doi:10.1006/jmbi.2001.5139 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **314**, 217–232

JMB

Identification of a High Affinity Nucleocapsid Protein Binding Element within the Moloney Murine Leukemia Virus Ψ-RNA Packaging Signal: Implications for Genome Recognition

Victoria D'Souza, Judy Melamed, Dina Habib, Kristi Pullen Keith Wallace and Michael F. Summers*

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

**Corresponding author*

Murine leukemia virus (MLV) is currently the most widely used gene delivery system in gene therapy trials. The simple retrovirus packages two copies of its RNA genome by a mechanism that involves interactions between the nucleocapsid (NC) domain of a virally-encoded Gag polyprotein and a segment of the RNA genome located just upstream of the Gag initiation codon, known as the Ψ-site. Previous studies indicated that the MLV Ψ-site contains three stem loops (SLB-SLD), and that stem loops SLC and SLD play prominent roles in packaging. We have developed a method for the preparation and purification of large quantities of recombinant Moloney MLV NC protein, and have studied its interactions with a series of oligoribonucleotides that contain one or more of the Ψ-RNA stem loops. At RNA concentrations above ~ 0.3 mM, isolated stem loop SLB forms a duplex and stem loops SL-C and SL-D form kissing complexes, as expected from previous studies. However, neither the monomeric nor the dimeric forms of these isolated stem loops binds NC with significant affinity. Longer constructs containing two stem loops (SL-BC and SL-CD) also exhibit low affinities for NC. However, NC binds with high affinity and stoichiometrically to both the monomeric and dimeric forms of an RNA construct that contains all three stem loops (SL-BCD; $K_d = 132(\pm 55)$ nM). Titration of SL-BCD with NC also shifts monomer-dimer equilibrium toward the dimer. Mutagenesis experiments demonstrate that the conserved GACG tetraloops of stem loops C and D do not influence the monomer-dimer equilibrium of SL-BCD, that the tetraloop of stem loop B does not participate directly in NC binding, and that the tetraloops of stem loops C and D probably also do not bind to NC. These surprising results differ considerably from those observed for HIV-1, where NC binds to individual stem loops with high affinity *via* interactions with exposed residues of the tetraloops. The present results indicate that MLV NC binds to a pocket or surface that only exists in the presence of all three stem loops.

Keywords: Moloney murine leukemia virus; ribonucleic acid (RNA); psi-site; nucleocapsid protein; nuclear magnetic resonance (NMR)

Introduction

Retroviruses have evolved highly efficient mechanisms for the specific packaging of their genetic material. As retroviruses assemble in infected cells, their genomes are selected from a cytosolic pool that contains a substantial excess (~100-fold) of cellular RNAs.[1] Only full-length, unspliced genomes are packaged, and spliced viral mRNAs that

Abbreviations used: BME, β-mercaptoethanol; CA, capsid protein; HIV-1, human immunodeficiency virus Type-1; HMQC, heteronuclear multiple quantum coherence; ITC, isothermal titration calorimetry; MA, matrix protein; MLV, murine leukemia virus; NC, nucleocapsid protein; NOE, nuclear Overhauser effect; PAGE, polyacrylamide gel electrophoresis; GST, glutathione-*S*-transferase.

E-mail address of the corresponding author: summers@hhmi.umbc.edu

0022-2836/01/020217–16 $35.00/0

encode for envelope and accessory proteins are ignored. In addition, retroviruses specifically package two copies of their RNA, even under conditions in which packaging efficiencies are diminished. This specificity is mediated predominantly by interactions between the nucleocapsid (NC) domains of the assembling Gag polyproteins and a segment of the viral genome called the Ψ-site, which is located between the 5′-Upstream Activator Sequence and the Gag initiation codon.[1–4]

Considerable effort has been made over the past 15 years to determine the mechanism by which retroviruses package their genomes. Although many recent studies have focused on the human immunodeficiency virus (HIV-1), there has also been substantial interest in the murine leukemia virus (MLV). MLV is currently the most widely used gene delivery system in gene therapy trials, and understanding the molecular basis for efficient genome selection and packaging is important for the design of more effective vectors with higher viral titers.[5] In addition, MLV possesses attributes that could potentially render biophysical and structural studies of genome packaging more tractable. Thus, MLV is a C-type retrovirus with a simple genome that lacks inhibitory elements, requires only one splicing event (to produce the *env* mRNA), and does not encode for accessory proteins. MLV contains a relatively simple NC protein that possesses only a single CCHC-type zinc knuckle, a miniglobular domain that is essential for genome selection and packaging ($C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$; C = Cys, H = His, X = variable or other conservatively substituted amino acid) (Figure 1).[6–8] Also, as discussed below, the portion of the Ψ-site that is critical for genome packaging is relatively short (~100 nucleotides), and is thus potentially amenable to high resolution solution-state structural studies.

In 1983, Mann and co-workers demonstrated that genome packaging by MLV could be ablated by deleting a 350 nucleotide segment of RNA located just downstream of the *env* splice donor site (nucleotides ~210-560)[9] (Figure 1). This segment is now commonly referred to as the MLV Ψ-site. Subgenomic *env* mRNA lacks this segment and is not packaged, providing a mechanism for discrimination between spliced and unspliced viral RNAs. The Ψ-site can be relocated to the 3′-end of the genome without severely affecting packaging, indicating that Ψ functions as an independent packaging element.[10] MLV constructs containing deletions that extend into the 5′ end of the Ψ-site (i.e. deletions beginning at nucleotide 400) are also packaged efficiently, indicating that principal elements of the Ψ-site that signal for packaging reside within residues 215-400.[11] However, packaging efficiency appears to be influenced by nucleotides both upstream[5,12] and downstream[13–15] of the Ψ-site, presumably *via* interactions that occur independent of, and possibly subsequent to, an initial Ψ-site mediated recognition event.

The secondary structure of the MLV Ψ-site has been studied using a combination of site-directed mutagenesis, chemical accessibility mapping, and phylogenetic and computer modeling.[16–24] These studies indicate that nucleotides 278 to 374 form three stem-loop structures (Figure 1(a)). The first stem loop, SL-B, contains a fully self-complementary tetraloop (AGCU), and several studies indicate that this stem loop facilitates RNA dimerization.[16,25–29] Note that the boundaries of the MLV dimer linkage site (DLS) have not been precisely identified, and it is possible that dimerization may also be promoted by sequences that overlap with the 5′-end of the Ψ-site (nucleotides 204-277),[29] and also by stem loops SLC and SLD.[27,30] In fact, recent NMR studies have revealed that isolated stem loop SL-D is capable of forming a stable ''kissing complex'' that is mediated by intermolecular hydrogen bonds between G and C residues in the conserved GACG tetraloop, and it has been suggested that stem loops SL-C and SL-D may facilitate packaging indirectly by promoting RNA dimerization.[27,30]

Recent mutagenesis experiments designed to delete or disrupt base-pairing in stem loops B, C and D led to substantial reductions in packaging efficiency, but generally did not completely eliminate packaging.[18,21,22,24] This suggested to us that these stem loops might play redundant roles in genome selection. A similar view is emerging from studies of HIV-1 genome recognition.[31] Thus, the HIV-1 Ψ-site contains four stem loops (SL1-SL4; see[32] and references therein), and the NC protein is capable of binding with high affinity to individual stem loops that contain a conserved GGNG tetraloop (G = guanosine, N = any ribonucleotide).[31–33] Solution-state structural studies revealed that tight binding of HIV-1 NC to SL2 and SL3 is mediated, at least in part, by specific hydrogen bonding, electrostatic and hydrophobic interactions between the zinc knuckles and exposed guanosine bases of the RNA tetraloops. Since the Ψ-sites of both HIV-1 and MLV are composed of stem loops with G-containing tetraloops, and since the NC proteins of both retroviruses contain a conserved CCHC-type zinc knuckle, we and others[18] concluded that MLV genome recognition might occur *via* similar NC-tetraloop interactions.

As a first step towards identifying the atomic determinants of MLV genome recognition and packaging, we prepared the recombinant NC protein and a series of oligoribonucleotides corresponding to segments of the Ψ-site for solution-state biophysical and NC-binding studies. Surprisingly, RNAs containing sequences of either single- or double-stem loop fragments of the MLV Ψ-site (SL-B, SL-C, SL-D, SL-BC, SL-CD) bind the NC protein weakly, as determined by native gel electrophoresis and ^{1}H NMR spectroscopy. Tight NC binding was only observed for constructs containing all three stem loops (SL-BCD). We also found that the RNA dimerization equilibrium constant of SL-BCD was unaffected by point mutations in the

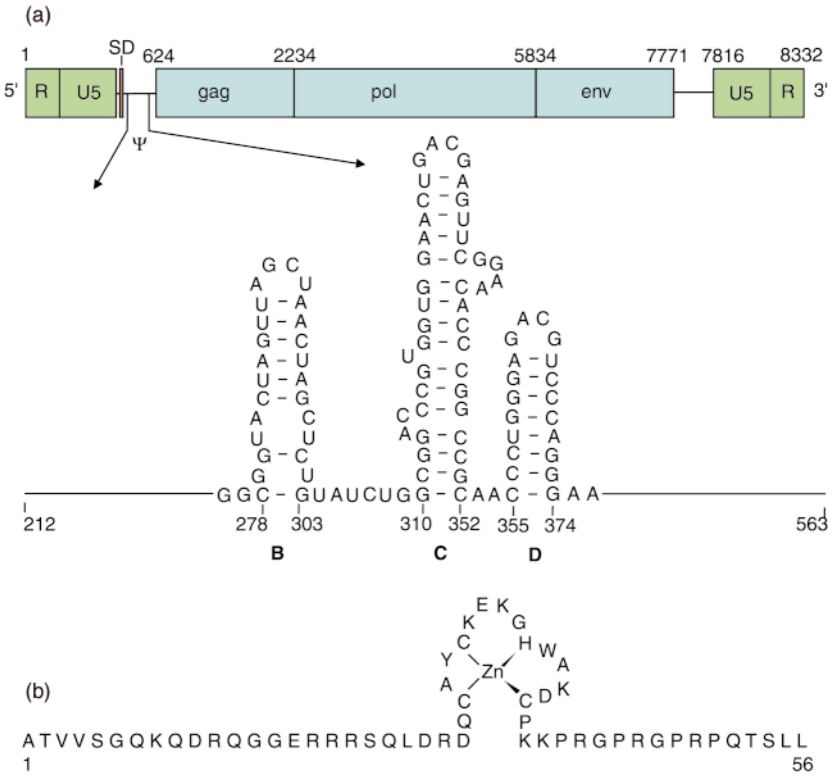

Figure 1. (a) Representation of the Moloney murine leukemia virus (MLV) genome showing the relative positions of the splice donor (SD) and Ψ-site, as well as the nucleotide sequence and predicted secondary structure of the portion of the Ψ-site that is essential for genome encapsidation. Nucleotides are numbered using the first residue after the 5′-cap as position 1. (b) Amino acid sequence and zinc binding mode of the NC protein.

conserved GACG tetraloops of stem loops C and D, and that NC binding was unaffected by point mutations in all three tetraloops. These and other results implicate a packaging mechanism that differs substantially from that utilized by HIV-1, in which the initial recognition event involves the stoichiometric binding of NC to an RNA binding site (possibly a pocket) that is defined by all three stem loops.

