doi:10.1016/j.jmb.2004.01.037 J. Mol. Biol. (2004) 337, 427–442

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

NMR Structure of the 101-nucleotide Core Encapsidation Signal of the Moloney Murine Leukemia Virus

Victoria D'Souza, Anwesha Dey, Dina Habib and Michael F. Summers*

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

The full length, positive-strand genome of the Moloney Murine Leukemia Virus contains a "core encapsidation signal" that is essential for efficient genome packaging during virus assembly. We have determined the structure of a 101-nucleotide RNA that contains this signal (called mΨ) using a novel isotope-edited NMR approach. The method is robust and should be generally applicable to larger RNAs. mΨ folds into three stem loops, two of which (SL-C and SL-D) co-stack to form an extended helix. The third stem loop (SL-B) is connected to SL-C by a flexible, four-nucleotide linker. The structure contains five mismatched base-pairs, an unusual C·CG base-triple platform, and a novel "A-minor K-turn," in which unpaired adenosine bases A340 and A341 of a GGAA bulge pack in the minor groove of a proximal stem, and a bulged distal uridine (U319) forms a hydrogen bond with the phosphodiester of A341. Phylogenetic analyses indicate that these essential structural elements are conserved among the murine C-type retroviruses.

*Corresponding author

Keywords: MLV; NMR; A-minor K-turn; stem loop

Introduction

All retroviruses package two copies of their full length, 5′-capped and 3′-polyadenylated RNA genomes *via* mechanisms that are highly efficient but poorly understood. As the viral Gag and Gag-Pol proteins assemble in infected cells, their genomes are selected from a cytosolic pool that contains an ~100-fold excess of cellular and spliced viral mRNAs.[1] Packaging specificity is mediated by interactions between the nucleocapsid (NC) domains of the assembling Gag polyproteins and a segment of the viral genome called the Ψ-site, located between the 5′-upstream activator sequence and the Gag initiation codon.[1–4] The packaging signals generally overlap with elements responsible for splicing and dimerization, providing potential mechanisms for selectively packaging two copies of the full length genome and ignoring spliced viral messages.[1,2] Conserved secondary structural elements have been identified within the Ψ-sites of several retroviruses that are important for genome packaging and viral infectivity. However, information on such fundamental issues as the number of Gag molecules that mediate sequence (or structure)-specific binding, the location of the NC binding sites within the Ψ packaging signals, and the roles of the conserved RNA elements (e.g. to promote NC binding, dimerization, reverse transcription, or other functions), remains limited, at best.[1,2,5]

The Moloney Murine Leukemia Virus (MLV) is a simple retrovirus that has received considerable attention as a model for virus assembly,[6–21] and is currently the most widely used vector in human gene therapy trials.[22] Early studies identified a 350-nucleotide segment of the MLV genome that can function as an independent packaging signal (Figure 1).[6] Nucleotides 278–374 are conservatively substituted among all murine C-type retroviruses and were predicted to form three stem loop structures[7–15] (Figure 1). The first stem loop (SL-B) contains a palindromic AGCU tetraloop that promotes dimerization,[7,16–20] and stem loops SL-C and SL-D contain conserved GACG tetraloops[10] that are capable of forming "kissing interactions." The GACG tetraloops enhance the rate of dimerization

Abbreviations used: MLV, Moloney Murine Leukaemia Virus; NC, nucleocapsid; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; RDC, residual dipolar coupling.

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

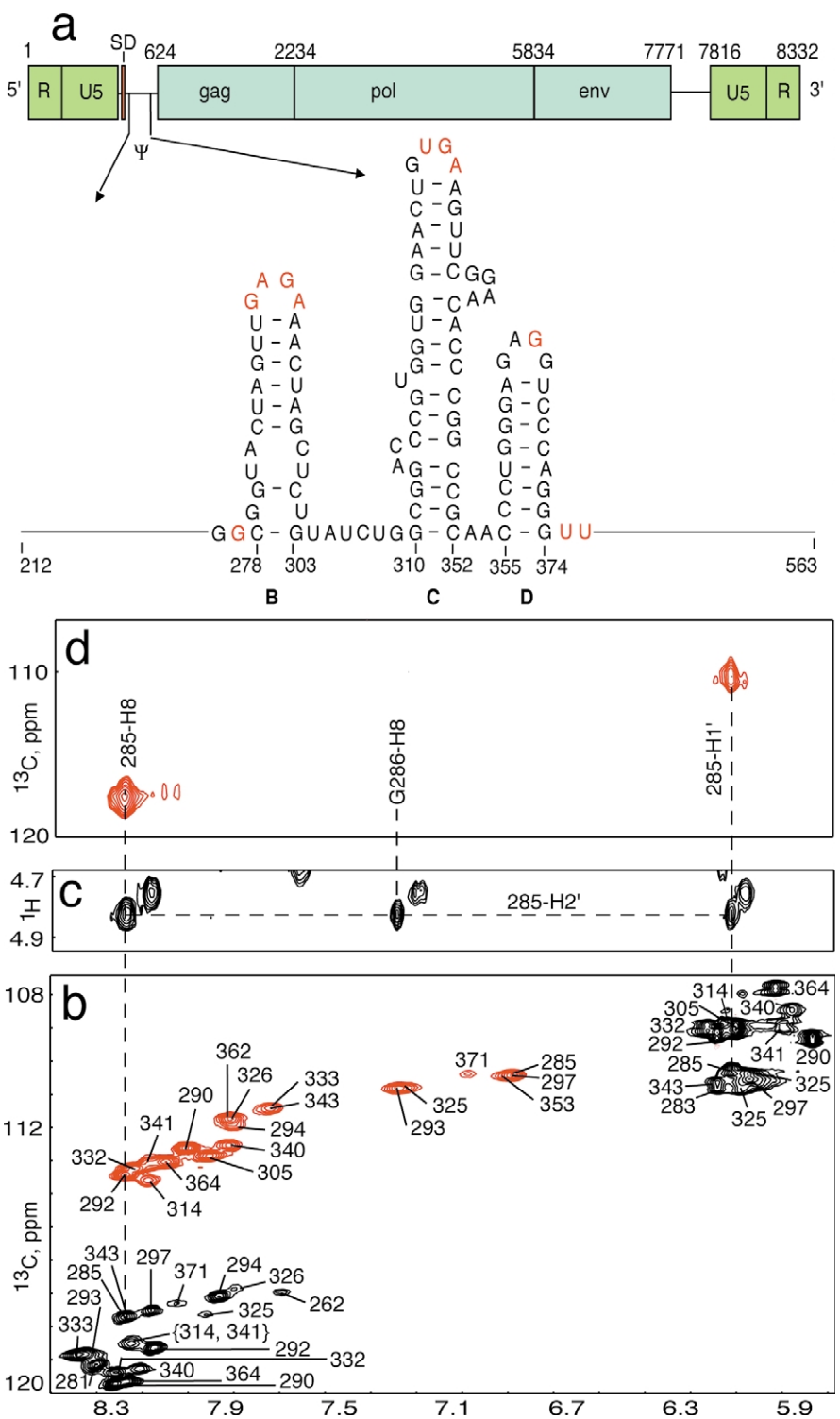

Figure 1. (a) Representation of the MLV genome showing the relative positions of the splice donor (SD) and Ψ-site, as well as the nucleotide sequence and predicted secondary structure of the portion of the Ψ-site that is essential for genome encapsidation. Nucleotides are numbered using the first residue after the 5′-cap as position 1. Non-native nucleotides that were added or modified to facilitate *in vitro* transcription, prevent oligomerization and stabilize the secondary structure are shown in red (see the text for details). Portions of the 2D $^{1}H-^{13}C$ HMQC (b), 3D ^{13}C-edited NOESY (c), and 4D $^{13}C,^{13}C$-edited NOESY (d) data obtained for $^{15}N,^{13}C$-adenosine labeled sample (A^{CN}-Ψ). These data illustrate a key element of the assignment strategy, in which NOEs involving un-like (un-labeled) residues are observed in the 3D spectrum but absent in the 4D data.

but do not affect the overall stability of the dimer.[18,21] Recent mutagenesis studies indicate that the secondary structures (but not necessarily the sequences) of stem loops C and D are important for genome packaging,[9,12,13,15] and it has been suggested that these stem loops function as a double hairpin motif.[10] Indeed, insertion of RNAs containing either stem loops C through D or B through D into non-packaged heterologous RNAs was sufficient to direct their packaging into virus-like particles at levels of ca 50% and 70% of wild-type levels, respectively,[13] providing strong evidence that stem loops C and D function as a "core encapsidation signal".[13]

In previous studies, we screened for interactions between the MLV NC protein and a series of oligoribonucleotides corresponding to segments of the Ψ-site.[23] Native gel electrophoresis data consistent with tight binding were only obtained for constructs containing all three stem loops (SL-BCD). Significantly, NC bound to both monomeric and dimeric forms of SL-BCD with similar affinities, and binding was insensitive to mutations in the three tetraloops. These and other results suggest a genome recognition mechanism that differs substantially from that utilized by HIV-1,[23] in which the NC domain of Gag binds to RNA tetraloops.[24,25] As a first step toward identifying the determinants of high-affinity NC binding, we have determined the structure of an SL-BCD construct containing tetraloop mutations that inhibit dimerization and aggregation, but that do not affect the affinity of the RNA for NC (mΨ, Figure 1).[23]

