nature chemical biology

RNA flexibility in the dimerization domain of a gamma retrovirus

Christopher S Badorrek & Kevin M Weeks

Retroviruses are the causative agents of serious diseases, such as acquired immunodeficiency syndromes and several cancers, and are also useful gene therapy vectors. Retroviruses contain two sense-strand RNA genomes, which become linked at their 5′ ends to form an RNA dimer. Understanding the molecular basis for dimerization may yield new approaches for controlling viral infectivity. Because this RNA domain is highly conserved within retrovirus groups, it has not been possible to define a consensus structure for the 5′ dimerization domain by comparative sequence analysis. Here, we defined a 170-nucleotide minimal dimerization active sequence (MiDAS) for a representative gamma retrovirus, the Moloney murine sarcoma virus, by stringent competitive dimerization. We then analyzed the structure at every nucleotide in the MiDAS monomeric starting state with quantitative selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) chemistry. Notably, SHAPE analysis demonstrated that the RNA monomer contains an extensive flexible domain spanning 50 nucleotides. These findings support a structural model in which RNA flexibility directly facilitates retroviral genome dimerization by reducing the energetic cost of disrupting pre-existing base pairings in the monomer.

Retroviruses selectively package two sense-strand RNA genomes in the infectious viral particle. The RNA genomes are linked together near their 5′ ends[1] by a precise, but poorly understood, set of noncovalent interactions. These interactions probably involve a mixture of base pairing and tertiary interactions. The structure of this RNA 'dimer' is important in several stages of the retroviral infectivity cycle, including RNA encapsidation into nascent viral particles[2–5] and recombination during reverse transcription[3,6,7]. Retroviruses are valuable biotechnology tools (as gene therapy vectors) and are also the causative agents of serious diseases, including acquired immunodeficiency syndromes and several cancers. Understanding the mechanism of retroviral dimerization at a molecular level thus represents an important opportunity both to enhance vector function and to disrupt the infectivity of pathogenic viruses.

Because the retroviral dimerization sequences within a virus family are often very similar[8–11], a consensus secondary structure cannot be inferred by phylogenetic covariation analysis, which is the most robust method to determine the secondary structure for a large RNA[12–14]. Secondary structure models for retroviral RNA dimerization domains are still provisional and probably only partially encompass the biological function of these RNAs. Determining the biologically relevant structure of the dimerization domain for any retroviral RNA thus represents a common problem in biology. The challenge is to understand an RNA secondary structure in enough detail to be able to formulate hypotheses about biological function, even though only one or a few highly similar sequences are known.

Algorithms for predicting an RNA secondary structure from a single sequence identify roughly 50–70% of known helices correctly[15,16]. However, prediction accuracy for a single RNA or for any helix within a larger structure is not known in advance. Incorrect prediction of even a few helices in a functionally important region makes it difficult or impossible to develop robust biological models.

Among the gamma retroviruses, several sequences have been consistently proposed as important for dimerization of the RNA genome[8–10] (summarized in **Fig. 1a**). PAL1 (also known as the 204–227 stem-loop[17], DIS1(ref. 18) or SL-B′ (ref. 5)) and PAL2 (also known as DIS2 (ref. 18), SL-B[19] or H1 (ref. 20)) are postulated to form hairpin loops (in green, **Fig. 1**). PAL1 and PAL2 span self-complementary ('palindromic') sequences and are conventionally proposed to interact through loop-loop interactions with PAL1 and PAL2 sequences from a second RNA, eventually forming extended duplexes in the dimer[10,17–19]. Highly conserved GACG tetraloops[9] (**Fig. 1a**) in stem-loops 1 and 2 (SL1 and SL2, also known as SL-C[19] or H2 (ref. 20) and SL-D or H3, respectively), have the potential to form stable loop-loop interactions through cross-loop G-C base pairs[19,21] and seem to be important for packaging via interactions with the viral Gag protein[22].

We developed a generalizable approach for obtaining a well-constrained secondary structure for a retroviral dimerization domain and for many other classes of RNA. We first used competition experiments to rigorously define a contiguous minimal dimerization active sequence (MiDAS) for a representative gamma retrovirus, the Moloney murine sarcoma virus (MuSV). We then used a new chemical approach, selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE)[23,24], to obtain comprehensive, quantitative, nucleotide-resolution and model-independent constraints for the secondary structure of the monomeric starting state of the retroviral dimerization domain.

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA. Correspondence should be addressed to K.M.W. (weeks@unc.edu).

Published online 5 June 2005; doi:10.1038/nchembio712

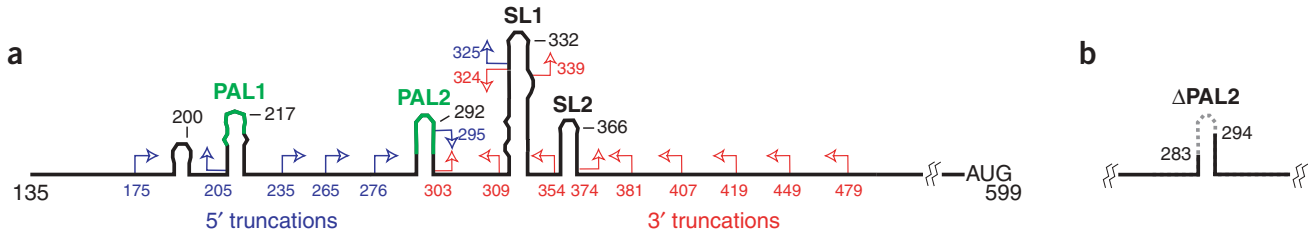


Figure 1 5′-untranslated region of MuSV. Conserved sequences and conventionally proposed secondary structures are illustrated schematically. The 5′ genomic RNA cap is position 1. (**a**) 5′ and 3′ truncation mutants are shown in blue and red, respectively. (**b**) ΔPAL2 mutant.

The results of these experiments emphasize that existing structural models for the dimerization domain in gamma retroviruses require reinterpretation. Most notably, a large region in the dimerization domain is conformationally flexible, which may facilitate retroviral RNA dimerization by decreasing the energetic cost of disrupting base pairing or other interactions in the monomer before the formation of functional structures specific to the dimer.

RESULTS

Rigorous definition of a MiDAS

In vitro studies using synthetic RNA transcripts have been essential for identifying candidate structures that contribute to dimerization in the gamma retroviruses[8,10,17–20,25–29]. A significant challenge in interpreting these experiments is that most RNAs containing a stem-loop structure will dimerize if dimerization reactions are performed under conditions of sufficiently high RNA concentration or ionic strength[30,31].

In exploratory experiments, we identified a roughly physiological ion environment—50 mM HEPES, 200 mM potassium acetate, and 5 mM $MgCl_2$ (pH 7.5)—that yields well-behaved single-conformation monomer and dimer complexes for an RNA spanning most of the MuSV 5′ untranslated region (**Fig. 1a**). This RNA spans all structures previously proposed to participate in dimerization in the gamma retroviruses. We used this RNA to impose a functional threshold in competitive dimerization experiments that, by design, strongly discriminates against promiscuous self-dimerization (**Fig. 2**).