Results and Discussion

NC protein expression and purification

The native NC gene (obtained from the pNCA Genome Library[34]) could not be readily expressed due to the presence of several Arg and Pro codons that are used rarely in *Escherichia coli*. A sequence-optimized insert obtained by site-directed mutagenesis was cloned into the pGEX expression vector as a GST fusion gene, and the sequence of the clone confirmed by nucleotide sequencing (Rockefeller Protein/DNA Technology Center). This optimized clone over-expressed efficiently in *E. coli*. Treatment of the resulting protein with PreScission protease to remove the GST tag afforded a 61-residue protein (referred to throughout this paper as NC) that contains five non-native residues at the N terminus due to cleavage specificity of the PreScission protease enzyme: NH_3^+- G P L G S A ^{1}T V V S G Q K Q D ^{10}R Q G G E R R R S Q ^{20}L D R D Q ^{25}C A Y C K E K G H W A K D C P ^{40}K K P R G P R G P R ^{50}P Q T S L ^{55}L (the zinc knuckle is underlined); yield ~10 mg purified protein per liter of culture media. Coomasie-stained SDS-PAGE of the purified protein showed a single band at about 7.0 kDa. The molecular mass of the purified ^{15}N-labeled apoprotein (6880.2(+4.5) Da, determined by ion-spray mass spectrometry) was as expected ($M_{r\ calc} = 6883.6$ Da).

Preparation of RNA samples

Oligoribonucleotides SL-B, SL-C, SL-D, SL-BC, SL-CD, SL-BCD, SL-BC_mD_m and SL-$B_mC_mD_m$ (m = mutant; see Figure 2 for definitions) were prepared using standard *in vitro* transcription methods.[35] Templates for SL-BCD (and its mutants) were prepared from cDNA clones, whereas the templates for the remaining stem loop constructs were obtained synthetically (KECK Foundation Biotechnology Research Center) and purified by denaturing polyacrylamide gel electrophoresis (PAGE). Reactions were optimized to give maximum yields by incorporating non-native guanosine bases at the 5′ terminus (Figure 2(a)). RNA samples were purified by preparative scale denaturing PAGE,[36,37] with yields typically in the range of 1.0-3.0 μmoles of purified RNA per 90 ml reaction solution (see Materials and Methods for details). Denaturing polyacrylamide gels showing relative electrophoretic mobility and sample purity are shown in Figure 2(b).

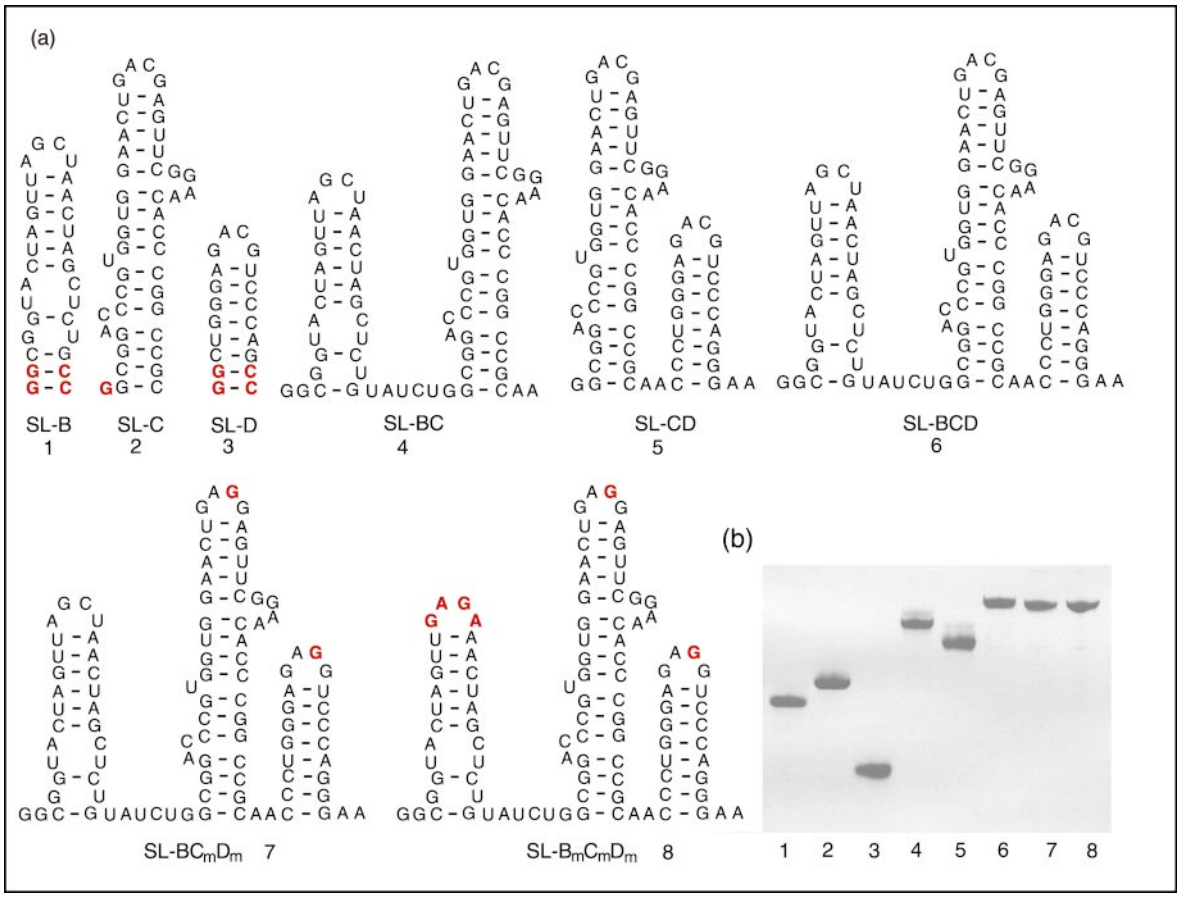

Figure 2. (a) RNA constructs used in the present studies. Non-native nucleotides that were added or modified to facilitate *in vitro* transcription (SL-B, SL-C, SL-D) or test oligomerization and NC binding modes (SLBC_mD_m, SL-$B_mC_mD_m$) are shown in red. (b) Denaturing PAGE results showing relative electrophoretic migration and sample purity for the RNA constructs used in the present studies.

Solution behavior and NC-binding studies of isolated stem loops SL-B, SL-C and SL-D

Native PAGE experiments were conducted to determine the oligomerization states of the isolated stem loop constructs. As shown in Figure 3(a), stem loop SL-B exhibits concentration-dependent changes typical of the expected hairpin-duplex equilibrium. Thus, a single band is observed at low concentration that migrates with the electrophoretic mobility expected for a monomer. At higher concentrations (>0.3 mM), a second band appears that migrates at a rate expected for a dimer. Concentration-dependent changes were also observed in the 1D ^{1}H NMR spectra obtained in H_2O and ^2H_2O, with several well resolved signals that are diagnostic of the SL-B dimer appearing at RNA concentrations above 0.5 mM (Figure 3(b)). Analysis of 2D NOESY and natural-abundance ^{1}H-^{13}C HMQC NMR data obtained for both dilute and concentrated solutions of SL-B afforded complete assignment of the aromatic and H1′ protons, as well as a portion of the remaining ribose protons. A portion of the 2D NOESY spectrum obtained for SL-B at high concentration (2.0 mM) showing sequential NOE connectivities for residues in the center of the duplex form of SL-B is given in Figure 3(c). Although a detailed 3D structural analysis of the hairpin and duplex forms of SL-B is beyond the scope of this paper, it is important to point out that the NOE cross-peak patterns and intensities observed for the dimeric form of SL-B are consistent with the expected duplex structure (Figure 3(d)). In particular, the upfield-shifted aromatic proton signals at 6.718 and 6.452 ppm, assigned to the H2 protons of A293 and A289, respectively, of the dimeric form of SL-B, exhibit long-range NOEs diagnostic of an A-helical RNA duplex (including A293-H2 to A289-H1′, A293-H2 to A294-H1′, A289-H2 to G290-H1′ and A289-H2 to A293-H1′). In addition, sequential H8/6(i) to H1′/2′/3′($i-1$), H8/6(i) to H5($i+1$) and H8/6(i) to H8/6($i\pm1$) NOEs were observed for residues C283 to G298.

