Structure of SL4 RNA from the HIV-1 Packaging Signal[†,‡]

Deborah J. Kerwood, Michael J. Cavaluzzi, and Philip N. Borer*

Department of Chemistry, Graduate Program in Structural Biology, Biochemistry, and Biophysics, Syracuse University, Syracuse, New York 13244-4100

Received June 8, 2001

ABSTRACT: The NMR-based structure is described for an RNA model of stem−loop 4 (SL4) from the HIV-1 major packaging domain. The GAGA tetraloop adopts a conformation similar to the classic GNRA form, although there are differences in the details. The type II tandem G·U pairs have a combination of wobble and bifurcated hydrogen bonds where the uracil 2-carbonyl oxygen is hydrogen-bonded to both G,H1 and G,H2. There is the likelihood of a Na^+ ion coordinated to the four carbonyl oxygens in the major groove for these G·U pairs and perhaps to the N7 lone pairs of the G bases as well. A continuous stack of five bases extends over nearly the whole length of the stem to the base of the loop in the RNA 16mer: C15|U14|G13|G5|C6. There is no evidence for a terminal G·A pair; instead, G1 appears quite unrestrained, and A16 stacks on both C15 and G2. Residues G2 through G5 exhibit broadened resonances, especially G3 and U4, suggesting enhanced mobility for the 5′-side of the stem. The structure shows G2|G3|U4 stacking along the same strand, nearly isolated from interaction with the other bases. This is probably an important factor in the signal broadening and apparent mobility of these residues and the low stability of the 16mer hairpin against thermal denaturation.

The HIV-1 RNA 5′-leader participates in the control of many viral functions including selectivity in packaging the genomic RNA, forming the dimer linkage structure, initiating reverse transcription, and in stable integration of the proviral DNA (*1*). Figure 1a shows the four stem−loops thought to be most critical in the process of packaging unspliced RNA into newly forming virions (*2*, *3*). High-resolution structures have been published for SL1[1] (*4*−*8*), SL2 (*9*, *10*), and SL3 (*11*−*13*), and now for SL4. The majority of specificity for packaging is conferred by SL3 (*2*, *3*, *14*−*19*), by virtue of its interaction with the 55-residue nucleocapsid (NC) domain of the *gag* and *gag-pol* precursor proteins. However, the NC domain also binds SL1, SL2, and SL4 (*2*, *10*), and SL3 knockouts are still packaged, albeit with much lower efficiency (*16*, *18*). A recent study showed that SL3 and SL2 have similar affinity for the mature 55mer NCp7 (*10*). Unpaired G residues are a common characteristic of each of the known sites for NC interaction (*20*). All of the published NC nucleic acid structures have a similar interaction between G residues and zinc "knuckles" from NC (*9*, *12*, *21*). About 1500 NC domains are packaged with two RNA strands (*22*), so it is likely that multiple RNA−protein interactions govern optimal packaging efficiency.

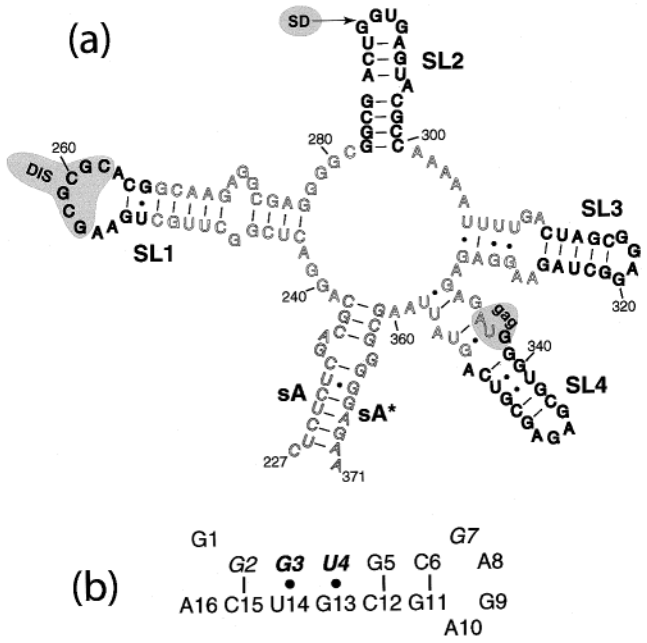

FIGURE 1: (a) Components of the major packaging domain of HIV-1 RNA. Regions for which high-resolution structures have been determined are shown in heavy letters. (b) The 16mer SL4 construct analyzed in this work. Add 337 to the construct's numbering to obtain base positions relative to Cap = 1. Residues 1−4 and 7, especially G3 and U4, appear to be more mobile than the others. A noncanonical base pair involving G1 and A16 is not present in the NMR-based structure.

While SL1−3 lie within the untranslated region, SL4 occurs at the beginning of the *gag* genes. Thus, its sequence is governed by protein coding as well as structural constraints. SL4 contains tandem G·U pairs in a low-stability "type II" configuration that is rarely found in rRNA (*23*). It also contains a GNRA tetraloop (*24*) and the potential to

[†] Supported in part by NIH Grant GM32691 and Syracuse University.

[‡] An ensemble of 14 structures has been deposited with the Protein Data Bank, Accession Number 1JTW.