We constructed an extensive series of viral sequences containing systematic truncations from their 5′ and 3′ ends (blue and red arrows, respectively, **Fig. 1a**). Competitive dimerization experiments were performed at 60 °C in the presence of the full-length transcript and visualized by the selective detection of the radiolabeled, truncated RNA variants in nondenaturing gels (**Fig. 2**). Both the radiolabeled mutant-mutant homodimer and mutant–full-length heterodimer are visualized directly. Full-length RNA homodimers also form but are not radiolabeled and thus are not observed.

We scored as structurally competent dimerization active sequences only those mutant RNAs that quantitatively competed with homodimerization by the (unlabeled) full-length RNA. RNAs that only homodimerized or only heterodimerized (see middle two lanes in **Fig. 2**) were scored as structurally deficient.

A minimal sequence active in dimerization

Truncation mutants were identified by the 5′ or 3′ nucleotide (nt) at which the mutant sequence terminates (**Fig. 1a**). Competitive dimerization experiments were performed with ~1.5 nM radiolabeled, truncated RNA and 1, 5 or 15 nM full-length RNA (**Fig. 3**). Markers for the mutant monomer and for the mutant homodimer were obtained by omitting the heating step or omitting FL RNA, respectively.

The progressively larger 3′ truncations, 3′-479, 3′-419 and 3′-374 (**Fig. 3a**), yielded native-like RNAs that both homo- and heterodimerized efficiently with the full-length RNA. In contrast, the 3′-339 truncation heterodimerized efficiently with the full-length RNA, but formed almost no homodimer (3′-339 panel, **Fig. 3a**). Thus, the 3′-339 RNA is deficient in dimerization in a way that can be rescued by the full-length RNA. This RNA also forms several monomeric conformations. Similarly, the 3′-354 truncation forms heterogeneous monomers and also forms heterodimers inefficiently (**Supplementary Fig. 1** online). The 3′ boundary for the MiDAS lies between nt 354 and 374; the largest fully functional deletion spans position 374 (3′-374 panel, **Fig. 3a**).

Truncations from the 5′ end through position 276 yield RNAs that correctly homo- and heterodimerized (**Fig. 3c**). In contrast, truncation through position 295 yields an RNA that homodimerized well but was incompetent at forming heterodimers with the full-length RNA (5′-295 panel in **Fig. 3c**). Further truncation through 5′-325 yields an RNA that formed neither homo- nor heterodimers. The 5′

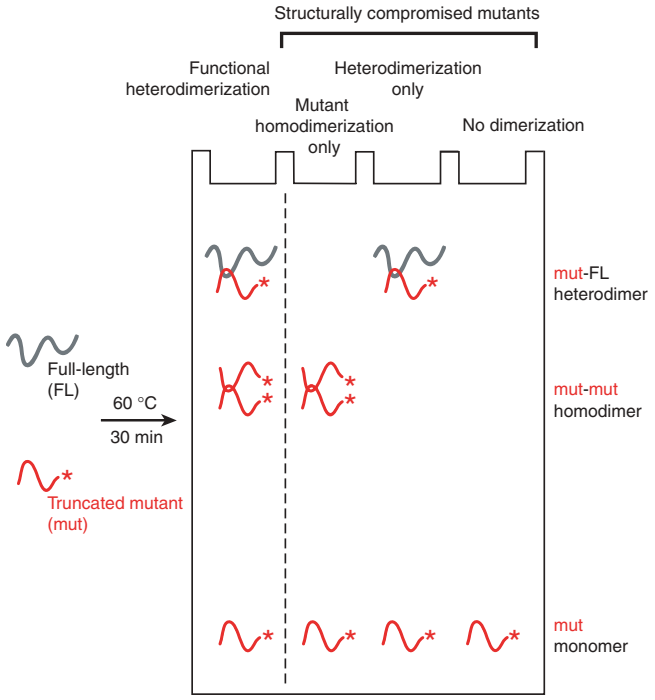


Figure 2 Competitive-dimerization assay for stringent definition of RNA structures essential for dimerization. Assay uses radiolabeled mutant RNA (mut; in red, with asterisk) and unlabeled full-length RNA (FL; gray). Only species containing the mutant RNA are visualized in the nondenaturing gel.

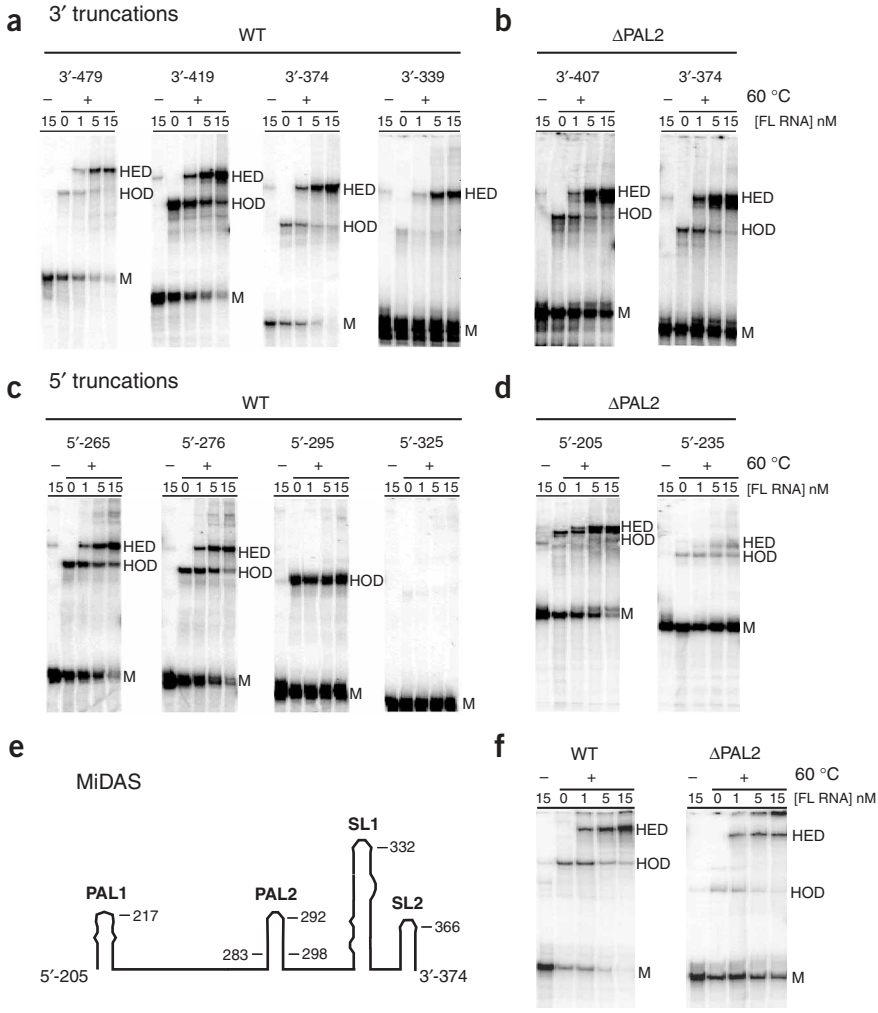


Figure 3 The MiDAS for MuSV defined by competitive dimerization. (**a**,**b**) 3′-end truncations in the native and ΔPAL2 contexts. (**c**,**d**) 5′-end truncations. (**e**) Schematic structure for the MiDAS in the context of the conventional secondary structure. (**f**) Efficient dimerization of the MiDAS in both native and ΔPAL2 contexts. FL, full-length; M, mutant monomer; HED, mutant-FL heterodimer; HOD, mutant-mutant homodimer.

boundary for the minimum dimerization active sequence was thus set at position 276.