Tinoco and colleagues recently demonstrated that SL-D can form a dimer that migrates with the same rate as the monomeric species.[30] In an elegant study, these investigators used ^{13}C-isotope editing NMR experiments to show that the SL-D dimer is actually a ''kissing'' species that is stabilized by two intermolecular G·C base-pairs involving Cyt365 and Gua366 of the conserved G363-A364-

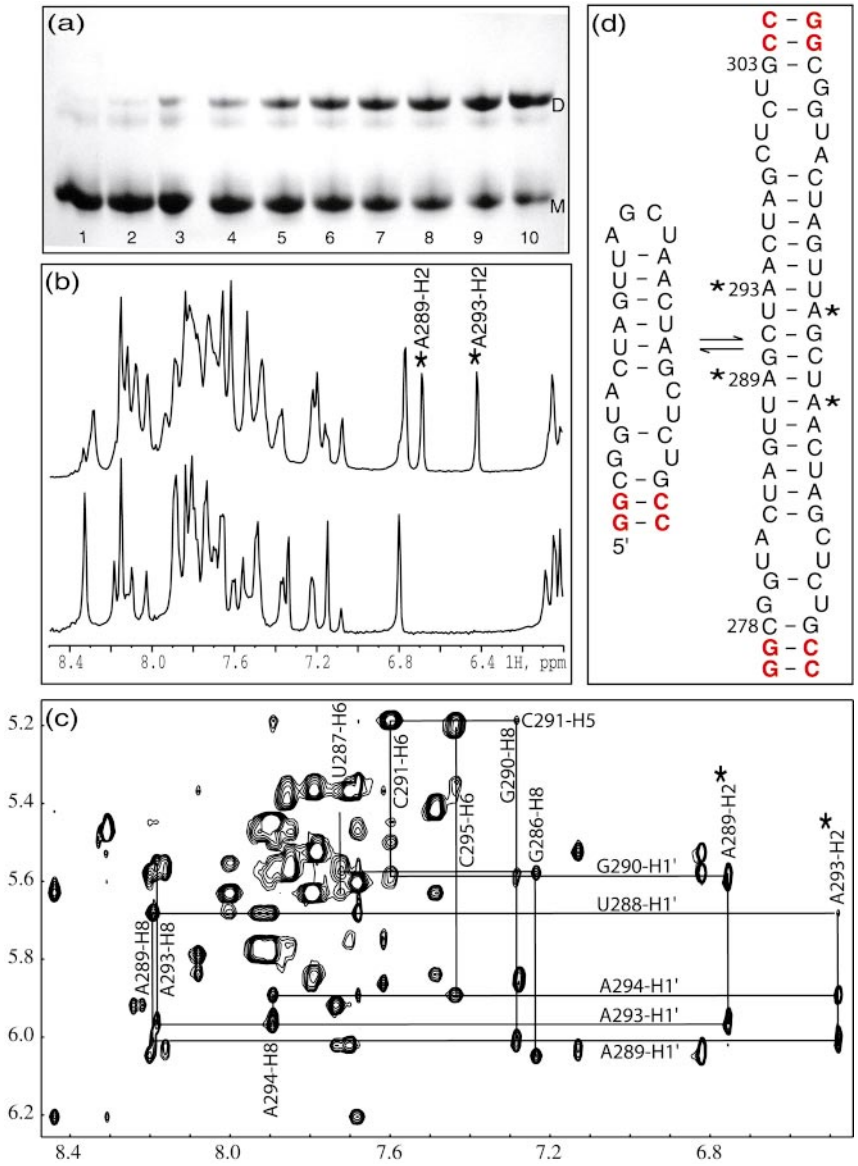

Figure 3. (a) Native PAGE data obtained for isolated stem loop SL-B as a function of concentration; M, monomer (hairpin); D, dimer (duplex). Lanes 1-10 correspond to RNA concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 mM, respectively. (b) Portions of 1D ^{1}H NMR spectra obtained in $^{2}H_2O$ at 2.0 mM and 0.5 mM SL-B concentration. Peaks marked * are due to the duplex form of SL-B. (c) Portion of the 2D NOESY spectrum obtained for the SL-B duplex form in $^{2}H_2O$ showing inter- and intra-residue anomeric-to-aromatic proton NOE connectivities. (d) Diagram showing the SL-B stem loop-duplex equilibrium.

C365-G366 tetraloop.[30] Although the monomer-kissing equilibrium was not detectable in the electrophoresis data (Figure 4(a)), it is readily observed in the concentration-dependent 1D ^{1}H NMR spectra. An unusually upfield-shifted ^{1}H NMR signal at 2.8 ppm appears in the SL-D ^{1}H NMR spectra obtained at concentrations greater than 0.3 mM (Figure 4(b)), and this signal (assigned to C368-H4′) is diagnostic of the kissing species,[30] (Figure 4(c)). The NMR spectra obtained in our laboratory for SL-D are fully consistent with the NMR data and atomic-level structure that were recently reported.[30] Similar results were obtained for SL-C, which contains the same tetralooop sequence as in SL-D and is also capable of forming a kissing species.

To determine if SL-B, SL-C or SL-D are capable of binding to the MLV NC protein with significant affinity, native gel-shift assays were performed. As shown in Figure 5(a), titration of the individual stem loop RNAs with NC resulted in a slight smearing of the electrophoretic bands, but did not lead to distinct new bands that would be characteristic of tight binding. In addition, titration of SL-B, SL-C and SL-D with NC under conditions that favored either the monomeric or dimeric species resulted in a broadening of the RNA NMR signals, but did not lead to discrete new signals as had been seen previously for HIV-1 NC binding to SL-3[33] and SL2.[31] Thus, the NMR data are consistent with a weak binding interaction that is fast on the NMR chemical shift timescale (milliseconds).

Solution behavior and NC-binding studies of SL-BC and SL-CD

Since tight NC binding was not observed for the isolated stem loops, RNA constructs were prepared that contain two adjacent stem loops (SL-BC and SL-CD). At concentrations below 0.1 mM, both

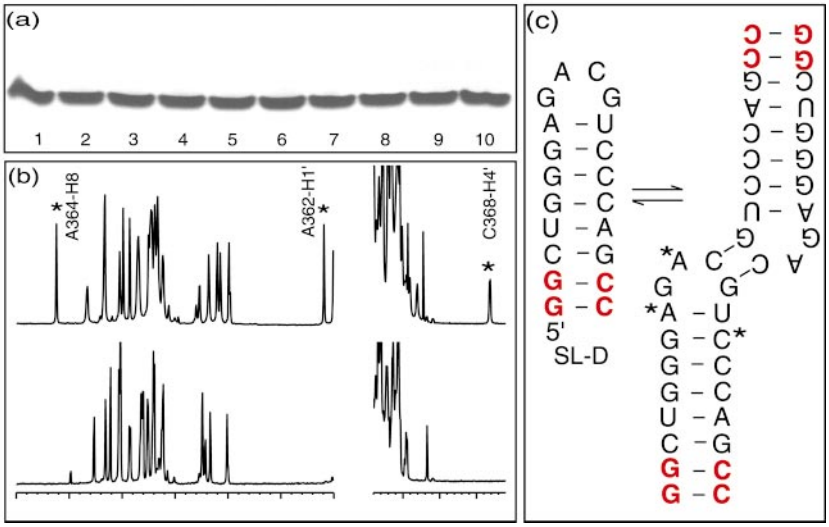

Figure 4. (a) Native PAGE data obtained for isolated stem loop SL-D as a function of concentration. Lanes 1-10 correspond to RNA concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 mM, respectively. (b) Portions of 1D ^{1}H NMR spectra obtained in $^{2}H_2O$ at 2.0 mM and 0.2 mM SL-D concentration. Peaks marked * are due to the kissing dimer form of SL-D. (c) SL-D stem loop - kissing dimer equilibrium, as identified originally by Tinoco and co-workers.[30]

constructs give rise to single electrophoretic bands with mobilities expected for monomeric structures (Figure 6). Whereas the electrophoretic behavior of SL-CD is independent of RNA concentration (Figure 6(b)), the electrophoresis data obtained for SL-BC exhibits a second band at concentrations above 0.1 mM that is consistent with an equilibrium shift to a dimeric state (Figure 6(a)). The apparent monomer-dimer equilibrium constant ($K_d \sim 0.5$ mM) is very similar to the value of $K_d \sim 0.8$ mM determined for the isolated SL-B stem loop (Figure 3(a)). In addition, 1D ^{1}H NMR spectra obtained for SL-BC exhibit concentration-dependent changes that are very similar to those observed for SL-B, with new signals appearing at 6.72 and 6.45 ppm at concentrations above 0.3 mM. Although the ^{1}H NMR spectrum of SL-BC has not been independently assigned, the similarity of the electrophoresis and ^{1}H 1D NMR spectra indicate that, at concentrations above 0.3 mM, stem loop-B of the SL-BC construct forms a duplex as observed in the isolated SL-B construct. It is noteworthy that no upfield NMR signal diagnostic for the kissing species (2.8 ppm) was observed under any conditions, which indicates that kissing interactions do not contribute to the structure of the SL-BC dimer. The fact that the dimerization equilibrium constant for SL-BC matches that observed for SL-B supports this conclusion. In contrast, the 1D ^{1}H NMR spectra observed for SL-CD at high concentrations did exhibit an upfield signal at 2.8 ppm, indicating that this construct is capable of forming a kissing species.

As observed for isolated stem loops SL-B, SL-C and SL-D, neither of the double-stem loop constructs exhibited significant affinity for the NC protein as measured by native PAGE NC titrations (Figure 5(b)).

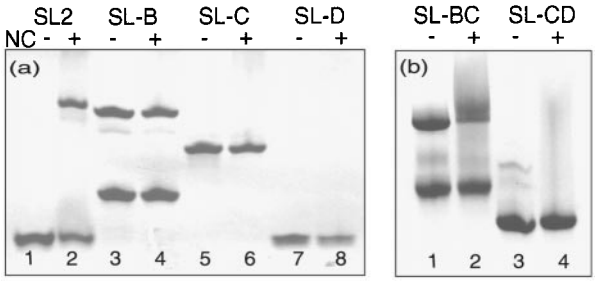

Figure 5. Native PAGE results obtained for single- and double-stem loop RNA constructs upon titration with NC. (a) Titration of SL-B, SL-C and SL-D with NC (lanes 3-4, 5-6 and 7-8, respectively). For comparison, data obtained upon titration of the HIV-1 stem loop SL-3 with the HIV-1 NC protein are also shown (lanes 1-2). (b) Titration of SL-BC and SL-CD with NC (lanes 1-2 and lanes 3-4, respectively). Under the conditions employed ([RNA] = 0.5 mM, 5 mM Tris-HCl, pH 7.0), both the monomer and dimer forms of the stem loops are present. Similar results were obtained at lower RNA concentrations that favor the RNA monomers, and at higher concentrations where the SL-B duplex and SL-C, SL-D kissing complexes are favored (not shown).

Solution behavior and NC-binding studies of SL-BCD

In view of the above findings that the MLV NC protein binds with relatively poor affinity to single-

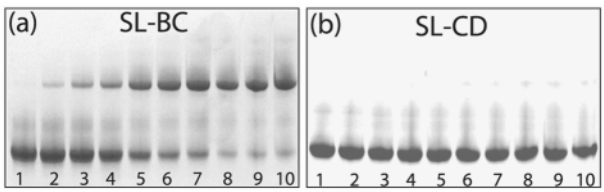

Figure 6. Native PAGE data obtained for SL-BC (a) and SL-CD (b) as a function of concentration. Lanes 1-10 correspond to RNA concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 mM, respectively.

