

Architecture of a Gamma Retroviral Genomic RNA Dimer[†]

Christopher S. Badorrek and Kevin M. Weeks*

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received March 15, 2006; Revised Manuscript Received August 22, 2006

ABSTRACT: Retroviral genomes contain two sense-strand RNAs that are noncovalently linked at their 5' ends, forming a dimer. Establishing a structure for this dimer is an obligatory first step toward understanding the fundamental role of the dimeric RNA in retroviral biology. We developed a secondary structure model for the minimal dimerization active sequence (MiDAS) for the Moloney murine sarcoma virus in the final dimer state using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE). In this model, two self-complementary, or palindromic, sequences (PAL1 and PAL2) form extended intermolecular duplexes of 10 and 16 base pairs, respectively. The monomeric starting state was shown previously to contain a flexible domain in which nucleotides do not form stable interactions with other parts of the RNA. In the final dimer state, portions of this initial flexible domain form stable base pairs, while previously base-paired elements lie in a new flexible domain. Thus, partially overlapping and structurally well-defined flexible domains are prominent features of both monomer and dimer states. We then used hydroxyl radical cleavage experiments to characterize the global architecture of the dimer state. Extensive regions, including portions of both PAL1 and PAL2, are occluded from solvent-based cleavage indicating that the MiDAS domain does not function simply as a collection of autonomous secondary structure elements. Instead, the retroviral dimerization domain adopts a compact architecture characterized by close packing of its constituent helices.

Retroviruses package exactly two copies of their RNA genomes into each nascent virion (1–4) in a dimeric state (5). Retroviral RNA genomes dimerize via conserved non-covalent interactions near their 5' ends. Formation of this dimeric RNA state is a fundamental feature of retroviral biology and functions to direct packaging of two genomic RNAs into nascent virions (5–7) and to facilitate strand switching and RNA-based recombination during reverse transcription of the viral genome into double-stranded DNA (8, 9). Thus, the retroviral dimer structure appears to be essential for enabling retroviruses to function as the simplest diploid entities in biology.

The closely related Moloney murine sarcoma and leukemia viruses (MuSV and MuLV, respectively) have served as important model systems for retroviral biology and, more specifically, for the role of RNA structure and conformational changes in establishing the dimer state (1–4). In the early 1990s, the structure of MuLV was analyzed using genomic RNA extracted from virions (10) and using refolded *in vitro* transcripts (11) using chemical mapping reagents. In addition, a comparative sequence analysis supported conservation of four sequence elements among the gamma retroviruses (12). These conserved elements include two self-complementary, or palindromic, sequences, PAL1 and PAL2, and two stem-loop motifs, SL1 and SL2 (Figure 1A). Both of the original chemical mapping studies supported similar foldings in the SL1 and SL2 structures that were also consistent with the

comparative sequence analysis. Biochemical experiments have since also supported a role for formation of extended intermolecular duplexes involving PAL1 and PAL2 as important contributors to the structure of the gamma retroviral dimerization domain (13–16).

We have recently used competitive dimerization experiments to define a minimal dimerization active sequence (MiDAS) for MuSV (17). This domain spans nucleotides (nts) 204–374 (Figure 1A). The MiDAS domain, defined *in vitro*, corresponds closely to the minimal sequences required to direct packaging of heterologous nonviral RNAs, as dimers, into nascent virions in cell culture (1, 5, 18) and to sequences sufficient to form a high-affinity interaction with the viral Gag protein as judged by a yeast three-hybrid assay (19). The MiDAS domain spans all of the sequence elements discussed above that were previously proposed to be important for forming the final dimer in gamma retroviral RNA genomes. Thus, there appears to be an emerging consensus that a limited region of ~170 nts is sufficient to direct dimerization both *in vitro* and *in vivo*.

Four developments suggest that the field is now in a strong position to propose unified models for this dimerization domain in the monomer and dimer states. First, the emerging consensus that a short region of ~170 nts spans the sequences necessary for dimer formation (5, 17) simplifies structural analysis of the dimerization domain because nonessential regions can be excluded from analysis. Simplification is especially helpful when interpreting chemical modification experiments. Second, RNA folding algorithms have improved dramatically in recent years due, in part, to two advances. The thermodynamic models that underlie RNA folding

[†] This work was supported by a grant from the U.S. National Institutes of Health (GM64803 to K.M.W. and Andrew Kaplan).

* To whom correspondence should be addressed. E-mail: weeks@unc.edu.

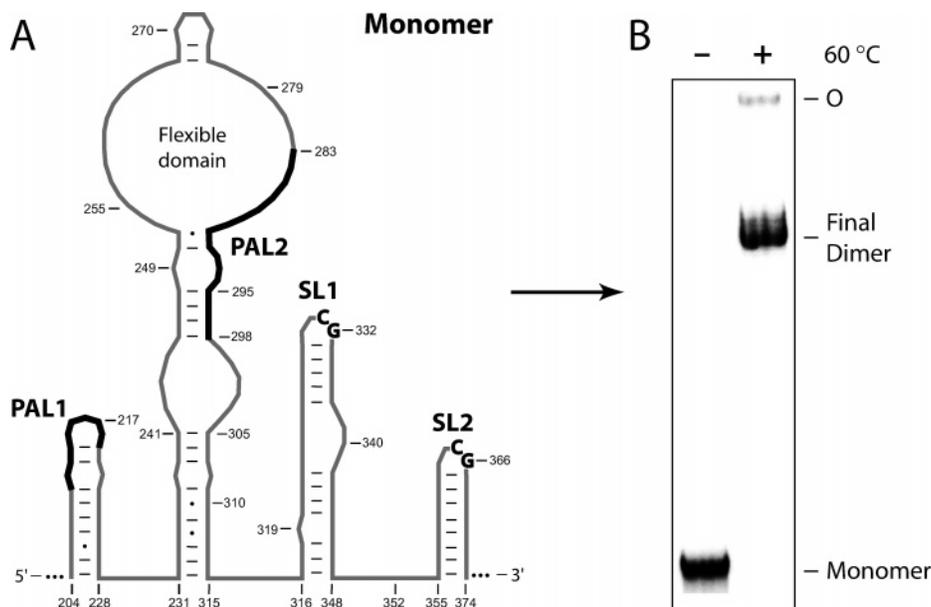


FIGURE 1: RNA dimerization of the MiDAS domain for a gamma retrovirus. (A) Secondary structure of the MiDAS in the monomeric starting state. RNA structure is shown with gray lines. Conserved PAL1 and PAL2 sequences (12) are shown with heavy black lines. Nucleotides in the SL1 and SL2 tetraloops that form cross-strand G–C pairs (27, 28) are emphasized in bold. Many naming conventions are in use for the PAL1, PAL2, SL1, and SL2 motifs; we use a system derived from the initial identification of these elements (12). (B) Dimerization of MiDAS domain visualized by nondenaturing gel electrophoresis. O, gel origin.