[1] Abbreviations: SL, stem−loop; NOE, nuclear Overhauser effect; COSY, correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence; rmsd, root-mean-square deviation. Protons are specified by N#,$, where N = type of base, # = sequence number, and $ = proton number within the residue. The nuclei contributing the parent frequencies for an NMR cross-peak are enclosed in braces, e.g., {G3,2′-U4,6}. H2′,3′,4′,5′,5″ are sometimes referred to as the F protons. The bridging phosphate group in X-^{31}P-Y is considered to be part of residue Y, viz., pY. The symbol "|" denotes base stacking or a step between base pairs.

10.1021/bi0111909 CCC: $20.00
Published on Web 11/08/2001

form noncanonical G−A pairs. Figure 1b shows the sequence of the 16mer construct described in this paper as a model for SL4.

NMR-based structure determination of the SL4 16mer is described here. Using natural abundance methods, we have assigned nearly all of the carbon-bound and slowly exchanging nitrogen-bound protons. The structure reveals several interesting features. The tetraloop adopts a conformation similar to the classic GNRA form (*24*), although there are differences in the details. The tandem G·U pairs cause the stem to be quite unstable, and there is no evidence for a terminal G·A pair.

Marginal thermodynamic stability and flexibility may aid the efficient production of the *gag* and *gag-pol* precursor proteins, easing ribosome entry to sequences lying between SL3 and SL4, followed by opening SL4 to initiate protein synthesis. The same stability and flexibility factors may also predispose the GAGA-loop sequence toward interaction with nucleocapsid domains of the *gag* precursor for enhancing packaging selectivity. In the structure of isolated SL3 RNA (*11*) all of the donor and acceptor sites are unoccupied that are ultimately used to make hydrogen bonds between the GAG-loop sequence and NCp7 (*12*). The corresponding H-bonding sites are also open in the SL4 structure reported here, so it is expected that NC can interact with these G residues in a very similar manner as in the SL3 and SL2 loops (*10*). Another role for SL4 may be in establishing a 3D fold for the HIV-1 RNA major packaging domain. There are many good precedents for GNRA tetraloops participating in tertiary interactions (*25*−*30*).

EXPERIMENTAL PROCEDURES

RNA Purification. The SL4 16mer was purchased from Dharmacon Research Inc. using the 2′-ACE protecting strategy (*31*). After 2′-deprotection, the SL4 powder was resuspended in 1 mL of HPLC solvent A, heated to 90 °C for 4 min, and snap cooled on ice prior to purification by HPLC. All water used in these procedures was freshly purified using a Milli-Q system (Millipore). Glass-distilled 99.8% and sealed ampules of 99.96% D_2O were purchased from Isotec. All procedures utilized baked glassware (8 h, 270 °C) or sterile plastic ware.

Deprotected SL4 was purified by HPLC on a Zorbax Oligo (Hewlett-Packard) anion-exchange column (9.4 × 250 mm), 2 mL/min, 32 min gradient from 14% to 60% solvent B. SL4 eluted between 18 and 19 min with clear resolution from shorter chains. Solvent A was 0.05 M PO_4^{3-} and 7.5 M urea (National Diagnostics); solvent B = A + 0.40 M Na_2SO_4. Solvents were filtered through 0.2 μm cellulose acetate (Corning) and stored in amber bottles. The pH of both solutions was adjusted daily to pH 6.5 with phosphoric acid and degassed by sonication.

The SL4/urea fractions were desalted on a 5 × 30 cm column of Bio-Gel P6DG (Bio-Rad). The column was eluted with a fresh solution of 3 drops of concentrated NH_3/2 L of H_2O at ∼3 mL/min. The high concentration fractions were dried by rotary evaporation and transferred to a Centricon-3 (Amicon) in <2 mL, washed once with 0.5 M NaCl, 2× with 0.1 M NaCl, and 3× with water. The Centricon steps were essential to eliminate small organic impurities that often copurify with RNA samples.

NMR Data Collection. NMR experiments were performed on a Brüker DRX-500 equipped with a Nalorac 5 mm inverse probe (^{1}H sensitivity = 850:1 for 0.1% ethylbenzene) or a DRX-600 with equivalent sensitivity. All samples were brought to a final volume of approximately 700 μL. Spectra were acquired on samples at ∼0.7 mM strand concentration in NMR buffer: 25 mM NaCl, 5 mM $Na_{1.4}H_{1.6}PO_4$ (pH 7.0), 0.1 mM EDTA, and 0.01% NaN_3. D_2O samples were prepared by lyophilizing two times from glass-distilled 99.8% D_2O and two times from 99.96% D_2O, releasing the vacuum under argon. NMR samples were stored at 4 °C. Before each series of NMR experiments, the SL4 sample was heat cycled to ensure hairpin formation by equilibrating at 60, 90, 60, and 0 °C for 2−3 min at each temperature. Chemical shifts of protons are reported relative to DSS (Supporting Information).

H_2O samples were prepared in the identical NMR buffer, but adjusted to pH 5.5 with HCl. To record spectra at −6 °C, dust and particulate matter must be removed (*32*). The samples were clarified through a prerinsed 0.2 μm cellulose acetate filter, washed with water, dried in a SpeedVac, reconstituted with 90% H_2O/10% D_2O to the appropriate volume for an NMR sample (usually 700 μL), and centrifuged for 1 h in a standard Eppendorf microcentrifuge. Without disturbing the bottom of the sample, ∼90% of the liquid was transferred to a clean NMR tube. The temperature in the NMR probe was reduced at 1 °C/h below 0 °C.