NMR Signal Assignments

NMR spectral assignment of relatively large RNAs can be complicated by chemical shift degeneracy and, in ^{13}C-labeled samples, severe line broadening resulting from strong one-bond ^{1}H–^{13}C dipolar coupling. Except for a very recent study of a 77-nucleotide IRES domain,[26] all RNA structures determined by NMR prior to this work comprise less than 45 nucleotides.[27,28] To assign the 101-nucleotide mΨ RNA, we developed an assignment strategy based on analysis of filtered nuclear Overhauser effect (NOE) data obtained for partial isotopically labeled samples. First, NOE-correlated signals were grouped according to nucleotide type from 3D ^{13}C-edited NOE spectroscopy (NOESY) data obtained for samples containing nucleotide-specific isotope labels (G^{CN}-Ψ, A^{CN}-Ψ, U^{CN}-Ψ, C^{CN}-Ψ, containing ^{13}C,^{15}N-labeled G, A, U, and C, respectively), as done previously for a 36-nucleotide RNA.[29] Inter- and intra-nucleotide NOEs were then differentiated by comparing the 3D spectra with 4D ^{13}C,^{13}C-NOESY spectra obtained for the same samples. This approach readily identifies stretches of sequential nucleotides of a given type, since intermolecular NOEs involving un-like nucleotides will be present in the 3D NOESY spectra but absent in the 4D spectra. In the case of the MLV Ψ-site, there are relatively few examples of such sequentially repeating residues: GGG (2), GG (8), AAA (1), AA (4), CCC (4), CC (1), UU (2). Thus, a stretch of three sequentially correlated adenosine bases is uniquely identified by this approach, and assignment options for stretches of other sequentially correlated nucleotides are limited. Representative ^{1}H,^{13}C-HMQC data obtained for A^{CN}-Ψ, and corresponding 3D ^{13}C-edited NOESY and 4D ^{13}C,^{13}C-edited NOESY data showing NOEs associated with A285, are presented in Figure 1(b–d). Sequential assignments for all other nucleotide pairs were based on ^{1}H NMR chemical shift matching of the inter-residue NOEs observed in the 3D ^{13}C-edited NOESY data, and portions of the spectra showing connectivities associated with the H2′ protons of G276–A282 are shown in Figure 2a. For systems with overlapping H2′ protons, correlations *via* the H1′ and/or H3′ protons often enabled or confirmed sequential signal assignments. NOEs in the 3D spectra associated with adeonsine-H2 protons provided both sequential and long-range connectivities (Figure 2b). Using this approach, sequential connectivities were established for more than 80% of the residues of mΨ.

Spin systems that could not be unambiguously assigned from the ^{13}C-edited data were sequentially correlated using 2D NOESY data obtained for nucleotide-specifically protonated mΨ samples (G^{H}-mΨ, A^{H}-mΨ, U^{H}-mΨ, C^{H}-mΨ which contained fully protonated G, A, U and C, respectively, with the remaining nucleotides being 90% per-deuterated). Partial deuteration of the adenosine and guanosine bases was advantageous and gave rise to weak but observable internucleotide (but not intranucleotide) NOE cross-peaks. A portion of the 2D NOESY spectrum obtained for G^{H}-mΨ showing NOEs involving both the fully protonated guanosine bases and partially protonated adenosine bases is given in Figure 2c. Spectra obtained for these samples exhibited much sharper lines and improved resolution compared with the ^{13}C-labeled samples, due mainly to the absence of ^{1}H–^{13}C (and to a lesser extent, ^{1}H–^{1}H) dipole-induced relaxation. With this approach, stretches of like nucleotides were unambiguously identified, and in combination with the ^{13}C-edited NOESY data, 100% of the aromatic C–H protons and carbons, 100% of the H1′, 91% of C1′, 80% of C2′–H2′, C3′–H3′, C4′–H4′ groups, and 75% of the C5′–H5′,5″ group of mΨ were assigned. Finally, signals for slowly exchanging imino protons (and nitrogen atoms) of base-paired nucleotides were assigned from 2D Watergate-NOESY and 3D Watergate-NOESY-HSQC spectra[30,31] obtained for an mΨ sample containing ^{15}N-labeled G and U ($G^{N}U^{N}$-mΨ) (Figure 3).

Analysis of the NOE Data

NOE cross-peak patterns and intensities

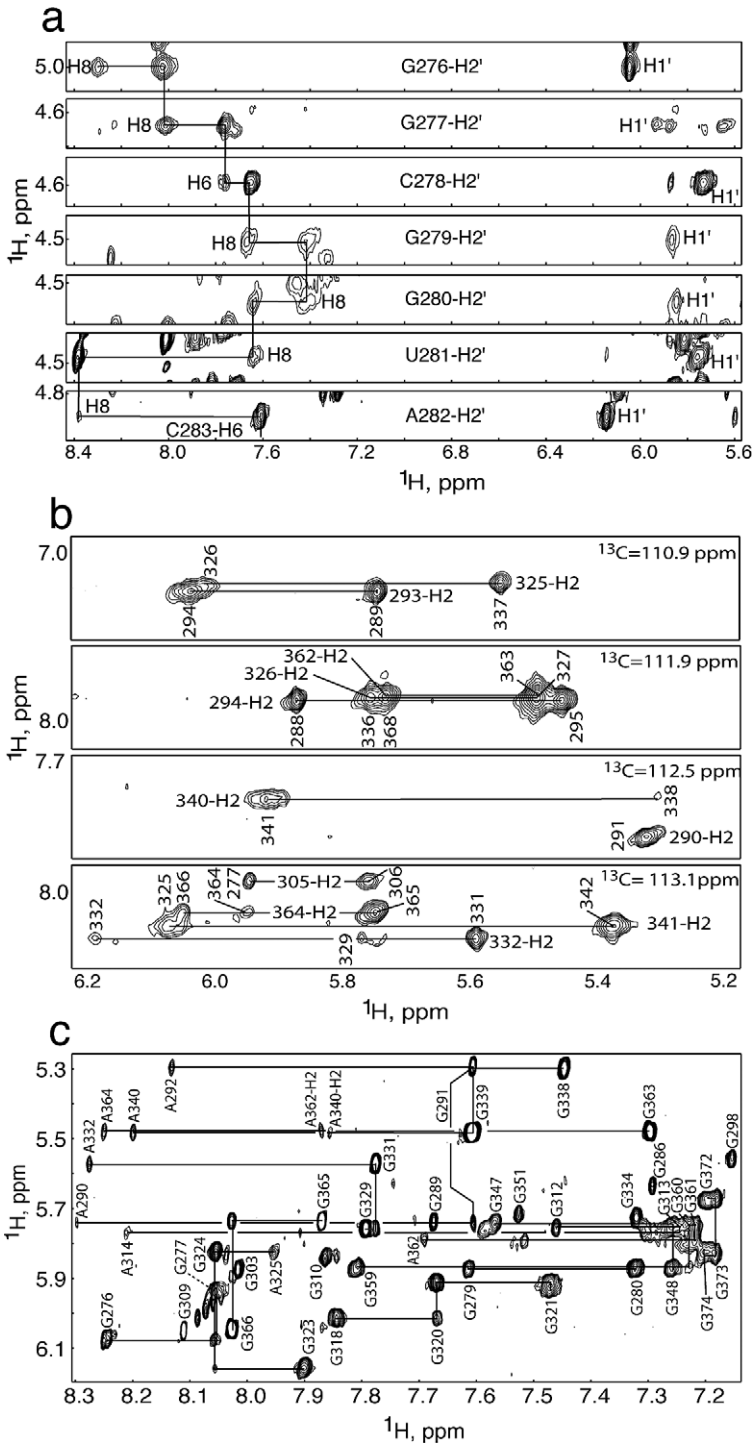

Figure 2. (a) Strips from 3D $^{1}H-^{13}C$ HMQC-NOESY spectra obtained for four different mΨ samples (A^{CN}-mΨ, U^{CN}-mΨ, G^{CN}-mΨ, C^{CN}-mΨ) showing sequential NOE connectivities involving the H2′ protons of G276–A282. (b) Strips from the 3D $^{1}H-^{13}C$ HMQC-NOESY spectrum of A^{CN}-mΨ showing sequential and long range connectivities associated with adenosine H2 protons. The NOE between A341–H2 and A325–H1′ proton is consistent with A-minor-like packing of A341 into the upper stem of SL-C. (c) Portion of the 2D NOESY spectrum obtained for G-protonated mΨ (G^{H}-mΨ) showing NOE cross-peaks associated with the fully protonated guanosine and partially protonated adenosine residues.

observed for the non-exchangeable and exchangeable imino protons are consistent with the secondary structure shown in Figure 1.[32] The imino proton of G276 was not detected, but the observation of standard sequential G276-to-G277, U304-to-A305, and A305–H2-to-G277–H1′ NOEs are consistent with a G276·A305 base-pair. Standard, sequential NOEs observed for the U281–A282–C283 and G298–C299–U300 stretches, and a cross-strand NOE between the A282–H2 and U300–H1′ protons, indicate that A282 and C299 are both stacked within the stem of SL-B. The A282–H2 and –H8 signals were sensitive to pH, consistent with a $A282^{+}$·C299 mismatched base-pair.[33–36] Residues G289–A292 of SL-B and G329–A332 of SL-C exhibit NOEs and chemical shifts typical of GNRA tetraoops.[37]

In SL-C, A314–H2 exhibits standard A helical NOEs to C315–H1′ and C349–H1′, respectively, and an unusual NOE to C315–H5. Unusual, strong-intensity C315–H1′/H4′-to-C316–H5/6 and very weak-intensity C315–H2′/H3′ to C316–H5/6 NOEs were observed, along with standard NOEs for the G348–C349 step that is

