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

Cristina Gherghe[a,1], Tania Lombo[b,1], Christopher W. Leonard[a,1], Siddhartha A. K. Datta[b], Julian W. Bess, Jr.[c], Robert J. Gorelick[c,1], Alan Rein[b,2], and Kevin M. Weeks[a,2]

[a]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290; [b]HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702-1201; and [c]AIDS 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 Zn^{2+}-coordinating cysteine and histidine residues. Mutations that abolish Zn^{2+} 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. 1*A*) (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

Author contributions: C.G., C.W.L., R.J.G., A.R., and K.M.W. designed research; C.G., T.L., C.W.L. and R.J.G. performed research; S.A.K.D. and J.W.B. contributed new reagents/analytic tools; C.G., T.L., C.W.L., A.R., and K.M.W. analyzed data; and C.G., A.R., and K.M.W. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]C.G., T.L., C.W.L., and R.J.G. contributed equally to this work.

[2]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.org/lookup/suppl/doi:10.1073/pnas.1006897107/-/DCSupplemental.

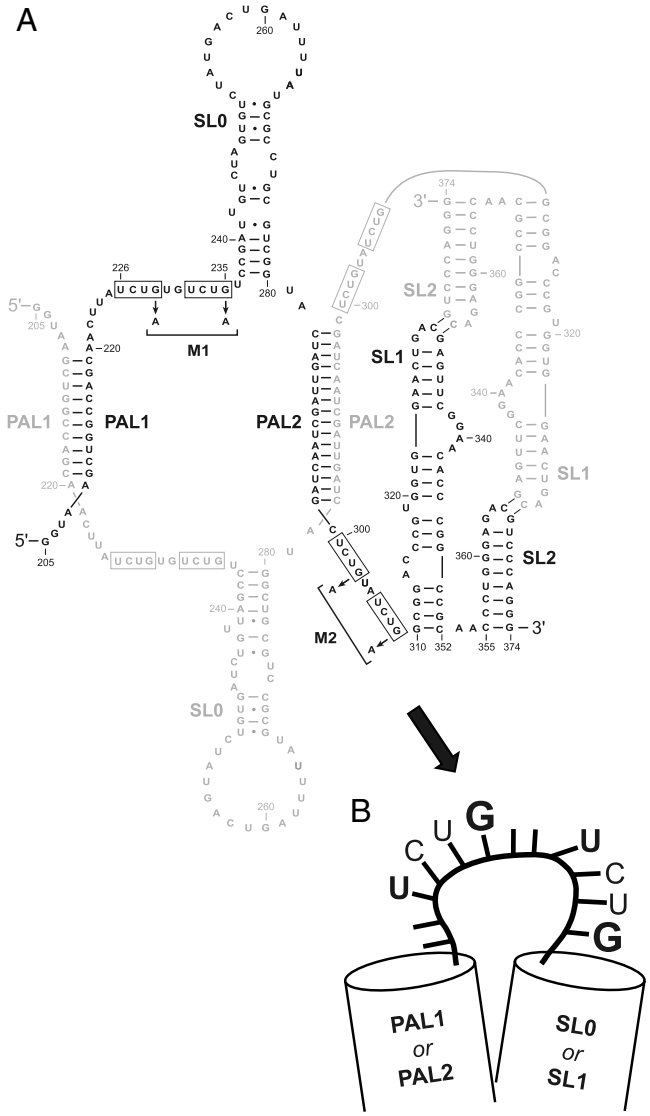

Fig. 1. Structure of the MuLV dimerization domain in the dimer state. (*A*) Secondary structure. The two RNA strands are shown in black and gray. Major structural elements are labeled. Protein-binding sites at UCUG tandem sequences identified in this work are boxed. (*B*) Local structure at each recognition site based on binding experiments (Fig. 5) and on prior work showing extensive helix packing and tertiary structure in this region of the MuLV RNA (38, 45).

by two distinct flexible elements, between PAL1 and SL0 and between PAL2 and SL1, respectively (Fig. 1*A*).