2D NOE spectra in H_2O were acquired with 4 K* × 512* hypercomplex points, 128 scans per t_1 increment, 3 s relaxation delay time, and 200 ms mixing time. 2D NOE spectra in ^2H_2O were acquired with 4 K* × 512* hypercomplex points, 64 scans, 8 s delay time, and 100 and 300 ms mixing times. Other details for acquiring and processing 2D NOE, TOCSY, natural abundance ^{1}H−^{13}C HMQC, and HP-COSY spectra are the same as described previously (*11*).

NMR-Based Restraints. NOE cross-peaks were integrated and categorized as described previously (*11*, *33*, *34*). Target distances were calculated to correct for spin diffusion using the RANDMARDI (*35*) implementation of the MARDIGRAS (*36*) program for nonoverlapping peaks and for peaks with light to moderate overlaps. Target ranges for DYANA (*37*) and AMBER (*38*) refinement were ±0.5 Å when the target distance was <4.2 Å. When the target from the RANDMARDI calculation was >4.2 Å, the range was set to 4−6 Å. Deconvolution of heavily overlapped peaks often gives a clear indication of the presence of an NOE, even though the peak cannot be reliably integrated. These peaks were estimated as strong (1.8−3 Å), medium (2−4 Å), or weak (3−6 Å). RANDMARDI occasionally rejects peaks for which there is no appropriate solution to the system of coupled equations that describes the cross-relaxation network; these peaks were given a 4−6 Å range if they were very weak or a 3−6 Å range if they were of medium intensity.

Stereospecific assignments were not made for most H5′ and H5″, so they were not included in the RANDMARDI calculations. A 3−6 Å range was assigned to distances from the base H8/6 to both H5′ and H5″ within a residue if both cross-peaks were weak in the 2D NOE spectrum. A 2−6 Å range was applied if either cross-peak had a medium intensity. If one of the cross-peaks was strong, preliminary structures were inspected to make a stereospecific assignment. A 1.8−3 Å range applied for that proton to H8/6; the

other proton was assigned a weak restraint (3−6 Å). Only one assignment is consistent with the ensemble of other NMR-based restraints in our experience; we have not encountered situations where both cross-peaks are strong.

Cross-peaks involving exchangeable protons were assigned broader ranges to account for uncertainties in the experimental intensities introduced by exchange and the excitation profile used for water suppression. All were classified as weak (3−6 Å), except for intense peaks within a single base pair, which were assigned an upper limit of 4 Å and no lower limit.

Hydrogen bonds were assigned a 0.3 Å range, centered about 1.95 Å for H···O and 1.99 Å for H···N (*39*). Additionally, the distances between heavy atoms involved in H-bonds were restrained in DYANA calculations since the program does not provide H-bond angle limits; we used a 0.3 Å range, centered at 2.95 Å for N···O and 2.90 Å for N···N (*39*). Only the heavy atom restraints were used for AMBER calculations, in addition to an angle restraint of 180 (±10)° for each H-bond. No H-bond restraints were used for the G·U pairs, and no restraints were included to force H-bonded amino groups to lie in the planes of the bases.

Dihedral angle restraints were obtained from three-bond ^{1}H−^{1}H and ^{1}H−^{31}P coupling constants. The procedures for extracting the restraints are as described earlier (*11*), with modifications for restraints upon the torsion angles, γ and ϵ (see Results). A list of restraints is assembled in the Supporting Information.

Structure Determination. Restraint lists were introduced to the torsion angle dynamics program, DYANA (*37*), to produce 50 or 400 preliminary structures. The 50-structure calculations required ∼20 min on a 2-processor 270 MHz R12k Silicon Graphics Octane computer; ∼3 h was required for 400 structures. Nearly all of these structures "converge" in the sense of older metric matrix distance geometry calculations. The structures are ranked in order of their target function value, which increases with increasing severity of restraint violations. DYANA has no associated force field and often produced structures that violate normal angles for NH···N and NH···O hydrogen bonds. Therefore, we subjected the DYANA structures to AMBER (*38*) refinement. The procedure gave a family of 15 structures with minimal violations of the NMR-based restraints.

In preliminary stages, the first few structures of a 50-structure DYANA run were examined for the effect of changes in the restraint list due to new assignments, changes in torsional restraints, etc. In later stages of refinement 400-structure ensembles were calculated and 20 structures were examined. Occasionally one or two of these were rejected because of large rmsd with respect to the others. The remaining structures were carried forward for AMBER refinement.

A typical AMBER trajectory consisted of 1000 steps of conjugate gradient energy minimization, a 10 ps molecular dynamics trajectory at 300 K, and another 1000 steps of minimization. A few simulated annealing trajectories of 35 ps were also calculated for some DYANA starting structures, under conditions matching those used in the SL3 RNA refinement (11, the maximum temperature was 1000 K in those trajectories). The NMR restraints were present at full strength throughout. Details were the same as for the implicit solvent model we have used previously (*11*, *33*, *34*). In this

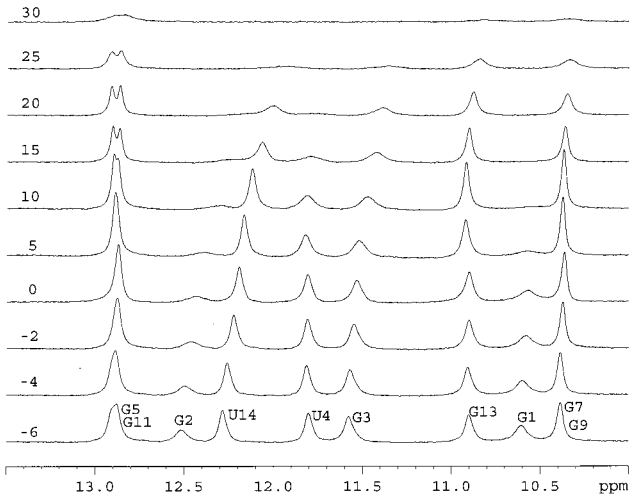

FIGURE 2: 1D imino proton signals of the SL4 construct at several temperatures at pH 5.5.

model, hexahydrated Na^+ ions were placed 4.5 Å away from the phosphates in order to simulate the effect of solvent on the system. Restraints were imposed to keep the sodium ion near its parent phosphate, as before (*11*).