The behavior of the 5′-295 mutant illustrates the stringency of the competitive dimerization assay. Although this RNA would have been scored as dimerization competent under less stringent conditions, it clearly lacks key elements required to dimerize competitively with the full-length RNA.

Deletion of PAL2 unmasks the contribution of PAL1

The minimal dimerization active region defined by this initial analysis includes the PAL2 sequence, which several groups[8,18–20,25–27,32] have proposed plays a role in dimerization. To explore whether any other accessory sequences contribute to dimerization but are masked by PAL2, we compromised PAL2 by removing nt 283–294 (ΔPAL2 mutant, **Fig. 1b**) and retested the panel of 5′ and 3′ truncations by competition with a full-length RNA also harboring the ΔPAL deletion. All ΔPAL2 3′ truncations through nt 374 formed both homo- and heterodimers efficiently (**Fig. 3b**); this indicated that the 3′ boundary of the MiDAS remains at position 374.

When the 5′ series was analyzed, truncations through nt 205 yielded fully functional RNAs. In contrast, both homo- and heterodimer formation was significantly impaired when the RNA was truncated through 5′-235 in the ΔPAL2 context (5′-235 panel, **Fig. 3d**). In addition, time-resolved dimerization experiments showed that a construct spanning the PAL2 through SL2 sequences dimerized only to about 80% (at concentrations up to 50 nM). In contrast, RNAs that also included the additional 5′ 205–275 sequences dimerized to completion (**Fig. 3f** and data not shown). We concluded that the role of sequences containing the PAL1 region (nt 205–217) is partially masked if PAL2 is present, and we assigned the 5′ boundary of the minimal dimerization domain to position 205 (**Fig. 3d**).

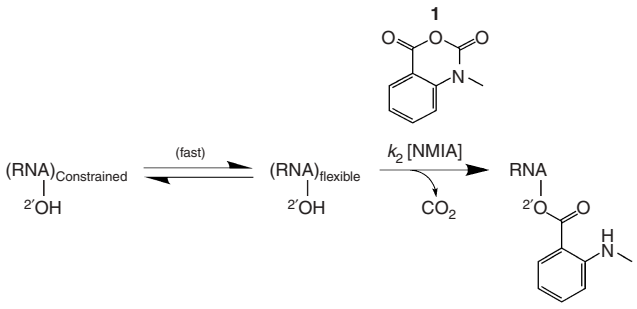
A minimal dimerization domain

We constructed a minimal RNA spanning MuSV sequences from 205 to 374 and tested the ability of this RNA to function in the competitive dimerization assay, both in the native PAL2 and ΔPAL2 contexts (**Fig. 3f**). Both RNAs formed single-conformation monomers, homodimers and heterodimers identical to those observed for native-like truncations. We infer that nt 205–374 span the minimal dimerization domain for MuSV and, potentially, for most gamma retroviruses.

Our MiDAS (**Fig. 3e**) incorporates RNA elements that have been proposed to either contribute or be a primary determinant for retroviral dimerization[8,10,17–21,25,33]. The competitive dimerization assay supports proposals[17,18] that the role of the PAL1 sequence may be especially important if the PAL2 sequence is compromised. Unique to the competitive dimerization assay is the strong inference that the MiDAS spans all RNA sequences stringently required for dimerization. The dimerization sequences defined here *in vitro* correspond closely to those from analogous experiments designed to define a MiDAS *in vivo*: an RNA spanning positions 215 to 404 from the Moloney murine leukemia virus (MuLV) is sufficient to increase packaging of a nonviral RNA, in dimeric form, by 50-fold[5]. To the extent that other viral components such as Gag or the nucleocapsid protein augment, but do not fundamentally alter, an RNA-centered process, the MiDAS represents a rigorously evaluated minimal domain for retroviral dimerization.

Domain structure analyzed by RNA SHAPE chemistry

Our laboratory has recently developed a single-nucleotide-resolution approach to examine, quantitatively, the local environment at every nucleotide in an RNA[23,24]. RNA ribose 2′-hydroxyl groups react with *N*-methylisatoic anhydride (NMIA, **1**) to form the nucleotide 2′-ester. 2′-Hydroxyl reactivity is gated by whether or not a given nucleotide is constrained by base pairing or tertiary interactions[23,34]. Flexible nucleotides react preferentially because they are better able to reach a conformation that facilitates nucleophilic attack of the 2′-hydroxyl



Scheme 1 Structure-selective reaction of RNA 2′-hydroxyl groups with NMIA (**1**).

on NMIA (**Scheme 1**). Formation of the bulky 2′-*O*-adduct is readily detected as a stop to reverse transcriptase–mediated primer extension: the complete experiment involves selective 2′-hydroxyl acylation followed by primer extension.

SHAPE experiments were performed in the context of a MiDAS RNA with 30- and 5-nt native sequence extensions, respectively, at the 5′ and 3′ ends to facilitate analysis of the entire domain by primer extension. We also appended an RNA cassette[23] to the 3′ end that contains an efficient DNA primer binding site.

Refolded MiDAS RNA was treated with NMIA, and sites of 2′-*O*-adduct formation were detected by primer extension, resolved on sequencing gels (**Fig. 4a**). The monomeric state of the RNA was confirmed by native gel analysis. Comparison of reactions performed in the presence of NMIA with reactions omitting the reagent reveals selective formation of 2′-*O*-adducts at a subset of sites in the RNA (compare + and − NMIA lanes in the MiDAS panel, **Fig. 4**). Individual band intensities were integrated[35] and absolute reactivities were computed for every position in the MiDAS RNA construct.

The nucleotide-resolution SHAPE experiment provides a large number of constraints that must be accommodated in any secondary structure prediction for the MiDAS RNA. We screened secondary structures for the MiDAS region (residues 205–374) by submitting positions whose calculated reactivity was at least 25% of the strongest observed reactivities (47 nt total) as chemical modification constraints to the RNAStructure program[15]. The quantitative data are shown superimposed on a secondary structure consistent with the entire body of SHAPE reactivity in **Figure 5**.

Residues with high and moderate reactivity (red and orange, **Fig. 5**) toward NMIA are located in single-stranded loops and connecting structures. Positions with low or undetectable reactivity (blue and black, **Fig. 5**) lie largely in base-paired helices. Because SHAPE is sensitive to any interaction that constrains a nucleotide[23], including noncanonical interactions, reactive positions should fall cleanly in flexible RNA structures, whereas some unreactive nucleotides may reflect tertiary structure constraints that are yet to be defined at this stage of analysis. As expected[23,24], the 3′ RNA structure cassette (**Fig. 5**, left) has a reactivity pattern exactly consistent with its designed fold, indicating that this appended structure does not interfere with folding of the MiDAS RNA.

SHAPE analysis (**Figs. 4** and **5**) strongly supports secondary structures for the PAL1 loop, SL2 and the upper portion of SL1 that are consistent with earlier proposals. In contrast, the PAL2 sequence (positions 283–303; see **Fig. 3e** for comparison), which has been almost universally assumed to form a stable stem-loop structure, is highly reactive toward NMIA. Moreover, the reactive PAL2 sequence resides in the middle of a larger flexible domain in which most nucleotides are reactive by SHAPE chemistry (**Fig. 5**).