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 (Aldrithiol-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. 2*A*). 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. 2*A*). 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. 2*B*). AT-2 treatment (Fig. 2*B*, 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. 2*B*). 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. 2*B*.

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

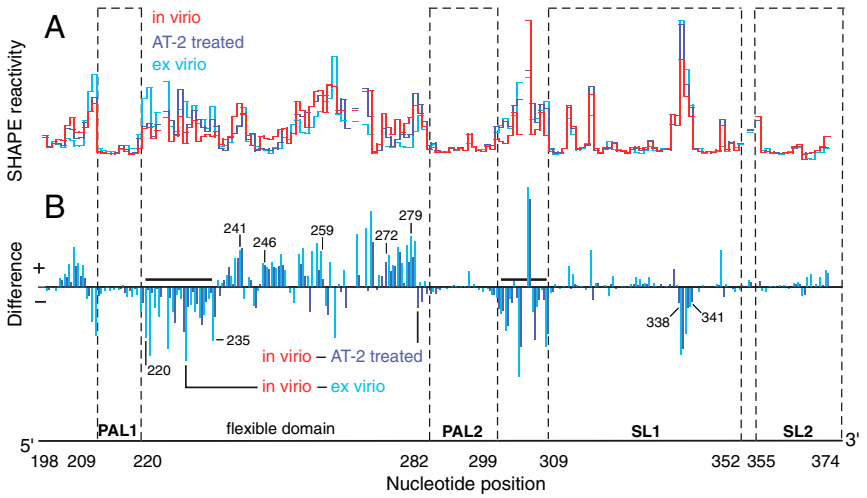

Fig. 2. Structural differences in the Ψ domain as a function of genomic RNA state. (*A*) SHAPE reactivity histograms for the intact in virio (red), AT-2 treated (dark blue), and ex virio (light blue) vRNA. Broken lines indicate a small number of nucleotides that were not analyzed due to high background. (*B*) Difference plot calculated by subtracting the ex virio (light blue columns) or AT-2 treated (dark blue columns) experiments from the in virio data. Positive and negative amplitudes indicate nucleotides that show greater or lesser flexibility, respectively, in virio as compared to the other two states. The two strongest sites of increased reactivity in the AT-2 treated and ex virio states are emphasized with bold lines.

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. 3*D*) 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 ($K_{app,1}$) of ~6 nM, and a second mode with $K_{app,2} \sim 500$ 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. 1*A*). 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 $(\text{lysine})_{25}$ 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

