Composition and Sequence-Dependent Binding of RNA to the Nucleocapsid Protein of Moloney Murine Leukemia Virus[†,‡]

Anwesha Dey, Danielle York, Adjoa Smalls-Mantey, and Michael F. Summers*

Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250

Received November 6, 2004; Revised Manuscript Received January 5, 2005

ABSTRACT: All retroviruses package two copies of their genomes during virus assembly, both of which are required for strand transfer-mediated recombination during reverse transcription. Genome packaging is mediated by interactions between the nucleocapsid (NC) domains of assembling Gag polyproteins and an RNA packaging signal, located near the 5′ end of the genome, called Ψ. We recently discovered that the NC protein of the Moloney murine leukemia virus (MLV) can bind with high affinity to conserved UCUG elements within the MLV packaging signal [D'Souza, V., and Summers, M. F. (2004) *Nature 431,* 586−590]. Selective binding to dimeric RNA is regulated by a conformational RNA switch, in which the UCUG elements are sequestered by base pairing in the monomeric RNA and do not bind NC, but become exposed for NC binding upon dimerization. Dimerization-dependent structural changes occur in other regions of the Ψ-site, exposing guanosine-containing segments that might also bind NC. Here we demonstrate that short RNAs containing three such sequences, ACAG, UUUG, and UCCG, can bind NC with significant affinity ($K_d = 94-315$ nM). Titration experiments with oligoribonucleotides of varying lengths and compositions, combined with NMR-based structural studies, reveal that binding is strictly dependent on the presence of an unpaired guanosine, and that relative binding affinities can vary by more than 1 order of magnitude depending on the nature of the three upstream nucleotides. Binding is enhanced in short RNAs containing terminal phosphates, indicating that electrostatic interactions contribute significantly to binding. Our findings extend a previously published model for genome recognition, in which the NC domains of assembling Gag molecules interact with multiple $X_{(i-3)}$-$X_{(i-2)}$-$X_{(i-1)}$-$G_{(i)}$ elements (X is a variable nucleotide) that appear to be preferentially exposed in the dimeric RNA.

As retroviruses assemble in infected cells, two copies of their full-length, 5′-capped, and 3′-polyadenylated genomes are selected from a cytosolic pool that contains a substantial excess (∼100-fold) of cellular and spliced viral mRNAs (*1*). Two RNA molecules are utilized during reverse transcription, enabling the virus to overcome otherwise deleterious breaks in the RNA strands and promoting genetic diversity and the evolution of drug resistant strains (*2*−*7*). Genome packaging is mediated by interactions between the nucleocapsid (NC)[1] domains of the assembling Gag polyproteins and a segment of the viral genome called the Ψ-site (*1, 8*−*11*). RNA packaging elements are generally located near the 5′ end of the genome, between the primer binding site (PBS) and Gag initiation codon (*12*−*23*). The Ψ-site often overlaps with the major splice donor site, providing a potential mechanism for discriminating between spliced and unspliced viral RNAs (*13*, *24*). The Ψ-sites also generally overlap with sequences that are important for RNA dimerization, which led to the suggestion that genome packaging and dimerization events may be closely coupled (*1*, *8*, *25*−*28*).

Much of what is currently known about retroviral genome packaging has been obtained from studies of the Moloney murine leukemia virus (MLV). MLV is a simple retrovirus that has been widely used in human gene therapy trials, and has recently been employed as a vector for the treatment of severe combined immunodeficiency (SCID) (*29*, *30*). Understanding the mechanism of RNA selection and packaging is important not only for the development of new antiviral strategies but also for facilitation of the design of more effective gene delivery vectors with higher viral titers (*31*). Our laboratory has focused on an ∼100-nucleotide component of the Ψ-site known as the "core encapsidation signal" (Ψ^{CES}), a segment that is independently capable of promoting both RNA dimerization in vitro and encapsidation in vivo (*32*−*34*). Ψ^{CES} contains three stem−loop motifs, one of which (DIS-2) promotes dimerization via duplex formation, and the other two (SL-C and SL-D) are capable of forming "kissing dimers" via intermolecular base pairing of their conserved GACG tetraloop nucleotides (*35*, *36*). DIS-2

[†] This research was supported by NIH Grant GM42561 and MARC U*STAR NIH Undergraduate Training Grant GM08663. D.Y. and A.S.-M. are MARC U*STAR and HHMI Fellow/Meyerhoff undergraduate scholars, respectively.

[‡] Atomic coordinates have been deposited as Protein Data Bank entries 1WWD, 1WWE, 1WWF, and 1WWG.

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

[1] Abbreviations: 2D, two-dimensional; BME, β-mercaptoethanol; HIV-1, human immunodeficiency virus type 1; ITC, isothermal titration calorimetry; MLV, murine leukemia virus; NC, nucleocapsid protein; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE correlated spectroscopy; PBS, primer binding site; SCID, severe combined immunodeficiency; TOCSY, total correlation spectroscopy.

10.1021/bi047639q CCC: $30.25
Published on Web 02/18/2005

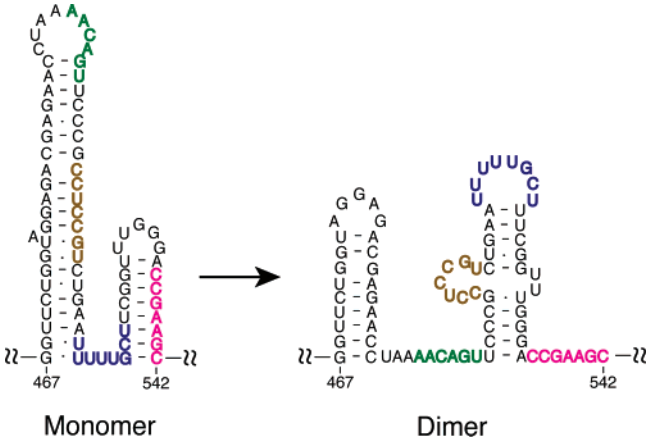

FIGURE 1: Secondary structure of a portion of the MLV Ψ-RNA packaging signal proposed to undergo a conformational change upon dimerization, based upon in vitro studies of the intact Ψ-site (*42*). Nucleotides are numbered using the first residue after the 5′ cap as position 1.

undergoes a register shift in base pairing upon dimerization, exposing a conserved UCUG element that binds NC with high affinity (*34*). This element is sequestered by base pairing in the monomeric RNA and does not bind NC. An additional pseudopalindromic segment of the MLV Ψ-site was identified recently that also promotes dimerization in vitro and genome packaging in vivo (DIS-1) (*37*, *38*), and this segment was similarly shown to undergo a dimerization-dependent conformational change that exposes a conserved NC-binding UCUG element in the dimer (*34*). Structural studies revealed that, for NC complexes with both DIS-1 and DIS-2, all four nucleotides of the UCUG element interact with the single CCHC zinc knuckle of NC. In both cases, the single guanosine is inserted into a hydrophobic cleft in a manner similar to that previously observed in HIV-1 NC−RNA complexes (*39*−*41*). On the basis of these findings, we proposed that genome packaging is regulated by a structural RNA switch mechanism, in which multiple protein binding sites are sequestered by base pairing in the monomeric RNA and become exposed upon dimerization to promote the specific packaging of a diploid genome (*34*).