PAL2 is unstructured in the MiDAS monomer

Because the observed structure is significantly different from conventional models for the dimerization domain, we analyzed the structure of two MiDAS RNAs carrying instructive mutations in PAL2 (**Fig. 4b**). The first mutant, PALSTB (PAL stabilization), was designed to stabilize PAL2 in the conventional stem-loop structure by increasing the G-C base-pair content at flanking helix positions (circled positions, **Fig. 4b**). Inspection of the SHAPE data shows that stabilizing the PAL2 duplex has the desired effect. Nucleotides located in the PAL2 loop are strongly reactive, whereas base-paired positions in the stem are now much less reactive than in the native sequence (compare the MiDAS and PALSTB lanes, **Fig. 4a**).

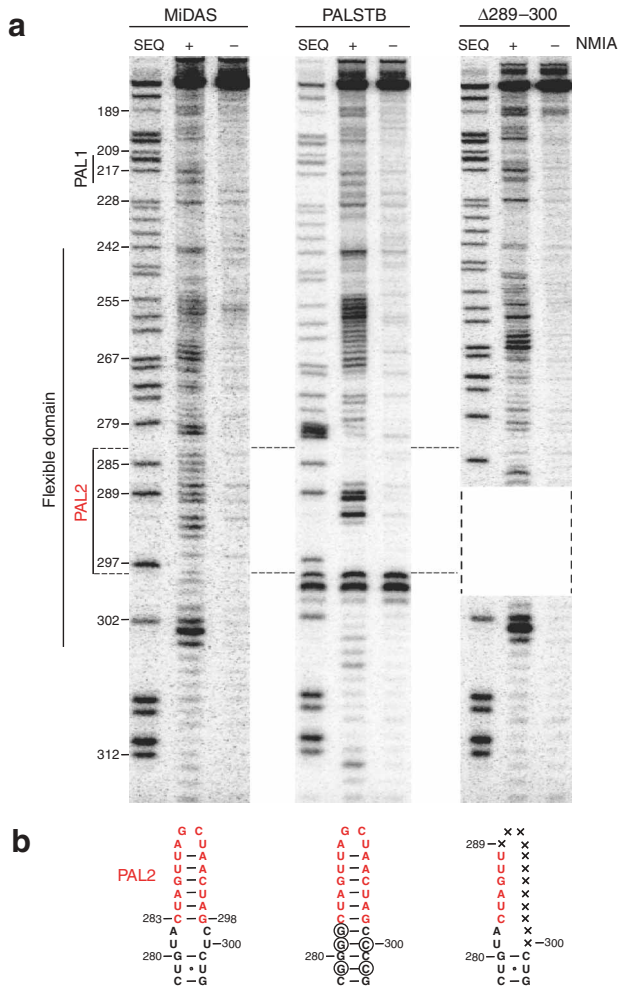


Figure 4 SHAPE analysis of the MuSV MiDAS RNA and of the PALSTB and Δ289–300 mutants. (**a**) 2′-*O*-adduct formation visualized in a sequencing gel. Reactions were performed in the presence (+) and absence (−) of NMIA. Sequencing lanes (seq) showing guanosine positions were generated by dideoxy nucleotide incorporation during primer extension. Extension products in the dideoxy sequencing ladder are exactly 1 nt longer than those in the corresponding NMIA lane; nucleotide positions are labeled with respect to NMIA lanes. (**b**) Sequences of (left to right) the native RNA and of the PALSTB and Δ289–300 mutants, drawn in the context of the conventional structure for PAL2.

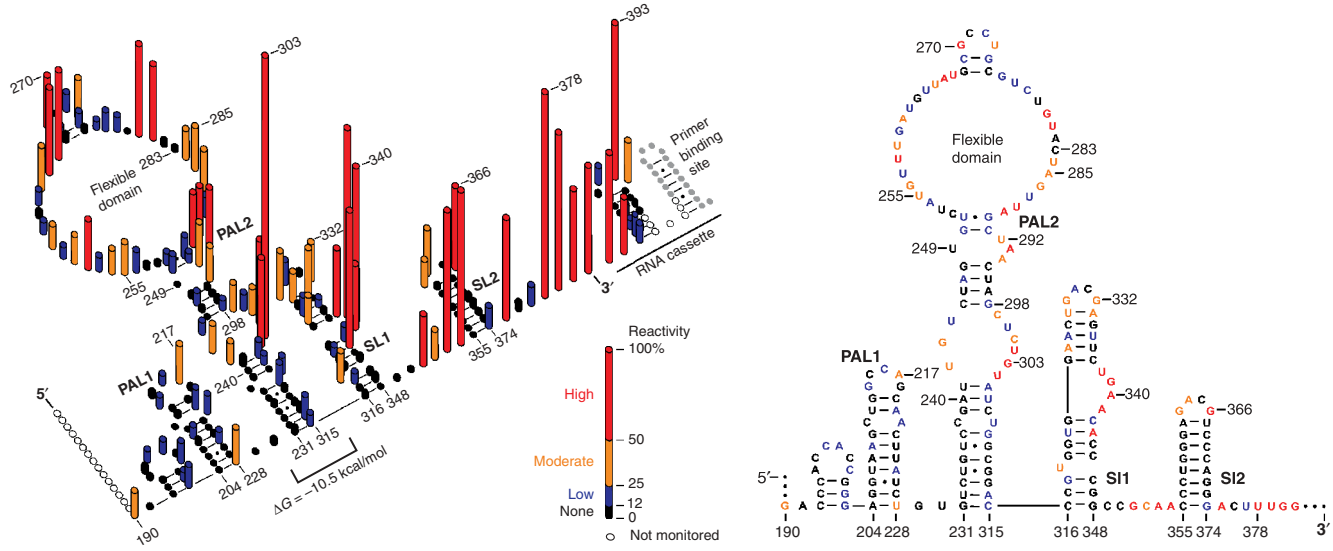


Figure 5 Secondary structure model of the MuSV MiDAS RNA. Left, quantitative NMIA reactivities, minus background, are superimposed as columns at each nucleotide position. Right shows the same secondary structure, but with residues labeled explicitly.

The experimental SHAPE reactivity data for the PALSTB mutant was subtracted from that for the native MiDAS RNA to create a quantitative difference map for every position in the PALSTB RNA (**Fig. 6a**). In the difference map, residues that are more reactive or more constrained in the mutant relative to the native MiDAS sequence are reported as positive and negative amplitudes, respectively (red and blue, **Fig. 6**). If PAL2 already existed as a hairpin in the monomeric native state, stabilizing this stem should have a minimal effect on global MiDAS RNA structure. In strong contrast to this expectation, stabilizing the PAL2 sequence as a stem-loop causes large changes to the SHAPE reactivity in the MiDAS domain.

Consistent with the design of this mutant, nucleotides in the loop at the apex of the PAL2 stem in the PALSTB mutant show an increase in reactivity, and nucleotides in the stabilized stem are much less reactive than in the native MiDAS sequence (see PAL2, **Fig. 6a**). More significantly, the PALSTB mutant shows very large changes in global structure that extend almost the entire length of the RNA and up to 80 nt away (see especially nt 252–268 in the flexible domain and nt 374–382 between SL2 and the 3′ end of the RNA, **Fig. 6a**). The peaks shown in the difference map are plotted on a scale comparable to that used in **Fig. 5**. Thus, the large positive peaks centered at positions 255, 312, 365 and 380 represent significant enhancements in absolute local nucleotide flexibility in these regions (for example, compare the 255 region in **Fig. 4** with the difference map in **Fig. 6a**).