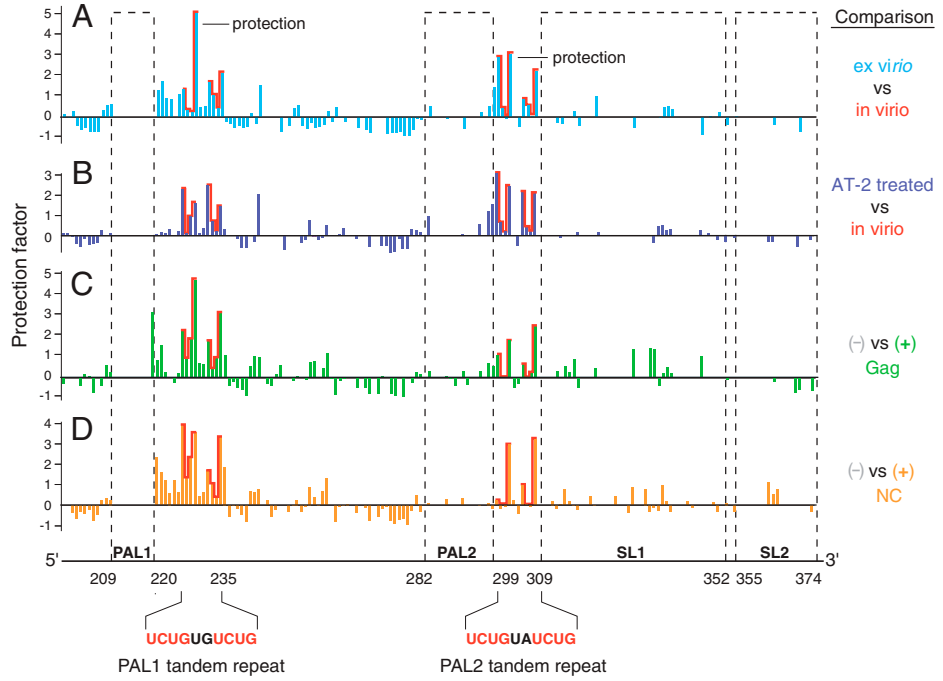

Fig. 3. Identification of specific protein binding sites in the MuLV Ψ region. Protection factors correspond to $(I_- - I_+)/I_+$, where I_+ and I_- are SHAPE reactivities in the presence and absence of protein, respectively. Conserved reactivity patterns at UCUG sequences are outlined in red. Nucleotides that were unreactive both before and after protein addition (reactivity less than 0.15 SHAPE units) are omitted.

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. 1*A*). 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 ($K_{app,1} \sim 6$ 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_4 positions in the PAL1-RS RNA weakened Gag binding by 3- to 5-fold (constructs 1–3 and 1–4, Fig. 5*C*), and the equivalent change in PAL2-RS RNA reduced affinity 23-fold (construct 2-3). Replacing the first G_4 in PAL2-RS had a larger effect than changing the second G_4 (constructs 2-4a and 2-4b, Fig. 5*D*). 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_4 was not significantly different from that of replacing G_4 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).

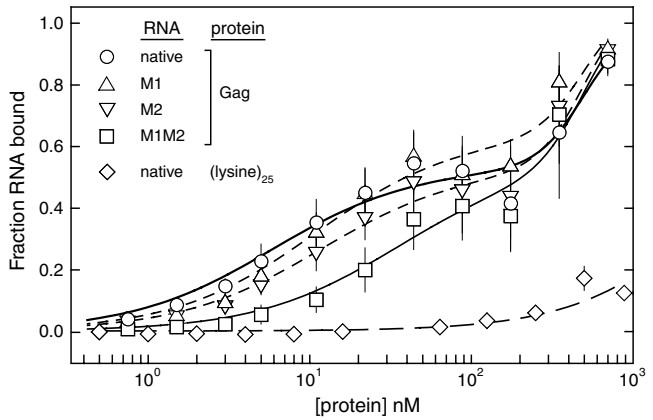

Fig. 4. Gag binding to the full-length MuLV dimerization domain. Data points, shown as the mean and standard deviation for four replicates, were fit to an equation for two independent binding events; R^2 is ≥0.97 in all cases. $K_{app,1}$ for the native, M1, M2, and M1M2 RNAs are 6 ± 2, 11 ± 3, 12 ± 3, and 35 ± 12 nM, respectively.

Effect of Mutation of UCUG Motifs upon Packaging of Viral RNA. The in virio and in vitro 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 an 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. 1*A*). 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. 1*B*). 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