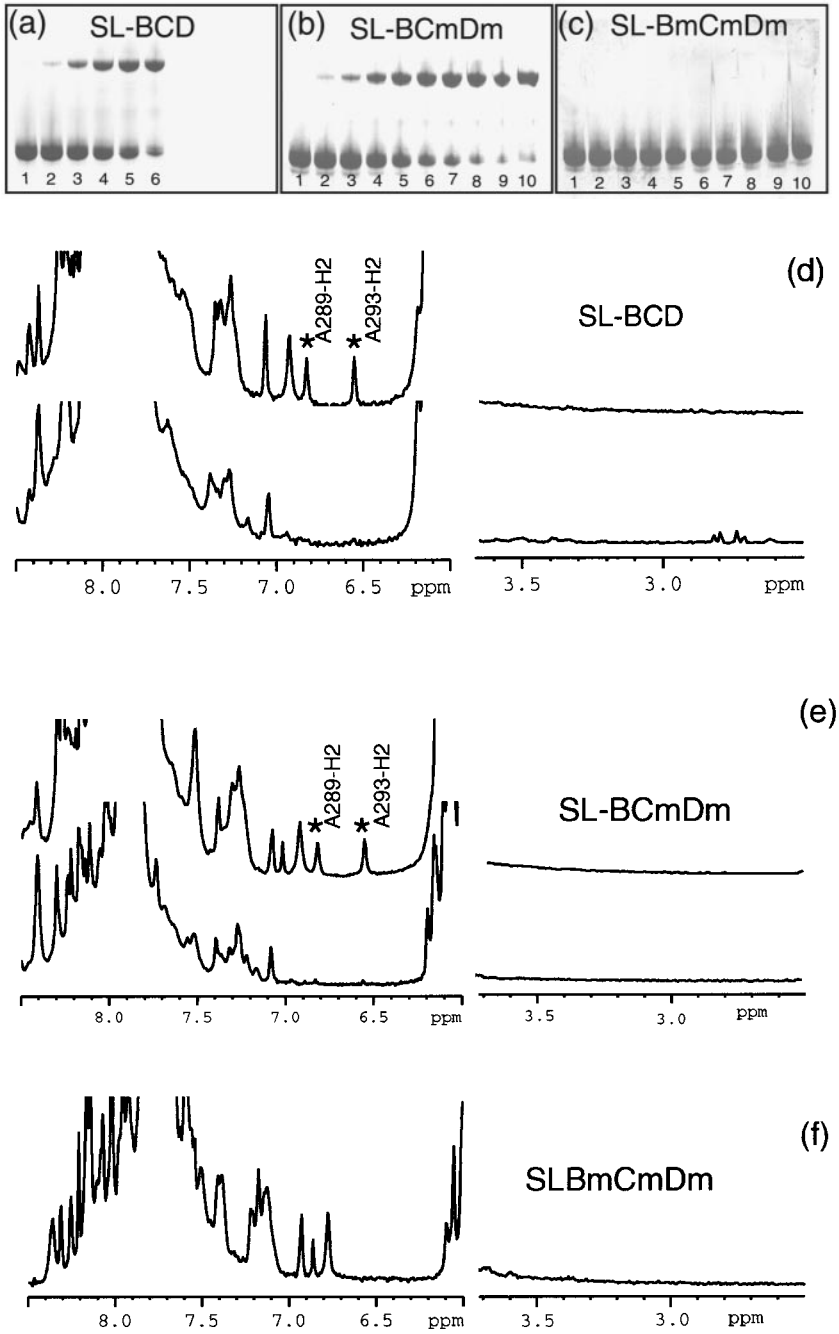

Figure 7. Native PAGE data obtained for SL-BCD (a) SL-BCmDm (b) and SL-BmCmDm (c) as a function of concentration. Lanes 1-10 correspond to RNA concentrations of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 mM, respectively. (d)-(f) Portions of 1D ^{1}H NMR spectra obtained for (d) SL-BCD in $^{2}H_2O$ at 0.2 mM (bottom) and 0.6 mm (top) concentration; (e) SL-BCmDm at 0.2 mM (bottom) and 2.0 mM (top) concentration; and (f) SL-BmCmDm at 2 mM concentration. Peaks denoted with a * are due to the duplex form of SL-B.

and double-stem loop constructs of its cognate Ψ-site, we initiated studies of the 102 nucleotide SL-BCD segment (Figure 2(a)). At RNA concentrations below 0.1 mM, SL-BCD migrates as a single band at a rate consistent with a monomeric species (Figure 7(a)). At RNA concentrations in the range 0.1-0.6 mM, a second band appears that is consistent with a dimeric species (Figure 7(a)), and the apparent monomer-dimer equilibrium constant ($K_d \sim 0.5$ mM) is essentially identical to those observed for SL-B and SL-BC ($K_d \sim 0.8$ and 0.5 mM, respectively; for comparison see Figures 3(a) and 6(a)). Interestingly, at RNA concentrations above 0.6 mM and temperatures at or below 35 °C, solutions of SL-BCD form a solid gel, precluding electrophoresis and NMR studies at these higher concentrations. Solutions of SL-BCD prepared with additional NaCl or $MgCl_2$ salts gelled at even lower concentrations.

^{1}H NMR spectra obtained for SL-BCD at a RNA concentration of 0.2 mM (which favors the monomeric form) exhibit relatively narrow ^{1}H NMR signals, and new signals appeared at 6.73 and 6.45 ppm as the concentration was increased (Figure 7(d)). The new signals matched those observed for the duplex forms of SL-B and SL-BC (Figure 3(b)). No evidence for a kissing species involving stem loops SL-C or SL-D was present in

the 1D ^{1}H NMR spectra obtained for SL-BCD at concentrations where signals attributable to the B-duplex were clearly visible. These findings are consistent with *in vitro* MLV Ψ-RNA dimerization studies, which showed that SL-C and SL-D do not influence the dimerization equilibrium constant of MLV Ψ-RNA (but do accelerate the equilibration process).[27] A sample of SL-BCD prepared at a concentration of 0.6 mM and 37 °C (which eventually gelled) exhibited a broad signal at 2.8 ppm that matched the characteristic signal observed for the kissing forms of SL-C, SL-D and SL-CD. Since the signal that is characteristic of the kissing species is only observed under conditions that lead to gel formation, we attribute the kissing interactions to intermolecular interactions between RNA dimers. Thus, the electrophoresis and NMR data indicate that SL-BCD dimerizes *via* conversion of stem loop B to an intermolecular duplex, that kissing interactions involving stem loops C and D do not contribute to the stability of the dimer, and that the kissing interactions that are detected at high sample concentrations are due to inter-dimer interactions that lead to aggregation.

Native gel electrophoresis/NC-titration assays were conducted with SL-BCD at concentrations in which both monomeric and dimeric species are present. Unlike for all of the single- and double-stem loop constructs, addition of sub-stoichiometric amounts of NC led to discrete new bands with retarded mobilities relative to the SL-BCD monomer and dimer bands (Figure 8(a)). The fact that band shifts were observed for both the monomeric and dimeric forms of SL-BCD in the presence of sub-stoichiometric amounts of NC indicates that both forms of SL-BCD bind NC with similar affinities. Further additions of NC resulted in a complete shift of the monomer-dimer equilibrium to the dimer species, as well as to a complete shift from NC-free RNA to NC-bound RNA (Figure 8(a)). ^{1}H NMR spectra were obtained for SL-BCD (0.2 mM) in the absence and presence of one equivalent of NC. Signals at 6.73 and 6.45 ppm that are diagnostic for duplex formation by stem loop-B were not present in the absence of NC, but appeared upon titration with NC (and incubation at 35 °C for ten minutes) (Figure 9). In addition, no upfield signals diagnostic of a kissing species were observed either in the presence or absence of NC at the low RNA concentrations employed. These data are consistent with the gel titration data, and indicate that the NC-induced SL-BCD dimer is stabilized by SL-B duplex formation. Sensitivity problems associated with sample gelling prohibited NMR studies at higher RNA concentrations, even for samples prepared in the presence of NC, and this has thus far precluded high resolution structural studies of SL-BCD either free in solution or bound to NC.

The affinity of MLV NC for SL-BCD was determined quantitatively by isothermal titration calorimetry (ITC). Titration of SL-BCD RNA with NC gave rise to negative binding enthalpies that fit to

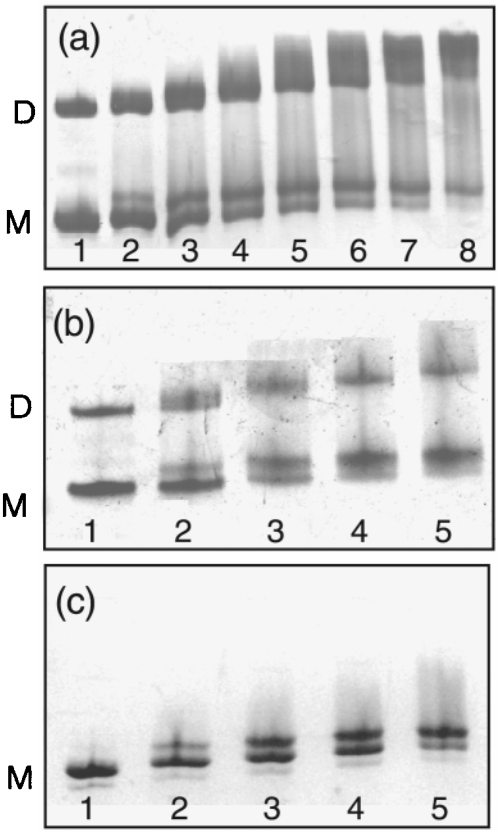

Figure 8. Native PAGE data obtained after titration of NC into (a) SL-BCD, (b) SL-BCmDm and (c) SL-BmCmDm. Lane 1 corresponds to the above RNA constructs incubated without NC, and lanes 2-8 correspond to the RNA solutions incubated with increasing amount of NC (see Materials and Methods for details).

1:1 binding isotherms (Figure 10(a)). As a control, ITC data obtained for titration of NC to SL-CD, which does not bind tightly, are also shown in Figure 10(d). The dissociation constant for the NC-SL-BCD complex, determined as the mean ± standard deviation from three independent experiments, is 132(±55) nM. This value is similar to the dissociation constants observed for HIV-1 NC binding to isolated stem loops SL2 and SL3 (110(±50) nM and 170(±65) nM, respectively).[31] Thus, the electrophoresis, NMR and ITC data indicate that SL-BCD is capable of binding one NC molecule stoichiometrically and with high affinity, that the binding of NC to SL-BCD shifts the RNA equilibrium toward the dimeric species, and that the $(NC:SL\text{-}BCD)_2$ dimer is stabilized by SL-B duplex formation.