After the AMBER refinements, a reference structure was distinguished, which had the lowest average of its rmsd to the other structures based on the heavy atoms of residues 3−14. The other structures were then ranked by rmsd to the reference structure (residues 3−14).

Structures are presented as drawings from InsightII (Molecular Simulations, San Diego, CA). Hydrogen bonds were identified in the InsightII program using default parameters: these require the donor-H···acceptor angle to be >120° and (1) the N−N or N−O donor−acceptor distance to be <3.0 Å or (2) the H···acceptor distance to be <2.0 Å. Helix twist angles were calculated with the program CURVES (*40*). NMR spectra were saved as postscript files, converted to .pdf format using Adobe Acrobat Professional, and annotated in Adobe Illustrator (Adobe Systems, Mountain View, CA). Figure 1 was created in Illustrator from RNAstructure (*41*, *42*).

RESULTS

Exchangeable Proton Spectra and Melting. Figure 2 presents the imino proton region of the SL4 16mer spectra at several temperatures. There are ten G and U residues in the RNA, and all are represented in the −6 °C spectrum at pH 5.5. Correlating cross-peaks in the 2D NOE spectral regions in H_2O (not shown) allowed unequivocal assignments to be made for all but G9,1 and G1,1. At all temperatures G9,1 and G7,1 overlap (near 10.4 ppm); the G5,1 and G11,1 signals also overlap at 0−5 °C (near 12.9 ppm), but one varies more with temperature (probably G11,1). At −6 °C, signals from the G·U pairs show full intensity, while broad signals are observed for G1,1 and G2,1. As temperature is increased, the signals broaden and some shift upfield substantially. G1,1 and G2,1 broaden at lowest temperature. Of the G·U peaks, U4,3 broadens first, followed closely by G3,1 on the same strand, then U14,3, and finally G13,1 on the other. The temperature-dependent broadening of the combined G9,1/G7,1 peak in the loop is nearly the same as for G13,1, while the last to disappear are G5,1 and G11,1 from the base pairs nearest the loop.

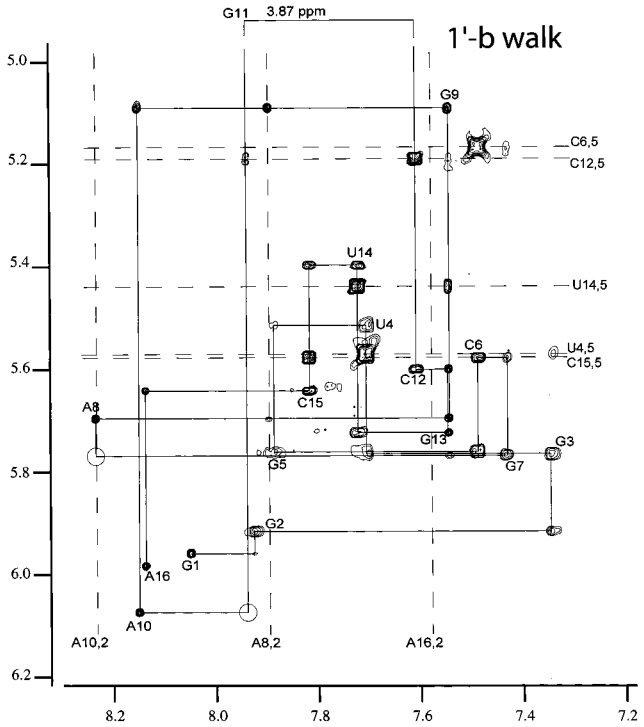

FIGURE 3: The H1′-H8/6 walk for the SL4 construct from the 2D NOE spectrum at 15 °C is shown (−6 Hz deconvolution). Cross-peaks involving protons in single residues are indicated along the walk, and dashed lines denote AH2 or C/UH5 frequencies. Sequential cross-peaks that can be seen at a lower contour level are shown as empty circles.

The A_{260} vs T profiles (not shown) exhibit single transitions, with T_m = 43 °C in the NMR buffer (32 mM Na^+). At 30 °C, where most of the exchangeable proton peaks have broadened into the baseline, the optically observed transition indicates that 15−20% of the hairpins are melted.