The segment between the PBS and Gag initiation codon of the gammaretroviruses (which includes MLV) is enriched in UCUG sequences, which occur at a frequency of approximately one in 50 nucleotides (compared to one in 225 nucleotides in the coding and LTR regions of the genome) (*34*). This raised the possibility that the MLV Ψ-site might contain additional NC binding sites. Chemical accessibility mapping experiments indicated that at least three additional guanosine-containing segments of the Ψ-site with sequences similar to UCUG are sequestered in the monomeric RNA and become exposed upon dimerization, including AACAGU, CCUCCGU, UUUUGCU, and CCGAAGC sequences (Figure 1) (*42*, *43*). To determine if these related elements are capable of binding NC, isothermal titration calorimetry (ITC) and NMR-based structural studies have been conducted with a series of short oligoribonucleotides corresponding to these and related RNA sequences. Our findings reveal that NC can bind to a variety of $X_{(i-3)}$-$X_{(i-2)}$-$X_{(i-1)}$-$G_{(i)}$ sequences (X is any ribonucleotide), and that the relative affinity depends significantly on the nature of the X nucleotides. The binding modes of several oligonucleotides that were examined, including those that appear to be exposed in the dimeric Ψ-site, are surprisingly similar to that observed for the NC complex with UAUCUG, with the guanosine nucleobase inserted into a hydrophobic pocket, and the nucleotides at positions $i - 2$ and $i - 3$ packing against the exposed side chain of Tyr 28. The primary determinants of NC binding thus appear to involve specific binding of an exposed guanosine to the hydrophobic pocket of the zinc knuckle domain, and additional hydrophobic and electrostatic interactions involving the three upstream nucleotides and the adjacent phosphodiester groups. The strict requirement for an exposed guanosine, and the influence of the nature and number of upstream nucleotides on binding affinity, have implications regarding the mechanism of MLV genome recognition and packaging.

MATERIALS AND METHODS

Sample Preparation. MLV NC protein was prepared and purified as described previously (*32*, *33*). Gel-purified RNA constructs were obtained from Dharmacon Inc. (Lafayette, CO). All samples for NMR and ITC experiments were prepared in Tris-HCl buffer [10 mM Tris (pH 7.0), 10 mM NaCl, 0.1 mM $ZnCl_2$, and 0.1 mM β-mercaptoethanol].

Isothermal Titration Calorimetry. Dissociation equilibrium constants for MLV NC binding to different RNA constructs were determined by standard ITC methods using a VP-isothermal titration microcalorimeter (VP-ITC) (MicroCal Corp., Northampton, MA) (*44*). NC concentrations varied from 45 to 100 μM, and RNA concentrations varied from 3 to 5 μM, as determined from UV−vis absorption measurements. Exothermic heats of reaction were measured at 30 °C for 25 injections of NC into 1.4 mL of RNA, and heats of dilution were measured by titrating NC into a buffer under identical conditions. Baseline corrections were performed by subtracting heats of dilution from the raw NC−RNA titration data, and binding curves were analyzed and dissociation constants determined by nonlinear least-squares fitting of the baseline-corrected data (*33*). Binding data for p-UCUG exhibited saturation behavior near 1:1 NC:RNA stoichiometries (at concentrations that afforded significant heats of binding). For this RNA, the dissociation constant was measured at NaCl concentrations of 10, 30, and 60 mM, and the value at no added NaCl was determined by extrapolation from a log[NaCl] versus log(K_a) plot.

NMR Spectroscopy. Two-dimensional NOESY ($\tau_m = 120$ ms) (*45*, *46*) and TOCSY ($\tau_m = 70$ ms) (*47*−*49*) data were obtained for NC-bound AACAGU, CCUCCGU, CCGAAGC, and UUUUGCU RNA samples using Bruker AVANCE 800 MHz and DMX 600 MHz spectrometers (15 °C). ^{1}H NMR signals for these small RNAs were readily assigned using standard sequential assignment procedures (*50*). NMR signals of the NC protein were readily assigned by analysis of one-dimensional (1D) ^{1}H RNA titration and two-dimensional (2D) NOESY NMR data, and by comparisons with the previously assigned NMR spectra obtained for the free and UAUCUG-bound NC protein (*34*). Chemical shift changes observed upon titration of NC with AACAGU, CCUCCGU, and UUUUGCU were very similar to those observed previously for UAUCUG. Significantly broader ^{1}H NMR spectra were observed for the NC complex with CCGAAGC, which precluded its structure determination. It was not

possible to readily determine if the broadening was due to exchange involving multiple binding modes (e.g., structures with the different guanosines inserted into the guanosine binding pocket) or conformational heterogeneity about a single binding mode. NMR data were processed with NMRPipe/NMRDraw (*51*) and analyzed with NMRView (*52*).

Structure Calculations. Structures were calculated and refined with CYANA (*53*) by parallel processing with a 28-node Linux cluster. Interproton distance restraints for intermolecular NOEs were determined by qualitative assessment of cross-peak intensities in the 2D NOESY data. Upper-limit distance restraints of 2.7, 3.3, and 5.0 Å were employed for direct NOE cross-peaks of strong, medium, and weak intensity, respectively, except for NOEs associated with the intraresidue H8/6−H2′ (3.2 Å) and H8/6−H3′ (4.3 Å) proton pairs (*32*). Distance restraints for pseudoatoms were adjusted as described previously (*53*). All of the intramolecular NOE cross-peak patterns and intensities observed for NC in the RNA complexes matched those observed previously for the NC−UAUCUG complex, and for this reason, the intramolecular NOE and hydrogen bond restraints employed for the NC−UAUCUG structure determination were used here without modification.

RESULTS AND DISCUSSION

Interactions of NC with Short RNA Oligomers. Previous studies of NC binding to the 101-nucleotide Ψ^{CES} core encapsidation signal identified an unpaired UCUG segment that connects stem−loop motifs DIS-2 and SL-C as a high-affinity NC binding site (*34*). A prominent feature of the NC−Ψ^{CES} structure is the binding of G309 [here termed G(*i*)] to a hydrophobic pocket, with its N1H, NH21, and O6 groups forming hydrogen bonds with backbone NH and O atoms located at the back of the pocket. Similar interactions were observed in previous NC−RNA (*40*, *41*) and NC−DNA (*39*, *54*) structures. In addition, the U(i − 1) nucleotide packs against the Leu 21 and Ala 27 side chains and the C(i − 2) and U(i − 3) nucleotides pack against the side chain of Tyr 28, and several Arg and Lys side chains appear to participate in electrostatic and/or hydrogen bonding interactions with phosphodiesters, ribose oxygen, and nucleobase groups (*34*). To assess the contributions of these interactions to NC binding, ITC studies of NC binding to a series of short RNAs were performed. No heats of binding were detected for homooligomers of C, U, or A (Figure 2). Titration of GGGGGG with NC resulted in a significant exothermic binding enthalpy that reached saturation at an NC:RNA ratio of approximately 1:1. Under these conditions, it was not possible to accurately calculate the binding constant, and poor sensitivity precluded attempts to obtain ITC data at lower sample concentrations. The susceptibility of guanosine homooligomers to formation of quartet structures in the presence of salts (*55*) precluded an estimation of the NC binding constant from NaCl-dependent binding plots (as was possible for 5′-p-UCUG).