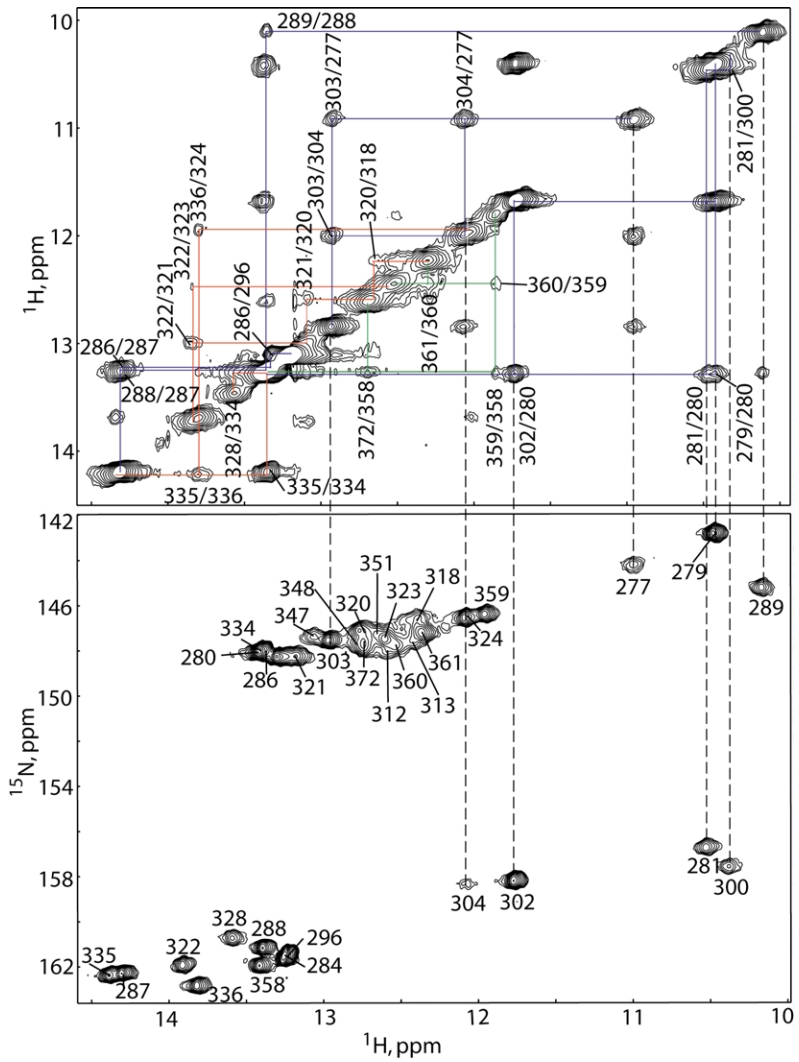

Figure 3. Portions of the 2D ^{1}H–^{15}N HSQC (lower) and ^{1}H–^{1}H NOESY (upper) spectra obtained for $G^{N}U^{N}$-mΨ. Signals of SL-B, SL-C and SL-D are correlated by blue, red and green lines, respectively. NOE signals for G312, G313, G347 and G348 were resolved in the 3D ^{15}N-edited NOESY data (not shown). Overlapping signals in the 2D NOESY spectrum were assigned from 3D ^{15}N-edited NOESY data.

adjacent to the A314–C315 bulge. These data indicate that G313, A314 and C315 are stacked in an A helix-like manner, and that C315 does not stack below C316, but instead is located adjacent to C316. Strong intensity sequential H2′/3′(*i*)-to-H8(*i* + 1) NOEs were observed between G318 and G320 (Figure 2c) and the aromatic protons of U319 did not exhibit NOEs to the neighboring nucleotides. Instead, the U319–H1′ and –H4′ protons exhibited moderate-intensity NOEs to G320–H8. These data indicate that G318 and G320 are stacked, and that U319 is bulged into the major groove of the SL-C stem. Additional unusual NOEs were observed for residues G338–A341, with G339–H8 exhibiting a strong intraresidue H8-to-H1′ NOE (consistent with a *syn* conformation) and only weak NOEs to the H2′ and H3′ protons of G338. The H8 proton of A340 exhibits unusual, moderate-intensity NOEs to the H1′ and H4′ protons of G339, and A340–H2 exhibits an NOE with G338–H1′. Regular A helical sequential NOEs were observed for the subsequent A340–A341–C342 steps. Most significantly, the A341–H2 proton exhibits long range NOEs to the H1′ and H4′ protons of A325, and very weak sequential NOEs were observed for the G323–G324 step (Figure 2c), indicating that the A341 nucleobase packs against the ribose of A325. Cross-strand A-H2 NOEs indicate that A353 and A354 are base-paired with U376 and U375, respectively. Significantly, A353–H2 exhibits NOEs to A354–H1′ and G310–H1′, and U376–H1′,H2′ and H3′ exhibit NOEs to G310–H8. However, no NOEs were observed between stem loop B and the other stem loops. In addition, the sequential inter-residue NOEs for residues that link SL-B and SL-C (U306–G309) were weak, and no imino proton signals were observed for these residues. These data indicate that the lower stems of SL-C and SL-D stack together to form an extended duplex, and that SL-B forms an independent domain that is connected to SL-C *via* a flexible linker (U306–G309).

^{15}N NMR Relaxation Analysis

^{15}N *T*1 and *T*2 relaxation data were obtained for $G^{N}U^{N}$-mΨ to assess the relative mobilites of the stem loops. The *R*1 ($= 1/T1$) and *R*2 ($= 1/T2$) data for SL-B are generally clustered and have

values that differ from the average values observed for SL-C and SL-D (Figure 4). Although it is likely that mΨ tumbles anisotropically in solution, insufficient relaxation data were obtained to allow a quantitative fitting to anisotropic rotational diffusion models. Under the simplified assumption of isotropic tumbling and no chemical exchange, rotational correlation times of 13.7, 21.7 and 19.4 ns are calculated for SL-B, SL-C and SL-D, respectively. If residues adjacent to frayed or non-canonical base-pairs are eliminated from the fitting (which all exhibit higher *R*2 values compared to those of sequestered Watson–Crick base-pairs within a given stem loop), tighter clustering is observed, with average isotropic rotational correlation times of 12.6, 20.0 and 19.4 ns, respectively. Thus, SL-D appears to tumble in solution at a slower rate than SL-B, despite being the smallest of the three stem loops, and SL-C and SL-D tumble at essentially the same rate. These data support the conclusion, based on the NOE data, that SL-C and -D are co-stacked.

Structure Calculations

Except for NOEs associated with adenosine-H2 protons, all sequential NOEs in nucleic acid A helices involve protons in the major groove. The resulting asymmetric distribution of distance restraints in A helical segments, coupled with propagating uncertainty errors that accumulate in linear structures and the influence of non-bonded contacts,[38] can lead to nucleic acid models with abnormally narrow major grooves.[39] This is illustrated in Figure 5a, which shows the SL-B stem loop of the mΨ structures generated using only base-pair hydrogen bonds and $^{1}H-^{1}H$ NOE-derived distance restraints. The addition of conservative dihedral angle restraints to loosely restrain phosphodiester dihedral angles about A helical values (±50°) did not alleviate this problem (Figure 5b). The problem can be corrected by the use of energetic terms that limit the approach of the phosphate groups.[39] We employed an analogous approach, in which cross-helix P–P distances were restrained to 8–14 Å (with a 20% weighted square-well potential), which allows generous sampling about values observed in idealized A helices (~9.5–10.5 Å) and high resolution X-ray crystal structures (~8.5–14 Å).[40–42] These soft P–P restraints (totaling 28, 29, and 16 restraints for SL-B, SL-C and SL-D, respectively) were only employed during the initial stages of the structure calculations, prior to the application or residual dipolar couplings (RDC) restraints.[43,44] As shown in Figure 5c, these restraints can lead to substantially improved convergence and to P–P distances consistent with those typically found in A helices. Significantly, the target functions were unaffected by the use of these restraints, and none of the P–P distances observed in the structures were at or very near the limits of the distance bounds.

RDCs can accurately establish helical parameters in elongated nucleic acid structures.[45–47] Unfortunately, intact mΨ precipitated in the presence of

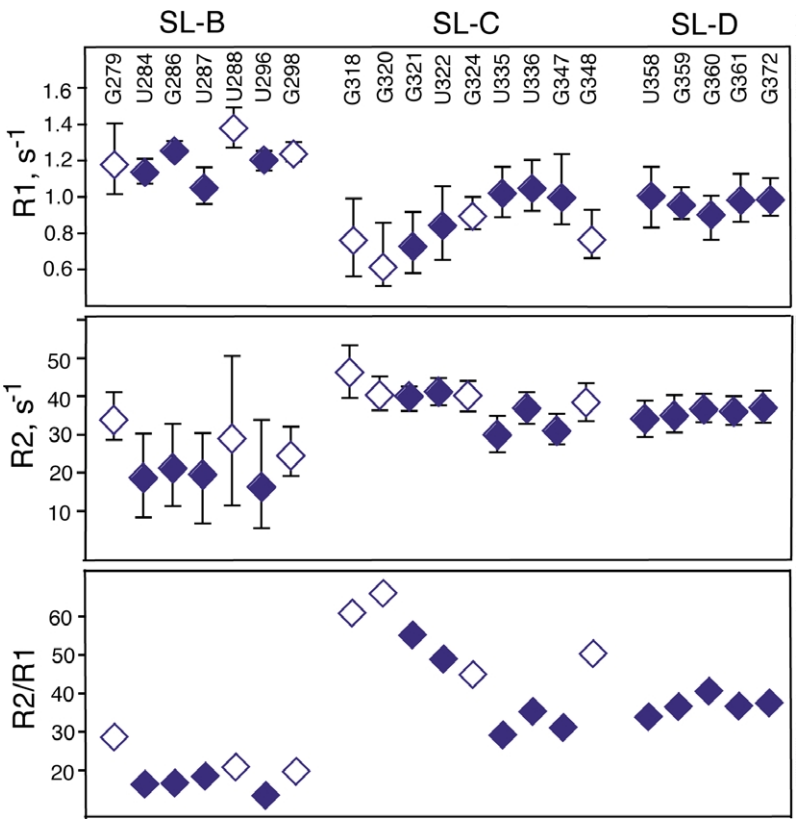

Figure 4. ^{15}N *R*1 (= 1/*T*1), *R*2 and *R*2/*R*1 data obtained for $G^{N}U^{N}$-mΨ. Open diamonds indicate residues adjacent to terminal or non-canonical base-pairs. These data were used to estimate rotational correlation times for SL-B, SL-C and SL-D, which indicate that SL-B tumbles in solution with a greater rate than SL-C and SL-D (see the text).