algorithms have improved significantly, and experimental constraints can be used to constrain these algorithms in increasingly sophisticated ways (20). Third, a new chemical approach for analyzing local nucleotide flexibility has been developed, involving selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE), that allows quantitative information to be obtained for almost every nucleotide in an RNA in a single experiment (17, 21, 22). Finally, evidence that dimerization involves formation of intermolecular duplexes at PAL1 and PAL2 (12–16) means that these two structures can be used to anchor global models for the dimer state.

We recently analyzed the structure of the MiDAS domain in the monomer starting state (17). Using RNA SHAPE chemistry and structural analysis of RNA mutants, we found that the structures of the PAL1, SL1, and SL2 regions were consistent with elements of earlier proposals. In contrast, the central region of the monomer state, including the PAL2 region, appears to fold into a structure that is different from conventional models. Specifically, the MiDAS monomer spans a large flexible domain of ~60 nts in which local nucleotide flexibility is very high (17). This flexible domain is linked to other structures in the MiDAS RNA via a stable anchoring helix (nts 231–241 and 305–315, Figure 1A).

In this work, we use SHAPE chemistry to analyze the structure of and to propose a unifying secondary structure model for the MuSV MiDAS domain in the dimer state. This model includes important elements emphasized in previous studies (10–16) and also contains significant new features.

A second feature of the dimerization domain, evaluated in this work, is whether the MiDAS domain functions primarily at the level of its base-paired secondary structure or, instead, might form long-range tertiary interactions. Using solvent-based hydroxyl radical footprinting, we detect long-range interactions in the dimer state that indicate that the retroviral RNA dimerization domain does not function simply as a collection of autonomous secondary structure elements

but instead forms a true three-dimensional entity with closely packed RNA helices.

EXPERIMENTAL PROCEDURES

Retroviral RNA Transcripts. The MiDAS RNA construct contained flanking 5' and 3' viral sequence extensions of 30 and 5 nucleotides, respectively. RNAs were generated by T7 RNA polymerase-mediated transcription [500 μ L, 37 °C, 5 h, containing 80 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (Hepes) (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (v/v) Triton X-100, 2 mM spermidine, 10 mM MgCl₂, 2 mM each nucleoside triphosphate, ~25 μ g of DNA template, 20 U of SUPERase-In (Ambion), and 0.1 mg/mL polymerase] using PCR-derived templates. RNAs were purified by denaturing gel electrophoresis (5% polyacrylamide, 7 M urea), excised from the gel, eluted overnight into $\frac{1}{2}\times$ TBE (45 mM Tris-borate, 1 mM EDTA), concentrated by ethanol precipitation, and stored in 10 mM Hepes (pH 7.5), 1 mM EDTA at –20 °C.

SHAPE Analysis of MiDAS Monomer and Dimer. 2'-Hydroxyl modification with *N*-methylisatoic anhydride (NMIA), primer extension, and band quantification steps for both monomer and dimer were performed exactly as described previously (17); dimers were created by adding a heating step (30 min at 60 °C). The MiDAS RNA construct used in the SHAPE experiments contained a nonviral RNA cassette at its 3' end (21) to facilitate analysis of the entire sequence by primer extension. For quantitative comparison of monomer versus dimer reactivities, SHAPE data were normalized to nt 393 in the UUCG tetraloop of the nonviral RNA cassette. In a SHAPE analysis of long RNAs (>150 nts), a modest stochastic drop-off in apparent reactivity is observed for nucleotides near the 5' end of the RNA because, even under single-hit conditions, some RNA molecules contain more than one 2'-*O*-adduct and because the reverse transcriptase enzyme is not perfectly processive. The prob-

ability of extension in the 3' to 5' direction was modeled as $I_{\text{obs}} = Ap^n + C$, where I_{obs} is the integrated (23) band intensity, A is the amplitude of the drop-off, p is the probability of extension, n is the nucleotide position, and C is the intensity offset. Corrected SHAPE reactivities were obtained using $I_{\text{corr}} = (A + C)/(Ap^n + C)$. The reproducibility of the SHAPE experiments is very high; absolute single nucleotide reactivities for experiments performed months apart typically agree within $\pm 10\%$ or better. Absolute SHAPE reactivities reported in this work are provided in the Supporting Information.

Solvent-Based Hydroxyl Radical Cleavage. Hydroxyl radical cleavage reactions (10 μL , at 25 °C for 1 h) were performed using refolded, dimeric, and 5'-[^{32}P]-labeled MiDAS RNA [100 nM RNA in 50 mM Hepes (pH 7.5), 200 mM potassium acetate (pH 7.5), 5 mM MgCl_2] at final concentrations of 3 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 4.5 mM EDTA, 6 mM sodium ascorbate, and 15 mM DTT. Reactions were quenched by addition of a thiourea-containing stop solution. Cleavage products were resolved by denaturing electrophoresis and protection at individual nucleotides was quantified by integrating (23) bands for RNAs treated under native versus denaturing conditions (in which mono- and divalent ions were omitted). Protection from hydroxyl radical cleavage was calculated by subtracting integrated band intensities for the folded RNA from the denaturing control.

RESULTS

SHAPE Analysis of the Retroviral RNA Dimer. The MiDAS RNA forms well-defined, single conformation monomer and dimer species as visualized by nondenaturing gel electrophoresis (Figure 1B). The monomer migrates as a single band and can be converted to the dimer upon heating at 60 °C for 30 min (100 nM RNA in 200 mM potassium acetate, 5 mM MgCl_2 , pH 7.5). The MiDAS RNA in the final dimer state migrates as a discrete single species on the time scale evaluated in a gel electrophoresis experiment (Figure 1B). These data support the interpretation that our structural analysis reports on a single conformation without interfering contributions from the initial monomer state or from stably folded alternative RNA states.