The above studies indicate that at least one exposed guanosine is required for tight NC binding. This finding was confirmed by studies of NC binding to a series of oligoribonucleotides containing only U, A, or C nucleotides and a single guanosine. As shown in Figure 3a, NC binds relatively

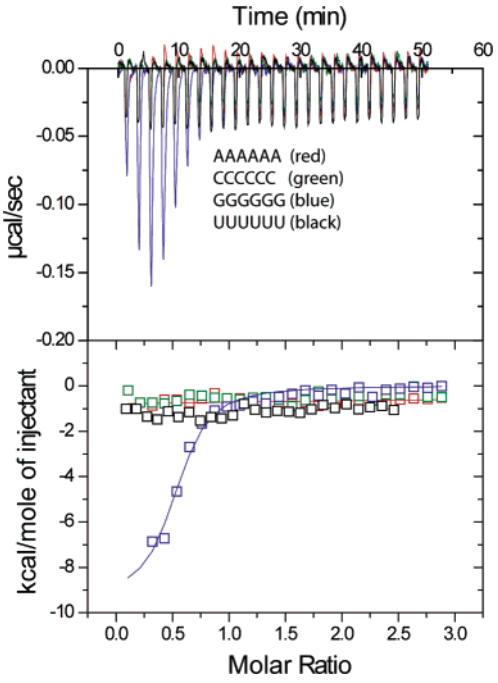

FIGURE 2: Representative ITC data obtained upon titration of homooligomeric RNAs (shown) with the MLV NC protein. Significant binding enthalpies were only observed for GGGGGG; however, the near-stoichiometric binding observed at useful concentrations, and the tendency of GGGGGG to form G-quartet structures in the presence of NaCl, precluded a quantitative determination of binding affinity and stoichiometry.

poorly to AAAAGA (K_d = 5.0 ± 0.6 μM), but exhibits increased affinity for CCCCGC (K_d = 620 ± 50 nM) and UUUUGU (K_d = 370 ± 40 nM). Thus, NC binding appears to be favored by oligonucleotides containing a single guanosine and upstream pyrimidines. Note, however, that the affinity of NC for UUUUGU is still more than 3 times weaker than for UAUCUG (K_d = 95 ± 25 nM; Table 1), which contains a single upstream adenosine.

To test for position dependence of the guanosine, binding studies were conducted with RNA constructs containing five uridines and a single guanosine (Figure 3b). Tightest NC binding was observed for UUUUUG and UUUUGU (K_d ∼ 400 nM) (Table 1). Constructs with fewer upstream nucleotides exhibited reduced affinities, with UUUGUU, UUGUUU, UGUUUU, and GUUUUU yielding dissociation constants of approximately 0.73, 1.5, 2.8, and 3.6 μM, respectively (Table 1). These findings indicate that the affinity for NC is considerably enhanced by the presence of at least three upstream nucleotides relative to the guanosine.

To further evaluate the determinants of NC binding, variants of the native UAUCUG linker were prepared for ITC measurements. As shown in Figure 4a (and summarized in Table 1), the UCUG RNA that lacks 3′- and 5′-phosphates binds to NC with relatively poor affinity (K_d = 7.8 ± 0.6 μM). Addition of the upstream A(i − 4) nucleotide results in an ∼30-fold increase in affinity (K_d = 250 ± 30 nM). An additional ∼2.5-fold increase in affinity is observed upon addition of the U(i − 5) nucleotide (K_d = 95 ± 25 nM), and a further ∼2-fold increase is obtained upon addition of the downstream U(i + 1) (K_d = 56 ± 13 nM). Electrostatic interactions appear to play a significant role in NC binding, since addition of a 3′-phosphate to UAUCUG led to a more substantial ∼2.5-fold increase in affinity (K_d = 38 ± 7 nM).

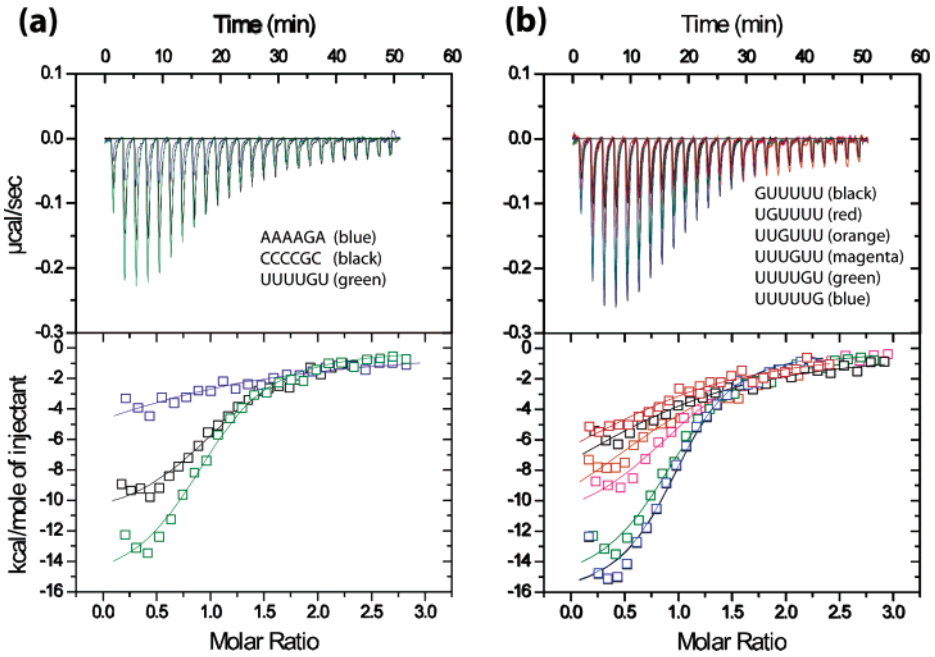

FIGURE 3: Representative ITC data obtained upon titration of short oligoribonucleotide homopolymers containing a single guanosine substitution with NC. (a) Dependence of binding affinity on nucleotide composition, which decreases in the following order: AAAAGA ($K_d = 5.0 \pm 0.6$ μM) < CCCCGC ($K_d = 620 \pm 50$ nM) < UUUUGU ($K_d = 370 \pm 40$ nM). (b) NC binding depends on the position of the guanosine in oligoribonucleotides containing only uridines and a single G nucleotide.

Table 1: Dissociation Equilibrium Data for NC−RNA Complexes[a]

RNA	K_d (nM)	*n*
AAAAAA	binding not detected	
CCCCCC	binding not detected	
UUUUUU	binding not detected	
GGGGGG	*b*	
AAAAGA	5000 ± 600	1.0
CCCCGC	620 ± 50	1.2
UUUUGU	370 ± 40	1.0
position of guanosine		
GUUUUU	3600 ± 800	1.0
UGUUUU	2800 ± 100	1.0
UUGUUU	1500 ± 400	1.1
UUUGUU	730 ± 30	1.0
UUUUGU	410 ± 40	1.0
UUUUUG	400 ± 30	1.1
context dependence of UCUG		
UCUG	7800 ± 600	1.0
AUCUG	250 ± 30	1.1
UAUCUG	95 ± 25	1.2
UAUCUGU	56 ± 13	1.1
UAUCUG-P	38 ± 7	1.0
P-UCUG	29 ± 16[c]	0.9
potential NC binding sites in Ψ		
AACAGU	250 ± 40	1.1
CCUCCGU	94 ± 24	1.0
UUUUGCU	170 ± 40	1.0

[a] Dissociation equilibrium constants determined by isothermal titration calorimetry and reported as the mean ± the standard deviation from two experiments. In cases where the standard deviation was less than the errors estimated from estimated uncertainties in the protein and RNA concentrations, error estimates based on the latter calculations are reported. The letter *n* denotes the number of NC binding sites calculated during data fitting. [b] Saturation binding at 1:1 NC:RNA ratios, and the tendency of the RNA to form G-quartets in the presence of added NaCl, precluded a quantitative determination of the dissociation equilibrium constant. [c] Determined by extrapolation from ITC data obtained in the presence of NaCl (10, 30, and 60 mM).

