Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome

Cristina Gherghe1,2, Tania Lobmo3, Christopher W. Leonard4, Siddhartha A. K. Datta5, Julian W. Bess, Jr., Robert J. Golelick3, Alan Rein3, and Kevin M. Weeks1,2

1Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290; 2HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702-1201; and 3AIDS and Cancer Virus Program, Science Applications International Corporation-Frederick, Inc., National Cancer Institute, Frederick, MD 21702-1201

Edited* by Stephen P. Goff, Columbia University College of Physicians and Surgeons, New York, NY, and approved August 30, 2010 (received for review May 17, 2010)

All retroviral genomic RNAs contain a cis-acting packaging signal by which dimeric genomes are selectively packaged into nascent virions. However, it is not understood how Gag (the viral structural protein) interacts with these signals to package the genome with high selectivity. We probed the structure of murine leukemia virus RNA inside virus particles using SHAPE, a high-throughput RNA structure analysis technology. These experiments showed that NC (the nucleic acid binding domain derived from Gag) binds within the virus to the sequence UCUG-UR-UCUG. Recombinant Gag and NC proteins bound to this same RNA sequence in dimeric RNA in vitro; in all cases, interactions were strongest with the first U and final G in each UCUG element. The RNA structural context is critical: High-affinity binding requires base-paired regions flanking this motif, and two UCUG-UR-UCUG motifs are specifically exposed in the viral RNA dimer. Mutating the guanosine residues in these two motifs—only four nucleotides per genomic RNA—reduced packaging 100-fold, comparable to the level of nonspecific packaging. These results thus explain the selective packaging of dimeric RNA. This paradigm has implications for RNA recognition in general, illustrating how local context and RNA structure can create information-rich recognition signals from simple single-stranded sequence elements in large RNAs.

retrovirus | RNA recognition code | RNA SHAPE chemistry

Expression of a single viral protein, termed Gag, is sufficient for assembly of retrovirus-like particles in mammalian cells. If present in the cell, the viral genomic RNA (vRNA) is selectively packaged into nascent particles; this selectivity is due to a cis-acting packaging signal in the RNA, termed Ψ (1, 2). Remarkably, when no Ψ-containing RNA is present, Gag still assembles efficiently, encapsidating cellular mRNAs nonselectively in place of the vRNA (3–5).

There are many indications that Ψ represents a high-affinity binding site for the Gag protein both in HIV-1 and in simpler retroviruses (6–14). However, the molecular mechanisms underly­ing selective encapsidation of vRNAs are incompletely understood, as are the features that enable Gag to bind preferentially to vRNA rather than to other cellular RNAs. Gag proteins contain several distinct domains, always including matrix (MA), capsid, and nucleocapsid (NC). vRNA packaging is mediated by the multidomain Gag protein, but Gag is cleaved following release of the virus from the cell. The NC domain plays a principal role in interactions with nucleic acids and is largely responsible for the specific interaction between Gag and its cognate viral RNA (12, 13). This domain of Gag is highly basic and contains one or more “zinc knuckles” with a conserved spacing of Zn2+—coordinating cysteine and histidine residues. Mutations that abolish Zn2+ coordination impair selective encapsidation of vRNA during virus assembly (6, 15). In addition, MA domains of many retroviral Gag proteins interact with nucleic acids (16–21) and may also contribute to specific interactions between Gag and vRNA.

When the vRNA is extracted from virus particles, it is found to be a dimer, in which two molecules of the same (positive-strand) polarity are joined together by a limited number of base pairs. There is strong, albeit indirect, evidence that dimerization is linked to packaging (15, 22–25), so that only dimers of vRNA are selectively packaged. The selective packaging of dimers, but not monomers, of MuLV vRNA likely reflects, in part, the exposure of UCUG sequence elements that become specifically accessible in dimers (26). However, MuLV nucleocapsid binds to nearly any sequence of the form NNNG (26–28), and it is not clear how recognition of this simple RNA element with only a single conserved nucleotide might direct selective packaging. Moreover, the minimal sequence required to mediate packaging (23, 29), dimerization (30, 31), and interaction with Gag (32) spans ~170 nucleotides. It has proven to be very difficult to dissociate the contributions of direct protein-RNA interactions from interactions that are required to maintain RNA base pairing and tertiary interactions in this region.