Our second mutant, Δ289–300, was designed to delete a large region within PAL2 that is flexible as judged by SHAPE chemistry (mutant is shown in **Fig. 4b**; flexible region is labeled in **Fig. 5**). If the conventional stem-loop model for PAL2 were correct, this 12-nt deletion should have a strong effect on global MiDAS structure. On the other hand, if PAL2 is unstructured in the native monomer state, as indicated by SHAPE chemistry, then this extensive deletion may have only a minimal effect on global MiDAS structure. A difference map for the Δ289–300 mutant shows that this deletion, in fact, introduces very modest changes to local nucleotide flexibility in the MiDAS RNA and induces virtually no significant structural change over large regions of the sequence (**Fig. 6b**). The most significant effect is an increase in SHAPE reactivity in the (already flexible) 267 region of the flexible domain and a decrease in reactivity at the bulge in SL1 at nt 340. Thus, direct analysis of the MiDAS RNA (**Fig. 5**) and differential analysis of two mutants (**Fig. 6**) both strongly support a

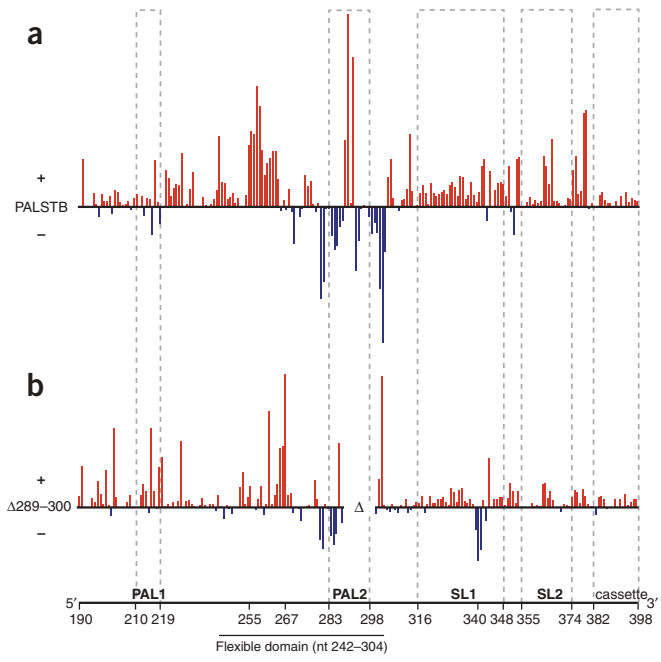


Figure 6 Quantitative difference maps for the effects of mutations on MiDAS structure. (**a**) PALSTB mutation. (**b**) Δ289–300 mutation. Vertical bars show absolute NMIA reactivities at each base position for the mutant RNA minus reactivity of the native RNA. Positive (red) and negative (blue) differences indicate increased versus reduced reactivity, respectively, in the mutants relative to the native MiDAS. Vertical scales are the same in upper and lower panels; the overall shorter bars in the lower panel reflect the more modest structural perturbation introduced by the Δ289–300 mutant.

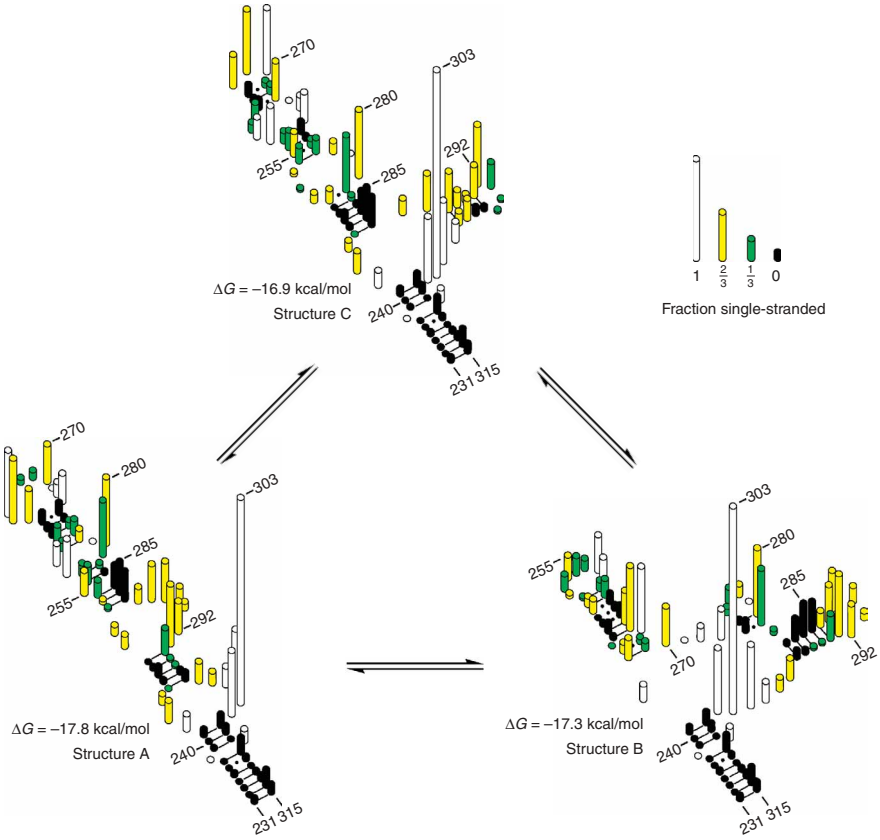


Figure 7 Structural model for overall flexibility in the 231–315 domain. Column heights indicate absolute NMIA reactivities (and are the same as reported in **Fig. 5**). Column colors illustrate the fraction of structures in which a position is single-stranded in this ensemble of three representative structures.

new model for the dimerization domain of MuSV in which the PAL2 sequence resides in an extensive flexible domain.

DISCUSSION

Our model for the minimal dimerization active structure of MuSV makes use of two innovations that are generalizable to any RNA structure prediction problem. First, we defined a minimal sequence for dimerization using an assay that requires the simplified RNA to functionally compete with a native-like sequence (**Fig. 2**). Second, SHAPE chemistry quantitatively interrogates every nucleotide in an RNA (**Scheme 1**), which means that secondary structure models can be evaluated with much greater confidence than when traditional chemical and enzymatic reagents are used.

Although the MiDAS secondary structure (**Fig. 5**) proposed here differs significantly from conventional models, this structure is consistent with two earlier sets of experimental information for dimerization domains in gamma retroviruses. The sequence of MuLV is almost identical to that of MuSV. The nucleotide resolution SHAPE information strongly supports the original MuLV model[8] for SL2 and the upper portion of SL1. In contrast, SHAPE does not support the earlier proposal that PAL2 forms a stable stem-loop structure. However, superposition of the chemical mapping information for MuLV on our secondary structure for MuSV shows that the previous information is exactly consistent with the present MiDAS proposal (**Supplementary Fig. 2** online). In particular, the PAL2 sequence is strongly reactive toward conventional single-strand–selective chemical reagents[8] and is thus consistent with the idea that PAL2 lies in a flexible domain.