Mutations in the conserved GACG tetraloops inhibit aggregation but do not affect dimerization or NC binding to SL-BCD

Since the NMR data suggest that kissing interactions involving SL-C and/or SL-D may be

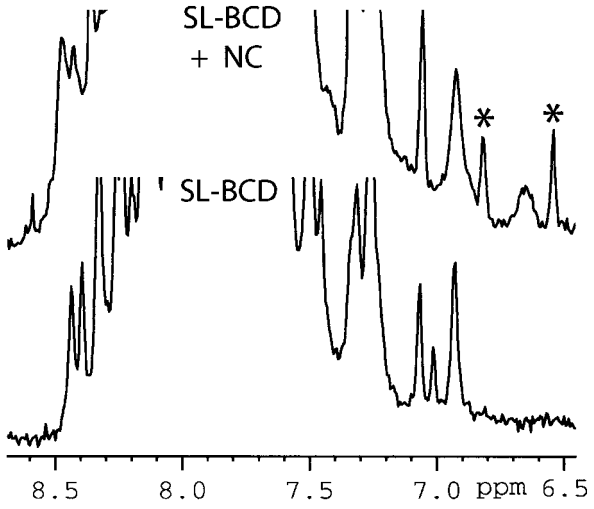

Figure 9. Portions of the 1D ^{1}H NMR spectra obtained for SL-BCD at low concentration (0.2 mM) in the absence (bottom) and presence (top) of one equivalent of NC. Peaks denoted by the * symbol are due to the duplex form of SL-B. These data demonstrate that NC binding shifts the SL-BCD monomer-dimer equilibrium toward the dimer species.

responsible for aggregation of the SL-BCD dimer, we hypothesized that mutating the conserved tetraloops should inhibit aggregation. Site-directed mutagenesis experiments were performed by PCR to mutate the GACG tetraloops of SL-C and SL-D to GAGG, which are not capable of forming kissing G·C base-pairs. The resulting mutated RNA, designated SL-BCmDm, was obtained in high yield and could be readily concentrated to 2 mM without aggregation or gelling. As shown in Figure 7(b), SL-BCmDm exhibits a single electrophoretic band under native conditions at concentratons below 0.1 mM that migrates at a rate consistent with a monomeric species, and a second band consistent with a dimer appears at RNA concentrations above 0.2 mM (apparent K_d for dimerization ~0.5 mM). These data are very similar to those obtained for SL-B, SL-BC and SL-BCD (see Figures 3(a), 6(a) and 7(a), respectively). In addition, the concentration-dependent 1D ^{1}H NMR spectra obtained for SL-BCmDm are essentially identical to those obtained for SL-BCD (Figure 7(e)), except that the ^{1}H NMR signals do not become broadened at high sample concentrations (up to 2.0 mM). Thus, the combined data confirm that the conserved GACG tetraloops of stem loops C and D promote aggregation, but do not affect the dimerization equilibrium of SL-BCD.

Gel-shift NC titration assays were conducted with SL-BCmDm to determine if the mutations in the conserved SL-C and -D tetraloops affect NC binding. As shown in Figure 8(b), NC binds to SL-BCmDm with affinity and stoichiometry that are similar to those observed for binding to native SL-BCD. Stoichiometric 1:1 binding isotherms were observed in the ITC data obtained for SL-BCmDm, and the resulting dissociation constant ($K_d = 133(\pm 47)$ nM) is essentially identical to that observed for NC binding to SL-BCD (Figure 10(b)). These findings indicate that the conserved tetraloops of stem loops C and D do not participate directly in tight NC binding.

Mutation of the B tetraloop favors the SL-BCD monomer but does not affect NC binding

The gel-shift data shown in Figure 8 indicate that NC is capable of binding with similar affinities to both the monomeric and dimeric forms of SL-BCD (and SL-BCmDm). Since the tetraloop nucleotides of stem loop B must exist in substantially different environments when the RNA is in the monomeric and duplex conformations, we hypothesized that these residues most likely do not interact directly with the NC protein upon binding. To test this hypothesis, and as an attempt generate a construct that would not readily dimerize for future high resolution solution-state structural studies, a mutant of SL-BCmDm was cloned in which the tetraloop of stem loop B was substituted by the GNRA-type tetraloop, GAGA. This class of tetraloops (R = purine; A = adenine) form highly stable structures and promote RNA hairpin formation.[38–40] As shown in Figure 7(c), this mutant, termed SL-BmCmDm, migrates on native gels at a rate consistent with a monomeric species, even at concentrations as high as 2.0 mM. As expected, this triple mutant did not form gels or exhibit signs of aggregation, even at the highest concentrations examined (2.0 mM). SL-BmCmDm gave rise to relatively sharp ^{1}H NMR signals, even at high NMR sample concentrations (Figure 7(f)), and did not exhibit slow-exchange concentration-dependent spectra over the concentration range examined (0.2 to 2.0 mM).

Addition of NC to SL-BmCmDm resulted in a gel-shift consistent with the shift observed for NC binding to the monomeric forms of SL-BCD and SL-BCmDm (Figure 8(c)). No additional bands indicative of dimer formation were observed in the gels upon addition of NC. Isothermal titration calorimetry measurements indicate that NC binds to SL-BmCmDm with the same affinity ($K_d = 157(\pm 20)$ nM) as observed for binding to SL-BCD and SL-BCmDm (Figure 10(c)). Thus, the native gel electrophoresis and NMR data indicate that the triple mutant forms a stable monomeric species that does not aggregate, as designed, and that the high-affinity NC binding site has been preserved.

Implications for Genome Recognition

As indicated above, deletion of stem loops B, C or D from the MLV Ψ-site leads to significant reductions in the levels of RNA packaged relative

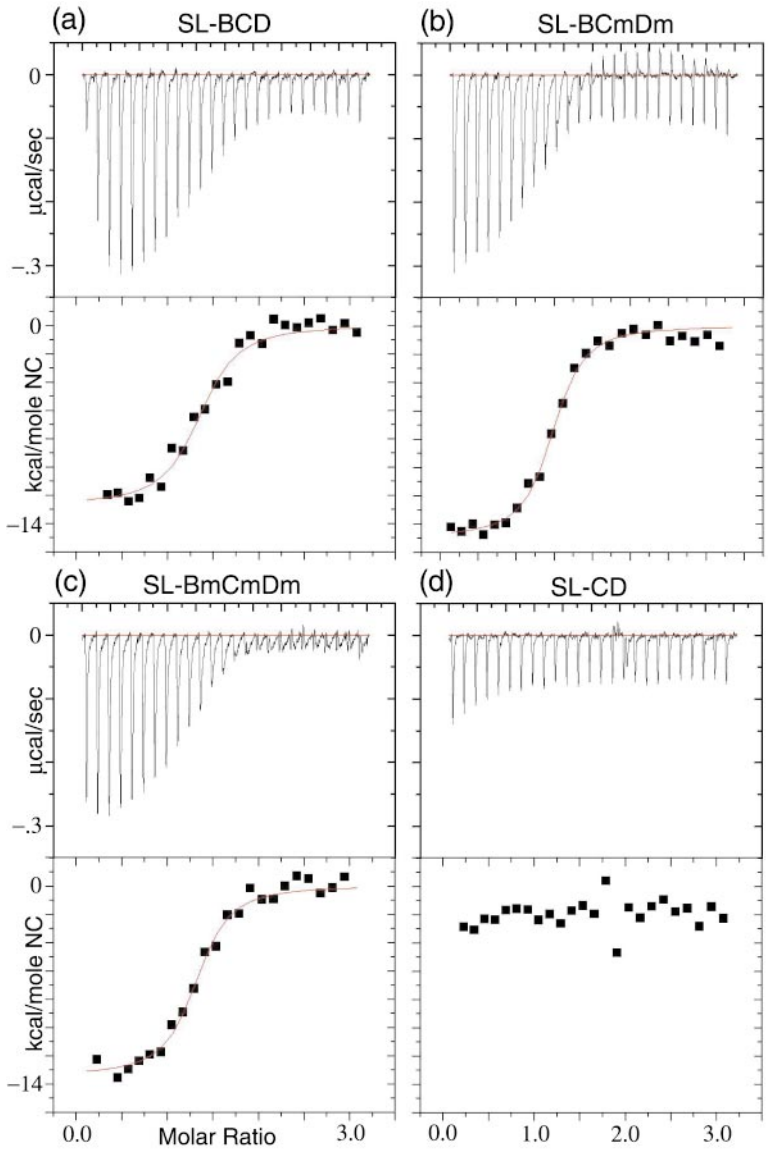

Figure 10. Representative isothermal titration calorimetry data obtained for NC binding to (a) SL-BCD, (b) SL-BCmDm and (c) SL-BmCmDm. As a control, ITC data obtained for titrations of NC to SL-CD, which does not bind tightly to NC, are shown in (d). Top panels: Raw data with each peak corresponding to the heat produced upon addition of a 10 µl aliquot of NC (80-90 µM). Bottom panels: Binding isotherms obtained after peak integration, normalization to molar concentration and subtraction of dilution enthalpies.

to wild-type levels, indicating that all three stem loops are important for efficient packaging.[18,21,22,24] Over the past 15 years, several laboratories have made efforts to define the roles of the individual stem loops. There is now very strong evidence that the mechanistic role of SLB in genome recognition is to facilitate RNA dimerization,[16,25–29] and it has been suggested that dimerization and packaging events are closely coupled.[16,41,42] Our data support the proposal that the earliest stages of genome recognition may involve the binding of the NC-domain of Gag to an equilibrium mixture of monomeric and dimeric Ψ-RNA, and that this binding results in an equilibrium shift towards the dimer form of the Ψ-site.[43,44]

The present data also indicate that stem loop B plays a role in NC binding, since RNA constructs that lack this stem loop exhibit relatively weak affinity towards NC. However, the data presented here indicate that the tetraloop residues of stem loop B in SL-BCD do not contribute to tight NC binding. Thus, NC binds with essentially identical affinities to SL-BCD and SL-BmCmDm, which contain different tetraloop-B sequences. In addition, NC binds with similar affinities to the monomeric and dimeric forms of SL-BCD, despite the fact that the tetraloop-B nucleotides adopt substantially different conformations (i.e. they exist as a tetraloop in the SL-BCD monomer and as an A-helix in the dimer).