In the present work, we outline a broadly useful approach for determining the protein recognition code for interactions involving simple sequence elements embedded in a large RNA structure. We show that NC binds to two specific UCUG-UR-UCUG motifs within mature MuLV particles; that recombinant MuLV Gag, as well as NC, binds specifically to these motifs in vitro; that Gag binding is notably more selective than NC binding; that highest affinity binding requires that these sequences be presented in a precise structural context; and that these motifs are crucial elements in the vRNA packaging signal.

Results

We previously used SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) to develop a model for the secondary structure of the MuLV dimerization domain, using authentic, dimeric genomic RNA gently extracted from virions (termed the ex virio RNA) (Fig. L4) (33). SHAPE uses a chemical reaction at the RNA 2′-hydroxyl position to measure local nucleotide flexibility (34, 35); flexibility can be reduced either by base pairing or by bound protein. We found that the two RNA strands in the dimer are held together by intermolecular base pairs in two palindromic stretches, termed PAL1 and PAL2, and by G-C base-pairing interactions in a highly conserved double stem-loop motif (SL1-SL2) (31, 33, 36–38). These elements are separated


The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.
†C.G., T.L., C.W.L., and R.J.G. contributed equally to this work.
‡To whom correspondence may be addressed. E-mail: reina@mail.nih.gov or weeks@unc.edu.

This article contains supporting information online at www.pnas.orglookup suppl/doi:10.1073/pnas.1006897107/DSSupplemental.

This research was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health (J.W.B., S.A.K.D., J.W.B., and K.M.W.) and by the National Institutes of Health (R.J.G., K.M.W., S.A.K.D., C.G., and T.L.) (grants CA112944 and GM059788). We thank R.J.G. and T.L. for providing antiretroviral and control express vectors and S.A.K.D. for performing SHAPE experiments. Additional reagents and analytic tools were provided by P. Enquist (C.R., C.W.L., R.J.G., and K.M.W.) and D. Lefkowitz (C.G., T.L., C.W.L., A.R., and K.M.W.). Additional publication data and computer programs are available on the SHAPE website (http://shapeserver.ucsd.edu/).
by two distinct flexible elements, between PAL1 and SL0 and between PAL2 and SL1, respectively (Fig. 1L).

We have now extended these experiments by performing SHAPE on vRNA inside intact virions (termed the in virio state), similar to previous experiments performed on HIV-1 virion RNA (39, 40). We also evaluated vRNA structure within viral particles after exposure of the virions to 2,2′-dithiodipyridine (Aldrichiol-2, AT-2) (39, 41). This compound, a mild oxidizing agent, penetrates the virus particle and compromises NC zinc knuckle structure by disrupting native cysteine-Zn\(^{2+}\) interactions (we term this the AT-2-treated state). NC is thought to bind tightly to RNA within the virion, due in part to this zinc knuckle motif (6–8, 26, 41, 42). Thus, AT-2 treatment disrupts or weakens NC-RNA interactions by oxidation of cysteine residues in NC but has no detectable effect on the exterior of HIV-1 particles (43).

**Structural Analysis of Authentic MuLV Genomic RNA Dimers.** Our analysis of the genomic RNA in virio yielded SHAPE reactivities for >95% of all positions in the dimerization domain (red histograms, Fig. 2A). As expected, each of the elements that stabilize the dimer state—PAL1, PAL2, and SL1-SL2—have low SHAPE reactivities. We then performed analogous experiments using virions that had been pretreated with AT-2, or using deproteinized RNA purified from virions (the AT-2 treated and ex virio states, respectively, Fig. 2A). The PAL1 and PAL2 intermolecular duplexes and the SL1-SL2 domain were unreactive in each of these states, indicating that their lack of reactivity is due to stable RNA–RNA interactions, and not protein–RNA interactions or other features of the intravirion environment.

In strong contrast, clear changes in nucleotide reactivity in the AT-2 treated and ex virio states, relative to the in virio state, occur throughout the regions that link the PAL1, PAL2, and SL1-SL2 structural elements. These differences are readily detected in SHAPE reactivity difference plots (Fig. 2B). AT-2 treatment (Fig. 2B, dark blue) rendered some regions more flexible (nts 220–235, immediately 3′ of PAL1; nts 299–309, just 3′ of PAL2; and nts 338–341, the GGAA bulge in SL1) and others less flexible (nts 246–259 and 272–279). The AT-2 treated profile was similar to that of deproteinized, ex virio, viral RNA (compare histograms in Fig. 2B). Thus, AT-2 treatment evidently disrupts most protein–vRNA interactions within the virus particle in this region of the genome. Sites of strongest protein–RNA interactions inside the virion correspond to regions of negative SHAPE reactivity differences: These are emphasized with bars in Fig. 2B.