Structure-mapping studies on the Harvey sarcoma virus (HaSV) also emphasize the importance of SL1- and SL2-like structures in the dimerization domain[28]. HaSV does not contain a PAL2 sequence[9,28], but the HaSV RNA can be folded into a secondary structure that is similar to that of MuSV and for which RNase-based cleavage data strongly support formation of an internal flexible domain (**Supplementary Fig. 2** online). Many sites in the HaSV domain are cleaved by both single- and double-strand–selective RNases[28] (**Supplementary Fig. 2** online). We infer that the HaSV RNA probably contains a flexible domain in which portions of the structure are alternately both paired and flexible in distinct conformations. That the MuLV and HaSV RNAs fold to similar monomeric starting structures provides a structural basis for the observation that these viruses readily heterodimerize[28], presumably via PAL1.

We folded the flexible domain, including its anchoring helix (spanning positions 231–315, **Fig. 5**), subject to the requirement that the 27 positions with high and moderate reactivities be single-stranded. The lowest-energy structure, which is compatible with all of the SHAPE information (**Fig. 5**), has a total calculated[15] folding free energy of only −10.5 kcal/mol. This single low-energy structure spans 84 nt and thus has a net stability comparable to that of a simple stem-loop structure containing roughly three base pairs. Moreover, although the entire flexible domain from nt 249 through 294 contains no instances in which there are more than two strongly constrained nucleotides in a row (black positions, **Fig. 5**), individual nucleotides vary significantly in their 2′-hydroxyl reactivity.

We therefore evaluated the alternative hypothesis that several more stable structures might be compatible with the SHAPE information. We submitted the 12 most highly reactive sites (red, **Fig. 5**) as chemical modification constraints for RNA structure prediction[15]. Four structures have calculated free energies within 10% of the most stable structure. Three of these structures have distinctive folds (**Fig. 7**), whereas the fourth (not shown) contains elements of the other structures. Absolute SHAPE reactivities were superimposed on the three most distinctive structures (**Fig. 7**). These intensity data are colored according to the fraction of structures in which they are single-stranded (always paired and always single-stranded are black and white, respectively). Each structure (A, B or C) is only partially consistent with the SHAPE data. Each of these structures, however, has a calculated folding free energy of approximately −17 kcal/mol, and thus is significantly more stable than the single consensus structure that incorporates all of the flexibility information.

This semiquantitative analysis, in which the large universe of possible structures is approximated by three low-energy structures (**Fig. 7**), supports a general model for RNA folding in which net, long-range flexibility in RNA can reflect contributions from several stable structures of similar energies.

Models for the genomic RNA retroviral dimerization domain, in which PAL2 forms a stable stem-loop structure (**Figs. 1a** and **4b**), have guided the gamma retrovirus field for over a decade. However, the

model-independent SHAPE intensity information (**Figs. 4a** and **5**) emphasizes that existing structural models merit careful reinterpretation.

For example, many mechanistic analyses of retroviral dimerization have used simplified RNAs or RNAs in which PAL2 was mutated to enforce the conventional structure (**Fig. 1a**) for the RNA. The RNA structure in such mutants should be roughly similar to that of the PALSTB mutant, which was also designed to artificially reinforce the conventional structure for PAL2. The PALSTB mutant yielded extensive changes to the global structure of the MiDAS, including at residues up to 80 nt distant from PAL2 (**Fig. 6a**). Thus, the structure of the retroviral dimerization domain with a native sequence can be quite different from that of RNAs containing mutations in PAL2 or that are shorter than the MiDAS. Moreover, the significant global changes that occur in the MiDAS domain upon introduction of mutations in PAL2 emphasize that the flexible domain communicates via long-range interactions with other regions of the RNA.

Because PAL2 (and PAL1) sequences are self-complementary, an attractive model for the noncovalent interactions that stabilize the retroviral dimer is for these sequences to form extended duplexes in the dimer[10,17–19]. Earlier models proposing that PAL2 initially exists as a stable stem-loop recognized that it might be energetically costly to disrupt the extensive pre-existing base pairing in this structure. These models thus generally proposed that dimerization proceeds stepwise, via base pairing between nucleotides in the loops of two PAL2 stem-loop structures followed by helix extension.

The nucleotide-resolution SHAPE experiments demonstrate that the PAL2 sequence instead lies in an RNA domain in which, on average, most nucleotides either are unconstrained by base pairing or are transiently in a single-stranded conformation (**Fig. 7**). Continuing SHAPE experiments do support formation of extended duplexes in the dimer (data not shown). Thus, formation of a flexible domain in the monomeric starting state has significant mechanistic implications for retroviral RNA genome dimerization. (i) Dimerization via PAL2, in the context of a flexible domain, is potentially much more thermodynamically favorable than previously thought, because fewer base-pair interactions in the monomer have to be disrupted to form extended duplexes between two PAL2 sequences in the dimer. (ii) Placing the PAL2 sequence in a flexible domain may also kinetically enhance retroviral RNA dimerization by lowering the activation barrier for extended duplex formation. (iii) The distinct conformations visualized for the flexible domain probably also have different dimerization activities. Retroviral dimerization could thus be modulated by interactions between these distinct conformations and other regions of the genomic RNA or retroviral proteins.

METHODS

Retroviral RNA transcripts. DNA templates for *in vitro* transcription of the full-length RNA, 5′ and 3′ truncations, and MiDAS constructs were generated by PCR from the pLNBS[26,27] plasmid. RNA constructs were generated with T7 RNA polymerase–mediated transcription (500 µl, 37 °C, 5 h) containing 80 mM HEPES (pH 7.4), 40 mM dithiothreitol (DTT), 0.01% (v/v) Triton X-100, 2 mM spermidine, 10 mM $MgCl_2$, 2 mM each nucleoside triphosphate, ~25 µg of PCR-generated template, 20 U of SUPERase-In (Ambion) and 0.1 mg/ml of polymerase. Internally labeled RNAs were synthesized with 20 µCi of α-[^{32}P]ATP and unlabeled ATP at 0.5 mM. RNAs were purified by denaturing gel electrophoresis (5% polyacrylamide, 7 M urea), excised from the gel, eluted overnight into 1/2× TBE (45 mM Tris-borate, 1 mM EDTA) and concentrated by ethanol precipitation. RNAs were resuspended in 10 mM HEPES (pH 7.5) and 1 mM EDTA and stored at −20 °C.

Competitive dimerization assay. Truncated RNA internally labeled with [^{32}P] (~1.5 nM) was incubated with unlabeled full-length RNA (at 1, 5 or 15 nM in 15 µl). Reactions were heated to 90 °C to eliminate pre-existing dimers, rapidly cooled on ice, treated with 5 µl 4× dimerization buffer (200 mM HEPES (pH 7.5), 800 mM potassium acetate (pH 7.5), 20 mM $MgCl_2$ at 25 °C), incubated at 60 °C for 30 min and placed on ice. Samples (3 µl) were mixed with 1 µl of 30% (v/v) glycerol (containing marker dyes) and resolved by nondenaturing electrophoresis at 4 °C. Gels (5% polyacrylamide in TBE) were pre-run for 15 min before sample loading and subsequently run for 2 h at 20 W.