In fact, addition of a 5′-phosphate to UCUG resulted in a dramatic ~250-fold increase in affinity ($K_d = 29 \pm 16$ nM, as determined by extrapolation of NaCl-dependent binding data; Figure 4b). It thus appears that binding is facilitated by the specific binding of a guanosine to the guanosine pocket of the zinc knuckle, by interactions between the nucleotides at positions $i - 1$, $i - 2$, and $i - 3$ and the hydrophobic surface of the zinc knuckle [with $C_{(i-2)}$ preferred over other nucleotides], and by electrostatic interactions involving the phosphodiesters of the UCUG and adjacent nucleotides.

Interactions of NC with Ψ-Site Fragments. Chemical accessibility mapping studies (*42*, *43*) revealed that residues 468−542 of the MLV leader undergo a conformational change upon dimerization, with at least four short guanosine-containing segments being sequestered by base pairing in the monomeric RNA and exposed in the dimer (AACAG^{497}U, CCUCCG^{509}U, UUUUG^{521}CU, and CCG^{538}AAG^{541}C) (Figure 1). The ITC studies described above suggested to us that these segments might also be able to bind NC. As shown in Figure 5 and summarized in Table 1, short RNA oligomers containing these sequences are, in fact, capable of binding NC with moderate to high affinity. In particular, the affinity of the CCUCCGU construct (94 ± 24 nM) is essentially identical to that observed for UAUCUG (95 ± 25 nM). NC binding to these RNAs was subsequently monitored by 1D ^{1}H NMR spectroscopy, with AACAGU, UUUUGCU, and CCUCCGU exhibiting slow chemical exchange between free and NC-bound RNA species on the NMR chemical shift time scale (milliseconds). Relatively broad spectra were observed upon titration of CCGAAGC with NC, which is indicative of conformational averaging between multiple bound structures (possibly with the different guanosines bound to the guanosine-binding pocket). As such, further NMR studies with this RNA were not possible.

^{1}H NMR chemical shift changes observed for the binding of NC to AACAGU, UUUUGCU, and CCUCCGU were very similar to the changes observed upon binding to UAUCUG (*34*), and representative portions of 2D NOESY

Table 2: Structural Statistics for MLV NC−RNA Complexes

	AACAGU	CCUCCGU	UUUUGCU	UAUCUG
NMR-derived restraints				
no. of NC intramolecular restraints				
intraresidue	22	22	22	22
sequential ($\|i-j\|=1$)	42	42	42	42
medium-range ($\|i-j\|=2, 3$)	46	46	46	46
long-range ($\|i-j\|\geq 4$)	51	51	51	51
H-bond restraints (4 per H-bond)	40	40	40	40
no. of RNA restraints				
intraresidue	30	35	35	21
sequential	3	8	4	4
intermolecular NOE	37	37	35	42
intermolecular H-bond	12	12	12	44
total no. of restraints per refined residue	16.1	17.6	17.2	17.9
target function (Å^2)				
mean ± SD	0.14 ± 0.04	0.18 ± 0.02	0.14 ± 0.03	0.32 ± 0.07
minimum	0.09	0.11	0.07	0.14
maximum	0.21	0.21	0.19	0.43
restraint violations				
average sum upper distance violation (Å)	0.8 ± 0.2	0.9 ± 0.1	0.7 ± 0.2	1.0 ± 0.2
average maximum upper distance violation (Å)	0.09 ± 0.03	0.12 ± 0.05	0.13 ± 0.04	0.11 ± 0.05
average sum of VDW violations (Å)	1.5 ± 0.3	1.5 ± 0.2	1.1 ± 0.2	2.3 ± 0.3
average maximum VDW violation (Å)	0.11 ± 0.04	0.13 ± 0.04	0.12 ± 0.03	0.27 ± 0.09
average sum of lower distance violations (Å)	0.2 ± 0.1	0.3 ± 0.1	0.2 ± 0.1	0.4 ± 0.1
average maximum lower distance violation (Å)	0.05 ± 0.02	0.06 ± 0.02	0.05 ± 0.01	0.08 ± 0.02
structure convergence (Å)				
NC (Leu 21−Pro 43)				
main chain atoms	0.34 ± 0.12	0.38 ± 0.11	0.26 ± 0.06	0.32 ± 0.11
all heavy atoms	0.74 ± 0.09	0.86 ± 0.12	0.73 ± 0.07	0.74 ± 0.10
RNA				
residues $X_{(i-3)}$-$G_{(i)}$	0.74 ± 0.17	0.77 ± 0.13	0.94 ± 0.25	0.79 ± 0.14
all RNA residues	1.21 ± 0.27	1.66 ± 0.37	2.22 ± 0.51	3.40 ± 0.57

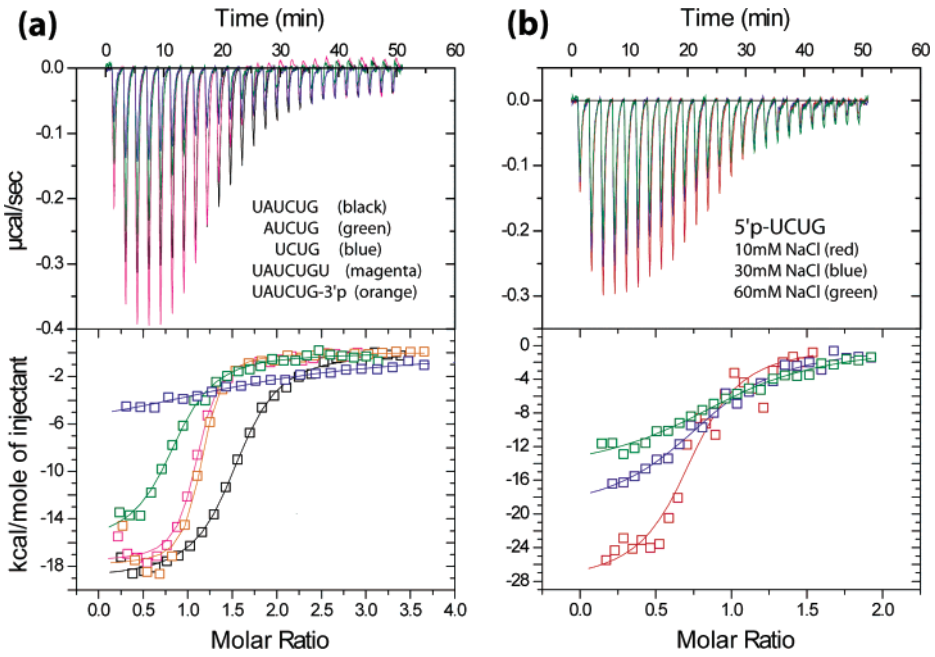

FIGURE 4: (a) Nucleotide extensions proximal and distal to the native NC-binding UCUG element are important for high-affinity binding (see Table 1 for a summary of dissociation constants). Addition of a non-native phosphate at the 3′ position significantly enhances binding. (b) ITC data obtained in the presence of added NaCl (concentrations shown) and used to estimate the dissociation constant of 5′-p-UCUG. NC binding is clearly influenced by electrostatic effects.