We analyzed the conformational changes that accompany the monomer-to-dimer transition in the MiDAS domain using RNA SHAPE chemistry. In a SHAPE experiment, 2'-hydroxyl groups at flexible nucleotide positions react preferentially with NMIA to form 2'-*O*-adducts. The 3'-phosphodiester anion strongly influences the nucleophilic reactivity of the adjacent 2'-ribose position (21, 24). Unconstrained nucleotides react preferentially because flexible nucleotides are better able to adopt conformations that facilitate formation of the 2'-oxyanion nucleophile. In principle, a nucleotide might also be rendered reactive by being constrained in an unusual reactive conformation. Any such conformation would involve a noncanonical local geometry and would likely be scored (correctly) as an unpaired position. SHAPE reactivity information is closely correlated with crystallographic temperature factors for tRNA^{Asp} (21, 25, 26), which emphasizes that a SHAPE experiment quantitatively monitors local RNA flexibility at single nucleotide resolution.

MiDAS RNAs in either the monomer or dimer state were treated with NMIA, and the sites of 2'-*O*-adduct formation

were detected as stops to primer extension (Figure 2A). Band intensities, corresponding to SHAPE reactivities, were quantified for almost every nucleotide in the MiDAS domain for RNAs in both monomer and dimer states. Absolute SHAPE reactivities were computed by subtracting background intensities observed in the absence of NMIA (compare + and - NMIA lanes, Figure 2A). Quantitative SHAPE histograms are shown in Figure 2B.

Upon formation of the final dimer state, SHAPE analysis reveals large structural changes throughout the MiDAS RNA (compare monomer and dimer experiments in Figure 2A,B). We computed a monomer-to-dimer difference plot by subtracting absolute SHAPE reactivities in the monomer state from those observed in the final dimer conformation. By this definition, positions that are more reactive in the dimer state are reported as positive differences and RNA regions that become more structured in the dimer show negative intensities (green and red bars, respectively, in the difference plot, Figure 2C).

SHAPE reactivities in both the PAL1 and PAL2 self-complementary sequences (nts 210–219 and 283–298, respectively) decrease significantly upon dimerization. These reduced SHAPE reactivities are consistent with formation of extended intermolecular duplexes in these regions in the final dimer (red bars, Figure 2C). The SL1–SL2 region constitutes an autonomous domain within the MiDAS region and forms a stable dimer via loop–loop base pairs mediated by GACG tetraloops (27, 28). The significant conformational changes observed in the SL1 and SL2 helices (compare nts 314–315, 319, and 338–343 in monomer and final dimer panels in Figure 2B) are consistent with a previously identified conformational switch in which SL1 becomes extended by four base pairs at its base (27).

Secondary Structure Model for the MiDAS Domain in the Final Dimer State. To develop candidate secondary structure models for the entire MiDAS domain in the final dimer state, we used SHAPE reactivities to constrain the output of secondary structure prediction programs that employ a thermodynamic model for base pair formation (29, 30). As shown in Figure 2, absolute SHAPE reactivities span a dynamic range of 20-fold or greater. To facilitate using SHAPE information to constrain secondary structure prediction algorithms, we simplified this information by dividing absolute SHAPE reactivities into three classes. Positions with absolute reactivities between 50% and 100% of the most highly reactive positions (red in Figure 3A) were algorithmically required to be single-stranded. Positions with reactivities in the range of 25–50% of the most reactive positions were constrained to be either single-stranded or adjacent to an unpaired nucleotide or a G–U base pair, as implemented in the RNAstructure algorithm (29). This latter, less stringent, requirement reflects that nucleotides at the ends of helices and adjacent to bulges, mismatches, or G–U pairs tend to be more dynamic than nucleotides in the center of uninterrupted helix. No folding constraints were imposed for nucleotides exhibiting the lowest (0–25%) reactivity.

Model-independent SHAPE reactivities are superimposed, as columns, on a secondary structure model for the MiDAS domain in the final dimer state (Figure 3A). Several features of this model are well defined by the SHAPE data. PAL1 and PAL2 sequences form extended intermolecular duplexes of 10 and 16 base pairs, respectively. The 5' and 3' termini