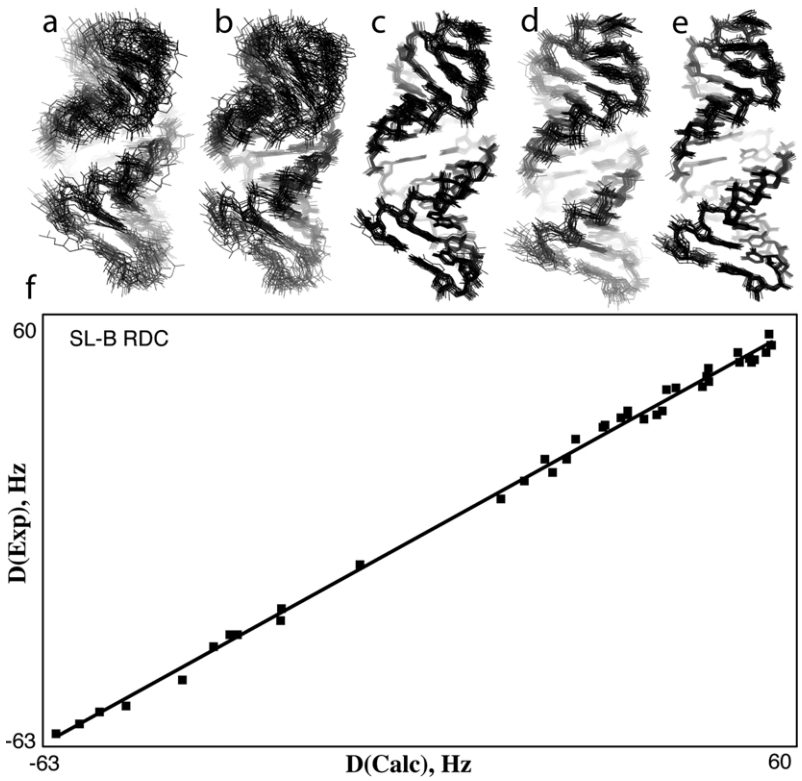

Figure 5. Structures of the SL-B stem loop of mΨ generated using the following restraints: (a) base-pair hydrogen bond (HB) and $^{1}H-^{1}H$ NOE-derived distance restraints; (b) HB, NOE and loose phosphodiester dihedral angle (PDA) restraints (restrained to values of an idealized A helix (±50°)); (c) HB, NOE, PDA, and loose cross-helix interphosphate (PP) restraints (8.0–14.0 Å, with a 20% weighting coefficient); (d) HB, NOE and residual dipolar coupling (RDC) restraints: and (e) HB, NOE, PDA and RDC restraints. (f) Plot of observed *versus* calculated RDCs (with MODULE[94]) for the structure in (e) with lowest target function.

the Pf1 phage alignment media,[45] and in minimally stretched 5% polyacrylamide gels,[48,49] severe broadening of the upfield multiplet components of $^{1}H-^{15}N$ IPAP spectra[50] precluded measurement of $^{1}H-^{15}N$ coupling constants. However, high-quality $^{1}H-^{13}C$ RDC data were obtained for three short oligoribonucleotides corresponding to the individual stem loops SL-B, SL-C and SL-D, and these were used as restraints to individually refine the three stem loops during the final stages of the structure calculations. This approach was previously employed in studies of a 77-nucleotide RNA,[26] and is valid because, except for the terminal nucleotides, the chemical shifts and NOE cross-peak patterns of the isolated stem loops exactly matched the data obtained for mΨ. As shown in Figure 5d, good convergence and A helical parameters can be obtained when only NOE and RDC restraints (40 aromatic C–H and ribose C1′–H1′ RDCs for SL-B) are employed, although better convergence is obtained when loose torsion angle restraints are employed during the structure calculations (Figure 5e). Back-calculated RDCs were in good agreement with the experimental values, and representative data for SL-B are shown in Figure 5f. No X-ray structural data are available for any of the stem loops of mΨ, which would have allowed a further assessment of the quality of the structures. However, in prevous NMR studies of an HIV-1 stem loop, helical parameters such as base-pair twist and rise in structures generated with H-bond, dihedral angle, NOE and RDC restraints were in very good agreement with values observed in X-ray structures.[51]

Structures of mΨ were determined in three steps. Initial structures were generated using the NOE-derived distance, base-pair H-bond, phosphodiester torsion angle, and soft inter-phosphate restraints. Structures were then minimized with orientation restraints applied individually to the three stem loops, using RDC data (primarily aromatic C–H and ribose C1′–H1′ dipolar couplings) collected for the isolated stem loop RNAs. P–P restraints were not employed during this and subsequent refinement steps. In a majority of structures (>85%) generated during this step, the following atom pairs were in sufficiently close proximity to suggest their participation in non-canonical hydrogen bonding: G338–H22 to A340–N7; A341–H61 to G324–O2′; U319–N3H to A341–O2P. Restraints added for these hydrogen bonds during the final stage of refinement did not affect the overall target functions of the final structures. A total of 944 NOE-derived interproton distance restraints were employed for the final structure calculations. Statistical information regarding restraint violations and structure convergence is provided in Table 1.

Description of the Structure and Comparison with other RNA Structures

The individual stem loops are structurally well defined by the NMR data, with best-fit superpositions of all heavy atoms of SL-B, SL-C and SL-D affording pairwise RMS deviations of 0.67(±0.17), 0.69(±0.18), and 2.09(±0.43)

Table 1. Statistics for 20 calculated mΨ structures

	mΨ	SL-B	SL-C	SL-D
NMR-derived restraints				
Interproton distance restraints	944	314	424	202
Intraresidue	464	150	211	103
Sequential ($\|i-j\|=1$)	340	118	156	66
Long range ($\|i-j\|>1$)	140	46	57	33
Hydrogen bond restraints	448	136	200	112
Torsion angle restraints	358	127	142	89
Dipolar coupling restraints	167	40	90	33
Average restraints/refined residue	29.4	30.5	27.0	23.3
Target function (Å^2)				
Mean ± SD	1.41 ± 0.16[a]	0.45 ± 0.03	2.00 ± 0.07	0.23±[b]
Minimum	1.09[a]	0.38	1.83	0.22
Maximum	1.77[a]	0.49	2.12	0.23
Restraint violations				
Av. sum of upper distance viol. (Å)	4.0 ± 0.8[a]	1.7 ± 0.2	7.5 ± 0.2	0.5±[b]
Av. max. upper distance viol. (Å)	0.27 ± 0.05[a]	0.17 ± 0.04	0.34 ± 0.04	0.04±[c]
Av. sum of VDW viol. (Å)	10.2 ± 0.5[a]	3.6 ± 0.1	8.8 ± 0.2	2.3±[b]
Av. max. VDW viol. (Å)	0.30 ± 0.07[a]	0.16 ± 0.02	0.24 ±[c]	0.14 ± 0.01
Av. sum of torsion angle viol. (deg.)	2.2 ± 0.9[a]	1.1 ± 0.4	8.5 ± 0.9	<0.1
Av max. torsion angle viol. (deg.)	0.38 ± 0.28[a]	0.78 ± 0.25	8.1 ± 0.5	0.03 ± 0.07
Av. sum of RDC violations (Hz)		0.1±[b]	3.8 ± 0.2	0.5 ± 0.1
Av. max. RDC violation (Hz)		0.01±[c]	0.25 ± 0.05	0.09 ± 0.01
Structure convergence (Å)[d]				
All heavy atoms	15.4 ± 2.7[a]	0.67 ± 0.17	0.69 ± 0.18	2.09 ± 0.43
Heavy atoms of stem segments		0.38 ± 0.10[e]	0.57 ± 0.17[f]	0.46 ± 0.11[g]
Heavy atoms of SL-C + SL-D[h]			With RDC:	0.88 ± 0.24
			Without RDC:	7.9 ± 3.4
Helical parameters				
Rise/base-pair (helical segments)		2.7 ± 0.6[i]	3.1 ± 1.2[j]	2.8 ± 0.8[k]
Twist/base-pair (helical segments)		33.2 ± 7.1[i]	31.8 ± 7.1[j]	31.8 ± 2.6[k]
Average bend, SL-C AC step			28.1 ± 6.3	
Average bend, SL-C GGAA step			53.5 ± 2.1	

[a] Target functions, violations and average pariwise RMS deviations for 20 structures generated without RDC restraints.
[b] < 0.1.
[c] < 0.01.
[d] Reported as average pairwise RMS deviations for the 20 refined structures.
[e] Heavy atoms of residues 278–287, 294–302.
[f] Heavy atoms of residues 310–314, 316–329, 331–352.
[g] Heavy atoms of residues 354–362, 367–375.
[h] Heavy atoms of residues 310–362, 367–376.
[i] Global fit for residues 277–288, 293–304.
[j] Global fit for residues 310–313, 316–318, 320–323, 324–328, 333–337, 342–345, 346–348, 349–352.
[k] Global fit for residues 353–362, 367–376.

(0.46(±0.11) excluding the disordered tetraloop) Å, respectively (Table 1 and Figure 6a–c). Somewhat poorer convergence was obtained upon superposition of all residues of SL-C and the stem residues of SL-D (0.88(±0.24) Å) due to the lack of RDC data for the intact mΨ construct. Residues of SL-B and the linker segment (U306–G309) were not experimentally restrained relative to SL-C and SL-D, and these segments therefore appear disordered (Figure 6d). Thus, mΨ consists of an elongated segment formed by the end-to-end stacking of SL-C and SL-D, which is connected to a shorter stem loop (SL-B) by a flexible four-nucleotide linker (U306–G309) (Figure 6e).