Nonexchangeable Proton and Phosphorus Spectra. Assignments followed procedures that are summarized in our earlier papers (*33*, *34*), relying mostly upon classic sequential assignment strategies (*43*−*45*). The most important information for proton assignments and structural restraints comes from 2D NOE spectra in D_2O, where the peaks have been sharpened by −6 Hz deconvolution, according to methods that preserve the relationships of peak integrals (*46*, *47*). The region containing the sequential H1′-H8/6 walk is shown in Figure 3 for the spectrum with a mixing time of 300 ms. Other spectra at shorter mixing time (100 ms) and different temperatures helped to resolve the remaining ambiguities imposed by peak overlap. The H1′ assignments were extended to the H2′ by comparison of F-H1′ and F-H8/6 regions of the 2D NOE spectra (the F protons = H2′,3′,4′,5′,5″). H3′ assignments were made by correlating the previous assignments with the F-H8/6 region and the HP-COSY spectrum. The H3′-H8/6 footprint is usually large in the 3′-dimension, due to coupling to H2′, H4′, and ^{31}P, which also helps to distinguish these signals. The H4′-H1′ cross-peaks are usually medium−strong in the F-H1′ region, with relatively small footprints in the H4′ direction. However, most of the ambiguities in the assignments are among the H4′ because NOE correlations to H8/H6 are usually missing or weak. The ambiguities in the H4′ assignments make little difference in the structure determination. Stereospecific assignments of the H5′/5″ have not been made, except for G11. However, many of the frequencies are easily distinguished in the HP-COSY and are correlated to signals in the F-H8/6 region. Weak H1′-H5′/5″ correlations are often seen in the F-H1′ region, as well. The TOCSY spectrum shows weak correlations to the H1′ for G1, G2, A8, G9, A10, and A16, which are helpful in confirming some of the F-proton assignments. In addition, a ^{1}H−^{13}C HMQC spectrum verified many of the F-region assignments. None of the sugars displays an H1′-H2′ coupling constant larger than the H1′ line widths, which puts an upper limit of about 3−4 Hz on $J_{1-2'}$.

Table 1: Restraint Statistics for the SL4 16mer Refinement

NMR-Based Restraints	
residue i to i	169
residue i to j, $j \neq i$	132
subtotal	301
H-bond	18
backbone dihedrals	72
total restraints	391
Average Violations per Structure	
average distance violation (Å)	0.03
distance violations >0.3 Å	4
distance violations >0.5 Å	0
average dihedral violation (deg)	1
dihedral violations >10°	2
dihedral violations >30°	0

Several features are apparent from Figure 3 that are reinforced in other spectral regions. Except for G1, none of the H8/6-H1′ cross-peaks have abnormally large intensities, implying anti configurations of the bases with respect to their furanose rings, including those in the loop. The pattern of NOE peaks is consistent with an A-family stem, with complete walks through each of the spectral regions. The loop has the characteristics of GNRA tetraloops (*24*), including G11,1′ resonating far upfield from the other anomeric protons, and extreme chemical shifts for G9,1′, A10,1′, and the ^{31}P signals in the loop. Signals in and near the G·U pairs also exhibit unusual properties. These include very broad signals for G3 and U4 and several other broadened peaks for G2, G5, and G7. Several cross-peaks from stem residues are very weak, including sequential (residue $i-1$ to i) peaks {G1,1′-G2,8}, {G1,2′-G2,1′}, {U4,1′-G5,8}, as well as intraresidue (i to i) peaks for {G3,2′-8} and {G13,2′-8}. These distances are all 4 Å or longer in the structures described in the Discussion. An unusually strong NOE is observed for {G11,8-5″}, which allowed stereospecific distinction of G11,5′/5″.

Doubly dispersive (DD) phasing, with display of only positive intensities, gathers most of the HP-COSY cross-peak intensity into a much smaller footprint than the normal antiphase presentation (*33*). At a single ^{31}P frequency, three-bond correlations with ($i-1$)H3′ and (i)H5′/5″ are found (by convention, residue i bears the 5′-phosphate). Occasionally, long-range ^{31}P couplings are seen to (i)H3′ or H4′. Thus, the HP-COSY gives important confirmation of sugar proton assignments and also provides ^{31}P chemical shifts for each residue.

Distance and Torsional Restraints. Statistics regarding the restraints and violations are collected in Table 1. The 2D NOE spectra were used to define 301 distance restraints, with both upper and lower limits estimated for 88 distances using RANDMARDI. About 35 NOE-based distances have restrictions that prevent them from contributing directly to the

structure; these distances are largely determined by covalent bonds and angles. Many of these contribute indirectly to the structure by normalizing the cross-relaxation rate equations in RANDMARDI. The intensities from the 15 °C spectrum were used to generate distance restraints for isolated peaks. Peaks visible only at other temperatures were used to generate range constraints in the strong, medium, or weak categories (see Experimental Procedures). We also imposed H-bond constraints for the three G·C pairs. Altogether, there were 391 − 35 = 356 effective structural restraints. This is about the same density of restraints as in the SL3 structure determination (*11*), providing 30−40 restraints upon a typical residue, and mean effective restraints per residue = 22. This should provide a sufficient set for structure determination (*48*). The Supporting Information includes a list of all restraints.

The glycosyl torsion angle, χ, was loosely restrained to the anti range (200 ± 50°), except for G1, where a medium intensity {G1,H8-H1′} cross-peak is evident. The χ angle was not restrained for G1.

The dominant sugar pucker for all of the residues falls in the N-range typical of A-family structures. However, there are indications of flexibility at the termini and loop. Therefore, δ was restrained to 86 ± 10° only for the stem residues 2−6 and 12−15.