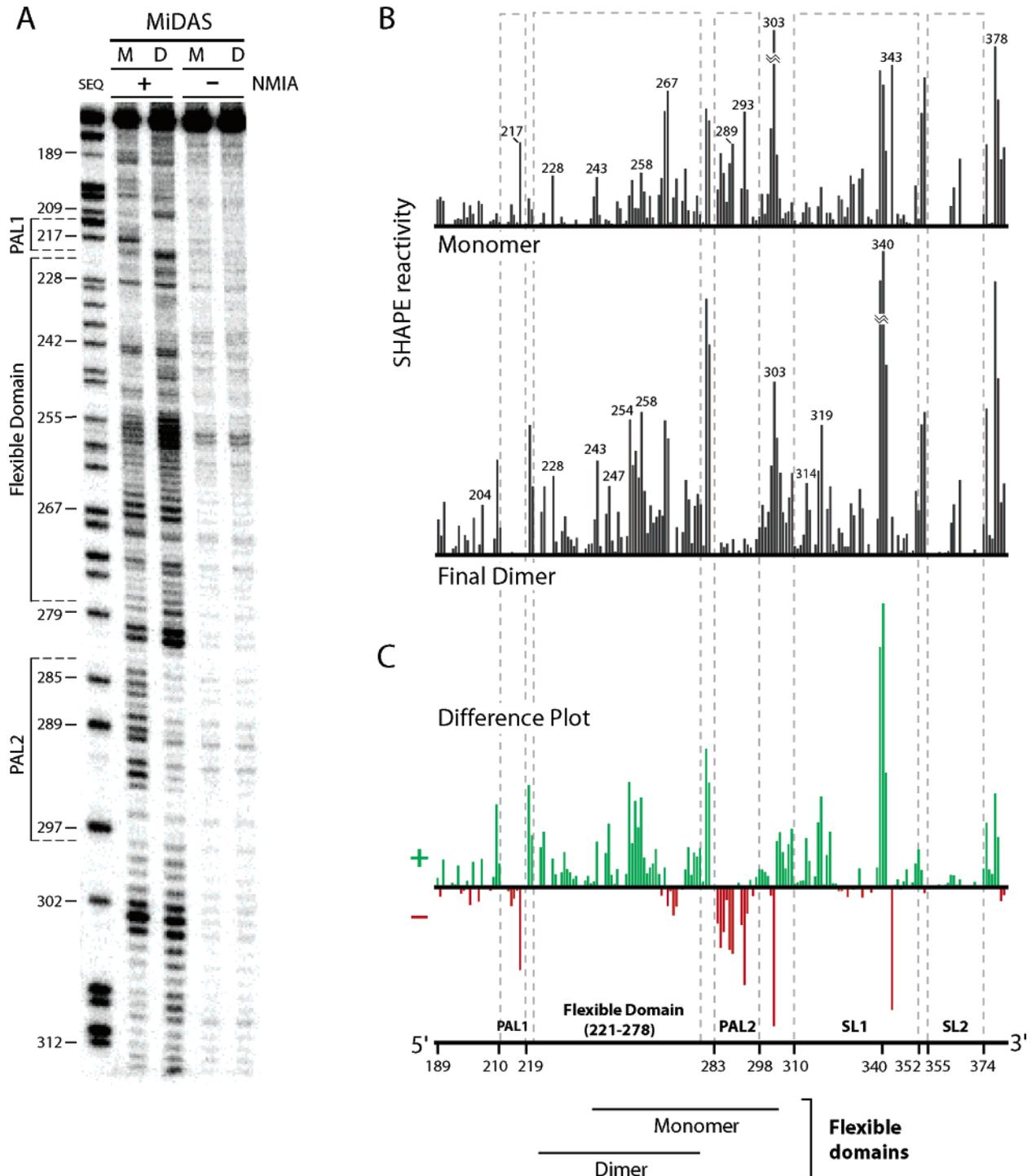


FIGURE 2: SHAPE analysis of MiDAS RNA in monomer (M) and dimer (D) states. (A) 2'-O-adduct formation detected by primer extension. Primer extension products are visualized in a sequencing gel. Experiments were performed in the presence (+) and absence (-) of the NMIA reagent. Sequencing lanes (SEQ) were generated by dideoxy nucleotide incorporation during the primer extension; nucleotide positions are labeled with respect to NMIA lanes. (B) Histograms of SHAPE reactivities. Monomer and final dimer panels show absolute levels of 2'-O-adduct formation minus background. Data for nucleotides 278–381 is taken from experiments reported previously (27). (C) Difference plot was calculated by subtracting monomer intensities from those of the final dimer; positive and negative (green and red) peaks indicate nucleotides that are more flexible versus more structured in the final dimer state, respectively.

of both intermolecular duplexes are adjacent to highly flexible nucleotides as judged by SHAPE reactivity. The PAL2 duplex is consistent with the original proposals (12, 13) that this structural region spans 16 nts but not with models in which the PAL2 duplex contains additional base-paired nucleotides at each end (31, 32). SHAPE-constrained predictions from RNAstructure strongly suggest that PAL1 forms a simple 10 base pair (12) extended duplex and is not

extended by additional base pairs as previously suggested (15). Instead, the SHAPE data suggest that the region immediately 5' to PAL1 (see Figure 1A) undergoes a conformational change during dimerization such that nucleotides 198–207 form a new stem-loop containing a stable (33) GNRA-type tetraloop at its apex (Figure 3A). The SL1–SL2 region, which forms an autonomous tertiary structure domain (27), is linked to the PAL2 intermolecular duplex