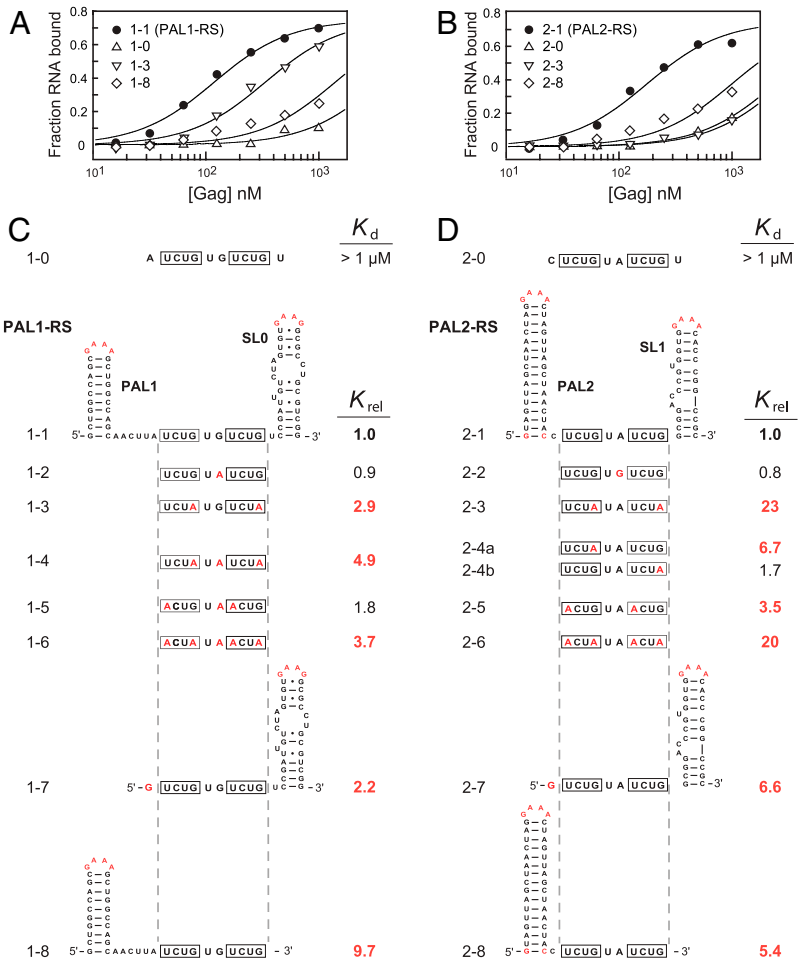

Fig. 5. Defining the minimal RNA binding motif for Gag. (*A* and *B*) Representative analysis of Gag binding to the PAL1-RS and PAL2-RS RNAs. K_d's for PAL1-RS and PAL2-RS (constructs 1-1 and 2-1) are 120 and 110 nm, respectively. (*C* and *D*) Effect of point mutations on Gag binding to PAL1-RS and PAL2-RS. Mutated sites relative to the native sequence are shown in red. Binding affinities for the mutants are reported as K_{rel}, equal to $K_d^{[mutant]}/K_d^{[(PAL(1/2)-RS]}$; larger values indicate weaker binding. K_{rel} values larger than 2 are emphasized in red.

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-UR-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 UCUG sequences (25, 26, 33). In fact, Gag actually binds weakly ($K_d \geq 1$ μM) 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. 1*B*).

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 Virio 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 (1M7, 2 μL; 2 mM in anhydrous DMSO) or with neat DMSO.

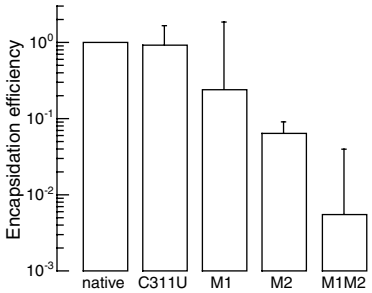

Fig. 6. Normalized encapsidation efficiencies for MuLV-derived pBabe-Luc RNAs containing native sequence and mutant Ψ domains. Geometric means and standard deviations are shown.

Detection of NMIA and 1M7 Modifications. Sites of 2′-*O*-adduct 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).

Gag Binding Affinities and Dimerization Controls for MiDAS RNA Constructs. Equilibrium dissociation constants were measured using a dual filter system in 50 mM Hepes (pH 7.6), 40 mM potassium acetate (pH 7.7), 0.8 mM $MgCl_2$, 0.2 mM DTT, 100 μg/mL BSA, and 0.01% (vol/vol) Triton X-100 and containing excess yeast $tRNA^{Phe}$ and trace (0.10 nM) [^{32}P]-labeled RNA. Binding data for the intact dimerization domain were fit assuming two independent sites; for the PAL(1 or 2)-RS constructs, data were fit to single-site equation.