The secondary structures of stem loops SL-C and SL-D are conserved among the C-type retroviruses (although the nucleotide sequences vary), and both contain a highly conserved GACG tetraloop. This conservation led to the proposal that stem loops C and D form a physiologically relevant ''double hairpin motif''.[19] Yang and Temin originally showed that mutations designed to disrupt base-pairing in either the upper portion of the SL-C stem or the entire SL-D stem of spleen necrosis

virus (SNV, a C-type retrovirus that is closely related to MLV) lead to significant reductions in viral titers (which correlate with reductions in genome packaging).[24] Wild-type packaging levels could be restored by compensatory mutations designed to restore base-pairing in the stems. However, a double-hairpin construct designed to be structurally similar to the native stem loops, and containing the native loop and bulge nucleotides but different nucleotides in base-pairing positions, was poorly packaged. This indicates that the nucleotide sequence of at least some portions of the stems is important for packaging, and is fully consistent with our findings. In addition, mutation of either of the conserved GACG loops individually to AUAU did not significantly affect packaging, but the simultaneous mutation of these loops, or the mutation of the SL-C loop to GGGG, led to significant reductions in viral titers. These and other results indicated that efficient packaging requires at least one of the conserved GACG tetraloops, that the two stem loops are not functionally equivalent, and that the nature of base-pairing in at least some portions of the stems is important.[24] More recently, Barklis and co-workers demonstrated that the simultaneous deletion of stem loops B and C in MLV reduces genome packaging to ca. 3% of wild-type levels, and that deletion of either individual stem loop leads to a ca. 80% reduction in packaging relative to wild-type levels.[21] In addition, insertion of RNAs containing either stem loops C through D or B through D into non-packaged heterologous RNAs was sufficient to direct their packaging into virions at levels of ca. 50% and 70% of wild-type levels, respectively.[22] Based on these studies, it was concluded that stem loops C and D function as a core encapsidation signal,[22] which is consistent with the proposal for a functional double hairpin motif.

Additional deletion, insertion, and mutagenesis studies are also consistent with the proposal that stem loops SL-C and SL-D play a prominent role in packaging.[18] Thus, mutations designed to disrupt the base-pairing in SL-D led to severe reductions in RNA encapsidation (<5% of wild-type levels) and poor replication kinetics, and compensatory mutations designed to restore base-pairing in the SL-D stem using non-native sequences restored replication kinetics to wild-type levels.[18] Interestingly, similar mutations of nucleotides in the upper portion of stem loop C had little or no effect on replication kinetics. These results indicated that SL-D is more sensitive to mutations than SL-C, and that base-pairing of the upper half of SL-C is not critical for packaging.[18]

The data presented here are generally consistent with the above *in vivo* mutagenesis results, and together can be used to derive a model for the early stages of MLV genome recognition and packaging. First, it is clear that SL-BCD can exist as an equilibrium mixture of monomeric and dimeric species, and it is likely that both monomeric and dimeric forms of the MLV genome are present in the cytosol of the infected cell (although the equilibrium may be shifted greatly toward the dimer). The NMR and mutagenesis data presented here provide strong evidence that the dimer is stabilized by an A-helical duplex formed by residues of stem loop B, and that kissing interactions observed for isolated stem loops C and D do not contribute to the stability of the dimer. These results are fully consistent with previous *in vitro* studies involving much larger RNA constructs, which showed stem loop B is required for the formation of stable MLV RNA dimers, and that mutations of the tetraloop residues of stem loops C and D do not affect the monomer-dimer equilibrium constant.[27] Our data clearly demonstrate that NC is capable of binding with similar affinity to both the monomeric and dimeric forms of SL-BCD, and they confirm previous observations that NC shifts the monomer-dimer equilibrium toward the dimer. Thus, the NC domain of Gag may bind to both monomeric and dimeric Ψ-sites *in vivo*, and thereby shift the equilibrium to the dimer species.

Second, since tight NC binding was observed only for constructs that contain all three stem loops, but not for any of the single- or double-stem loop RNA constructs, it seems likely that all three stem loops contribute to the NC binding site. Of course, it is also possible that NC might bind to only two of the stem loops in a site that is only formed when the third stem loop is present. For example, it is possible that stem loop B positions stem loops C and D in a manner necessary for them to bind tightly to NC. This latter model might explain why heterologous RNAs that contain stem loops C and D are packaged into virus-like particles.[22] The construct employed in those studies included a significant fraction of stem loop B (beginning with residue 285), and it is conceivable that tight NC binding requires only a portion of stem loop B, perhaps even only a few nucleotides, in order to properly orient the stems of SL-C and SL-D for NC binding.

Third, since mutations in the three tetraloops do not affect the affinity of NC for SL-BCD, the binding site probably comprises a pocket or surface that is formed by stem residues. The data argue against a binding mechanism that involves NC interactions with residues of the tetraloops, since binding is uaffected by the substitution of nucleotides in SL-B (AGCU to GAGA), SL-C (GACG to GAGG) and SL-D (GACG to GAGG). These findings are consistent with the previous observation that the C and D tetraloops can be independently mutated (to AUAU) without affecting genome packaging *in vivo*.[24]

Fourth, since SL-BCD binds stoichiometrically to a single NC protein, our data predict that the initial complex that forms *in vivo* and is stable towards packaging may consist of a dimeric Ψ-site with two bound Gag molecules. Additional NC domains of assembling Gag molecules would then bind non-specifically to other portions of the RNA as the virus forms, consistent with the observation

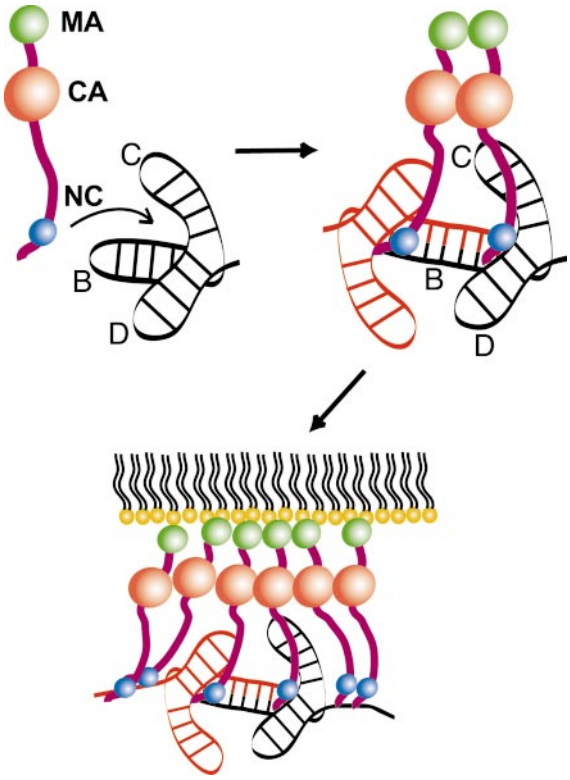

Figure 11. Proposed model for the early stages of MLV genome recognition and packaging that is consistent with the present data and previously published biochemical and *in vitro* mutagenesis results. In this model, the NC domain of the MLV Gag protein binds tightly and stoichiometrically to a high affinity site formed by stem loops -B, -C and -D of Ψ. NC may interact with all three stem loops, or may bind to a site on one or two of the stem loops that is only formed when all three stem loops are present. The bases of the tetraloop nucleotides probably do not contribute to the NC binding site. NC binding shifts the monomer-dimer equilibrium toward the dimer species, facilitating the packaging of two copies of the genome. Intermolecular interactions between the capsid (CA) and/or matrix (MA) domains of Gag, which forms oligomers both in the absence and presence of RNA, may facilitate this process. The condensation of additional Gag molecules likely leads to additional, non-specific and lower-affinity interactions between their NC domains and the RNA. Although the MLV Ψ-site can function as an independent unit in genome packaging, other regions of the genome may contain NC binding sites that contribute to high packaging efficiency.

that both viral and non-viral RNAs can promote particle assembly.[45,46] A diagram showing the general aspects of this packaging model is shown in Figure 11.

Finally, these studies indicate that the mechanisms used by different retroviruses to package their genomes may be substantially different, even though recognition is mediated, at least in part, by a common structural motif (i.e. the CCHC zinc knuckle). Earlier studies demonstrated that HIV-1 NC can bind tightly to isolated stem loops, and that the zinc knuckles of the HIV-1 NC protein interact specifically with exposed guanosine bases in the conserved GGNG tetraloops. In contrast, the MLV NC protein does not bind tightly to isolated stem loops, but instead exhibits high affinity for RNA constructs that contain stem loops B, C and D. In addition, the MLV NC protein appears to bind to a site that is formed by the stems or bulges of the three stem loops, and does not appear to interact with the tetraloops. These findings provide an explanation for the interesting genome packaging data obtained recently for chimeric virions that contain ''swapped'' NC domains. Thus, an HIV-1 mutant, in which the native NC domain was substituted by MLV NC, specifically packaged RNAs containing the MLV Ψ-packaging element.[6] Our present findings are fully consistent with this observation, since the NC proteins of HIV-1 and MLV appear to recognize and bind tightly to targets with completely different structures (i.e. exposed loops in HIV-1 Ψ-site and a stem-derived pocket in MLV). Substitution of the Rous sarcoma virus (RSV) NC domain by that of MLV also led to the preferential packaging of the MLV genome,[47] but the swapping of the NC domains of HIV-1 and mouse mammary tumor virus (MMTV) (in which both NC domains contain two zinc knuckle domains) did not lead to a reversal in genome packaging.[48] As mentioned earlier, the MLV NC domain contains a single CCHC zinc knuckle, whereas the NC proteins of HIV-1, RSV and MMTV all contain two zinc knuckles. It is therefore possible that retroviruses with two zinc knuckles per NC protein select their genomes by a common mechanism that involves NC binding to exposed tetraloops, whereas genome packaging by the C-type retroviruses proceeds *via* NC binding to a pocket formed by multiple stem loops. The present findings lay the groundwork for high-resolution structural studies of the minimal high-affinity NC binding site in the MLV packaging signal, and suggest new *in vivo* mutagenesis experiments to test aspects of the proposed packaging mechanism (underway).