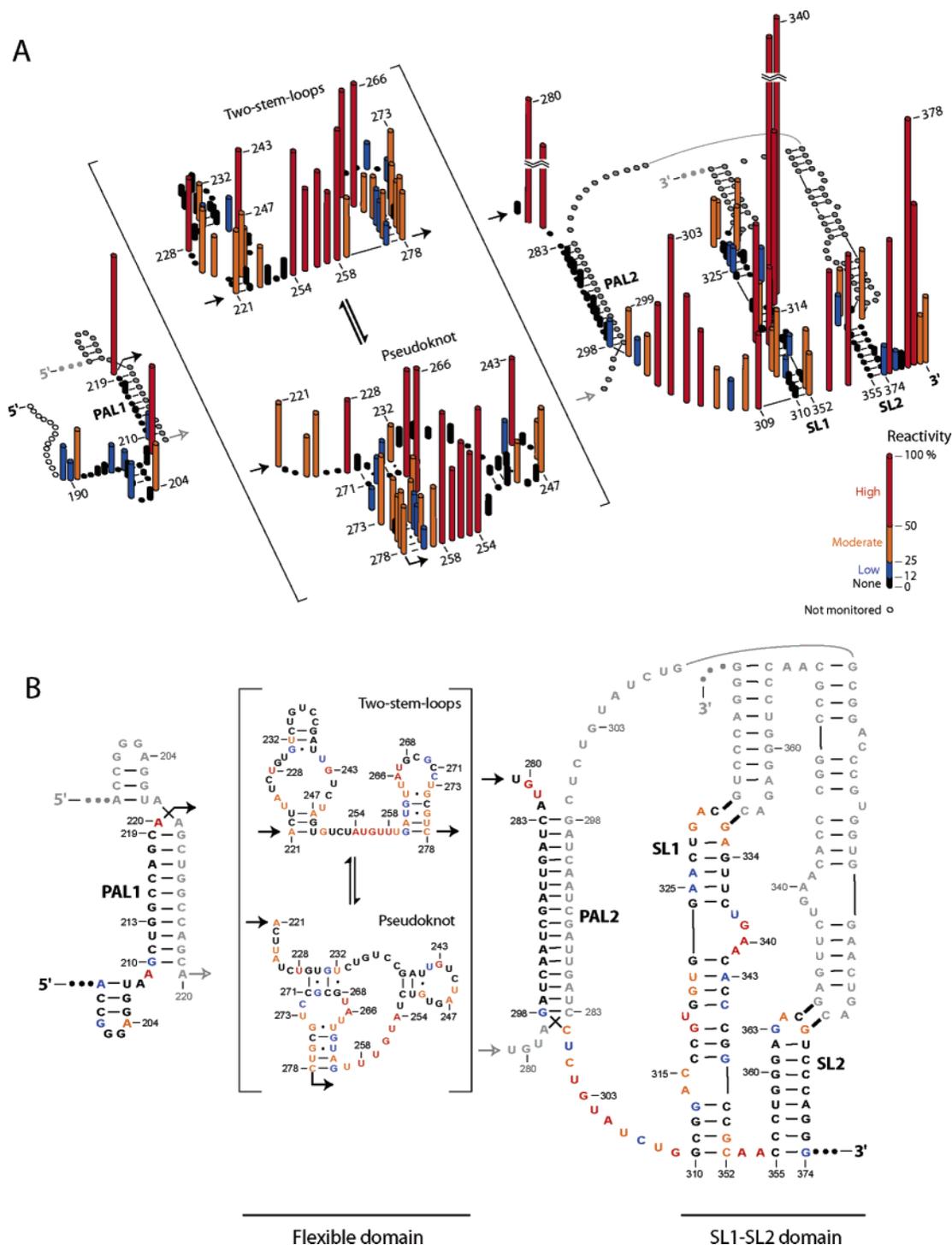


FIGURE 3: Secondary structure model for MiDAS RNA in the final dimer state. Nucleotides are represented as color-coded columns (A) or letters (B) indicating their absolute SHAPE reactivities. For clarity, only one of the two strands in the dimer is annotated; the second strand is shown in gray. Secondary structure assignments for the loop-loop interactions in the SL1-SL2 domain (involving G-C base pairs) are derived from NMR (28) and site-directed hydroxyl radical cleavage (27) experiments. In the flexible domain (in brackets), two possible conformations for either of the two strands are shown.

and the rest of the MiDAS dimer via a flexible junction (nts 299-309, Figure 3).

SHAPE reactivities indicate that sequences between the PAL1 and PAL2 extended duplexes do not form a single well-defined intramolecular secondary structure in the final dimer state. The thermodynamic model that underlies RNA-structure (29) suggests that the region between the PAL1 and PAL2 intermolecular duplexes may form a weakly stable two-stem-loop motif (-4 kcal/mol, or equivalent to a single

duplex with 2-3 base pairs) (Figure 3). RNAstructure is designed to identify low-energy structures that do not contain pseudoknots (29). Therefore, we searched for alternate secondary structures consistent with SHAPE constraints using the heuristic algorithm implemented in HotKnots (30). SHAPE reactivities are also compatible with formation of a pseudoknot between the PAL1 and PAL2 intermolecular duplexes. Both the two-stem-loop and pseudoknot structures are equally compatible with the pattern of SHAPE data;

overall flexibility in this region likely reflects contributions from multiple structures. The proposal that there may be multiple, rapidly exchanging structures is consistent with the native gel electrophoresis analysis (Figure 1B) because rapidly interchanging structures would yield a single conformation visible in the gel. There are also several possibilities for cross-strand intermolecular interactions involving nucleotides in this region. The most stable involve the self-complementary G²³⁵UCCGAU²⁴¹ and G²⁶⁸CGC²⁷¹ sequences.

The Dimer State Is More Flexible Than the Monomer State. SHAPE reactivity information supports two global differences between the monomer and dimer states. First, with the exceptions of PAL1 and PAL2 intermolecular duplexes and a few nucleotides in the SL1–SL2 domain, absolute SHAPE reactivity is higher in most regions of the dimer than in the monomer (Figure 2C). Thus, local nucleotide flexibility increases significantly during dimerization, overall.

Second, an important feature of the secondary structure in the initial monomer state was formation of an extensive flexible domain, spanning ~60 nts, between the PAL1 and SL1 regions. This flexible domain spans positions 242–304 in the monomer (see Figure 1A) and potentially interconverts between multiple structures, each with roughly similar net thermodynamic stability (17). SHAPE analysis of the dimer state supports a model in which the MiDAS dimer also contains a flexible domain, characterized by the ability to form multiple, interconverting structures (positions 221–278 in final dimer panel of Figure 2B and Figure 3). Multiple regions with no or low reactivity in the monomer state become reactive because they are incorporated into a new flexible domain in the final dimer state. The flexible domain in the dimer is anchored on each side by the PAL1 and PAL2 intermolecular duplexes (see nts 232, 243, 246–247, and 254–258; middle panel, Figure 2B). Flexible domains are thus prominent features of both monomer and dimer states for the MiDAS domain.

Tertiary Structure in the MiDAS Dimer. Our model for the secondary structure of the MiDAS domain emphasizes that the final dimer structure is built up from three major constituents: the PAL1 and PAL2 intermolecular duplexes and the SL1–SL2 domain (Figure 3). To evaluate whether these constituent structures form intermotif tertiary interactions, we detected RNA–RNA interactions that occlude the backbone from access by solvent in the final dimer state using hydroxyl radical cleavage. Protection from hydroxyl radical cleavage has been shown to correspond closely to regions of an RNA backbone with low solvent accessibility that results from formation of long-range tertiary interactions in an RNA (34–36). The MiDAS RNA was subjected to hydroxyl radical cleavage using the Fe(II)–EDTA reagent (34, 37) under solution conditions that either stabilize formation of the native final dimer (Figure 1B) or in the absence of added mono- and divalent ions (denaturing conditions) where the RNA backbone is expected to be broadly accessible to cleavage.

We computed a hydroxyl radical protection profile by subtracting integrated nucleotide reactivities toward the hydroxyl radical reagent for the RNA in the native dimer from those measured under denaturing conditions. Nucleotides that are protected or become more accessible in the final dimer relative to the denatured state are reported as

positive and negative amplitudes, respectively (Figure 4A). Protections 5-fold and 3-fold above the average background were assigned as strong and moderate.

Superposition of these strong and moderate protections on the secondary structure of the MiDAS dimer (Figure 4B) indicates that extensive regions of the backbone are inaccessible to solvent in the SL1–SL2 domain. The protection pattern in this domain in the context of the full-length MiDAS RNA, especially at the apexes of SL1 and SL2, coincides closely with previous work (27) that focused on the isolated SL1–SL2 domain (Figure 5). These data support the previously articulated model that the SL1–SL2 domain forms a compact and autonomous tertiary structure in the final dimer state.