**Identification of High-Affinity Binding Sites for MuLV Gag and NC.** Using purified components, we tested the possibility that the sites protected within the virion (Fig. 2) represent high-affinity binding sites for viral proteins. We prepared a 331-nt-long RNA, containing the ~170-nt minimal dimerization active sequence (MiDAS) domain (30, 38) that forms homogenous monomers and dimers and recapitulates the structure of the ex virio dimer [except that one structure, SL0, forms only in the context of the full-length genomic RNA (33)]. Dimers of this transcript were incubated with recombinant MuLV Gag or NC protein; the resulting ribonucleoprotein complexes were then analyzed by SHAPE for comparison with the naked dimeric transcript.

We initially evaluated binding by Gag and NC to the MiDAS RNA dimer under near-physiological ionic conditions (200 mM potassium acetate, 5 mM MgCl\(_2\),) but observed no NC-specific effects and only weak effects of Gag binding. However, at lower ionic strength (40 mM potassium acetate or NaCl, 0.8 mM MgCl\(_2\),), we observed clear structural changes upon addition of Gag and NC. We therefore first formed the dimer under the near-physiological ion condition and then diluted the RNAs to the lower ionic strength prior to protein addition. The dimer structure is retained upon dilution (see Methods). Gag-RNA interactions were readily detected at a ratio of only 5 Gag molecules per MuLV RNA, whereas detection of any effect of NC required addition of 250 NC molecules per each 331-nt RNA strand. Addition of Gag or NC to the MiDAS RNA induced large changes in SHAPE reactivities, and difference plots revealed that these changes occurred predominantly in the regions between PAL1 and PAL2 and between PAL2 and SL1 (Fig. S1).

Together with the data in Fig. 2, the SHAPE data enabled us to make four instructive comparisons: in virio versus (protein-free) ex virio; in virio versus AT-2 treated; and, for the dimer of the 331-nt transcript, the effects of Gag and NC binding. The first two comparisons reflect NC-dependent protections inside intact virions, whereas the latter two report the binding of Gag and NC to transcripts in vitro. We quantified each of these SHAPE-derived comparisons by calculation of a protection factor.

Remarkably, the protection patterns for all four comparisons are highly similar (Fig. 3). For every comparison, there are strong sites of protection immediately 3′ of PAL1 (nts 220–235) and PAL2 (nts 299–309). Both sites contain two copies of the UCUG sequence previously shown to interact with NC in short
RNAs (26, 28). Nucleotides within the tandem UCUG motifs exhibit a conserved pattern, such that the first U and the final G show the strongest protection from SHAPE in the presence of viral proteins (emphasized with red lines, Fig. 3). Although the overall patterns of protection are similar for the four comparisons, there are notable differences at selected positions. First, in addition to conserved protections at the tandem UCUG sequences, positions 220–223 also show protection, but predominantly in the comparison between the in virio and ex virio states and for the experiments employing purified NC (Fig. 3 A and D). This site apparently reflects binding by NC via a mechanism that does not depend upon the zinc knuckle motif because it is not affected by AT-2 treatment (compare panels, Fig. 3 A and B). Second, at the PAL2 tandem binding site (nts 299–309), the protection pattern induced by recombinant NC (Fig. 3D) has a different local pattern than observed for any of the other three comparisons: The first U in the UCUG sequence shows little or no protection with NC, whereas this nucleotide is strongly protected in the other three comparisons.

**Specific Gag Binding to the Dimerization Domain.** We next analyzed the binding affinity of Gag to the full-length, dimeric MiDAS RNA. We evaluated Gag binding affinities by nitrocellulose filter partitioning, using excess tRNA to suppress the nonspecific binding activity of Gag. There is a bimodal pattern of binding by Gag to the wild-type RNA (Fig. 4). Fitting this profile to a model postulating two consecutive, unlinked binding events gave a high-affinity mode with an apparent dissociation constant ($K_{app}$) of ~6 nM, and a second mode with $K_{app}$ ~ 500 nM (circles and solid lines, Fig. 4).