Materials and Methods

Cloning of recombinant MLV NC protein

The recombinant gene for NCp10 was optimized for codon usage by site-directed mutagenesis of pCNA template (kindly supplied by Goff) using a commercial antibiotic selection protocol (GeneEditor Kit; Promega). The mutagenic oligonucleotides, 5′-A A C A G G A T C G T C A G G G C G G C G A A C G T C G T C G T T C C C A A C T C G-3′, 5′-A A A C C A C G T G G C C C T C G T G G C C C A C G T C C G C A G A C C T C C C-3′ were obtained from Gibco BRL. The base alterations in the gene were checked by dideoxynucleotide sequencing. Oligonucleotides MLV-NCf (5′-G C A A G G G A T C C G C C A C T G T C G T T A G T-3′) and MLV-NCr (5′-G T T G T C C T C G A G T T A C A G G A G G G A G G T-3′), containing *Bam*HI and *Xho*I restriction sites,

respectively, were used in a PCR to amplify the NCp10 gene from the optimized pCNA DNA. The PCR was performed with *Ready to Go* PCR beads (Amersham Pharmacia Biotech) and consisted of 35 cycles at 94 °C for one minute, 55 °C for one minute, and 72 °C for 1.5 minutes followed by a final extension at 72 °C for ten minutes. The PCR-amplified fragment was gel-purified with Gel-extration Kit (Qiagen), digested with *Bam*HI and *Eco*RI (New England Biolabs), and ligated with likewise digested vector pGEX-6p-1 (Amersham Pharmacia Biotech) using T4 DNA ligase (New England Biolabs). Plasmid pMLV-NC was then transformed into DH5 *E. coli* using heat shock. Transformants were screened for presence of insert by plasmid mini-preps (*Quick- Spin* DNA Extraction kit, Qiagen) and restriction enzyme digestion. The clone used for overexpression of the fusion protein, pMLV-NC was verified by purifying the plasmid and sequencing the complete insert by the dideoxynucleotide method.

Expression and purification of the MLV NC protein

The fusion protein was expressed in *E. coli*, BL21pLysS and purified by affinity chromatography on glutathione-Sepharose (Amersham Pharmacia Biotech). GST-NC protein expression was induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) (four hours, 37 °C), bacteria were collected by centrifugation (8000 rpm, ten minutes) and were washed once in phosphate-buffered saline (PBS). All protein manipulations were performed at 4 °C. The bacterial pellet was re-suspended in PBS and lysed by sonication. The nucleic acids were precipitated using 10 % polyethylene imine, the debris pelleted by centrifugation (18,000 rpm, 20 minutes) and the resulting supernatant was passed through glutathione-Sepharose beads. The beads were washed with five column volumes of PBS and two column volumes of the cleavage buffer (50 mM Tris-HCl (pH 7.0), 150 mM NaCl, 1 mM DTT). Cleavage of N-terminal GST protein from GST-NCp10 fusion was done overnight using the enzyme PreScission protease (Amersham Pharmacia Biotech). Completion of digestion was monitored by sodium dodecyl sulfate-PAGE. The purified NC protein was exchanged into the buffer (5 mM Tris-HCl, 0.1 mM BME, pH 7) using centricon-3 (Millipore) and concentration was determined by optical absorbance at 280 nM.

Preparation of RNA samples

Plasmid construction

pMLV-BCD: Polymerase chain reaction (PCR) was performed on pCNA which contains the proviral DNA of MLV with oligonucleotide MLV-BCDf (5′-G T C G A A T T C *T A A T A C G A C T C A C T A T A* G G C G G T A C T A G T T A G C-3′), carrying an *Eco*RI site and a T7 RNA polymerase promoter, and oligonucleotide MLV-BCDr (5′-A A A G G A T C C G T T A A C C C T G G G A C G T A T C C C A-3′), carrying a *Bam*HI site and a *Hpa*I site. The *Hpa*I site was added to produce a blunt-end cut after the desired sequence with minimal incorporation of non-native sequences. The amplified product was inserted in the pUC19 plasmid after *Eco*RI and *Bam*HI digestions (Promega). *pMLV-BCmDm*: The plasmid pMLV-BCD (from above) was used as a template for making the mutations in stem loops C and D. Site-directed mutagenesis was carried out by the PCR method. The oligonucleotide MLV-BCmDmr (5′-A A A G G A T C C G T T A A C C C T G G G A C **C** T C T C C C A G G G T T G C G G C C G G G T G T T C C G A A C T C **C** T C A G-3′) was used to introduce the mutations G to C (bold) and also contained the *Bam*HI and *Hpa*I sites required for cloning the mutated SL-BCD insert into pUC19. The oligonucleotide for the 5′primer was MLV-BCDf (above). This fragment was inserted into pUC19 digested with *Eco*RI and *Bam*HI. The resulting plasmid was named pMLV-BCmDm. *pMLV-BmCmDm:* Site-directed muatgenesis using PCR was performed using pMLV-BCmDm as a template to mutate stem loop B. Oligonucleotides used were MLV-BmCmDmf: (5′- A G T G A A T T C T A A T A C G A C T C A C T A C T A T A G G C G G T A C T A G T T **G A G A** A A C T A G C-3′), carrying the mutations AGCU to GAGA (bold) , an *Eco*RI site and the T7 polymerase promoter, and MLV-BmCmDmr (5′-A G A G G A T C C G T T A A C C C T G G G A C-3′), carrying the *Bam*HI and *Hpa*I site. The plasmid derived after inserting the fragment into pUC19 digested with *Bam*HI and *Hpa*I named pMLV-BmCmDm.

Template preparation

All plasmids were amplified using DH5α. The plasmids were extracted form the cells using the *Plasmid Mega Kit* (Qiagen) and linearized with *Hpa*I enzyme (New England Biolabs) at 2 μg/unit for four hours at 37 °C. The cut DNA was then extracted twice with phenol-chloroform and precipitated with ethanol. The pellet was washed with 70 % (v/v) ethanol, and the DNA was dissolved in sterile distilled water. The DNA templates for making RNA SL-B, SL-C, SL-D, SL-BC and SL-CD were ordered from KECK oligonucleotide facility at Yale. The templates were designed with having the T7 promoter at the 3′of the complementary sequence of the transcript required. The templates were mixed with the 1.5 times of the top strand of T7 promoter (5′-T A A T A C G A C T C A C T A T A-3′), heated and then cooled to anneal the two strands so as to form the double-stranded promoter.

RNA synthesis

RNA oligonucleotides were synthesized by *in vitro* transcription using the method previously described[36] with T7 polymerase and purified to a single resolution by gel electrophoresis.[37] The DNA transcripts were transcribed *in vitro* by T7 RNA polymerase in 30 ml reactions containing the optimized amount of template, 20 mM MgCl2, 2 mM spermidine, 80 mM Tris-HCl (pH 8.1), 4 mM of each NTP, 2 mM DTT, and 0.3 mg T7 polymerase. Reactions were incubated for three hours at 37 °C, and quenched with 25 mM EDTA. The RNA was ethanol precipitated, air dried and re-suspended in water. After denaturation at 96 °C for five minutes, the RNA was purified by electrophoresis on urea-containing polyacrylamide denaturing gels. The concentration of each sample was determined by measuring the optical absorbance at 260 nM.

Native polyacrylamide gel electrophoresis

RNA dimerization studies

All RNA constructs except SL-BCD were prepared at a 1 mM stock solution in 5 mM Tris-HCl (pH 7.0). For dimerization the RNA was diluted to different concentrations ranging from 1 mM to 0.1 mM in the dimerization buffer to a final volume of 10 μl. The RNA was

heated at 90 °C for two minutes and chilled for two minutes on ice. The samples were then incubated at 37 °C for 30 minutes. At the end of the incubation time all samples were put on ice and mixed with 2 μl of 50 % glycerol. 4.0 μg of each sample was loaded on the gels, and electrophoresed at 200 V at 4 °C in 1× TB buffer. After staining with Stains-all the gels were photographed using the 1× Kodak camera. Peak areas, corresponding to the monomeric and dimeric forms, were quantified by using the NIH Image 1.62 program, and the fraction of the free monomer was defined as the area of the monomer peak divided by the sum of the areas of the monomeric and dimeric forms.

NC-RNA binding studies

All RNA samples were prepared at 0.5 mM stock solution in 5 mM Tris-HCl (pH 7.0), heated at 90 °C for two minutes and cooled on ice. 0.5 μl of a 2 mM stock NC solution was added to 5 μl of SL-B, SL-C, SL-D, SL-BC and SL-CD, to give a ratio of 1:0.5 RNA:NC. Increasing amounts of 2 mM NC; 0.25 μl, 0.5 μl, 0.75 μl, 1.0 μl, 1.25 μl, 1.5 μl and 1.75 μl were added to 5 μl of SL-BCD, SL-BCmDm and SL-BmCmDm in a final reaction mixture of 7 μl with 5 mM Tris-HCl (pH 7.0). 5 μl of 50 % glycerol was added to each mixture and 3 μg of each sample were loaded onto native gels and electrophoresed at 4 °C in 0.5× TB buffer. The gels were stained with Stains-all and photographed using a Kodak digital camera.

Isothermal titration calorimetry

Dissociation constants for MLV NC binding to SL-BCD, SL-BCmDm and SL-BmCmDm were determined by standard ITC methods using a VP-Isothermal Titration Calorimeter MicorCalorimeter (VP-ITC) (MicroCal Corp. Northampton, MA).[49] RNA and the NC samples containing buffer (5 mM Tris-HCl (pH 7.0), 10 mM NaCl, 0.1 mM BME) were degassed prior to titration. Protein and RNA concentrations were established using UV-Vis absorption measurements. Exothermic heats of reaction (μcal/second) were measured at 30 °C for 25 injections of NC protein (80-100 μM) into 1.41 ml of RNA (5 μM). The heats of dilution were obtained by titrating the identical protein sample into a cell containing sample buffer and subtracted from the raw data prior to analysis. Heats of dilution were typically 0.03 μcal/second and the maximum heats of binding were 0.3 μcal/second, giving an apparent signal-to-noise ratio of 10. The syringe mixing speed was 310 rpm as recommended by the vendor. Binding curves were analyzed by non-linear least squares fitting of the baseline-corrected ITC data to a single binding site model (ITC Origin program, V2.8, MicroCal, Northampton, MA). Binding constants are reported as the mean ± standard deviation obtained from three independent titration experiments.