No restraints on γ were imposed because it is difficult to estimate the ^3J-couplings between H4′-5′ and H4′-H5″ in natural abundance spectra, especially for signals with broad line widths. (Such estimates require analysis of passive couplings on {P-H5′}, {P-H5″}, {H8/6-H5′}, and {H8/6-H5″} cross-peaks or analysis of DQF-COSY spectra in crowded spectral regions.) The imposition of distance ranges from *i*,H8/6 to *i*,H5′/5″ provides an alternative to restraints on γ when the bases are in the anti configuration. Both H8/6-H5′/5″ distances are >3.5 Å when γ is in the usual g^+ staggered configuration, with H4′ lying between H5′ and H5″ in the Newman projection about the C4′−C5′ bond. In other rotamers, either H5′ or H5″ is <3 Å from H8/6, which would produce a high-intensity NOE cross-peak. This is the case for G11,8-G11,5″, for which a 1.8−3 Å restraint was imposed. A 3 Å lower distance limit was imposed for all of the remaining H8/6-H5′ and H8/6-H5″ distances because other strong {H8/6-H5′/5″} cross-peaks were never observed; the presence of a 2−3 Å cross-peak would have been obvious in the F-H8/6 region, even in crowded H8/6 lanes.

The most distinctive features of the HP-COSY spectrum are the intense {H3′-P} correlations. The largest factor controlling the intensity of the peaks is the active coupling of ∼9 Hz for regular RNA helices (*49*). (Of course, line width and passive couplings influence the shape and intensity of {H3′-P} cross-peaks, so it is unwise to predict specific *J*-values based upon relative peak heights.) In the antiphase presentation of the spectra, the intensity of the largest lobe varies over a considerable range in the spectrum of the SL4 16mer. The largest intensity is for the {G1,3′-pG2}, but several cross-peaks are at less than half this intensity: G3,3′ (0.3), U4,3′ (0.4), G7,3′ (0.2), A10,3′ (0.2), and G13,3′ (0.4). It is possible that unusual ϵ torsions occur for these residues; therefore, ϵ was not constrained. For the other 11 residues, ϵ was constrained to 210 ± 30°, where $^3J_{H3'-P}$ should have values of 4−11 Hz (*50*).

It is usually evident in HP-COSY spectra that both the {H5′-P} and {H5″-P} cross-peaks are weak at a given ^{31}P frequency. This implies a trans configuration for β where both of the $^3J_{H-P}$ are small (*49*, *51*), so we constrained β = 180 ± 40° for most of the residues. The main exception was ^{31}pG11, which has two large cross-peaks; here we restrained β to −110 ± 20°, for which both $^3J_{H-P}$ are in the range of 8−14 Hz (*50*). The restraint was always violated in structures which set β to alternative angles that satisfy the requirement for two large coupling constants. These alternatives must be inconsistent with the other NMR and holonomic restrictions. The β angle was left unrestrained for pA8, which has one large and one small cross-peak.

Most of the ^{31}P signals of the 16mer resonate within a 1 ppm range; this includes the 5′-phosphates of residues 2−6 and 12−16. Trans rotamers for the phosphates 5′ to these residues were excluded (allowed values were $\alpha = \zeta = 0 \pm 120°$). However, the signals for five residues in and near the loop have unusual ^{31}P chemical shifts. The following resonances are displaced from the center of the main band by the amount in parentheses: ^{31}pG7 (+0.9 ppm), pA8 (+2.4), pG9 (+1.2), pA10 (−0.7), and pG11 (+1.8). These values suggest a blend of backbone conformations that is different from the A-family stem; therefore, the α angles of these residues and the ζ angles of the preceding residues were not restrained. There is nothing unusual in the backbone conformations surrounding the G·U pairs, given that the ^{31}P resonances for residues 3−5 and 13−15 lie within the normal range.

3D Structure. The structure of the 16mer SL4 RNA construct is shown in Figure 4, leaving out the terminal bases, G1 and A16. The figure reveals that the loop (residues 7−10) follows the general outline of GNRA tetraloops (*24*), with residues 8−10 mutually stacked (magenta); the SL4 tetraloop appears to differ in some details from previously determined structures (see Discussion). The G7-base stacks weakly on 11|12 (green), and a five-base stack exists from 6|5|13|14|15 (blue). Note that residues 2−4 are stacked along the same strand (red), with a discontinuity in the stacking between U4-G5. The figure also shows that the tandem G3·U14|U4·G13 pairs are formed. There is no G1·A16 base pair at the end of the stem. Instead, as shown in Figure 5, A16 stacks on both G2 and C15, and there is a pronounced buckle in the G2·C15 pair.

DISCUSSION

Thermal Stability, Flexibility, and Overall Structure. The SL4 16mer is less stable toward thermal denaturation than most other RNA hairpins for which NMR structures have been determined. This stems from the presence of the tandem G·U pairs and the unpaired terminal G·A mismatch, which contribute only minimally toward stability. The standard free energy of forming the hairpin from the unpaired strands is predicted to be only −5.3 kcal/mol at 37 °C in 1 M NaCl (*42*), and the measured T_m = 43 °C in our 32 mM [Na^+] NMR buffer.

