Reson.* **101**, 114 (1993)], and 3D HCCH correlated spectroscopy (COSY) [M. Ikura, L. E. Kay, A. Bax, *J. Biomol. NMR* **1**, 299 (1991)]; 3D HCCH total correlation spectroscopy (TOCSY) [A. Bax, G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* **88**, 425 (1990)]; 3D ^{15}N-edited TOCSY (*36*) obtained with a 75-ms clean-MLEV-17 mixing period [C. Griesinger, G. Otting, K. Wüthrich, R. R. Ernst, *J. Am. Chem. Soc.* **110**, 7870 (1988)] and sensitivity-improved gradient coherence selection [O. Zhang, L. E. Kay, J. P. Olivier, J. D. Forman-Kay, *J. Biomol. NMR* **4**, 845 (1994)]; 2D rotating frame Overhauser effect (ROESY) [A. Bax and D. G. Davis, *J. Magn. Reson.* **63**, 207 (1985)] (120 ms, 6.0-kHz continuous-wave spin-lock mixing pulse); and constant-time heteronuclear single-quantum coherence (HSQC) [G. W. Vuister and A. Bax, *J. Magn. Reson.* **98**, 428 (1992)]. NOE data (mixing time τ_m = 50 to 200 ms) were obtained from 2D NOESY [J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **71**, 4546 (1979); S. Macura and R. R. Ernst, *Mol. Phys.* **41**, 95 (1980)], 3D ^{15}N-edited NOESY-HSQC (*36*), 3D ^{15}N,^{15}N-edited heteronuclear- multiple-quantum coherence (HMQC)–NOESY-HSQC [M. Ikura, A. Bax, G. M. Clore, A. M. Gronenborn, *J. Am. Chem. Soc.* **112**, 9020 (1990)], 3D ^{13}C-edited HMQC-NOESY [S. W. Fesik and E. R. P. Zuiderweg, *J. Magn. Reson.* **78**, 588 (1988)], and 4D ^{13}C,^{15}N-edited HMQC-NOESY-HSQC [L. E. Kay, G. M. Clore, A. Bax, A. M. Gronenborn, *Science* **249**, 411 (1990); D. R. Muhandiram, G. Y. Xu, L. E. Kay, *J. Biomol. NMR* **3**, 463 (1993)] data. Intermolecular NOEs were confirmed from ^{13}C-filtered,^{12}C-detected gradient-purged 3D HMQC-NOESY data [M. Ikura and A. Bax, *J. Am. Chem. Soc.* **114**, 2433 (1992); W. Lee, M. J. Revington, C. Arrowsmith, L. E. Kay, *FEBS Lett.* **350**, 87 (1994)].
20. Structures were calculated with the program DYANA [P. Güntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **273**, 283 (1997)]. Secondary structure elements were identified by analysis of NOE cross-peak patterns [K. Wuthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986)]. Upper-limit distance restraints of 2.7, 3.3, and 5.0 Å (with 0.5 Å added for NOEs involving methyl protons)—corresponding to strong-, medium-, and weak-intensity NOE cross-peaks, respectively—were used for both the NC protein and RNA. Loose torsion angle restraints were used for the RNA phosphodiester groups, which allowed sampling of both A- and B-helical conformations but prevented excessive fraying [D. S. Wutke, M. P. Foster, D. A. Case, J. M. Gottesfeld, P. E. Wright, *J. Mol. Biol.* **273**, 183 (1997)]. Only functional restraints were included in the calculations (for example, in cases where a proton exhibited NOEs of different intensities to geminal methylene protons, restraints were used only for the stronger of the two cross-peaks). Convergence statistics were calculated with MOLMOL [R. Koradi, M. Billeter, K. Wuthrich, *J. Mol. Graphics* **14**, 51 (1996)]. Color images were generated with MidasPlus [T. E. Ferrin, C. C. Huang, L. E. Javis, R. Langridge, *ibid.* **6**, 13 (1988)], Molscript [P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991)], GRASP (A. Nicholls, *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University), and Raster-3D [D. J. Bacon and W. F. Anderson, *J. Mol. Graphics* **6**, 219 (1988)].