data obtained for these complexes are given in Figure 6. In all cases, the $G_{(i)}$ H8 proton exhibited intermolecular NOEs to the Trp 35 aromatic, Ala 27 methyl, Ala 36 methyl, and Leu 21 side chain protons, and the $G_{(i)}$ H1′ proton exhibited intermolecular NOEs to the side chain protons of Leu 21, Trp 35, and Arg 23. For all residues at position $i-1$, intermolecular NOEs were observed between the nucleobase H8/6 protons and the side chain protons of Ala 27, Leu 21, and Arg 18. All residues at position $i-2$ exhibit strong intensity nucleobase H8/6 NOEs with the aromatic protons of Tyr 28, as well as moderate to weak intensity NOEs with the side chain protons of Ala 27, Ala 36, and Lys 42. Finally, residues at position $i-3$ exhibit strong nucleobase and ribose proton NOEs with the side chain protons of Tyr 28, as well as weak intermolecular NOEs with the side chain of Lys 30. All RNAs except AACAGU also exhibited weak

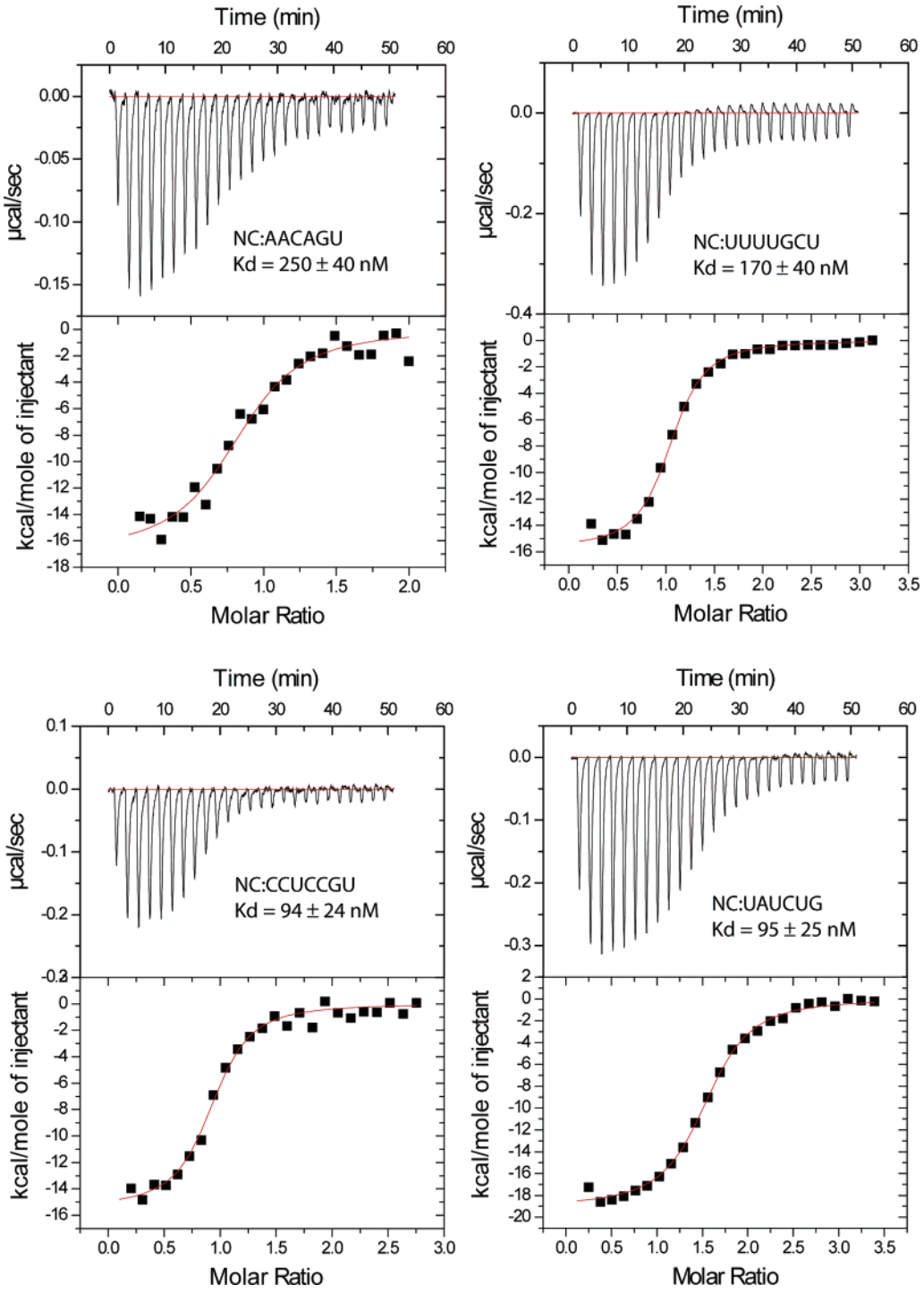

FIGURE 5: Representative ITC data obtained upon NC titration of short, G-containing segments of the MLV Ψ-site that are believed to be sequestered by base pairing in the monomeric genome and exposed in the dimer (*42*). All elements bind NC with significant affinity, although not as tightly as the conserved and unusually abundant UCUG (see Table 1 for dissociation constants).

NOEs between the $i - 3$ nucleobase and Lys 41 side chain protons. For AACAGU, the $A_{(i-3)}$ H2 proton exhibited intermolecular NOEs with the Tyr 28 aromatic, Cys 29 Hα, and Lys 30 side chain protons. Nucleotides at position $i + 1$ exhibited very weak NOEs with the Trp 35 aromatic protons, as well as relatively weak internucleotide NOEs, suggesting that these residues may interact transiently with the protein and probably do not adopt a single, well-defined conformation. Nucleotides that precede residue $i - 3$ also exhibited weak internucleotide NOEs and also are not conformationally defined by the NMR data. Overall, the NOE cross-peak patterns and intensities observed for these RNAs are very similar to those observed previously upon NC binding to UAUCUG (*34*).

In view of the similarities in the NOE cross-peak patterns and intensities observed for the NC complexes with AACAGU, UUUUGCU, CCUCCGU, and UAUCUG, it is not surprising that the NOE-based structure calculations afforded very similar three-dimensional structures (Figures 7 and 8). In all cases, the nucleobase of the guanosine at position *i* is inserted into a hydrophobic cleft formed by the side chains of Trp 35, Ala 27, Ala 36, and Leu 21; the