The lack of thermal stability is also reflected in the imino proton spectra of the SL4 16mer, displayed at several temperatures in Figure 2. The results show that the 5′-side of the stem allows exchange with water prior to the 3′-side. This includes G3 and U4 in the tandem G·U pairs broadening prior to their partners, G13 and U14. The stem region nearest

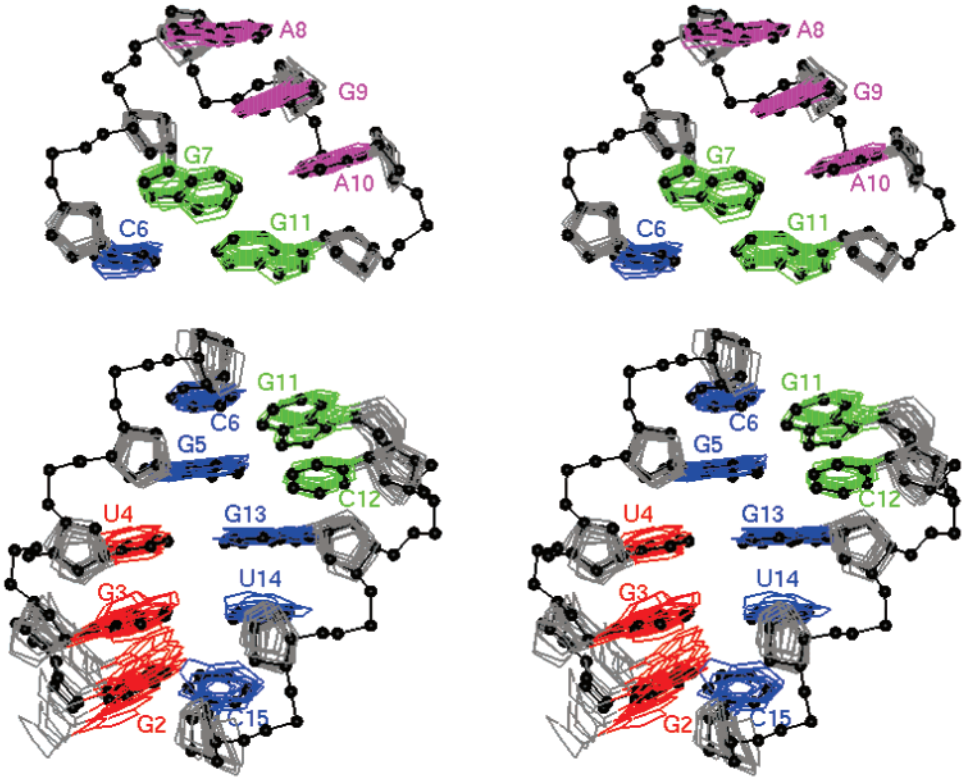

FIGURE 4: Two local superpositions of 15 NMR-based structures of the SL4 16mer RNA construct. The reference structure is displayed in black, with the base and sugar rings of the others superimposed upon it. Colors denote mutually stacked groups of bases. Bottom: the stem residues 3−6 and 11−14 (average rmsd = 0.8 Å from the reference structure). Top: residues 6−11 include the loop (average rmsd = 0.6 Å). The top view was generated by rotating the displayed residues 45° about the *y*-axis and then leveling C6·G11 by a small rotation about *z*. A viewer can imagine rotating the fragments to superimpose the C6·G11 base pair, which appears in both views. Least-squares superpositions in this and subsequent figures were based on the heavy atoms of the specified residues. All stereoviews are defocused representations.

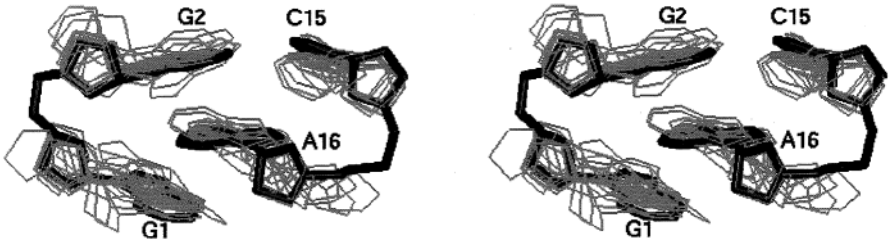

FIGURE 5: Terminal residues in the SL4 16mer. The least-squares superposition was based on residues 2, 15, and 16 (average rmsd = 1.5 Å from the reference structure, shown in black).

the tetraloop is most resistant to thermally induced melting. It was important to use procedures introduced by the Szyperski laboratory (*32*) for measuring NMR spectra in supercooled solutions to observe and assign many of the imino and amino proton signals.

It is well-known that long-range features are not faithfully represented in NOE- and ^3J-based structures of extended nucleic acids. Also, the C4′-C5′-O5′-P-O3′-C3′ backbone linkages are ill-defined in most NMR structures, especially in flexible molecules such as this one. Therefore, we have chosen to present structural details using local least-squares superpositions and, usually, to omit the complete backbone geometries. Figures 4−6 show the full details of the reference structure, along with the ring atoms of others having similarly low DYANA target functions and AMBER energies (see Experimental Procedures). Residues 1, 2, and 16 are quite disordered in the DYANA structures. The average rmsd = 0.9 Å for residues 3−14 in comparing the 14 other structures in the ensemble to the reference.

Main Stem. The bottom part of Figure 4 offers details of most of the interactions in the stem. Residues G2|G3|U4 are stacked along the same strand (red), with a discontinuity in the stacking between U4-G5. G2 is particularly disordered among the stem base pairs. The two sides of the main stem show only weak cross-strand stacking interactions: G7|G11 and G5|G13, with an additional G2|A16 interaction at the stem's terminus (see Figure 5). Thus, it should be easy to unzip the hydrogen bonds and denature the stem. This provides a natural explanation for the broadening observed for G2, G3, and U4 resonances in both D_2O and H_2O spectra. Motions that are intermediate on the chemical shift time scale would broaden carbon-bound protons and broaden the N-bound protons by exchange with solvent upon transient opening of the stem.