The SHAPE protection data (Fig. 3) suggest that the core recognition element for Gag and NC has the consensus UCUG-UR-UCUG. To test the role of the tandem UCUG motifs in the binding of Gag to these dimeric transcripts, we mutated both $G_4$ residues in the 5’ repeat (mutant M1), in the 3’ repeat (M2), or in both repeats (M1M2) (see Fig. 1A). The higher-affinity binding mode is only 2-fold weaker in M1 and M2 (triangle symbols, Fig. 4), but ~6-fold weaker for M1M2 RNA than for the wild-type control (squares, Fig. 4). Binding in the second, weaker phase is similar for all of the RNAs. As a further test of the specificity of Gag binding to the MiDAS dimer, we also monitored the binding of (lysine)$_{12}$ to the wild-type and mutant RNAs. We found (diamonds, Fig. 4) that this basic peptide binds poorly to all four RNAs, with $K_{app}$ > 1 μM. This polycation thus appears to
mimic the low-affinity but not the high-affinity mode of Gag binding; its binding is independent of the UCUG motif.

**Defining a Minimal Gag Binding Site.** We next evaluated Gag-RNA interactions in the context of simplified RNAs that limit opportunities for nonspecific interactions. Both instances of the UCUG-UR-UCUG motif occur in similar structural contexts in the authentic viral RNA. In each case, the motif is flanked by base-paired regions: The first motif is flanked on the 5′ side by the PAL1 intermolecular duplex and on the 3′ side by SL0. The second motif is flanked on the 5′ side by the PAL2 duplex and on the 3′ side by SL1 (Fig. 1A). To assess the contributions of these structural elements for specific recognition by Gag, we evaluated two recognition site (RS) RNAs. As a monomer, each RS RNA spans one of the tandem UCUG motifs and its flanking double-stranded regions (termed the PAL1-RS and PAL2-RS RNAs).

Gag binds to PAL1-RS and PAL2-RS with $K_d$'s of 120 and 110 nM, respectively (constructs 1–1 and 2–1, Fig. 5). This 20-fold reduction in affinity, relative to the dimeric MiDAS construct ($K_{app,1} ∼ 6$ nM), likely reflects that there are only one-fourth as many Gag binding sites in each RS RNA; the simplified RNAs may also be too short to support cooperative binding. The base-paired regions that flank the tandem UCUG elements are critical for high-affinity Gag interaction, as short single-stranded RNAs containing only the UCUG-UR-UCUG motif bind Gag very weakly ($K_d > 1$ μM, constructs 1–0 and 2–0, Fig. 5).

We evaluated the contributions of individual elements within the PAL1-RS and PAL2-RS RNAs using an instructive set of short RNAs derived from the two primary RS RNAs (Fig. 5 C and D). PAL1-RS and PAL2-RS have slightly different sequences in the 2-nT element that links the two UCUG motifs (UG and UA, respectively). The G residue in the 2-nT linker element does not contribute to Gag recognition (constructs 1–2 and 2–2, Fig. 5 C and D). Subsequent mutations in the PAL1-RS were tested in the context of a UA linker sequence.

Mutating both $G_1$ positions in the PAL1-RS RNA weakened Gag binding by 3- to 5-fold (constructs 1–3 and 1–4, Fig. 5C), and the equivalent change in PAL2-RS RNA reduced affinity 23-fold (construct 2–3). Replacing the first $G_1$ in PAL2-RS had a larger effect than changing the second $G_1$ (constructs 2–2a and 2–4b, Fig. 5D). Mutating $U_1$ in both UCUG sequences did not significantly reduce binding to PAL1-RS (construct 1–5), but produced a reproducible change of 3.5-fold in binding to PAL2-RS (construct 2–5). In each case, the effect of replacing both $U_1$ and $G_1$ was not significantly different from that of replacing $G_1$ alone (constructs 1–6 and 2–6, Fig. 5). Finally, we tested the role of the flanking base-paired duplexes. In PAL1-RS, eliminating SL0 decreased affinity ∼10-fold, whereas eliminating PAL1 had a smaller effect (constructs 1–7 and 1–8); in PAL2-RS, removing either PAL2 or SL1 reduced affinity ∼6-fold (constructs 2–7 and 2–8).