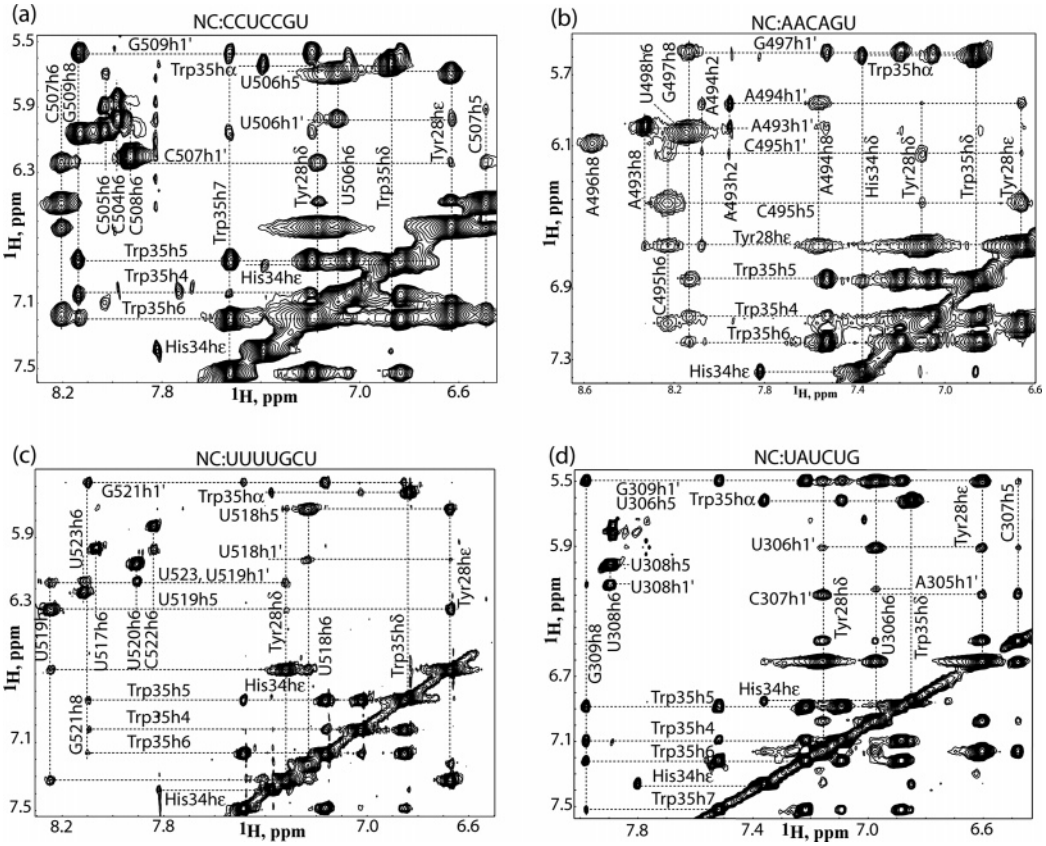

FIGURE 6: (a−c) Portions of 2D NOESY data obtained for NC−RNA complexes. The RNAs (labeled) correspond to segments of the MLV Ψ-site that bind NC with significant affinity and are believed to be exposed in the dimeric (but not the monomeric) genome (Figure 1). For comparison, the relevant portion of the 2D NOESY spectrum obtained previously for the NC complex with UAUCUG is shown in panel d.

nucleobase at position $i - 1$ packs against Ala 27 and Leu 21, and the nucleobases at positions $i - 2$ and $i - 3$ pack against the side chain of Tyr 28. Together with the ITC results, these structural data indicate that the following factors are important for RNA binding to NC. (1) An exposed guanosine is essential for tight binding. In all cases examined thus far, the guanosine nucleobase binds to the hydrophobic guanosine-binding pocket by a combination of hydrophobic and hydrogen bonding interactions (Figure 8). (2) The ribose and nucleobase moieties of the three upstream nucleotides make contacts with hydrophobic residues on the surface of the zinc knuckle. (3) Affinity is increased further by electrostatic interactions with the phosphodiester groups immediately upstream and downstream of the $X_{(i-3)}$-$X_{(i-2)}$-$X_{(i-1)}$-$G_{(i)}$ segment. (4) NC binding is sensitive to the position of the guanosine and nature of the X residues of $X_{(i-4)}$-$X_{(i-3)}$-$X_{(i-2)}$-$X_{(i-1)}$-$G_{(i)}$-$X_{(i+1)}$ sequences, with the tightest binding observed for the $X_{(-4)}$-UCUG-$X_{(+1)}$ and the weakest binding observed for the AAAAGA sequence.

Implications for Genome Recognition. Considerable effort has been made to identify the molecular determinants and mechanisms of retroviral genome recognition and packaging, with MLV receiving much attention (see refs *1*, *8*−*11*, *28*, and *56*−*58* and references therein). Early studies demonstrated that a 350-nucleotide segment of the viral RNA, located just downstream of the *env* splice donor site (nucleotides ∼210−560), is essential for genome packaging (*59*). This segment, which is commonly termed the Ψ-site, can be relocated to the 3′ end of the genome without severely affecting packaging, suggesting that Ψ functions as an independent packaging element (*13*). Fragments of Ψ as small as the ∼100-nucleotide core encapsidation signal can independently direct the packaging of heterologous RNAs into viruslike particles (*21*), although packaging efficiency is diminished relative to that of the full-length Ψ-site. Our recent studies indicate that NC can bind with high affinity to a UCUG element within the Ψ^{CES} RNA, and that binding is regulated by a conformational RNA switch that sequesters the UCUG element in the monomeric RNA and exposes it in the dimer (*34*). The RNA conformational change results from a register shift within two pseudopalindromic stem−loop motifs (DIS-1 and DIS-2) that re-optimizes base pairing upon dimerization. Earlier chemical accessibility mapping and free energy calculations indicated that these structural changes also occur in the intact Ψ-site (*42*, *43*). Residues 467−542 of the Ψ-site were also proposed to change conformation upon dimerization in a manner that exposes four additional short, guanosine-containing segments. Here, we have shown that short, unstructured oligoribonucleotides with sequences of three of these segments are capable of binding NC with significant affinity, and that all bind to the CCHC zinc knuckle in a manner that is very similar to that observed upon binding to UCUG-containing RNAs.

Perhaps the most biologically significant finding from this study is that tight binding by the MLV nucleocapsid protein does not explicitly require a UCUG sequence. Related sequences can bind NC with affinities approaching that observed for UCUG, depending on the nature of the residues upstream of the guanosine. This observation significantly increases the number of sites within the MLV genome that