Stem loop SL-B contains five mismatched base-pairs, all of which are located in the lower half of the stem (G276·A305, G277·U304, G279·U302, U281·U300, A282$^+$·C299), and residues G289–A292 form a GNRA-type tetraloop, as designed. These elements adopt conformations similar to those observed in previous high resolution X-ray and NMR structures (Amarasinghe *et al*. unpublished results).[37,52–57] Stem loop SL-C does not contain mismatched base-pairs, but does contain several unopposed bases that contribute to non-canonical structural elements. A314 and C315 are stacked above the G313–C349 base-pair, with A314–H2 in close proximity to the C315–H1′, C316–H5 and C349–H1′ protons (Figure 7). The C315 and C316 nucleobases are juxtaposed in the major groove, with C316–H42 poised to form either a direct or water-mediated hydrogen bond to C315–O2. The structure resembles a C315·C316–G348 base-triple platform observed previously in the 30 S ribosomal

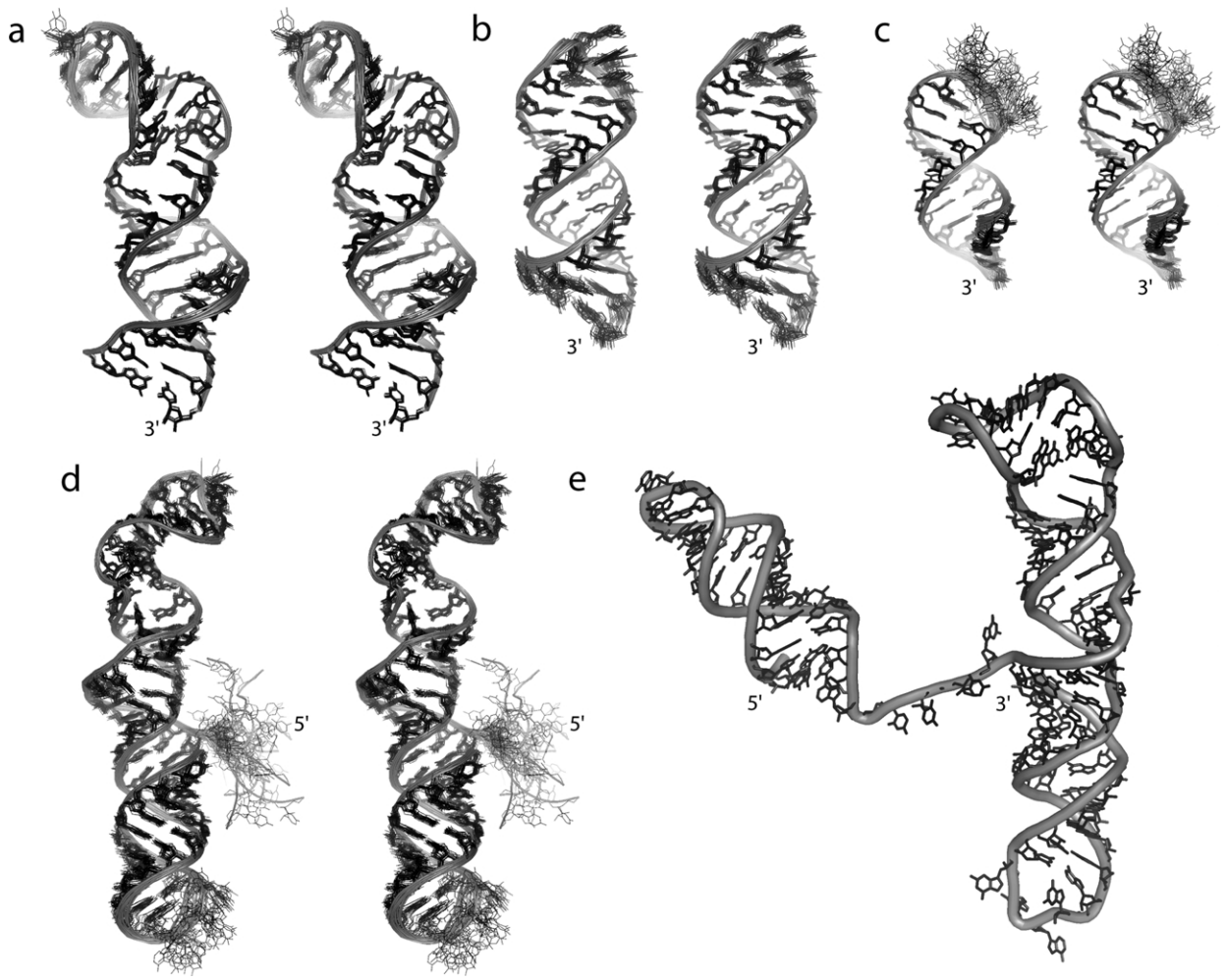

Figure 6. (a)–(d) Stereoviews of the best-fit superpositions of stem loops SL-C, SL-B, SL-D, and SL-CD, respectively, from the ensemble of 20 refined mΨ structures. Disordered nucleotides U306–G309 are also shown in (d). (e) Cartoon representation of a representative mΨ structure.

subunit, although the extra-helical C in the 30 S subunit was contributed *via* a long-range interaction.[58] The arrangement of the stacked, unopposed A314 nucleotide and the C315·C316–G348 base triple results in a 28(±6)° bend in the lower stem of SL-C (Figure 7).

Perhaps the most interesting structural element is that associated with the unpaired residues U319 and G338–A341. G338 adopts a normal, A helical conformation and stacks below the G324–C337 base-pair of the upper stem, and G339 exists in a *syn* conformation and does not stack against other nucleobases. Instead, the ribose of G339 packs against the nucleobase of A340, and both A340 and A341 stack in an A helix-like manner on top of the G323–C342 base-pair (Figure 7). In addition, the nucleobase of A340 packs against the ribose of G338, and the nucleobase of A341 packs against the ribose of A325 (Figure 7). In this arrangement, the A341–H62 proton is poised to form a hydrogen bond with the O2′ oxygen of G324, and the A340–NH2 is in the vicinity of A345–N3 and –O2′, and could make either direct or water-mediated hydrogen bonds to either or both of these groups. The structure is further stabilized by a hydrogen bond from the H3 proton of U319 to the phosphate group of A341 (Figure 7). U319 is bulged into the major groove of the lower stem, allowing the adjacent G318–C346 and G320–C345 base-pairs to co-stack. U nucleotide bulges have been observed in several RNA structures, although in most cases the U is located on the minor groove side of the helix and does not make long-range contacts.[59,60] However, in the crystal structure of the 23 S ribosomal subunit,[61] one bulged U (U835) makes a long range hydrogen bond with a phosphodiester similar to that exhibited by U319 of mΨ. The packing arrangement of these residues in mΨ results in a small helical displacement at the G318–U319–G320 step, and to a 54(±2) degree bend in the helix at the G323–G324 step (Figure 7d and e).

The GGAA structure is in some ways similar to folding elements observed recently in the 23 S subunit of the ribosome. First, the interactions of the A340 and A341 nucleobases with the minor groove of the upper stem are reminiscent of "A-minor"

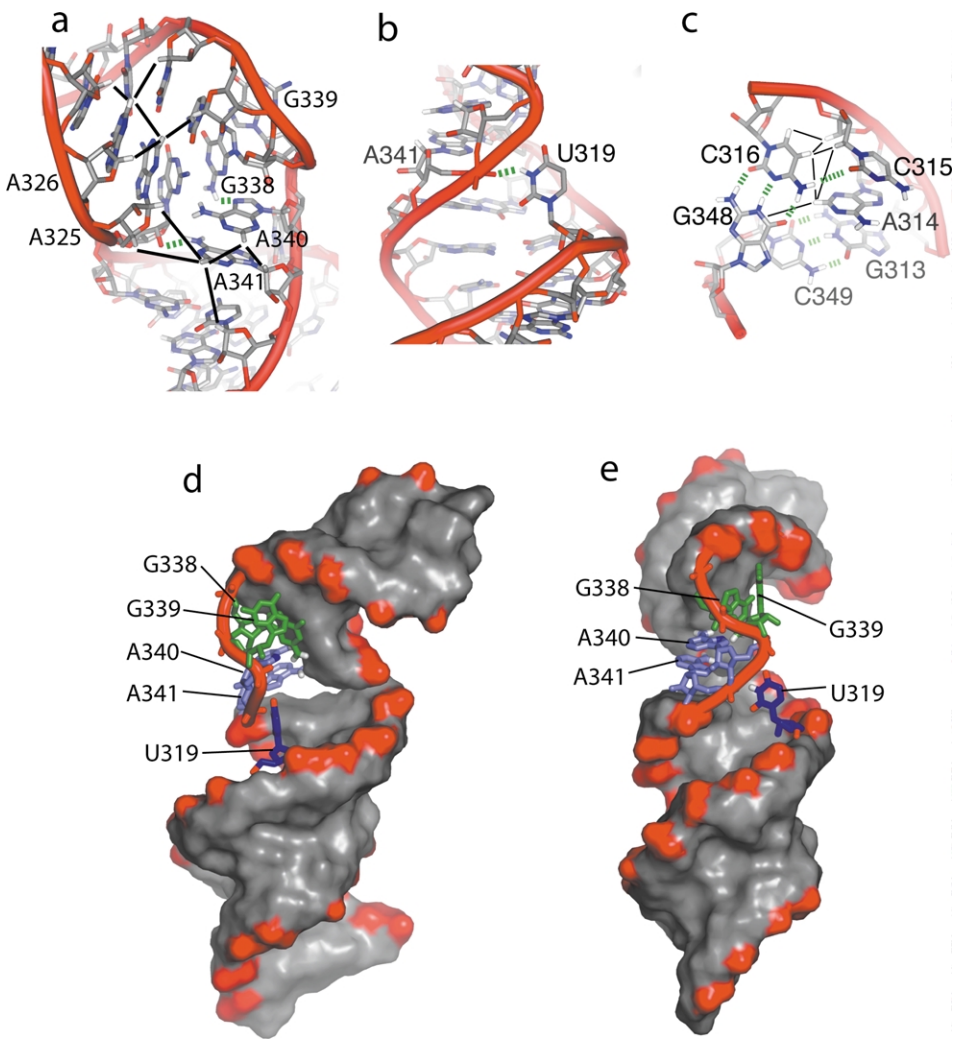

Figure 7. Stick representations showing the NOE cross-peaks (black lines) and hydrogen bonds (green broken lines) associated with the A-minor packing (a, b), and the C·CG base triple platform (c) of stem loop SL-C. (d) and (e) Surface representations showing the 54 degree bend between the upper and lower stems of SL-C induced by the A-minor K-turn motif.

motif packing.[62] Of the four major classes of the A-minor motif identified thus far, the participating adenosine basess generally interact with a G–C base-pair, and in all cases, the 2′-hydroxyl of A is involved in hydrogen bonding. In mΨ, the 2′-hydroxyls of both A340 and A341 are well removed from the minor groove of the upper stem and clearly do not participate in a classical A-minor-like hydrogen bond. Such packing, if it occurred, would have placed the A340 and/or A341 H1′ protons in close proximity to ribose protons within the minor groove, and no such NOEs were detected. Instead, the H2 proton of A341 is in close proximity to ribose protons of the minor groove, consistent with the structure and hydrogen-bonding pattern shown in Figure 7a. The structure also somewhat resembles those of the "kink-turn" (or K-turn) family, in which two, three and four nucleotide bulges adopt local structures that induce a kink in the phosphodiester backbone and a sharp turn in the RNA helix.[61] Although the sequence and packing arrangement of the GGAA bulge of mΨ is unlike those of previously characterized K-turns, its overall effect is to induce a similar, dramatic kink between adjoining helical segments. In view of the similarities with both A-minor and K-turn structural elements, we refer to the GGAA/U-bulge structure as an "A-minor K-turn" motif.