SHAPE analysis of MuSV monomers. SHAPE experiments were performed with a MiDAS RNA that contained flanking 5′ and 3′ extensions of viral sequence of 30 and 5 nt, respectively; a 3′ nonviral RNA cassette containing an efficient DNA primer binding site[23] was appended to the 3′ end. The MiDAS RNA construct (10 pmol) was heated at 90 °C for 3 min in 7.2 µl of water, cooled on ice, treated with 1.8 µl of 5× dimerization buffer (250 mM HEPES (pH 8.0), 1 M potassium acetate (pH 7.5), 25 mM $MgCl_2$), incubated at room temperature (~25 °C) for 30 s and returned to ice. The RNA solution was then equilibrated at 37 °C for 5 min, treated with NMIA (1 µl, 180 mM in anhydrous DMSO), allowed to react for 50 min (approximately five half lives[23,24]) at 37 °C, and placed on ice. Control reactions contained DMSO without NMIA.

Primer extension. Two DNA primers were used to analyze the MiDAS RNA construct. Primers were complementary to the 3′ end of the RNA structure cassette (5′-GAA CCG GAC CGA AGC CCG) and to SL1 (5′-CAG AAC TCG TCA GTT CCA CCA). We performed primer extension reactions by adding modified RNA (2 µl, 2 pmol) and 5′-[^{32}P]DNA primer (1 µl, 1 pmol) to 9 µl water and annealing by incubation at 95 °C (30 sec), 60 °C (6 min) and 35 °C (10 min). Reverse transcription buffer (7 µl; 143 mM Tris (pH 8.3), 214 mM KCl, 7.14 mM $MgCl_2$, 1.43 mM each dNTP, 14.3 mM DTT) was added, and subsequent primer extension steps were performed exactly as described[23], except that primer extension was performed at 48.5 °C. cDNA fragments were resolved on a series of 8% (w/v) polyacrylamide gels to achieve nucleotide resolution throughout the region analyzed.

Accession codes. BIND identifier (http://bind.ca/): 295526.

Note: Supplementary information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

This work was supported by a grant from the US National Institutes of Health (GM64803 to K.M.W. and A. Kaplan). We are indebted to A. Kaplan, C. Gherghe and A. Rein for many helpful discussions; to E. Merino and K. Wilkinson for assistance with SHAPE chemistry; and to D. Mathews for extensive advice with the RNAStructure program.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing interests.

Received 15 March; accepted 16 May 2005
Published online at http://www.nature.com/naturechemicalbiology/

1. Murti, K.G., Bondurant, M. & Tereba, A. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol.* **37**, 411–419 (1981).
2. Darlix, J.L., Lapadat-Tapolski, M., de Rocquigny, H. & Roques, B.P. First glimpses at structure-function relationships of the nucleocapsid protein of retroviruses. *J. Mol. Biol.* **254**, 523–537 (1995).
3. Paillart, J. *et al.* A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J. Virol.* **70**, 8348–8354 (1996).
4. Laughrea, M. *et al.* Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and proviral DNA synthesis. *J. Virol.* **71**, 3397–3406 (1997).
5. Hibbert, C.S., Mirro, J. & Rein, A. mRNA molecules containing murine leukemia virus packaging signals are encapsidated as dimers. *J. Virol.* **78**, 10927–10938 (2004).
6. Mikkelsen, J.G., Lund, A.H., Duch, M. & Pedersen, F.S. Recombination in the 5′ leader of murine leukemia virus is accurate and influenced by sequence identity with a strong bias toward the kissing-loop dimerization region. *J. Virol.* **72**, 6967–6978 (1998).
7. Mikkelsen, J.G., Lund, A.H., Duch, M. & Pedersen, F.S. Mutations of the kissing-loop dimerization sequence influence the site specificity of murine leukemia virus recombination in vivo. *J. Virol.* **74**, 600–610 (2000).
8. Tounekti, N. *et al.* Effect of dimerization on the conformation of the encapsidation psi domain of the Moloney murine leukemia virus. *J. Mol. Biol.* **223**, 205–220 (1992).

9. Konings, D.A.M., Nash, M.A., Maizel, J.V. & Arlinghaus, R.B. Novel GACG-hairpin pair motif in the 5′ untranslated region of type C retroviruses related to murine leukemia virus. *J. Virol.* **66**, 632–640 (1992).
10. Paillart, J., Marquet, R., Skripkin, E., Ehresmann, C. & Ehresmann, B. Dimerization of retroviral genomic RNAs: structural and functional implications. *Biochimie* **78**, 639–653 (1996).
11. D'Souza, V., Dey, A., Habib, D. & Summers, M.F. NMR Structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* **337**, 427–442 (2004).
12. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610 (1990).
13. Frank, D.N. & Pace, N.R. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153–180 (1998).
14. Gutell, R.R., Lee, J.C. & Cannone, J.J. The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.* **12**, 301–310 (2002).
15. Mathews, D.H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **101**, 7287–7292 (2004).
16. Dowell, R.D. & Eddy, S.R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 71 (2004).
17. Oroudjev, E.M., Kang, P.C.E. & Kohlstaedt, L.A. An additional dimer linkage structure in Moloney murine leukemia virus RNA. *J. Mol. Biol.* **291**, 603–613 (1999).
18. Ly, H. & Parslow, T.G. Bipartite signal for genomic RNA dimerization in Moloney murine leukemia virus. *J. Virol.* **76**, 3135–3144 (2002).
19. D'Souza, V. *et al.* 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 (2001).
20. De Tapia, M., Metzler, V., Mougel, M., Ehresmann, B. & Ehresmann, C. Dimerization of the Moloney murine leukemia virus genomic RNA: redefinition of the role of the palindromic stem-loop H1 (278–303) and new roles for stem-loops H2 (310–352) and H3 (355–374). *Biochemistry* **37**, 6077–6085 (1998).
21. Kim, C. & Tinoco, I. A retroviral RNA kissing complex containing only two GC base pairs. *Proc. Natl. Acad. Sci. USA* **97**, 9396–9401 (2000).
22. Rein, A., Harvin, D.P., Mirro, J., Ernst, S.M. & Gorelick, R.J. Evidence that a central domain of nucleocapsid protein is required for RNA packaging in murine leukemia virus. *J. Virol.* **68**, 6124–6129 (1994).
23. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
24. Wilkinson, K.A., Merino, E.J. & Weeks, K.M. RNA SHAPE chemistry reveals non-hierarchical interactions dominate equilibrium structural transitions in $tRNA^{Asp}$ transcripts. *J. Am. Chem. Soc.* **127**, 4659–4667 (2005).
25. Girard, P.M., Bonnet-Mathoniere, B., Muriaux, D. & Paoletti, J. A short autocomplementary sequence in the 5′ leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry* **34**, 9785–9794 (1995).
26. Ly, H., Nierlich, D., Olsen, J. & Kaplan, A. Moloney murine sarcoma virus genomic RNAs dimerize via a two-step process: a concentration-dependent kissing-loop interaction is driven by initial contact between consecutive guanosines. *J. Virol.* **73**, 7255–7261 (1999).
27. Ly, H., Nierlich, D., Olsen, J. & Kaplan, A. Functional characterization of the dimer linkage structure RNA of Moloney murine sarcoma virus. *J. Virol.* **74**, 9937–9945 (2000).
28. Rasmussen, S.V., Mikkelsen, J.G. & Pedersen, F.S. Modulation of homo- and hetero-dimerization of Harvey sarcoma virus RNA by GACG tetraloops and point mutations in palindromic sequences. *J. Mol. Biol.* **323**, 613–628 (2002).
29. Aagaard, L., Rasmussen, S.V., Mikkelsen, J.G. & Pedersen, F.S. Efficient replication of full-length murine leukemia viruses modified at the dimer initiation site regions. *Virology* **318**, 360–370 (2004).
30. Holbrook, S.R., Cheong, C., Tinoco, I., Jr. & Kim, S.H. Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature* **353**, 579–581 (1991).
31. Wild, K., Weichenrieder, O., Leonard, G.A. & Cusack, S. The 2 Å structure of helix 6 of the human signal recognition particle RNA. *Struct. Fold. Des.* **7**, 1345–1352 (1999).
32. D'Souza, V. & Summers, M.F. Structural basis for packaging the dimeric genome of Moloney murine leukemia virus. *Nature* **431**, 586–590 (2004).
33. Monie, T.P. *et al.* Identification and visualization of the dimerization initiation site of the prototype lentivirus, Maedi Visna virus: a potential GACG tetraloop displays structural homology with the α- and γ-retroviruses. *Biochemistry* **44**, 294–302 (2005).
34. Chamberlin, S.I., Merino, E.J. & Weeks, K.M. Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proc. Natl. Acad. Sci. USA* **99**, 14688–14693 (2002).
35. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D. & Altman, R.B. SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* **11**, 344–354 (2005).