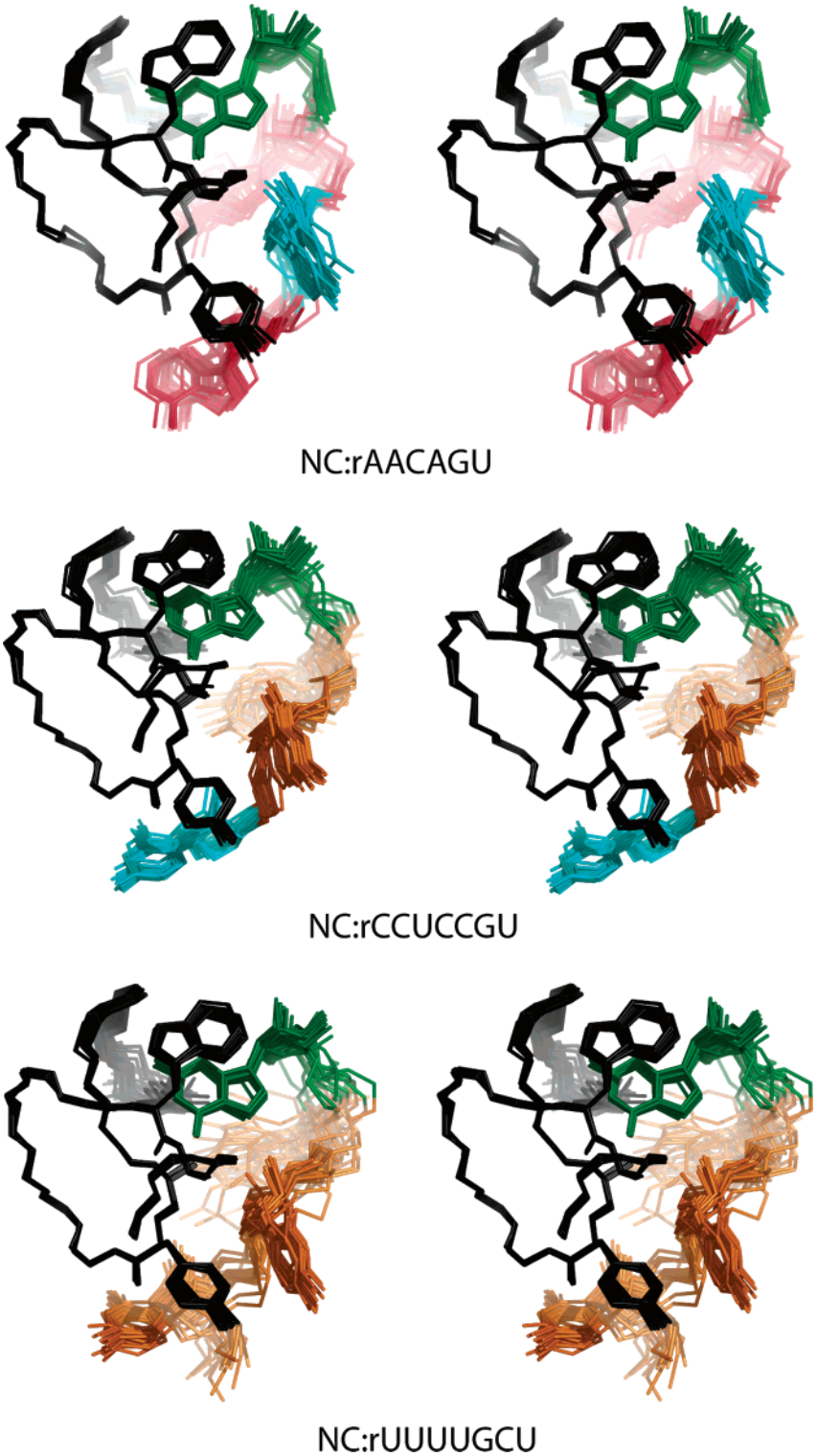

FIGURE 7: NMR structures determined for NC complexes with AACAGU, CCUCCGU, and UUUUGCU. This figure was generated by superpositioning the backbone heavy atoms of NC residues Leu 21−Pro 43.

can potentially function in NC binding. It is important to note that very few guanosines were predicted to exist in exposed, unstructured segments of the monomeric MLV 5′ leader (based on chemical accessibility mapping and free energy calculations), and that the number of such exposed segments increases significantly upon dimerization (*42*, *43*). In addition, RNA constructs that include Ψ and 462 additional downstream nucleotides (called Ψ+) are packaged more efficiently than Ψ-only RNAs, suggesting that additional downstream elements also play a role in packaging (*60*). Taken together, these results indicate that MLV genome packaging proceeds via interactions between multiple NC domains and multiple RNA binding sites, and suggest the following general mechanism. (i) Gag molecules do not bind in significant numbers to the monomeric Ψ-site, due to the sequestration of guanosine-containing segments by base pairing. (ii) The Ψ-site forms a dimer, exposing multiple UCUG and related guanosine-containing $X_{(i-3)}$-$X_{(i-2)}$-$X_{(i-1)}$-$G_{(i)}$ segments. Dimerization may be catalyzed by the NC domains of full-length Gag, or by processed NC proteins

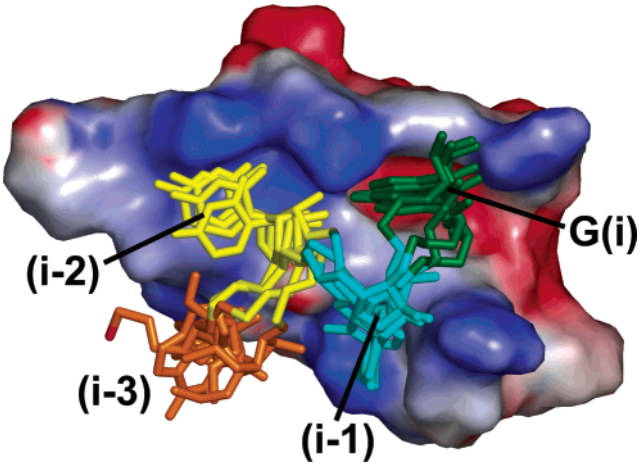

FIGURE 8: AACAGU, CCUCCGU, UUUUGCU, and UAUCUG RNAs bind the zinc knuckle of the MLV NC protein in a similar manner, with the nucleobase of $G_{(i)}$ inserted into the guanosine-specific hydrophobic pocket, the $i - 1$ nucleotide packed against Leu 21 and Ala 27, and the $i - 2$ and $i - 3$ nucleotides packed against the Tyr 28 side chain. This figure was generated by superpositioning the zinc knuckles of each of the independently determined structures. Only the surface of the zinc knuckle of the NC−UAUCUG complex is shown.

that appear to exist in the cytosol of infected cells. (iii) Gag molecules bind cooperatively to the unstructured, guanosine-containing segments. These elements appear to be closely spaced and highly abundant in the dimeric Ψ-site, allowing the Gag−Ψ complex to be stabilized by both Gag−RNA and Gag−Gag interactions. (iv) The MLV Ψ-site contains multiple dimer-promoting elements (DIS-1, DIS-2, SL-C, SL-D, and possibly additional elements), and it is conceivable that dimerization (and Gag binding) may proceed in a sequential, stepwise manner. Experiments aimed at testing this model using both in vivo and in vitro assays are underway.

ACKNOWLEDGMENT

Technical support from Robert Edwards and Chen Yu (Howard Hughes Medical Institute, University of Maryland Baltimore County) is greatly appreciated.