Stem loop SL-D does not contain any bulges or non-Watson–Crick base-pairs and adopts a classical A helical structure, and the GAGG tetraloop does not give rise to a self-consistent pattern of NOEs and thus appears to be flexible (Figure 6c).

Biological Relevance

Early phylogenetic studies of C-type retroviruses identified SL-B, SL-C and SL-D as conserved structural elements with both strict co-evolutionary and

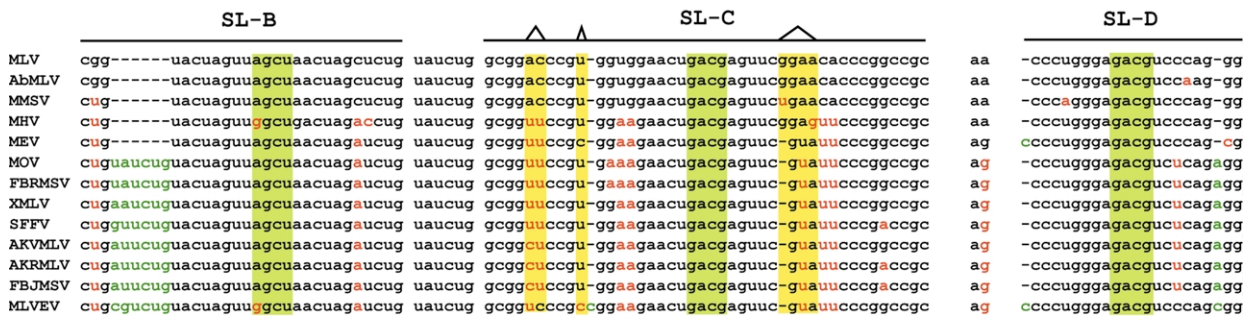

Figure 8. Sequence alignment of several murine retrovirus genomes guided by the secondary structure alignment by MFOLD. Sequences (and NCBI accession numbers) that are aligned include: MLV, Moloney murine leukemia virus (AF033811); AbMLV, Abelson murine leukemia virus (NC001499); MMSV, Moloney murine sarcoma virus (NC001502); MHV, *Mus hortulanus* virus (M26526); MEV, Mouse endogenous virus [10]; MOV, Murine osteosarcoma virus (NC001506); FBRMSV, FBR murine sarcoma virus (K02712); XMLV, xenotropic murine leukemia virus (K02730); SFFV, Spleen focus forming virus (NC001500); AKVMLV, AKV murine leukemia virus (J01998); AKRMLV, AKR murine leukemia virus;[14] FBJMSV, FBJ murine sarcoma virus (V01184); MLEV, Murine leukemia like endogenous virus (AF041383). Nucleotides in green represent insertions and nucleotides in red represent non-identical sequences when compared to the MLV genome. Green boxes denote the conserved tetraloops of SL-B, SL-C and SL-D and yellow boxes denote the bulges seen in SL-C.

conservative nucleotide variations.[8,10,14] We have obtained consistent results using more recent sequence data (Figure 8). Nucleotides 280–298 of SL-B are strictly conserved among the murine retroviral genomes, with a six-nucleotide insertion often found between residues 279 and 280. The inserted nucleotides could form a stretch of normal and mismatched base-pairs with the 3′-nucleotides of the SL-B lower stem and linker, analogous to the base-pairings observed for mΨ. There is now considerable evidence that the SL-B functions, at least in part, by stabilizing the dimeric form of the RNA genome,[7,16–20] and it has been suggested that dimerization and packaging events are closely coupled.[7,63–66] Consistent with this hypothesis, our previous studies demonstrated that NC binding to the native SL-BCD RNA promotes dimerization *via* conversion of SL-B from a hairpin to an intermolecular duplex.[23] The presence of conservatively substituted non-canonical base-pairs in the lower stem of SL-B is probably important for destabilization of the stem, enabling NC-mediated conversion from the stem loop to the duplex species.

The bulges of SL-C are also conserved among the C-type retroviruses (Figure 8), although the nature and number of residues that comprise the bugles can vary.[14] Substitution of C315 by U (which occurs in most of the murine sequences) would likely lead to a U315·C316–G348 base-triple, analogous to the C315·C316–G348 triple observed in mΨ since both U and C can hydrogen bond *via* their exocyclic O2 oxygen atoms. In this regard, U·C-G base-triples have been observed with higher frequency than C·C-G triples.[67,68] U319 is strictly conserved among the murine retroviruses, but is sometimes substituted by C in other C-type retroviruses. In mΨ, U319 stabilizes the A-minor K-Turn by forming a hydrogen bond with the phosphodiester of A341, and the exocyclic NH_2 group of the substituted cytidine bases could function in a similar manner.

Although all C-type retroviruses contain a bulge at position G338–A341, most murine retroviruses contain a GUA bulge (Figure 8). Three nucleotide bugles are commonly found in K-turns, and it is likely that these conserved bulges form kinked helices similar to those observed both here for the A-minor K-turn motif and for other classical K-turn structures.[61] K-turn elements observed in previous X-ray crystal structures often participate in both long range RNA–RNA and protein–RNA interactions.[61] Although the A-minor K-turn of mΨ might participate directly in NC binding, gel retardation and isothermal titration calorimetry assays indicate that elements in SL-B and/or the linker that connects SL-B and SL-C would also be required.[23] Protein–RNA interactions are often mediated by RNA segments that are flexible in the absence of protein but become ordered upon complex formation,[69] and in mΨ, the only residues that appear to be flexible (other than the tetraloop of SL-D, which does not appear to be important for NC binding[23]) are those of the linker. Thus, the linker and/or the mismatched base-pairs of SL-B could function alone or in concert with the A-minor K-turn to promote high-affinity NC binding.

We also cannot rule out the possibility that stem loops C and D might unstack upon binding NC, or that they may not be stacked in the context of the native MLV sequence. Non-native uridines at positions 375 and 375 were included to accomodate an Hpa 1 restriction site used to linearize the DNA template, and base pairing between these residues and A354 and A353 probably promotes end-to-end stacking of the stem loops. Regardless, NC binding does not appear to be influenced by the presence of these non-native bases.[23]

Summary

The strategy developed here to assign the NMR

signals of mΨ required the collection of 2D, 3D and 4D NMR datasets for eight different isotopically labeled RNA samples. Although the procedure is both labor and instrumentation-intensive, it is otherwise robust, and should be generally applicable to RNAs as large as 100 nucleotides. The most significant limitation to high quality structure determination of larger RNAs appears not to be resonance assignment or measurement of $^{1}H-^{1}H$ NOEs, but instead the inability to measure large numbers of residual dipolar couplings. Although promising new methods for obtaining RDCs for RNA have been developed,[46,70–74] we were unable to apply them to intact mΨ due to sample precipitation in the presence of phage, and severe line broadening in the presence of polyacrylamide gels. It thus remains to be determined whether significant numbers of dipolar couplings can be measured for RNAs of 100 nucleotides or more, in which signal degeneracy and NMR relaxation become severe. Nevertheless, our findings, and those reported recently by Puglisi and coworkers,[26] demonstrate the feasibility of applying RDCs obtained for isolated RNA fragments to the structure determination and refinement of much larger constructs. The combination of a significant number of RDCs obtainable for smaller fragments with a relatively small number of RDCs obtainable for the larger construct (for example, $^{1}H-^{15}N$ RDCs)[70] should be sufficient to accurately define both local and global structure.

The A-minor K-turn and C·CG base-triple of SL-C appear to be conserved among the murine packaging signals, and similarities with protein binding elements in the 23 S ribosomal subunit suggest that they could play roles in NC-mediated RNA binding during virus assembly. The conserved, flexible linker that connects SL-B may also be important for genome packaging. Information regarding the precise nature of the interactions that contribute to NC binding should now be accessible using the NMR methods described here.

Methods

Preparation of RNA samples

The native tetraloops were mutated as shown in Figure 1 to inhibit dimerization and higher-order aggregation, but not to interfere with NC binding, as described.[23] mΨ also contains two non-native U nucleotides at the 3′ end that resulted from enzyme processing of the DNA template. RNAs were synthesized by *in vitro* transcription[75] as described.[23] Nucleotide-specific labeled [^{15}N]RNA ($G^{N}U^{N}$-mΨ) and [^{15}N]- and [^{13}C]RNA (A^{CN}-mΨ, U^{CN}-mΨ, G^{CN}-mΨ, C^{CN}-mΨ) were prepared using labeled nucleotide triphosphates (NTPs, VLI Research Inc, Malvern, PA). Specifically protonated samples (A^{H}-mΨ, U^{H}-mΨ, G^{H}-mΨ, C^{H}-mΨ) were prepared using a combination of three deuterated NTPs (~90% perdeuteration) and the respective protonated NTP. Short, unlabeled and ^{15}N,^{13}C-labeled oligoribonucleotides corresponding to the three mutated stem loops of mΨ (for RDC measurements) were prepared from synthetic templates that encoded the following sequences: G276–G303 (SL-B); G309–C352 (SL-C); and C357–G372, with two additional guanosine bases appended to the 5′ end and two additional cytidine bases appended to the 3′ end (SL-D). The additional guanosine bases were needed to obtain reasonable transcription yields. All RNA samples were purified by denaturing gel electrophoresis.