**Effect of Mutation of UCUG Motifs upon Packaging of Viral RNA.** The in vitro and in vivo biochemical studies described above strongly suggest that a small number of interaction sites comprise key binding structures for Gag and NC. We therefore evaluated the role of the tandem UCUG motifs in encapsidation of viral RNAs. We generated mutations in these motifs in a MuLV-derived luciferase vector whose first ∼1,040 nt are nearly identical to MuLV (44) and compared the encapsidation efficiencies of the mutant and native sequence RNAs. The G nucleotides were replaced with A in the 5′ UCUG repeats (M1), in the 3′ UCUG repeats (M2), and in both tandem repeats (M1M2, see Fig. 1D). These mutations were introduced with a C311U change, designed to maintain native base pairing in the vRNA monomer. The C311U mutation has no effect on packaging (Fig. 6). Each of the mutant and wild-type vectors was transiently transfected into 293T cells, together with an infectious MuLV plasmid clone. Culture fluids and cells were harvested and luciferase RNA levels in the released particles and in the cells were quantified by real-time RT-PCR. For each culture, the encapsidation efficiency was calculated as the luciferase copies/ng RNA in the viral sample divided by the luciferase copies/ng RNA in the cells (5).

Changing the UCUG motifs in the 5′ tandem repeat (mutant M1) had a small, 4-fold, effect on encapsidation efficiency, whereas mutation of the 3′ repeat reduced encapsidation efficiency ∼12-fold (Fig. 6). However, changing all four UCUG elements drastically reduced encapsidation of the vector: The encapsidation efficiency of this mutant was ∼200-fold lower than that of the native sequence control (M1M2, Fig. 6). These results imply that G residues in the UCUG motifs are crucial elements in the MuLV packaging signal, but that there is some redundancy in this signal: The presence of either the 5′ or the 3′ motif is sufficient for partial encapsidation of vRNA.

**Discussion**

We have developed, and then applied, several unique experimental approaches to explore the signal that governs vRNA packaging in the prototypical gammaretrovirus, MuLV. Specifically, we analyzed the secondary structure of the vRNA within infectious virions, probed the effects of NC upon this structure inside the virion, and defined nucleotide–protein interactions in vitro (Figs. 2 and 3). These experiments all pointed to the motif, UCUG-UR-UCUG, as a specific binding site for Gag and NC and indicated that NC is bound to this motif within the virion.

We then analyzed the binding of recombinant Gag to short RNAs containing portions of the $Ψ$ region. This binding is a specific interaction between MuLV Gag and the RNA, because it is not seen with the control basic peptide (lysine)$_{25}$, and is diminished if individual G nucleotides in the motif are replaced by A residues (Fig. 4). High-affinity binding of Gag to the RNA requires the presence of the 10-base tandem repeat motif (which occurs twice within $Ψ$), which must also be flanked by base-paired regions (Fig. 5).

Prior work has shown that the MiDAS RNA dimer contains true long-range tertiary interactions involving PAL1, PAL2, and SL1-SL2 (38, 45). Thus, the simplest element in the packaging signal involves both the primary sequence and the tertiary architecture of the RNA. In particular, PAL2 likely packs against the SL1-SL2 domain, which would cause the UCUG-UR-UCUG motif to be presented to Gag as a loop (illustrated in Fig. 1B). This model is also supported by a prior yeast three-hybrid study that concluded that MuLV positions 212–354 are necessary for high-affinity binding by Gag (32).

We then tested the functional role of this motif in the specific packaging of vRNA. We found (Fig. 6) that when all four UCUG
elements within the two UCUG-UR-UCUG motifs were changed to UCUA, the efficiency with which genomic RNA was packaged was reduced by ∼200-fold. A comparable effect is obtained by deleting large regions of the packaging signal (150–350 nts) or by introducing mutations that compromise dimerization (29, 46). These results have important implications for the mechanism of vRNA packaging and for RNA recognition, in general.

First, mutating only four bases in a retroviral RNA leads to a drastic effect on encapsidation in a dimerization-competent RNA (Fig. 6). The large size of the region identified previously as necessary for packaging reflects both the requirement to maintain a specific three-dimensional architecture and to present single-stranded elements for recognition by Gag. Similar principles also appear to govern selective packaging of HIV-1 vRNA: The strongest SHAPE-detected effects of HIV-1 NC protein binding to the viral RNA occur at single-stranded elements flanked on both sides by base-paired elements (39).

Second, our results show that, in MuLV, the binding specificity of Gag for RNA differs from that of NC in at least two critical ways. Whereas a near-stoichiometric number of Gag molecules is sufficient to bind the UCUG-UCUG motif, a large excess of NC is required to induce a similar change in RNA structure as measured by SHAPE (Fig. S1). In addition, NC and Gag binding induce distinct changes in RNA structure, especially at positions 220–223 and 299–309 (Fig. 3). Overall, Gag binds much more selectively to the dimeric RNA than does NC, probably due to molecular cooperativity, the presence of MA, or both.