Viral Packaging Experiments. Encapsidation efficiencies for native sequence and Ψ region mutants were measured using pBabe-Luc (44), a derivative of a MuLV-based vector.

Additional details regarding the methods for vRNA isolation, SHAPE analysis and data processing, Gag binding experiments, and virus packaging are available in the *SI Text*.

ACKNOWLEDGMENTS. We thank Jane Mirro and Demetria Harvin for superb technical assistance. This work was supported by National Institutes of Health (NIH) Grant GM064803 (to K.M.W.); the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research and a grant from the Intramural AIDS Targeted Antiviral Program (to A.R.); and the National Cancer Institute under Contract N01-CO-12400 (to J.W.B and R.J.G.).

1. Berkowitz R, Fisher J, Goff SP (1996) RNA packaging. *Curr Top Microbiol Immunol* 214:177–218.
2. Vogt VM (1997) Retroviral virions and genomes. *Retroviruses*, eds JM Coffin, SH Hughes, and HE Varmus (Cold Spring Harbor Lab Press, Plainview, NY).
3. Aronoff R, Linial M (1991) Specificity of retroviral RNA packaging. *J Virol* 65:71–80.
4. 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.
5. Rulli SJ, Jr, et al. (2007) Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J Virol* 81:6623–6631.
6. Gorelick RJ, Henderson LE, Hanser JP, Rein A (1988) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci USA* 85:8420–8424.
7. Meric C, Goff SP (1989) Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the Cys-His box of the nucleocapsid protein. *J Virol* 63:1558–1568.
8. Gorelick RJ, et al. (1990) Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J Virol* 64:3207–3211.
9. Aldovini A, Young RA (1990) Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J Virol* 64:1920–1926.
10. Berkowitz RD, Luban J, Goff SP (1993) Specific binding of human immunodeficiency virus type 1 *gag* polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* 67:7190–7200.
11. Rein A, Harvin DP, Mirro J, Ernst SM, Gorelick RJ (1994) Evidence that a central domain of nucleocapsid protein is required for RNA packaging in murine leukemia virus. *J Virol* 68:6124–6129.
12. Berkowitz RD, Ohagen A, Hoglund S, Goff SP (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.
13. Zhang Y, Barklis E (1995) Nucleocapsid protein effects on the specificity of retrovirus RNA encapsidation. *J Virol* 69:5716–5722.
14. Clever J, Sassetti C, Parslow TG (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.
15. Rein A (1994) Retroviral RNA packaging: A review. *Arch Virol* 9(Suppl):513–522.
16. Lochrie M, et al. (1997) In vitro selection of RNAs that bind to the human immunodeficiency virus type-1 gag polyprotein. *Nucleic Acids Res* 25:2902–2910.
17. Purohit P, Dupont S, Stevenson M, Green MR (2001) Sequence-specific interaction between HIV-1 matrix protein and viral genomic RNA revealed by in vitro genetic selection. *RNA* 7:576–584.
18. Campbell S, et al. (2001) Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc Natl Acad Sci USA* 98:10875–10879.
19. Shkriabai N, et al. (2006) Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry* 45:4077–4083.
20. Alfadhli A, Still A, Barklis E (2009) Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *J Virol* 83:12196–12203.
21. Chukkapalli V, Oh SJ, Ono A (2010) Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc Natl Acad Sci USA* 107:1600–1605.
22. Sakuragi J, Shioda T, Panganiban AT (2001) Duplication of the primary encapsidation and dimer linkage region of human immunodeficiency virus type 1 RNA results in the appearance of monomeric RNA in virions. *J Virol* 75:2557–2565.
23. Hibbert CS, Mirro J, Rein A (2004) mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J Virol* 78:10927–10938.