Reproducible protection from hydroxyl radical cleavage is also observed outside the SL1–SL2 domain, in the 5' GNRA tetraloop (nts 198–207) and in the PAL1 and PAL2 extended duplexes (Figure 4B). Protection in PAL1 is confined to a compact region in the center of the helix, whereas PAL2 shows extensive protection spanning almost the entire 16 base pair duplex. In contrast, nucleotides in the flexible domain and in the linker between PAL2 and SL1 are broadly accessible to cleavage. Thus, PAL1, PAL2, and the SL1–SL2 domain form a structure that features closely packed helices and other interactions in the context of the MiDAS domain. These three motifs are then connected by flexible and solvent-accessible linkages (Figures 3 and 4).

DISCUSSION

A Well Constrained Secondary Structure for the MiDAS Domain. SHAPE analysis of local nucleotide flexibility (Figures 2 and 3), taken together with information from prior studies (10–13, 15, 16), supports a consensus model for the dimerization domain of MuSV in the dimer state. This model provides a working framework for future studies designed to understand the roles played by protein binding and conformational changes in retroviral genomic RNA dimerization and packaging.

First, the SHAPE reactivity information strongly supports previous proposals (12–16) that both the PAL1 and PAL2 sequences form intermolecular duplexes in the dimer state. There have been several distinct proposals for the extent of base pairing in these two intermolecular duplexes. PAL1 and PAL2 are characterized by 10 and 16 consecutive nucleotides, respectively, with near-zero SHAPE reactivities (Figure 3). For both duplexes, nucleotides immediately adjacent to these core double-stranded regions show high reactivities.

Second, SHAPE reactivity information emphasizes that both the monomeric starting state (17) and the final dimer state (Figure 3A) contain well-delimited flexible domains. For both monomer and dimer states, the flexible domains span ~60 nts in length. The flexible domains overlap but do not occupy the same regions in the two states (Figure 2C). In the initial monomer state, the flexible domain is linked to the rest of the RNA via an anchoring helix (nts 231–241/305–315, see Figure 1A). In the final dimer state, the flexible domain lies between the stable PAL1 and PAL2 duplexes (Figure 3). Thus, for both states these flexible domains also share the feature that their borders with the rest of the RNA are clearly delineated by stable flanking structural elements.

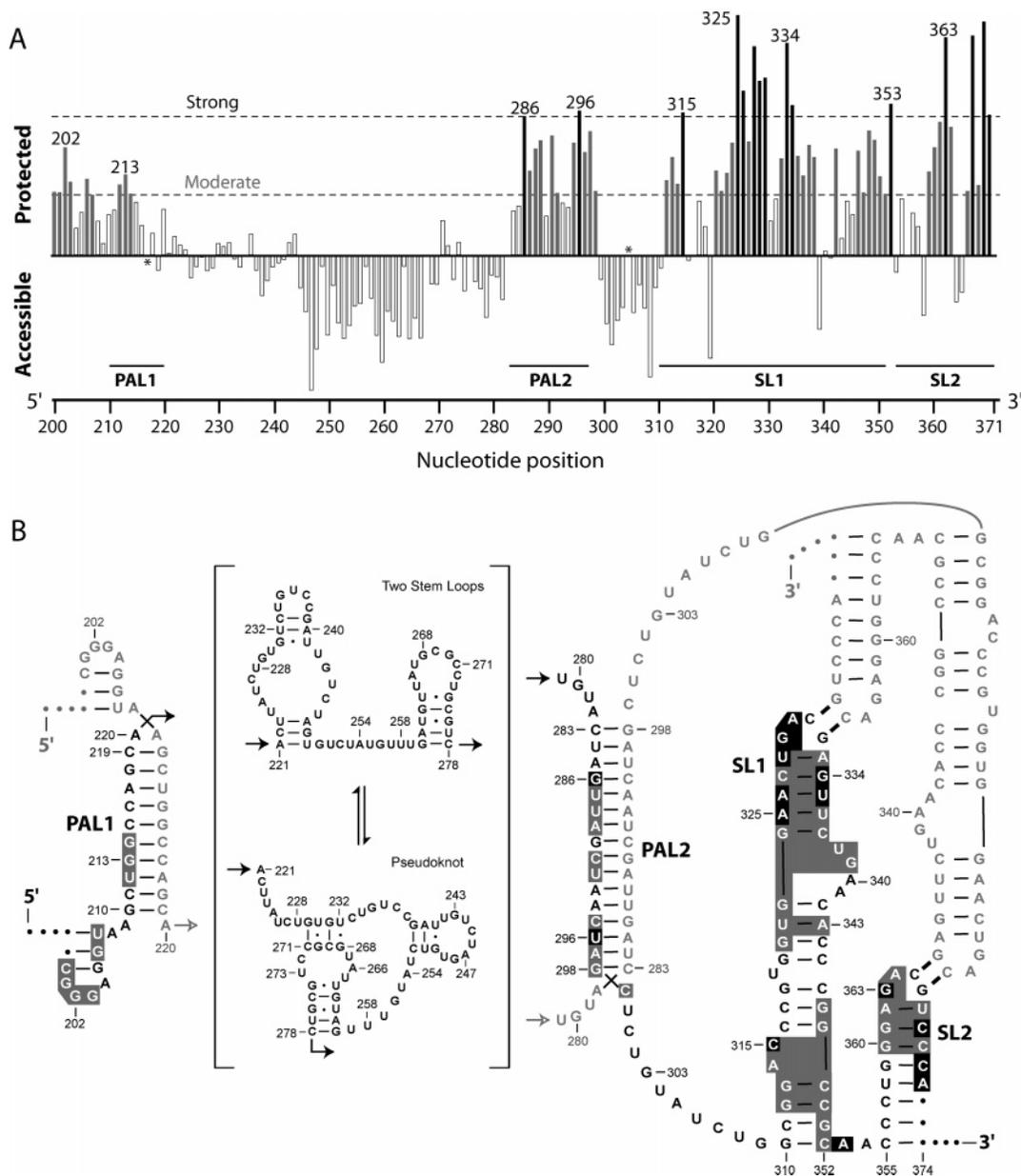


FIGURE 4: Tertiary structure in the final dimer. (A) Quantitative histogram of hydroxyl radical protection. Positive and negative peaks indicate positions that become protected or become more accessible in the natively folded dimer. Strong and moderate protection are emphasized with black and gray bars, respectively. Asterisks indicate positions where high background prevented quantitative analysis. (B) Summary of observed protections, superimposed on a secondary structure model for the final MiDAS dimer. For clarity, protection is illustrated for only one of the two strands in the dimer.