Third, this work rationalizes how recognition of a simple 4-nt sequence element (26, 28), with a low information content, could mediate specific packaging of an entire viral RNA genome. It was proposed many years ago that dimeric RNAs are selectively packaged (15, 47, 48), and an important initial proposal emphasized that dimerization of MuLV vRNA constitutes a switch that exposes UCUA sequences (25, 26, 33). In fact, Gag actually binds weakly ($K_d > 1 \mu M$) to short single-stranded RNAs containing only this motif (Fig. 5, constructs 1-0 and 2-0). Thus, the full RNA recognition code is an information-rich structure involving tandemly repeated motifs positioned between flanking base-paired elements (Fig. 1B).

Diverse cellular regulation processes are mediated by RNA binding proteins that, like NC, recognize short, often degenerate, sequence elements (49). This work illustrates how the full recognition code for this class of proteins is linked to the underlying RNA structure and also outlines broadly applicable functional tools for dissecting this code.

**Methods**

**In Vitro Probing of MuLV Genome Dimer Structure by SHAPE.** Moloney MuLV particles were resuspended in HFS buffer (2 mL; 50 mM Hepes (pH 8.0), 200 mM NaCl, 0.1 mM EDTA, 10% (vol/vol) fetal bovine serum), divided into two equal aliquots, and treated with either Aldrithiol-2 (AT-2, 2,2’-dithioldipyridine) in DMSO (2 μL of 0.5 mM stock) or DMSO alone and incubated overnight at 4 °C. Virions were then purified over a sucrose cushion, resuspended in 1 mL HFS buffer, divided into two aliquots, and treated either with N-methyl isatoic anhydride (NMIA, 50 μL of 100 mM in DMSO) or neat DMSO.

**In Vitro SHAPE Analysis of MIDAS Dimers in the Presence of Gag and NC.** MuLV dimers were formed from a 331-nt RNA (33). Purified recombinant MuLV Gag and NC were incubated with MuLV dimers (18 μL in 40 mM NaCl or potassium acetate for Gag and NC, respectively, and 0.8 mM MgCl$_2$) and were treated with 1-methyl-7-nitroisatoic anhydride (1MM, 2 μL; 2 mM in anhydrous DMSO) or with neat DMSO.
Fig. 6. Normalized encapsidation efficiencies for MuLV-derived pBabe-Luc RNAs containing native sequence and mutant \( \Psi \) domains. Geometric means and standard deviations are shown.

Detection of NMIA and 1M7 Modifications. Sites of 2'-O-actylation formation in the authentic MuLV genome or in simplified transcript RNAs were analyzed by capillary electrophoresis using fluorescently labeled DNA primers and reverse transcriptase-mediated primer extension (33).


Supporting Information

Gherghe et al. 10.1073/pnas.1006897107

SI Methods.

Virus Preparation and Source of the Ex Virio and In Vitro RNAs. Virus was obtained from clarified cell cultures (10 L) using an NIH3T3-derived cell line constitutively producing MuLV, as described (1, 2). Purified virus was resuspended in 0.01 M Tris-HCl (pH 7.2), 0.1 M NaCl, and 1 mM EDTA at final concentration of 1000× relative to the starting culture volume; aliquots were stored at −70°C. The ex virio RNA was extracted from virions as described (2). The 331-nt in vitro MuLV RNA structure spanned the ~170-nt minimal dimerization active (MiDAS) region plus flanking 5′ and 3′ viral sequences of 46 and 115 nucleotides, respectively (2).