Badorrek and Weeks
Nature Chem. Biol.
Supplementary Figures 1 & 2

3' Truncations

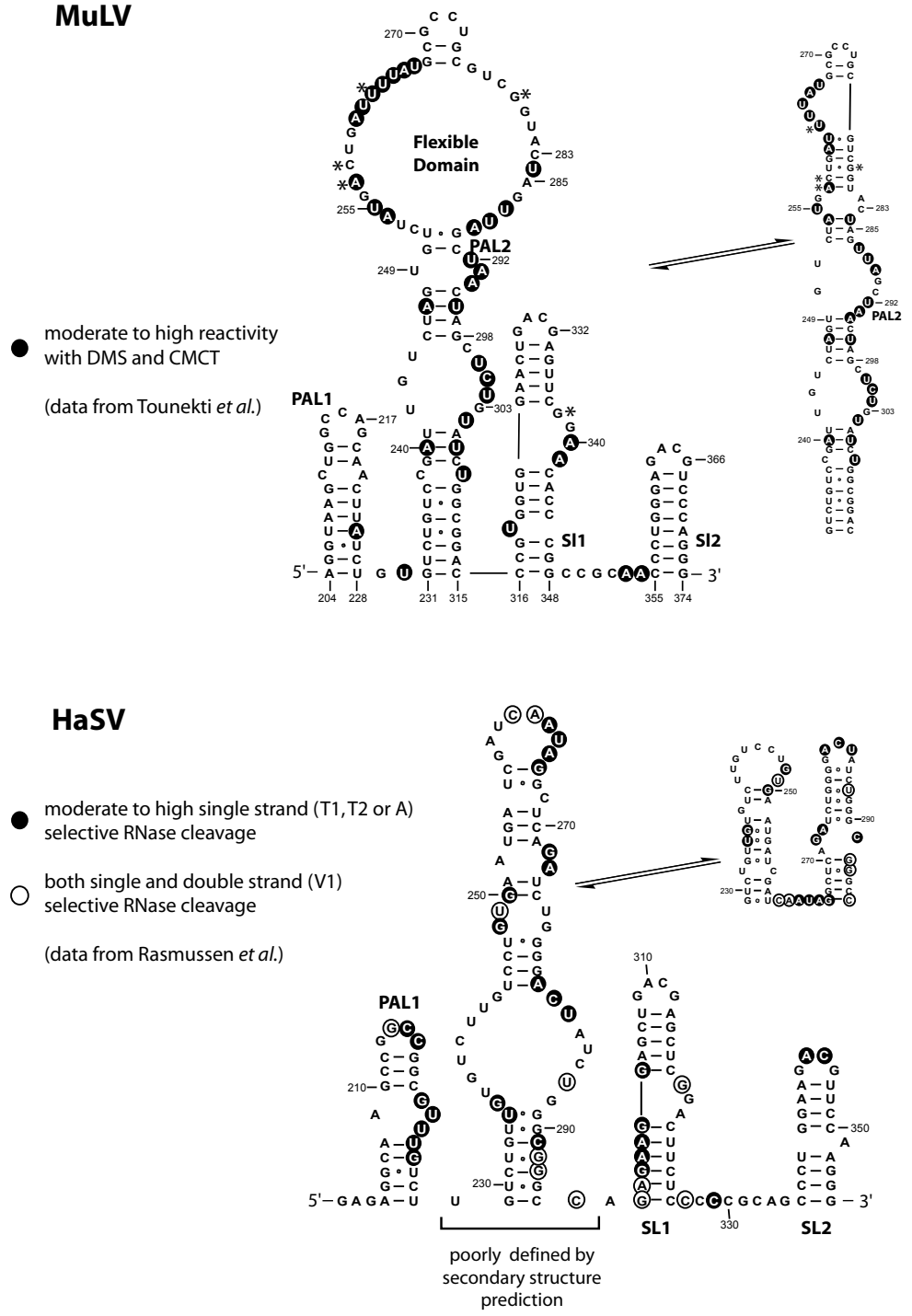
	native			ΔPAL2		
	HOD	HED	Native?	HOD	HED	Native?
3'-569	+	+	✓			
3'-539	+	+	✓			
3'-509	+	+	✓	+	+	✓
3'-479	+	+	✓	+	+	✓
3'-449	+	+	✓	+	+	✓
3'-419	+	+	✓	+	+	✓
3'-407	+	+	✓	+	+	✓
3'-381	+	+	✓	+	+	✓
3'-374	+	+	✓	+	+	✓
3'-354	+	+	M*			
3'-339	–	+	M*			
3'-324	–	+/–	x			
3'-309	+/–	+/–	x			
3'-303	+/–	+/–	x			

5' Truncations

	native			ΔPAL2		
	HOD	HED	Native?	HOD	HED	Native?
5'-175	+	+	✓	+	+	✓
5'-205	+	+	✓	+	+	✓
5'-235	+	+	✓	+/–	+/–	x
5'-265	+	+	✓	+/–	+/–	x
5'-276	+	+	✓			
5'-295	+	–	x			
5'-325	–	–	x			

+, native-like dimerization activity; **+/–**, dimerization occurs but is substantially compromised. ✓ and x, overall native-like versus compromised dimerization, respectively. M*, forms multiple monomer conformations.

Supplementary Figure 1. Dimerization activity of 3' and 5' truncation mutants.



Supplementary Figure 2. Proposed secondary structures for MuLV and HaSV dimerization domains. Mapping data are from refs. 8 and 28. Solid circles indicate positions reactive towards single-strand selective chemical reagents (DMS and CMCT; upper panel) or enzymes (RNasesT1, T2 and A; lower panel). Open circles (lower panel) indicate positions cleaved by both single- and double-strand (RNase V1) selective enzymes. Alternate structures for the flexible domains are shown as inserts. Asterisks (top panel) indicate minor sequence differences between MuLV and MuSV.