Two short regions within the flexible domain in the dimer state are unreactive toward SHAPE chemistry yet are not predicted to form stable intramolecular secondary structures by current models for base pairing (nts 235–241 and 268–271, Figure 3). Both of these regions potentially have the ability to form short self-complementary helices analogous to PAL1 and PAL2.

Alford et al. (10) previously performed a comprehensive analysis of MuLV genomic RNA, likely in the dimer state (5, 6), that had been extracted from virions. These *ex vivo* data superimpose closely with our model for the MiDAS domain in the dimer state (see the Supporting Information). In particular, the *ex vivo* mapping experiments strongly support the formation of a flexible domain between PAL1 and PAL2.

Retroviral Genomic RNA Dimerization Domains as Tertiary Structure Motifs. Retroviral dimerization domains have

been almost universally analyzed in terms of their constituent secondary structure elements. This view has led to the important consensus that formation of extended intermolecular duplexes is a fundamental and conserved feature of retroviral biology (1, 3, 4, 14). In contrast, higher order tertiary interactions have been tacitly assumed to play a minor role. This view has been recently reinforced by NMR studies in which an RNA containing the PAL1, SL1, and SL2 elements was found to fold into a T-shaped structure in which no long-range tertiary interactions form between the constituent helices (31, 32). To create an RNA amenable to analysis by NMR, it was necessary to introduce non-native mutations into the dimerization domain and to create a unimolecular analogue of the dimer state (31).

As judged by hydroxyl radical footprinting experiments on the intact MiDAS domain, extensive regions of the RNA backbone are protected from the solvent-based reagent

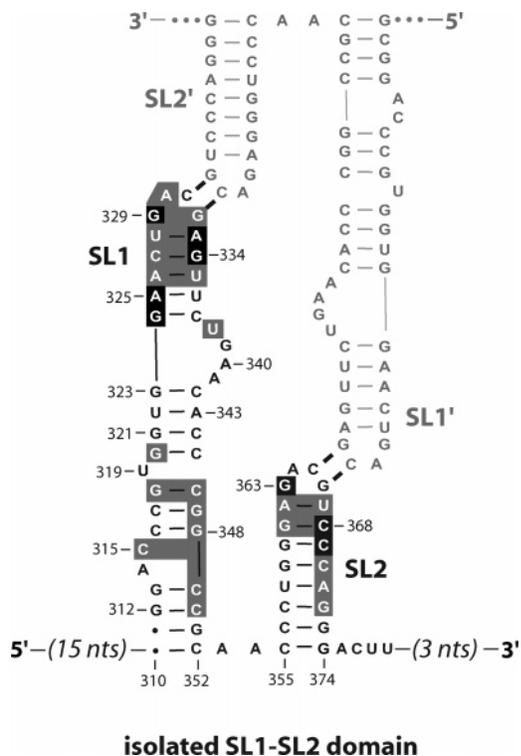


FIGURE 5: Hydroxyl radical protection for the isolated SL1-SL2 domain (27). Symbols are the same as those used in Figure 4B.

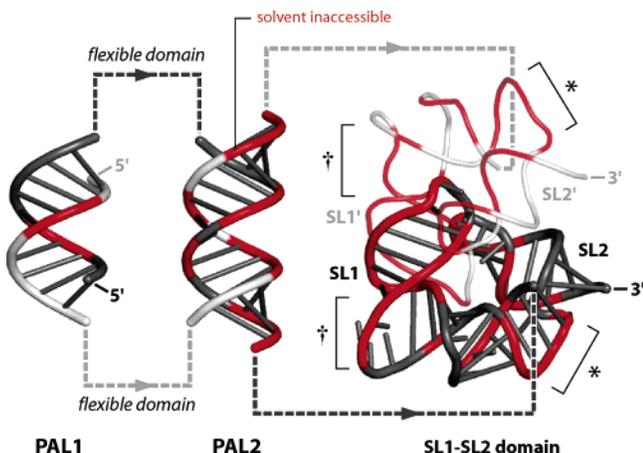


FIGURE 6: Architecture of a gamma retroviral genomic RNA dimer. The two RNA strands in the dimer are dark gray and white. Solvent-inaccessible regions, as judged by hydroxyl radical cleavage, are emphasized with a red backbone. PAL1 and PAL2 extended duplexes are shown as A-form helices; the refined structure of the SL1-SL2 domain is from ref 27. The * and † symbols indicate positions in the SL1-SL2 domain that are selectively protected only in the context of the intact MiDAS RNA.

(Figure 4). Thus, like many other RNAs, including the large ribozymes (38), the MiDAS domain dimer with a native sequence has a true three-dimensional architecture. Regions that are inaccessible to solvent in the final dimer state include the short 5' stem-loop at nts 198-207, PAL1, PAL2, and the SL1-SL2 domain (Figure 4B). We visualized the overall three-dimensional architecture of the MiDAS domain by superimposing the experimental hydroxyl radical protection pattern (i) on A-form helix models for PAL1 and PAL2 and (ii) on our previously refined (27) model for the SL1-SL2 domain (Figure 6).

PAL2 shows a clear pattern in which both sides of this extended duplex are protected from solvent (Figure 6). We infer that this duplex lies sandwiched between other RNA elements in the dimer. We previously refined a three-dimensional structure for the SL1-SL2 domain that spans 65 nts of the 170 nt MiDAS domain (27). By comparing the solvent-inaccessible regions in the short SL1-SL2 domain (Figure 5) with the protections observed in the SL1-SL2 region in the context of the entire MiDAS RNA (Figure 4), we can characterize regions in the SL1-SL2 domain likely to participate in interactions with PAL1 or PAL2. The new protections from hydroxyl radical cleavage occur at the base of SL1 and centered at the 340 bulge in SL1 (see * and † symbols, respectively, Figure 6). The protections at the 340 bulge, characteristic of the intact MiDAS sequence, lie on one face of the SL1-SL2 domain model.