In Virio and Aldriothiol-2 (AT-2)-Treated RNA Dimer Structures. For analysis of the MuLV genomic RNA inside intact virions, the virus concentration (10 mL) was initially treated with subtilisin (3) (total volume 20 mL; 25°C, overnight) to remove copurifying membranous particles (4). Subtilisin-treated virions were pelleted by ultracentrifugation (Beckman SW-41 rotor; 37,000 rpm, 4°C, 1.5 h) through a 20% (wt/vol) sucrose cushion in PBS. Pellets were resuspended in HFS buffer [2 mL; 50 mM HEPES (pH 8.0), 200 mM NaCl, 0.1 mM EDTA, 10% (vol/vol) fetal bovine serum], divided into two equal aliquots that were treated with either Aldriothiol-2 (AT-2, 2.2-‘dithioldipyridine) in DMSO (2 μL of 0.5 μM stock) or DMSO alone (2 μL) and incubated overnight at 4°C. Virions were then pelleted through a 2-ML 20% (wt/vol) sucrose cushion (Beckman SW-60 rotor; 54,000 rpm, 1.5 h), and each pellet was resuspended in 1 mL HFS buffer and divided into two aliquots. Samples were then treated either with 1 mM isatoic anhydride (NMA, 50 μL of 100 mM in DMSO) or neat DMSO. RNA was recovered from all samples by lysis with proteinase K (190 μg/mL) in 50 mM Tris (pH 7.5), 1 mM EDTA, 10 mM DTT, 120 μg/mL glycerol, and 1% (wt/vol) SDS (25°C, 30 min), followed by extractions with equal volumes of phenol:chloroform:isoamyl alcohol (25:24:1, 10 times), and chloroform (twice). RNA was precipitated in 70% ethanol with 0.3 M sodium acetate and stored at −20°C. RNA amounts were quantified by real-time reverse transcription PCR (2).

In Vitro SHAPE Analysis of Minimal Dimerization Active (MiDAS) Dimers in the Presence of Gag and NC. MuLV dimers were formed from a 331-nt RNA spanning the ~170-nt MiDAS region plus flanking 5′ and 3′ viral sequences of 46 and 115 nucleotides, respectively (2). Recombinant MuLV Gag and NC were purified as described (5, 6), with the exception that MuLV Gag was initially precipitated from the bacterial lysate by addition of 0.66 vol saturated ammonium sulfate and final purification of Gag included an additional size exclusion chromatography step (Superose 12 column, GE). Proteins were stored frozen at −80°C (storage buffers are given below). The in vitro RNA construct (4 μL, 4 pmol) was denatured at 95°C for 3 min, snap cooled on ice for 2 min, and treated with 3× dimerization buffer [2 μL: 150 mM HEPES (pH 7.5), 600 mM potassium acetate (pH 7.5), 15 mM MgCl₂]. The sample was incubated at 60°C for 30 min, then diluted with 25 μL 50 mM HEPES (pH 8). The resulting RNA dimer was equilibrated with either Gag [0.5 μL; 40 μM (one protein per 66 nts)] or NC [5 μL; 200 μM (one protein per 1.3 nts)] at 37°C for 20 min. Control reactions contained the Gag [20 mM Tris-HEPES (pH 7.5), 0.5 M NaCl, 1 mM PMSF, 5 mM DTT] or NC [50 mM HEPES (pH 8), 33.3 mM potassium acetate (pH 8), 1 mM tris(2-carboxyethyl)phosphine, 100 mM ZnCl₂] storage buffers. Final concentrations of monovalent and divalent ions are similar for both protein-RNA complexes (40 mM NaCl versus potassium acetate for Gag and NC, respectively, and 0.8 mM MgCl₂); the slightly different buffer compositions do not cause detectable differences in RNA dimer structure. Protein-RNA dimer complexes (18 μL) were treated with 1-methyl-7-nitroisatoic anhydride (IM7, 2 μL; 2 mM in anhydrous DMF) (7) or with neat DMSO (37°C, 2 min). Modified RNA (20 μL) was diluted in H₂O (to 100 μL) and treated with EDTA (4.67 mM) and AT-2 (2 mM, 25°C, 10 min). Samples were incubated with proteinase K (0.53 mg/mL) in SDS (0.55% [wt/vol], 25°C, 30 min), followed by phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation.

Analysis of NMIA- and IM7-Modified RNA. To map the in virio RNA, we used the SHAPE reagent NMIA (N-methylisatoic anhydride), previously shown to penetrate virion membranes (8). NMIA reacts with RNA over 30 min. For in vitro analyses, we used IM7 (1-methyl-7-nitroisatoic anhydride) (2, 7), a more reactive electrophile (reaction complete in 70 s). IM7 and NMIA give essentially identical results in vitro. Sites of 2′-O-ADDuct formation in the authentic MuLV genome or in the simplified transcript RNA were analyzed using fluorescently labeled DNA primers and reverse transcriptase-mediated primer extension, exactly as described (2). Raw sequence traces were corrected for variation in dye intensities and signal decay, peak intensities were integrated using ShapeFinder (9). Absolute SHAPE reactivities were normalized to a scale spanning 0 to −1.5 (2), where 0 is defined as the mean intensity of highly reactive nucleotides.