REFERENCES

1. Berkowitz, R., Fisher, J., and Goff, S. P. (1996) RNA packaging, *Curr. Top. Microbiol. Immunol. 214*, 177−218.
2. Hu, W.-S., and Temin, H. M. (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination, *Proc. Natl. Acad. Sci. U.S.A. 87*, 1556−1560.
3. Hu, W. S., and Temin, H. M. (1990) Retroviral recombination and reverse transcription, *Science 250*, 1227−1233.
4. Temin, H. M. (1991) Sex and recombination in retroviruses, *Trends Genet. 7*, 71−74.
5. Panganiban, A. T., and Fiore, D. (1988) Ordered interstrand and intrastrand DNA transfer during reverse transcription, *Science 241*, 1064−1069.
6. Jones, J. S., Allan, R. W., and Temin, H. M. (1993) Alteration of location of dimer linkage sequence in retroviral RNA: Little effect on replication or homologous recombination, *J. Virol. 67*, 3151−3158.
7. Mikkelsen, J. G., and Pedersen, F. S. (2000) Genetic reassortment and patch repair by recombination in retroviruses, *J. Biomed. Sci. 7*, 77−99.
8. Rein, A. (1994) Retroviral RNA packaging: A review, *Arch. Virol. 9*, 513−522.
9. Berkhout, B. (1996) Structure and function of the human immunodeficiency virus leader RNA, *Prog. Nucleic Acid Res. Mol. Biol.*, 1−34.
10. Jewell, N. A., and Mansky, L. M. (2000) In the beginning: Genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly, *J. Gen. Virol. 81*, 1889−1899.
11. Swanstrom, R., and Wills, J. W. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., Eds.) pp 263−334, Cold Spring Harbor Laboratory Press, Plainview, NY.
12. Stoker, A. W., and Bissell, M. J. (1988) Development of avian sarcoma and leukosis virus based vector packaging cell lines, *J. Virol. 16*, 1161−1170.
13. Mann, R., and Baltimore, D. (1985) Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNA, *J. Virol. 54*, 401−407.
14. Watanabe, S., and Temin, H. M. (1982) Encapsidation sequences for spleen necrosis virus, an avian retrovirus, are between the 5′ long terminal repeat and the start of the *gag* gene, *Proc. Natl. Acad. Sci. U.S.A. 79*, 5986−5990.
15. Lever, A. M. L., Göttlinger, H. G., Haseltine, W. A., and Sodroski, J. G. (1989) Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions, *J. Virol. 63*, 4085−4087.
16. Aldovini, A., and Young, R. A. (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus, *J. Virol. 64*, 1920−1926.
17. Clavel, F., and Orenstein, J. M. (1990) A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology, *J. Virol. 64*, 5230−5234.
18. Poznansky, M., Lever, A. M. L., Bergeron, L., Haseltine, W., and Sodroski, J. (1991) Gene transfer into human lymphocytes by a defective human immunodeficiency virus type 1 vector, *J. Virol. 65*, 532−536.
19. Fisher, J., and Goff, S. P. (1998) Mutational analysis of stem-loops in the RNA packaging signal of the moloney murine leukemia virus, *Virology 244*, 133−145.
20. Mougel, M., Zhang, Y., and Barklis, E. (1996) cis-Active structural motifs involved in specific encapsidation of Moloney Murine leukemia virus RNA, *J. Virol. 70*, 5043−5050.
21. Mougel, M., and Barklis, E. (1997) A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions, *J. Virol. 71*, 8061−8065.
22. Mansky, L. M., Krueger, A. E., and Temin, H. M. (1995) The bovine leukemia virus encapsidation signal is discontinuous and extends into the 5′ end of the gag gene, *J. Virol. 69*, 3282−3289.
23. Mansky, L. M., and Wisniewski, R. M. (1998) The bovine leukemia virus encapsidation signal is composed of RNA secondary structures, *J. Virol. 72*, 3196−3204.
24. McBride, M. S., and Panganiban, A. T. (1996) The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures, *J. Virol. 70*, 2963−2973.
25. Prats, A.-C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C., Paoletti, C., and Darlix, J.-L. (1990) Cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA, *J. Virol. 64*, 774−783.
26. Housset, V., De Rocquigny, H., Roques, B. P., and Darlix, J.-L. (1993) Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity, *J. Virol. 67*, 2537−2545.
27. Levin, J. G., Grimley, P. M., Ramseur, J. M., and Berezesky, I. K. (1974) Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D, *J. Virol. 14*, 152−161.
28. Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: An inseparable pair, *Nat. Rev. Microbiol. 2*, 461−472.
29. Kay, M. A., Liu, D., and Hoogerbrugge, P. M. (1997) Gene therapy, *Proc. Natl. Acad. Sci. U.S.A. 94*, 12744−12746.
30. Marcel, T., and Grausz, J. D. (1997) The TMC worldwide gene therapy enrollment report, end 1996, *Hum. Gene Ther. 8*, 775−800.
31. Yu, S. S., Kim, J.-M., and Kim, S. (2000) The 17 nucleotides downstream from the env gene stop codon are important for Murine Leukemia Virus packaging, *J. Virol. 74*, 8775−8780.

32. D'Souza, V., Dey, A., Habib, D., and Summers, M. F. (2004) NMR structure of the 101 nucleotide core encapsidation signal of the Moloney Murine Leukemia Virus, *J. Mol. Biol. 337*, 427−442.
33. D'Souza, V., Melamed, J., Habib, D., Pullen, K., Wallace, K., and Summers, M. F. (2001) Identification of a high-affinity nucleocapsid protein binding site within the Moloney Murine Leukemia Virus Ψ-RNA packaging signal. Implications for genome recognition, *J. Mol. Biol. 314*, 217−232.
34. D'Souza, V., and Summers, M. F. (2004) Structural basis for packaging the dimeric genome of Moloney Murine Leukaemia Virus, *Nature 431*, 586−590.
35. De Tapia, M., Metzler, V., Mougel, M., Ehresmann, B., and Ehresmann, C. (1998) Dimerization of MoMuLV genomic RNA: Redefinition of the role of the palindromic stem-loop H1 (278−303) and new roles for stem-loops H2 (310−352) and H3 (355−374), *Biochemistry 37*, 6077−6085.
36. Kim, C.-H., and Tinoco, I., Jr. (2000) A retroviral RNA kissing complex containing only two G−C base pairs, *Proc. Natl. Acad. Sci. U.S.A. 97*, 9396−9401.
37. Oroudjev, E. M., Kang, P. C. E., and Kohlstaedt, L. A. (1999) An additional dimer linkage structure in Moloney Murine Leukemia Virus RNA, *J. Mol. Biol. 291*, 603−613.
38. Ly, H., and Parslow, T. G. (2002) Bipartite signal for genomic RNA dimerization in the Moloney Murine Leukemia Virus, *J. Virol. 76*, 3135−3144.
39. South, T. L., and Summers, M. F. (1993) Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the ψ-site analog, dACGCC, *Protein Sci. 2*, 3−19.
40. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element, *Science 279*, 384−388.
41. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K., Wu, Z.-R., and Summers, M. F. (2000) NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal, *J. Mol. Biol. 301*, 491−511.
42. Tounekti, N., Mougel, M., Roy, C., Marquet, R., Darlix, J.-L., Paoletti, J., Ehresmann, B., and Ehresmann, C. (1992) Effect of dimerization on the conformation of the encapsidation ψ domain of Moloney murine leukemia virus RNA, *J. Mol. Biol. 223*, 205−220.
43. Mougel, M., Tounekti, N., Darlix, J.-L., Paoletti, J., Ehresmann, B., and Ehresmann, C. (1993) Conformational analysis of the 5′ leader and the *gag* initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization, *Nucleic Acids Res. 21*, 4677−4684.
44. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal. Biochem. 179*, 131−137.
45. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy, *J. Chem. Phys. 71*, 4546−4553.
46. Macura, S., and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy, *Mol. Phys. 41*, 95−117.
47. Bax, A., and Davis, D. G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy, *J. Magn. Reson. 65*, 355−360.
48. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy, *J. Magn. Reson. 53*, 521−528.
49. Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R. R. (1988) Clean TOCSY for ^{1}H spin system identification in macromolecules, *J. Am. Chem. Soc. 110*, 7870−7872.
50. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
51. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
52. Johnson, B. A., and Blevins, R. A. (1994) NMRview: A Computer Program for the Visualization and Analysis of NMR Data, *J. Biomol. NMR 4*, 603−614.
53. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for protein structure calculations with a new program, DYANA, *J. Mol. Biol. 273*, 283−298.
54. Schuller, W., Dong, C.-Z., Wecker, K., and Roques, B.-P. (1999) NMR structure of the complex between the zinc finger protein NCp10 of Moloney Murine Leukemia Virus and the single-stranded pentanucleotide d(ACGCC): Comparison with HIV−NCp7 complexes, *Biochemistry 38*, 12984−12994.
55. Davis, J. T. (2004) G-Quartets 40 years later: From 5′-GMP to molecular biology and supramolecular chemistry, *Angew. Chem. 43*, 668−698.
56. Greatorex, J. (2004) The retroviral RNA dimer linkage: Different structures may reflect different roles, *Retrovirology 1*, Aug 18 issue.
57. Russell, R. S., Liang, C., and Wainberg, M. A. (2004) Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology 1*, Sept 2 issue.
58. Nisole, S., and Saib, A. (2004) Early steps of retrovirus replicative cycle, *Retrovirology 1*, May 14 issue.
59. Mann, R., Mulligan, R. C., and Baltimore, D. (1983) Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus, *Cell 33*, 153−159.
60. Bender, M. A., Palmer, T. D., Gelinas, R. E., and Miller, A. D. (1987) Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region, *J. Virol. 61*, 1639−1646.

BI047639Q