NMR spectroscopy

NMR data were collected with Bruker AVANCE (800 MHz, ^{1}H) and DMX (600 MHz, ^{1}H) spectrometers, processed with NMRPipe/NMRDraw,[76] and analyzed with NMRView.[77] Non-exchangeable ^{1}H and ^{13}C NMR assignments were obtained from 2D NOESY ($\tau_m = 120$ ms),[78,79] 2D ROESY,[80] 2D $^{1}H-^{13}C$ HMQC,[81,82] 3D $^{1}H-^{13}C$ HMQC-NOESY[83] ($\tau_m = 120$ ms) and 4D $^{1}H-^{13}C$ HMQC-NOESY-HMQC[84] ($\tau_m = 120$ ms) and 2D natural abundance $^{1}H-^{13}C$ HMQC[82] data collected at 35 °C. The exchangeable protons were assigned from 2D Watergate-NOESY[85] ($\tau_m = 120$ ms) and $^{1}H-^{15}N$ HSQC[30] spectra collected at 5, 10, 20 and 25 °C and 3D $H-^{15}N$ HSQC-NOESY[86] data collected at 5 °C and 25 °C. ^{15}N relaxation data were obtained[87] for $G^{N}U^{N}$-mΨ samples at 25 °C with the following delays: T_1 ($1/R_2$), 10.051, 130.663, 512.601, 643.264, 1286.528, 1547.854, 2060.455 and 2573.056 ms; $T2$ ($1/R2$), 0, 14.692, 29.38, 44.07, 58.76, 73.45, 88.14 and 102.8 ms. Rotational correlation times were estimated from $R2/R1$ ($= T1/T2$) ratios for spins with resolved imino signals using the program Quadric (A.G. Palmer, Columbia University), which was modified to incorporate imino nitrogen chemical shift anisotropies of −130 ppm and −100 ppm for guanine-N_1 and uridine-N_3, respectively, and imino proton bond lengths (1.01 A).[88] $^{1}J_{CH}$ coupling constants were measured in the presence and absence of Pf1 phage using a constant-time HSQC experiment[89] modified to collect both the in-phase (IP) and anti-phase (AP) doublets in the indirect dimension in interleaved mode.[50]

Structure calculations

Structures were calculated and refined with CYANA[90] using the AMBER[91] residue library. Upper-limit distance restraints of 2.7 Å, 3.3 Å and 5.0 Å were employed for direct NOE cross-peaks of strong, medium and weak intensities, respectively, for all cross-peaks except those associated with the intra-residue H8/6 to H2′ and -H3′ interactions. In an idealized A helix[92] and in high resolution X-ray crystal structures of short oligoribonucleotides,[40–42] these distances range from 3.83 Å to 4.13 Å and 2.33 Å to 3.16 Å, respectively, and upper distance limits of 4.2 Å and 3.2 Å were therefore employed for intra-residue H8/6 to H3′ (strong) and H2′ (medium) NOEs, respectively. This is functionally equivalent to relaxing the upper distance limits of sequential NH-to-NH interproton restraints in protein structure calculations from 2.7 Å to 2.9 Å. Cross-helix P–P distance restraints of 8.0–14.0 Å (with 20% weighting coefficient; see above) were only employed during the first stage of the structure calculations, and were not employed during subsequent refinement stages when RDC restraints were employed. Torsion angle restraints for A helical stem residues were centered around

published A-form RNA values ($\alpha = -68°$, $\beta = 178°$, $\gamma = 54°$, $\varepsilon = -153°$, $\zeta = -71°$)[93] with allowed deviations of $\pm 50°$. Four restraints per hydrogen bond were employed to enforce approximately linear NH–N and NH–O bond distances of 1.85($\pm$0.05) Å. Axial and rhombic components of the orientation tensors for the isolated stem loops were determined from grid search calculations: SL-B, −70.0 Hz and 4.3%; SL-C, −28.0 Hz and 20.4%; SL-D, −26.0 Hz and 7.7%. Plots of experimental *versus* calculated dipolar couplings were obtained with Module,[94] structural statistics and molecular images were generated with PyMOL,[95] and helical parameters were calculated with CURVES.[96,97]

Atomic coordinates

Coordinates and restraints for structure calculations and NMR signal assignments have been deposited in the RCSB (PDB ID code 1S9S), and BMRB (accession number 6094) data banks, respectively.

Acknowledgements

Supported by NIH grant GM42561. D.H. was supported by a MARC U*STAR grant (NIH GM08663). Technical support from Yu Chen (HHMI, UMBC) is greatly appreciated.

References

1. Berkowitz, R., Fisher, J. & Goff, S. P. (1996). RNA packaging. *Curr. Top. Microbiol. Immun.* **214**, 177–218.
2. Rein, A. (1994). Retroviral RNA packaging: a review. *Arch. Virol.* **9**, 513–522.
3. Swanstrom, R. & Wills, J. W. (1997). Synthesis, assembly and processing of viral proteins. In *Retroviruses* (Coffin, J. M., Hughes, S. H. & Varmus, H. E., eds), pp. 263–334, Cold Spring Harbor Press, Plainview, NY.
4. Berkhout, B. (1996). *Structure and function of the human immunodeficiency virus leader RNA*. Prog. Nucl. Acid Res. and Mol. Biol., vol. 54, Academic Press, New York pp. 1–34.
5. Jewel, N. A. & Mansky, L. M. (2000). In the beginning: genome recognition, RNA encapsidation and the initiation of complex retrovirus assembly. *J. Gen. Virol.* **81**, 1889–1899.
6. Mann, R. & Baltimore, D. (1985). Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNA. *J. Virol.* **54**, 401–407.
7. Prats, A.-C., Roy, C., Wang, P., Erard, M., Housset, V., Gabus, C. *et al.* (1990). *Cis* elements and *trans*-acting factors involved in dimer formation of murine leukemia virus RNA. *J. Virol.* **64**, 774–783.
8. Alford, R. L., Honda, S., Lawrence, C. B. & Belmont, J. W. (1991). RNA secondary structure analysis of the packing signal for moloney murine leukemia virus. *Virology*, **183**, 611–619.
9. Fisher, J. & Goff, S. P. (1998). Mutational analysis of stem-loops in the RNA packaging signal of the moloney murine leukemia virus. *Virology*, **244**, 133–145.
10. Konings, D. A. M., Nash, M. A., Maizel, J. V. & Arlinghaus, R. B. (1992). Novel GACG-hairpin pair motif in the 5′ untranslated region of type C retrovirus related to murine leukemia virus. *J. Virol.* **66**, 632–640.
11. Mougel, M., Tounekti, N., Darlix, J.-L., Paoletti, J., Ehresmann, B. & Ehresmann, C. (1993). Conformational analysis of the 5′ leader and the gag initiation site of Mo-MuLV RNA and allosteric transitions induced by dimerization. *Nucl. Acids Res.* **21**, 4677–4684.
12. Mougel, M., Zhang, Y. & Barklis, E. (1996). *cis*-Active structural motifs involved in specific encapsidation of moloney murine leukemia virus RNA. *J. Virol.* **70**, 5043–5050.
13. Mougel, M. & Barklis, E. (1997). A role for two hairpin structures as a core RNA encapsidation signal in murine leukemia virus virions. *J. Virol.* **71**, 8061–8065.
14. Tounekti, N., Mougel, M., Roy, C., Marquet, R., Darlix, J.-L., Paoletti, J. *et al.* (1992). Effect of dimerization on the conformation of the encapsidation psi domain of moloney murine leukemia virus RNA. *J. Mol. Biol.* **223**, 205–220.
15. Yang, S. & Temin, H. M. (1994). A double hairpin structure is necessary for the efficient encapsidation of spleen necrosis virus retroviral RNA. *EMBO J.* **13**, 713–726.
16. Roy, C., Tounekti, N., Mougel, M., Darlix, J.-L., Paoletti, C., Ehresmann, C. *et al.* (1990). An analytical study of the dimerization of *in vitro* generated RNA of moloney murine leukemia virus MoMuLV. *Nucl. Acids Res.* **18**, 7287–7292.
17. Girard, P.-M., Bonnet-Mathoniere, B., Mauriaux, D. & Paoletti, J. (1995). A short autocomplementary sequence in the 5′ leader region is responsible for dimerization of MoMuLV genomic RNA. *Biochemistry*, **34**, 9785–9794.
18. De Tapia, M., Metzler, V., Mougel, M., Ehresmann, B. & Ehresmann, C. (1998). Dimerization of MoMuLV genomic RNA: redefinition of the role of the palindromic stem-loop H1 (278–303) and new roles for stem-loops H2 (310–352) and H3 (355–374). *Biochemistry*, **37**, 6077–6085.
19. Ly, H., Nierlich, D. P., Olsen, J. C. & Kaplan, A. H. (1999). Moloney murine sarcoma virus genomic RNAs dimerize *via* a two-step process: a concentration-dependent kissing-loop interaction is driven by initial contact between consecutive guanines. *J. Virol.* **73**, 7255–7261.
20. Oroudjev, E. M., Kang, P. C. E. & Kohlstaedt, L. A. (1999). An additional dimer linkage structure in moloney murine leukemia virus RNA. *J. Mol. Biol.* **291**, 603–613.
21. Kim, C.-H. & Tinoco, I., Jr (2000). A retroviral RNA kissing cokmplex containing only two G-C basepairs. *Proc. Natl Acad. Sci. USA*, **97**, 9396–9401.
22. Yu, S. S., Kim, J.-M. & Kim, S. (2000). The 17 nucleotides downstream from the env gene stop codon are important for murine leukemia virus packaging. *J. Virol.* **74**, 8775–8780.
23. D'Souza, V., Melamed, J., Habib, D., Pullen, K., Wallace, K. & Summers, M. F. (2001). Identification of a high-affinity nucleocapsid protein binding site within the moloney murine leukemia virus Y-RNA packaging signal. Implications for genome recognition. *J. Mol. Biol.* **314**, 217–232.
24. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein

bound to the SL3 Ψ-RNA recognition element. *Science*, **279**, 384–388.
25. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K., Wu, Z.-R. & Summers, M. F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. *J. Mol. Biol.* **301**, 491–511.
26. Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. (2003). Structure of HCV IRES domain II determined by NMR. *Nature Struct. Biol.* **10**, 1033–1038.
27. Kolk, M. H., van der Graaf, M., Wijmenga, S. S., Pleij, C. W. A., Heus, H. A. & Hilbers, C. W. (1998). NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. *Science*, **280**, 434–438.
28. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D. & Varani, G. (2000). The NMR structure of the 38 kDa U1A protein–PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nature Struct. Biol.* **7**, 329–335.
29. Dieckmann, T. & Feigon, J. (1997). Assignment methodology for larger RNA oligonucleotides: application to an ATP-binding RNA aptamer. *J. Biomol. NMR*, **9**, 259–272.
30. Saudek, V., Piotto, M. & Sklenar, V. (1994) WATERGATE: *WATER* suppression by *Gr*Adient-*T*ailored *Ex*citation. In *Bruker Report*, vol. 140/94, pp. 6–9.
31. Grzesiek, S. & Bax, A. (1993). The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurement. *J. Am. Chem. Soc.* **115**, 12593–12593.
32. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
33. Legault, P. & Pardi, A. (1997). Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc.* **119**, 6621–6628.
34. Ravindranathan, S., Butcher, S. E. & Feigon, J. (2000). Adenine protonation in domain B of the hairpin ribozyme. *Biochemistry*, **39**, 16026–16032.
35. Cai, Z. & Tinoco, I., Jr (1996). Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, **35**, 6026–6036.
36. Huppler, A., Nikstad, L. J., Allmann, A. M., Brow, D. A. & Butcher, S. E. (2002). Metal binding and base ionization in the U6 RNA intramolecular stem-loop structure. *Nature Struct. Biol.* **9**, 431–435.
37. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. (1996). A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264**, 968–980.
38. Clore, G. M. & Kuszewski, J. (2003). Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipplar coupling cross-validation. *J. Am. Chem. Soc.*, **125**, 1518–1525.
39. Rife, J. P., Stallings, S. C., Correll, C. C., Dallas, A., Steitz, T. A. & Moore, P. B. (1999). Comparison of the crystal and solution structures of two RNA oligonucleotides. *Biophys. J.* **76**, 66–75.
40. Leonard, G. A., McAuley-Hecht, K. E., Ebel, S., Lough, D. M., Brown, T. & Hunter, W. N. (1994). Crystal and molecular structure of r(CGCGAAUU AGCG): an RNA duplex containing two G(anti). A(anti) base-pairs. *Structure*, **2**, 483–494.
41. Klosterman, P. S., Shah, S. A. & Steitz, T. A. (1999). Crystal structures of two plasmid copy control related RNA duplexes: an 18 base base-pair duplex at 1.20 Å resolution and a 19 base-pair duplex at 1.55 Å resolution. *Biochemistry*, **38**, 14784–14792.
42. Trikha, J., Filman, D. J. & Hogle, J. M. (1999). Crystal structure of a 14 base-pair RNA duplex with non-symmetrical tandem G–U wobble base-pairs. *Nucl. Acids Res.* **27**, 1728–1739.
43. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995). Nuclear magnetic dipolar interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl Acad. Sci. USA*, **92**, 9279–9283.
44. Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science*, **278**, 1111–1114.
45. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature Struct. Biol.* **5**, 1065–1074.
46. Trantirek, L., Urbasek, M., Stefl, R., Feigon, J. & Sklenar, V. (2000). A method for direct determination of helical parameters in nucleic acids using residual dipolar couplings. *J. Am. Chem. Soc.* **122**, 10454–10455.
47. Wu, Z.-R., Delaglio, F., Tjandra, N., Zhurkin, V. B. & Bax, A. (2003). Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and 31P chemical shift anisotropy. *J. Biomol. NMR*, **26**, 297–315.
48. Tycko, R., Blanco, F. J. & Ishii, Y. (2000). Alignment to biopolymers in strained gels: a new way to create detectable dipole–dipole couplings in high-resolution biomolecular NMR. *J. Am. Chem. Soc.* **122**, 9340–9341.
49. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. (2000). Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR*, **18**, 303–309.
50. Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378.
51. Lawrence, D. C., Stover, C. C., Noznitsky, J., Wu, Z.-R. & Summers, M. F. (2003). Structure of the intact stem adn bulge of HIV-1 Ψ-RNA stem loop SL1. *J. Mol. Biol.* **326**, 529–542.
52. Pan, B., Mitra, S. N. & Sundaralingam, M. (1998). Structure of a 16-mer RNA duplex r(GCAGACUUA AAUCUGC)2 with wobble C*A + mismatches. *J. Mol. Biol.* **283**, 977–984.
53. Hoogstraten, C. G., Legault, P. & Pardi, A. (1998). NMR solution structure of a lead-dependent ribozyme: evidence for dynamics in RNA catalysis. *J. Mol. Biol.* **284**, 337–350.
54. Lietzke, S. E., Barnes, C. L., Berglund, J. A. & Kundrot, C. E. (1996). The structure of an RNA dodecamer shows how tandem U–U base-pairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure*, **4**, 917–930.
55. Wang, Y.-X., Huang, S. & Draper, D. E. (1996). Structure of a U–U pair within a conserved ribosomal RNA hairpin. *Nucl. Acids Res.* **24**, 2666–2672.
56. Shi, K., Wahi, M. & Sundaralingam, M. (1999). Crystal structure of an RNA duplex r(GGGCG CUCC)2 with non-adjacent G–U base-pairs. *Nucl. Acids Res.* **27**, 2196–2201.
57. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A.

(1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell*, **91**, 705–712.
58. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr, Morgan-Warren, R. J., Carter, A. P., Vonrhein, C. *et al.* (2000). Structure of the 30s ribosomal subunit. *Nature*, **407**, 327–339.
59. Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus tat peptide–TAR RNA complex. *Chem. Biol.* **2**, 827–840.
60. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E. *et al.* (1996). Crystal structure of a group I ribozyme domain: principles of RNA packaging. *Science*, **273**, 1678–1685.
61. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221.
62. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.
63. Bieth, E., Gabus, C. & Darlix, J.-L. (1990). A study of the dimer formation of Rous sarcoma virus RNA and of its effect on viral protein synthesis *in vitro*. *Nucl. Acids Res.* **18**, 119–127.
64. Darlix, J.-L., Gabus, C., Nugeyre, M.-T., Clavel, F. & Barre-Sinoussi, F. (1990). *Cis* elements and *trans*-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol.* **216**, 689–699.
65. Bonnet-Mathoniere, B., Girard, P.-M., Muriaux, D. & Paoletti, J. (1996). Nucleocapsid protein 10 activates dimerization of the RNA of moloney murine leukaemia virus *in vitro*. *Eur. J. Biochem.* **238**, 129–135.
66. Girard, P.-M., de Rocquigny, H., Roques, B.-P. & Paoletti, J. (1996). A model of Psi dimerization: destabilization of the C278-G303 stem-loop by the nucleocapsid protein (NCp10) of MoMuLV. *Biochemistry*, **35**, 8705–8714.
67. Nagaswamy, U., Voss, N., Zhang, Z. & Fox, G. E. (2000). Database of non-canonical base-pairs found in known RNA structures. *Nucl. Acids Res.* **28**, 375–376.
68. Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q. & Fox, G. E. (2002). NCIR: a database of non-canonical interactions in known RNA structures. *Nucl. Acids Res.* **30**, 395–397.
69. Patel, D. J. (1999). Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* **9**, 74–87.
70. Mollova, E. T., Hansen, M. R. & Pardi, A. (2000). Global structure of RNA determined with residual dipolar couplings. *J. Am. Chem. Soc.* **122**, 11561–11562.
71. Hennig, M., Carlomagno, T. & Williamson, J. R. (2001). Residual dipolar coupling TOCSY for direct through space correlations of base protons and phosphorus nuclei in RNA. *J. Am. Chem. Soc.* **123**, 3395–3396.
72. Vallurupalli, P. & Moore, P. B. (2002). Measurement of H2′–C2′ and H3′–C3′ dipolar couplings in RNA molecules. *J. Biomol. NMR*, **24**, 63–66.
73. Miclet, E., O'Neil-Cabello, E., Nikonowicz, E. P., Live, D. & Bax, A. (2003). 1H–1H dipolar couplings provide a unique probe of RNA backbone structure. *J. Am. Chem. Soc.* **125**, 15740–15741.
74. O'Neil-Cabello, E., Bryce, D. L., Nikonowicz, E. P. & Bax, A. (2004). Measurement of five dipolar couplings from a single 3D NMR multiplet applied to the study of RNA dynamics. *J. Am. Chem. Soc.* **126**, 66–67.
75. Milligan, J. F. & Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**, 51–62.
76. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
77. Johnson, B. A. & Blevins, R. A. (1994). NMRview: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
78. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Chem. Phys.* **71**, 4546–4553.
79. Macura, S. & Ernst, R. R. (1980). Elucidation of cross relaxation in liquids by two-dimensional NMR spectroscopy. *Mol. Phys.* **41**, 95–117.
80. Bothner-By, A., Stephens, R. L., Lee, J.-M., Warren, C. D. & Jeanloz, R. W. (1984). Structure determination of a tetrasaccharide: transient nuclear Overhauser effects in the rotating frame. *J. Am. Chem. Soc.* **106**, 811–813.
81. Bax, A., Griffey, R. H. & Hawkins, B. L. (1983). Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR. *J. Magn. Reson.* **55**, 301–315.
82. Bax, A. & Summers, M. F. (1986). 1H and 13C assignments from sensitivity enhanced detection of heteronuclear multiple bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.* **108**, 2093–2094.
83. Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588–593.
84. Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R. & Bax, A. (1993). Increased resolution and improved spectral quality in four-dimensional 13C/13C-separated HMQC-NOESY-HMQC spectra using pulsed field gradients. *J. Magn. Reson., ser. B*, **101**, 210213.
85. Piotto, M., Saudek, V. & Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
86. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H–^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry*, **28**, 6150–6156.
87. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
88. Mikael Akke, R. F., Jiang, F., Patel, D. & Palmer, A. G., III (1997). Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *RNA*, **3**, 702–709.
89. Vuister, G. W. & Bax, A. (1992). Resolution enhancement and spectral editing of uniformly ^{13}C-enriched proteins by homonuclear broadband ^{13}C decoupling. *J. Magn. Reson.* **98**, 428–435.
90. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997).

Torsion angle dynamics for protein structure calculations with a new program, DYANA. *J. Mol. Biol.* **273**, 283–298.
91. Case, D. A. (2002). AMBER 7 Users Manual, University of California, San Francisco.
92. Dickerson, R. E. (1989). Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dynam.* **6**, 627–634.
93. Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY.
94. Dossett, P., Hus, J.-C., Marion, D. & Blackledge, M. (2001). A novel interactive tool for rigid-body modelling of multi-domain macromolecules using residual dipolar couplings. *J. Biomol. NMR*, **20**, 223–231.
95. DeLano Scientific (2002). The PyMOL Molecular Graphics System, DeLano, W. L.
96. Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655–667.
97. Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91.

Edited by P. Wright

(Received 6 January 2004; received in revised form 19 January 2004; accepted 20 January 2004)