Binding Affinities and Dimerization Controls for MiDAS RNA Constructs. Equilibrium dissociation constants were measured using a dual filter system (10) in 50 mM HEPES (pH 7.6), 40 mM potassium acetate (pH 7.7), 0.8 mM MgCl₂, 0.2 mM DTT, 100 μM MBSA, and 0.01% (vol/vol) Triton X-100 and containing excess yeast tRNA⁰⁰⁰ (900 nM, Sigma-Aldrich) as a nonspecific competitor. Reactions contained trace (0.10 mM) [³²P]-labeled RNA and 1–1000 nM Gag. Binding data for the intact dimerization domain were fit assuming two independent sites, fraction RNA bound = A[Gag]/(1+[Kd,app,1] + B[Gag]²/(1+[Kd,app,2]²), where Kd,app,1 and Kd,app,2 are the two apparent dissociation constants (Kd,app,2 was approximated as 500 nM), n ≥ 2.5, and A and B are the amplitudes for each transition and sum to 1.0. The M1, M2, and M1M2 dimers were all indistinguishable from the native sequence dimer with respect to mobility as analyzed by nondenaturing gel electrophoresis (11, 12). For the PAL(1 or 2)-RS constructs, data were fit to a simple single-site equation: fraction RNA bound = A[Gag]/(1+[Kd,app]), where Kd is the equilibrium dissociation constant and A is the fraction RNA bound at saturating Gag concentrations.

Viral Packaging Experiments. Encapsulation was measured using pBabe-Luc (13), a derivative of a pBabe MuLV-based vector (14). Although the Ψ region of this vector is partially derived from murine sarcoma virus (14), it is nearly identical to MuLV over the 5′ 1,040 nt and differs from MuLV at two positions within MiDAS. Mutants in the Ψ region of the pBabe-Luc plasmid were generated by site-directed mutagenesis (Stratagene QuikChange). Mutations were introduced in the context of a reference vector that contained a C311U change to maintain the stability of the ‘anchoring helix’ in the M2 mutant (see Fig. S2). This mutation has no effect on packaging. Five-μg mutant or native sequence pBabe-Luc plasmid were mixed with 5-μg plasmid DNA of an infectious MuLV clone, and the mixture was cotransfected into
293T cells (using TransIT-293, Mirus). Twenty-four-hour culture supernatants were collected 48 and 72 h after transfection. Culture fluids were filtered through 0.45-μm filters and virus particles were isolated by pelleting through 20% (wt/vol) sucrose in TNE [10 mM Tris-HCl (pH 7.4), 100 mM NaCl, 1 mM EDTA]. The pellets were resuspended in TNE and RNA was isolated by incubation at 55°C for 1 h in lysis buffer [50 mM Tris-HCl (pH 7.4), 100 mM NaCl, 10 mM EDTA, 1% SDS, and 100 μg/ml Proteinase K], followed by phenol-chloroform extraction, ethanol precipitation, and DNase I treatment (Fermentas; 37°C, 30 min).

In parallel, RNA was extracted from the transfected 293T cells (Ambion RiboPure), and the purified RNA was suspended in TE [10 mM Tris-HCl (pH 8.0), 1 mM EDTA]. RNAs were quantified using a Nanodrop spectrophotometer (ThermoFisher). Luciferase RNA copies were quantified by real-time reverse transcription-PCR [primer-probe set: LucF: CACTG AGACT ACATC AGCTA TTTCGT; LucR: GCCTC TCTTG TAAC GCCCA GGC; and LucProbe, 6FAM-TGGT AAGGT TGTTC CATT TTGTGA AGCGA AGTT G-TAMRA] using a TaqMan RNA-to-Ct kit (Applied Biosystems) in a DNA Engine Opticon 2 Real-Time Cycler (MJ Research). Encapsulation efficiencies were calculated as the copies/ng RNA in the viral RNA divided by copies/ng RNA in the cellular RNA. Encapsulation efficiencies for the mutants were divided by that for native sequence pBabe-Luc to generate normalized encapsidation efficiencies; data from two to four experiments were summarized as the geometric mean and standard deviation.

Fig. S2. Structural changes that accompany the monomer to dimer transition in the MuLV RNA genome (2). Sequences comprising the tandem UCUG elements identified in this work are boxed. The C311U mutation, introduced to maintain the structure of the “anchoring helix” (11, 12) in the monomer is circled.