NMR spectroscopy

NMR data were collected with Bruker AVANCE 800 MHz and DMX 600 MHz spectrometers, processed with NMRPipe/NMRDraw[50] and analyzed with NMRView.[51] NMR data were collected at 35 °C unless otherwise noted. Proton resonance assignments for free and bound RNA constructs were obtained from 2D NOESY ($\tau_m = 200$ ms),[52,53] 2D TOCSY ($\tau_m = 70$ ms)[54–56] and 2D natural abundance ^{1}H-^{13}C HMQC.[57]

Acknowledgments

Supported by NIH grant GM42561. D.H. and K.W. are MARC U*STAR undergraduate scholars (NIH GM08663). Technical support from Dr Rossi Gitti, Robert Edwards, Cindy Finch and Carrie Stover (HHMI, UMBC) is appreciated.

References

1. Berkowitz, R., Fisher, J. & Goff, S. P. (1996). RNA packaging. *Curr. Top. Microbiol. Immunol.* **214**, 177-218.
2. Rein, A. (1994). Retroviral RNA packaging: a review. *Arch. Virol.* **9**, 513-522.
3. Swanstrom, R. & Wills, J. W. (1997). Synthesis, assembly and processing of viral proteins. In *Retroviruses* (Coffin, J. M., Hughes, S. H. & Varmus, H. E., eds), pp. 263-334, Cold Spring Harbor Laboratory Press, Plainview, NY.
4. Berkhout, B. (1996). Structure and function of the human immunodeficiency virus leader RNA. In *Progress in Nucleic Acid Research and Molecular Biology*, vol. 54, pp. 1-34, Academic Press, Inc.
5. Yu, S. S., Kim, J.-M. & 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.
6. Berkowitz, R. D., Ohagen, A., Hoglund, S. & Goff, S. P. (1995). Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric Gag polyproteins during RNA packaging *in vivo*. *J. Virol.* **69**, 6445-6456.
7. Rein, A., Harvin, D. P., Mirro, J., Ernst, S. M. & Gorelick, R. J. (1994). Evidence that a central domain of nucleocapsid protein is required for RNA packaging in murine leukemia virus. *J. Virol.* **68**, 6124-6129.
8. Zhang, Y. & Barklis, E. (1995). Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J. Virol.* **69**, 5716-5722.
9. Mann, R., Mulligan, R. C. & Baltimore, D. (1983). Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell*, **33**, 153-159.
10. Mann, R. & 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.
11. Schwartzberg, P., Colicelli, J. & Goff, S. P. (1993). Deletion mutants of Moloney murine leukemia virus which lack glycosylated gag protien are replication competent. *J. Virol.* **46**, 538-546.
12. Murphy, J. E. & Goff, S. P. (1989). Construction and analysis of deletion mutations in the U5 region of Moloney murine leukemia virus: effects on RNA packaging and reverse transcription. *J. Virol.* **63**, 319-327.
13. Bender, M. A., Palmer, T. D., Gelinas, R. E. & Miller, A. D. (1987). Evidence that the packaging signal of Moloney murine leukemia virus extends into the *gag* region. *J. Virol.* **61**, 1639-1646.

14. Adam, M. A. & Miller, A. D. (1988). Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* **62**, 3802-3806.
15. Armentano, D., Yu, S.-F., Kantoff, P. W., von Ruden, T., Anderson, W. F. & Gilboa, E. (1987). Effect of internal viral sequences on the utility of retroviral vectors. *J. Virol.* **61**, 1647-1650.
16. Prats, A.-C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C., Paoletti, C. & Darlix, J.-L. (1990). Cis elements and trans-acting factors involved in dimer formation of murine leukemia virus RNA. *J. Virol.* **64**, 774-783.
17. Alford, R. L., Honda, S., Lawrence, C. B. & Belmont, J. W. (1991). RNA secondary structure analysis of the packing signal for Moloney murine leukemia virus. *Virology,* **183**, 611-619.
18. Fisher, J. & Goff, S. P. (1998). Mutational analysis of stem-loops in the RNA packaging signal of the moloney murine leukemia virus. *Virology,* **244**, 133-145.
19. Konings, D. A. M., Nash, M. A., Maizel, J. V. & Arlinghaus, R. B. (1992). Novel GACG-hairpin pair motif in the 5′ untranslated region of type C retrovirus related to murine leukemia virus. *J. Virol.* **66**, 632-640.
20. Mougel, M., Tounekti, N., Darlix, J.-L., Paoletti, J., Ehresmann, B. & Ehresmann, C. (1993). Conformational analysis of the 5′ leader and the *gag* initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucl. Acids Res.* **21**, 4677-4684.
21. Mougel, M., Zhang, Y. & Barklis, E. (1996). cis-Active structural motifs involved in specific encapsidation of moloney murine leukemia virus RNA. *J. Virol.* **70**, 5043-5050.
22. Mougel, M. & 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.
23. Tounekti, N., Mougel, M., Roy, C., Marquet, R., Darlix, J.-L., Paoletti, J. *et al*. (1992). Effect of dimerization on the conformation of the encapsidation psi domain of Moloney murine leukemia virus RNA. *J. Mol. Biol.* **223**, 205-220.
24. Yang, S. & Temin, H. M. (1994). A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *EMBO J.* **13**, 713-726.
25. Roy, C., Tounekti, N., Mougel, M., Darlix, J.-L., Paoletti, C., Ehresmann, C. *et al*. (1990). An analytical study of the dimerization of *in vitro* generated RNA of Moloney murine leukemia virus MoMuLV. *Nucl. Acids Res.* **18**, 7287-7292.
26. Girard, P.-M., Bonnet-Mathoniere, B., Mauriaux, D. & Paoletti, J. (1995). A short autocomplementary sequence in the 5′ leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry,* **34**, 9785-9794.
27. De Tapia, M., Metzler, V., Mougel, M., Ehresmann, B. & 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.
28. Ly, H., Nierlich, D. P., Olsen, J. C. & Kaplan, A. H. (1999). Moloney murine sarcoma virus genomic RNAs dimerize *via* a two-step process: a concentration-dependent kissing-loop interaction is driven by initial contact between consecutive guanines. *J. Virol.* **73**, 7255-7261.
29. Oroudjev, E. M., Kang, P. C. E. & Kohlstaedt, L. A. (1999). An additional dimer linkage structure in Moloney murine leukemia Virus RNA. *J. Mol. Biol.* **291**, 603-613.
30. Kim, C.-H. & Tinoco, I., Jr (2000). A retroviral RNA kissing cokmplex containing only two G-C base-pairs. *Proc. Natl Acad. Sci. USA,* **97**, 9396-9401.
31. Amarasinghe, G. K., De Guzman R., N., Turner, R. B., Chancellor, K., Wu, Z.-R. & Summers, M. F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Y-RNA packaging signal. *J. Mol. Biol.* **301**, 491-511.
32. Clever, J., Sassetti, C. & Parslow, T. G. (1995). RNA secondary structure and binding sites for gag gene products in the 5′ packaging signal of human Immunodeficiency virus type 1. *J. Virol.* **69**, 2101-2109.
33. De Guzman R., N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Y-RNA recognition element. *Science,* **279**, 384-388.
34. Colicelli, J. & Goff, S. P. (1988). Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**, 47-59.
35. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucl. Acids Res.* **15**, 8783-9798.
36. Milligan, J. F. & Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51-62.
37. Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991). Synthesis and purification of large amounts of RNA oligonucleotides. *BioTechniques,* **11**, 764-769.
38. Varani, G., Cheong, C. & Tinoco, I., Jr (1991). Structure of an unusually stable RNA hairpin. *Biochemistry,* **30**, 3280-3289.
39. Woese, C. R., Winder, S. & Gutell, R. R. (1990). Architecture of ribosomal RNA: constraints on the sequence of ''tetra-loops''. *Proc. Natl Acad. Sci. USA,* **87**, 8467-8471.
40. Murphy, F. L. & Cech, T. R. (1994). GAAA tetraloop and conserved bulge stabilize tertiary structure of a group-I intron domain. *J. Mol. Biol.* **236**, 49-63.
41. Bieth, E., Gabus, C. & Darlix, J.-L. (1990). A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis *in vitro*. *Nucl. Acids Res.* **18**, 119-127.
42. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. & Barre-Sinoussi, F. (1990). *Cis* elements and *trans*-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**, 689-699.
43. Bonnet-Mathoniere, B., Girard, P.-M., Muriaux, D. & Paoletti, J. (1996). Nucleocapsid protein 10 activates dimerization of the RNA of Moloney murine leukaemia virus *in vitro*. *Eur. J. Biochem.* **238**, 129-135.
44. Girard, P.-M., de Rocquigny, H., Roques, B.-P. & Paoletti, J. (1996). A model of Psi dimerization: destabilization of the C278-G303 stem-loop by the nucleocapsid protein (NCp10) of MoMuLV. *Biochemistry,* **35**, 8705-8714.
45. Yeager, M., Wilson-Kubalek, E. M., Weiner, S. G., Brown, P. O. & Rein, A. (1998). Supramolecular organization of immature and mature murine leukemia virus revealed by electron cryo-microscopy:

implications for retroviral assembly mechanisms. *Proc. Natl Acad. Sci. USA,* **95**, 7299-7304.
46. Muriaux, D., Mirro, J., Harvin, D. & Rein, A. (2001). RNA is a structural element in retrovirus particles. *Proc. Natl Acad. Sci. USA,* **98**, 5246-5251.
47. Dupraz, P. & Spahr, P.-F. (1992). Specificity of Rous sarcoma virus nucleocapsid protein in genomic RNA packaging. *J. Virol.* **66**, 4662-4670.
48. Poon, D. T. K., Li, G. & Aldovini, A. (1998). Nucleocapsid and matrix protein contributions to selective human immunodeficiency virus type 1 genomic RNA packaging. *J. Virol.* **72**, 1983-1993.
49. Wiseman, T., Williston, S., Brandts, J. F. & Lin, L.-N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-137.
50. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR,* **6**, 277-293.
51. Johnson, B. A. & Blevins, R. A. (1994). NMRview: a computer program for the visualization and analysis for NMR data. *J. Biomol. NMR,* **4**, 603-614.
52. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546-4553.
53. Macura, S. & Ernst, R. R. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* **41**, 95-117.
54. Bax, A. & Davis, D. G. (1985). MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* **65**, 355-360.
55. Braunschweiler, L. & Ernst, R. R. (1983). Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521-528.
56. Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988). Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110**, 7870-7872.
57. Bax, A. & Summers, M. F. (1986). ^{1}H and ^{13}C assignments from sensitivity enhanced detection of heteronuclear multiple bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.* **108**, 2093-2094.

Edited by I. Tinoco

(Received 6 August 2001; received in revised form 20 September 2001; accepted 26 September 2001)