24. Russell RS, Liang C, Wainberg MA (2004) Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? *Retrovirology* 1:23.
25. D'Souza V, Summers MF (2005) How retroviruses select their genomes. *Nature Rev Microbiol* 3:643–655.
26. D'Souza V, Summers MF (2004) Structural basis for packaging the dimeric genome of Moloney murine leukemia virus. *Nature* 431:586–590.
27. D'Souza V, et al. (2001) Identification of a high affinity nucleocapsid protein binding element within the Moloney murine leukemia virus Ψ-RNA packaging signal: Implications for genome recognition. *J Mol Biol* 314:217–232.
28. Dey A, York D, Smalls-Mantey A, Summers MF (2005) Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus. *Biochemistry* 44:3735–3744.
29. Mann R, Mulligan RC, Baltimore D (1983) Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* 33:153–159.
30. Badorrek CS, Weeks KM (2005) RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol* 1:104–111.
31. Gherghe C, Weeks KM (2006) The SL1-SL2 (stem loop) domain is the primary determinant for stability of the gamma retroviral genomic RNA dimer. *J Biol Chem* 281:37952–37961.
32. Evans MJ, Bacharach E, Goff SP (2004) RNA sequences in the moloney murine leukemia virus genome bound by the gag precursor protein in the yeast three-hybrid system. *J Virol* 78:7677–7684.
33. Gherghe C, Leonard CW, Gorelick RJ, Weeks KM (2010) Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. *J Virol* 84:898–906.
34. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM (2005) RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 127:4223–4231.
35. Gherghe CM, Shajani Z, Wilkinson KA, Varani G, Weeks KM (2008) Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S^2) in RNA. *J Am Chem Soc* 130:12244–12245.
36. Konings DAM, Nash MA, Maizel JV, Arlinghaus RB (1992) Novel GACG-hairpin pair motif in the 5′ untranslated region of type C retroviruses related to murine leukemia virus. *J Virol* 66:632–640.
37. Kim C, Tinoco I (2000) A retroviral RNA kissing complex containing only two GC base pairs. *Proc Natl Acad Sci USA* 97:9396–9401.
38. Badorrek CS, Weeks KM (2006) Architecture of a gamma retroviral genomic RNA dimer. *Biochemistry* 45:12664–12672.
39. Wilkinson KA, et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol* 6:e96.
40. Watts JM, et al. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460:711–716.
41. Rein A, et al. (1996) Inactivation of murine leukemia virus by compounds that react with the zinc finger in the viral nucleocapsid protein. *J Virol* 70:4966–4972.
42. Fisher RJ, et al. (2006) Complex interactions of HIV-1 nucleocapsid protein with oligonucleotides. *Nucleic Acids Res* 34:472–484.
43. Chertova E, et al. (2003) Sites, mechanism of action and lack of reversibility of primate lentivirus inactivation by preferential covalent modification of virion internal proteins. *Curr Mol Med* 3:265–272.
44. Rulli SJ, Jr, et al. (2006) Mutant murine leukemia virus Gag proteins lacking proline at the N-terminus of the capsid domain block infectivity in virions containing wild-type Gag. *Virology* 347:364–371.
45. Badorrek CS, Gherghe CM, Weeks KM (2006) Structure of an RNA switch enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci USA* 103:13640–13645.
46. Miyazaki Y, et al. (2010) An RNA structural switch regulates diploid genome packaging by Moloney murine leukemia virus. *J Mol Biol* 396:141–152.
47. Fu W, Rein A (1993) Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* 67:5443–5449.
48. Torrent C, Bordet T, Darlix JL (1994) Analytical study of rat retrotransposon VL30 RNA dimerization in vitro and packaging in murine leukemia virus. *J Mol Biol* 240:434–444.
49. Auweter SD, Oberstrass FC, Allain FH (2006) Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res* 34:4943–4959.