21. The guanosine binding mode is essentially identical to that observed previously for a complex formed between an 18-residue F1 peptide and the single-stranded DNA oligonucleotide, d(ACGCC) [T. L. South and M. F. Summers, *Protein Sci.* **2**, 1 (1993)].
22. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
23. M. P. Foster *et al.*, *Nature Struct. Biol.* **4**, 605 (1997).
24. G. Meyers *et al.*, *Human Retroviruses and AIDS* (Los Alamos National Laboratory, Los Alamos, NM, 1995).
25. M. Ottmann, C. Gabus, J. Darlix, *J. Gen. Virol.* **69**, 1778 (1995).
26. ^{15}N NMR relaxation studies indicate that, in the absence of RNA, the NH_2- and COOH-terminal tails and the linker segment of NC are flexible. The two knuckles exhibit different rotational correlation times and thus do not form a single globular domain (M. Summers, unpublished data). Weak interfinger NOEs observed in the absence of RNA [N. Morellet *et al.*, *EMBO J.* **11**, 3059 (1992)] are consistent with an equilibrium between transiently interacting and noninteracting knuckle domains.
27. A. Ramos, C. C. Gubser, G. Varani, *Curr. Opin. Struct. Biol.* **7**, 317 (1997).
28. J. L. Battiste *et al.*, *Science* **273**, 1547 (1996); X. Ye, A. G. Gorin, A. D. Ellington, D. J. Patel, *Nature Struct. Biol.* **3**, 1026 (1996).
29. J. D. Puglisi, L. Chen, S. Blanchard, A. D. Frankel, *Science* **270**, 1200 (1995); X. Ye, R. A. Kumar, D. J. Patel, *Chem. Biol.* **2**, 827 (1995); J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, *Science* **257**, 76 (1992).
30. C. Ouvridge, N. Ito, P. R. Evans, C.-H. Teo, K. Nagai, *Nature* **372**, 432 (1994); F. H.-T. Allain *et al.*, *ibid.* **380**, 646 (1996).
31. H. A. Heus and A. Pardi, *Science* **253**, 191 (1991); U. Schmitz *et al.*, *RNA* **2**, 1213 (1996).
32. K. Valegard, J. B. Murray, P. G. Stockley, N. J. Stonehouse, L. Liljas, *Nature* **371**, 623 (1994).
33. L. Pappalardo and P. N. Borer, in preparation.
34. W. G. Rice *et al.*, *Nature* **361**, 473 (1993); W. G. Rice *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9721 (1993); X. Yu *et al.*, *Chem. Res. Toxicol.* **8**, 586 (1995); W. G. Rice *et al.*, *Science* **270**, 1194 (1995); W. G. Rice and J. A. Turpin, *Rev. Med. Virol.* **6**, 187 (1996); W. G. Rice *et al.*, *Nature Med.* **3**, 341 (1997).
35. N. B. McDonnell, R. N. De Guzman, W. G. Rice, J. A. Turpin, M. F. Summers, *J. Med. Chem.* **40**, 1969 (1997); M. Vandevelde *et al.*, *AIDS Res. Hum. Retroviruses* **12**, 567 (1996).
36. D. Marion *et al.*, *Biochemistry* **28**, 6150 (1989).
37. Supported by NIH grants GM42561 (M.F.S.) and GM32691 (P.N.B.). C.C.S. is an Undergraduate Meyerhoff Scholar at UMBC and L.P. is supported by a Syracuse University Graduate Scholarship. We thank B. M. Lee for collecting some of the NMR data, R. Wall for assistance in RNA purification, and R. Edwards and M. Zawrotny for technical support.

23 October 1997; accepted 11 December 1997