These data suggest that the extensive protection pattern observed in the PAL2 intermolecular duplex reflects close packing interactions between this helix and the SL1-SL2 domain (Figure 6). The observed protections on one face of PAL1 are then consistent with formation of additional packing interactions between PAL1 and the PAL2-SL1-SL2 motif. Finally, the flexible domain in the final dimer is accessible to solvent, consistent with this region projecting out from core interactions involving PAL1 and PAL2.

We conclude that the dimerization domain for the MuSV gamma retrovirus, and likely for many other retroviruses, functions not as a collection of autonomous stem-loop and helical structures but, instead, as a true three-dimensional entity with a compact and roughly globular core. In this structural context (Figure 6), the specific orientations of the flexible domain and the linker region between PAL2 and SL1 then have the potential to modulate interactions with the retroviral Gag protein during packaging.

ACKNOWLEDGMENT

We are indebted to Cristina Gherghe and Andrew Kaplan for many helpful discussions, to David Mathews for assistance with RNAstructure, and to Jihong Ren and Anne Condon for helpful interactions regarding the HotKnots program.

SUPPORTING INFORMATION AVAILABLE

One figure showing chemical modification information, obtained ex virio (from ref 10), superimposed on a secondary structure model for the MiDAS RNA in the dimer state and one table providing the absolute SHAPE reactivities reported in this work. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Berkowitz, R., Fisher, J., and Goff, S. P. (1996) RNA packaging. *Curr. Top. Microbiol. Immunol.* 214, 177-218.
- Rein, A. (2004) Take two. *Nat. Struct. Mol. Biol.* 11, 1034-1035.
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: an inseparable pair. *Nat. Rev. Microbiol.* 2, 461-472.
- D'Souza, V., and Summers, M. F. (2005) How retroviruses select their genomes. *Nat. Rev. Microbiol.* 3, 643-655.
- Hibbert, C. S., Mirro, J., and Rein, A. (2004) mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J. Virol.* 78, 10927-10938.

6. Fu, W., and Rein, A. (1993) Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J. Virol.* 67, 5443–5449.
7. Sakuragi, J., Shioda, T., and Panganiban, A. T. (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.
8. Rasmussen, S. V., Mikkelsen, J. G., and Pedersen, F. S. (2002) Modulation of homo- and hetero-dimerization of Harvey sarcoma virus RNA by GACG tetraloops and point mutations in palindromic sequences. *J. Virol.* 76, 613–628.
9. Rasmussen, S. V., and Pedersen, F. S. (2004) Complementarity between RNA dimerization elements favors formation of functional heterozygous murine leukemia viruses. *Virology* 329, 440–453.
10. Alford, R. L., Honda, S., Lawrence, C. B., and Belmont, J. W. (1991) RNA secondary structure analysis of the packaging signal for Moloney murine leukemia virus. *Virology* 183, 611–619.
11. Tounekti, N., Mougel, M., Roy, C., Marquet, R., Darlix, J., Paoletti, J., Ehresmann, B., and Ehresmann, F. (1992) Effect of dimerization on the conformation of the encapsidation psi domain of the Moloney murine leukemia virus. *J. Mol. Biol.* 223, 205–220.
12. Konings, D. A. M., Nash, M. A., Maizel, J. V., and Arlinghaus, R. B. (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.
13. Girard, P. M., Bonnet-Mathoniere, B., Muriaux, D., and Paoletti, J. (1995) A short autocomplementary sequence in the 5' leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry* 34, 9785–9794.
14. Paillart, J., Marquet, R., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1996) Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* 78, 639–653.
15. Oroudjev, E. M., Kang, P. C. E., and Kohlstaedt, L. A. (1999) An additional dimer linkage structure in Moloney murine leukemia virus RNA. *J. Mol. Biol.* 291, 603–613.
16. Ly, H., and Parslow, T. G. (2002) Bipartite signal for genomic RNA dimerization in Moloney murine leukemia virus. *J. Virol.* 76, 3135–3144.
17. Badorrek, C. S., and Weeks, K. M. (2005) RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat. Chem. Biol.* 1, 104–111.
18. Adam, M. A., and 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.
19. Evans, M. J., Bacharach, E., and Goff, S. P. (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.
20. Mathews, D. H., and Turner, D. H. (2006) Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.* 16, 270–278.
21. Merino, E. J., Wilkinson, K. A., Coughlan, J. L., and Weeks, K. M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223–4231.
22. Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA^{Asp} transcripts. *J. Am. Chem. Soc.* 127, 4659–4667.
23. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D., and Altman, R. B. (2005) SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* 11, 344–354.
24. Chamberlin, S. I., Merino, E. J., and Weeks, K. M. (2002) Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14688–14693.
25. Westhof, E., Dumas, P., and Moras, D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* 184, 119–145.
26. Chamberlin, S. I., and Weeks, K. M. (2000) Mapping local nucleotide flexibility by selective acylation of 2'-amine substituted RNA. *J. Am. Chem. Soc.* 122, 216–224.
27. Badorrek, C. S., Gherge, C. M., and Weeks, K. M. (2006) Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13640–13645.
28. Kim, C., and Tinoco, I. (2000) A retroviral RNA kissing complex containing only two GC base pairs. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9396–9401.
29. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7287–7292.
30. Ren, J., Rastegari, B., Condon, A., and Hoos, H. H. (2005) HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* 11, 1494–1504.
31. D'Souza, V., Dey, A., Habib, D., and Summers, M. F. (2004) Structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* 337, 427–442.
32. D'Souza, V., and Summers, M. F. (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. *Nature* 431, 586–590.
33. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. U.S.A.* 86, 7706–7710.
34. Latham, J. A., and Cech, T. R. (1989) Defining the inside and outside of a catalytic RNA molecule. *Science* 245, 276–282.
35. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273, 1678–1685.
36. Krasilnikov, A. S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature* 421, 760–764.
37. Tullius, T. D., and Dombroski, B. A. (1986) Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci. U.S.A.* 83, 5469–5473.
38. Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds. (1999) *The RNA World*, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

BI060521K