Instability of the stem may be significant for efficient translation of the *gag* genes beginning at the AUG start codon (see Figure 1a; G1 of the target 16mer is G338 of the HIV-1 5′-leader). The G/A-rich region, just upstream from SL4,

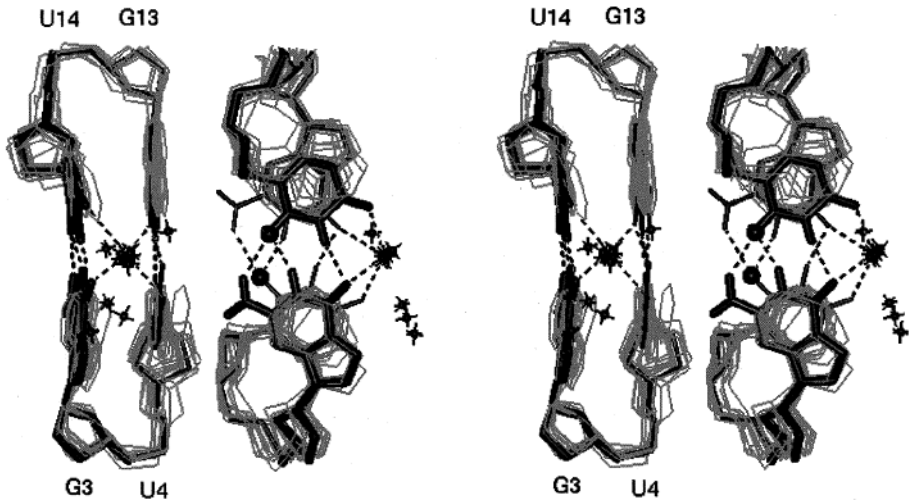

FIGURE 6: Details of the G·U pairs. Side and stacked views related by 90° rotation about *y* (G3·U14 are shown in heavier lines for the reference structure). The U,O2 carbonyl oxygens are rendered as balls. Dashed lines indicate H-bonds detected by InsightII. Hexahydrated sodium ions, rendered as spiked balls, were attracted to the major groove in the AMBER simulations. Average rmsd = 0.7 Å from the reference structure for residues 3−4 and 13−14.

includes likely Shine−Dalgarno sequences for ribosome entry; the secondary structure proposed in Figure 1a suggests that this region can be opened easily, as well.

A five-base stack exists in the stem: C6|G5|G13|U14|C15 (blue in Figure 4). Others have recognized such extended stacks involving tandem G·U pairs (*52*), and we have specified sequence contexts that can lead to patterns of four-base and three-base extended stacks (*53*). However, to our knowledge it is unclear whether extended stacking arrays serve real functions or are simply accidents of geometry. Perhaps they rigidify the bases involved and reduce thermal fluctuations in the associated sugar and backbone atoms. This helps to explain the persistence of G13,1, and U14,3 in the H_2O spectra at higher temperatures even though their pairing partners broaden first. It also explains the absence of intermediate rate broadening for most of the residues in the long stack. Intermediate motional rates for U4 could affect chemical shifts at G5, even though the G5·C12 base pair is one of the most stable as evidenced by the persistence of the G5,1 signal after most of the others have broadened.

Terminal Residues. A common mismatch in RNA structures involves G opposite A. There is strong evidence for hydrogen bonding in G·A pairs embedded in a helix (*54*−*58*). Coaxial stacking between helices may also be promoted by a G·A mismatch at one end of a helix segment (*59*−*61*). It is also clear that 3′-dangling A residues provide substantial stabilization of RNA and DNA helices (*62*, *63*).

The G1·A16 opposition is present in the wild-type SL4 secondary structure (Figure 1a), so it was included at the termini of our construct. Our results indicate that G·A is unstable at the end of a helix when it is not sandwiched between two helical domains, at least in the present sequence context. Figure 5 shows that G1 is not in position to hydrogen bond with A16. Although G1 appears to stack on A16 in the AMBER structures, it is highly disordered in the DYANA structures. Figure 5 shows that the A16 base stacks upon both G2 and C15 and may be the source of a substantial buckle in the G2·C15 pair. The buckled structure is driven largely by nine NOEs that deviate by 0.2−0.6 Å from corresponding distances in an A-form stem. The G2|A16 stack probably provides some additional stabilization, but the tandem G·U pairs at the next step weaken this part of the stem. This allows helix opening to occur fast enough to broaden G2,1, G3,1, and U4,1.

Other indications of mobility at the termini are present in the spectra. In H_2O the G1,1 signal was impossible to securely assign because it is broad and lacks NOEs to other protons. G2,1 is also broad, but NOEs to neighboring protons are observed. The fact that G2,1 is also broad indicates that the G2·C15 pair opens frequently and is not protected by a reasonably stable G1·A16 pair. In D_2O spectra the {G1,1′-G1,8} NOE is 2−3 times more intense than for (*i* to *i*) NOEs between similar protons; this suggests a substantial population of syn conformations. Thus, no restraint was placed upon χ for G1. The {G1,1′-G2,8} sequential connectivity is weak, although G1,2′ and G1,3′ are near enough to G2,8 to produce strong and medium cross-peaks, respectively. All of the nonexchangeable proton signals from G1 are sharp, suggesting fast conformational averaging. The F-protons for G1, G2, and A16 show TOCSY correlations to H1′. These sugars undergo rapid pseudorotational jumps that include S-puckered states, another indication of flexibility.

Tandem G·U Pairs. Westhof (*23*), Gutell (*64*), and coworkers have described the three tandem G·U motifs, and they have been the subject of several structural investigations by NMR and X-ray crystallography (*52*, *65*−*72*). The type II motif occurring in the SL4 16mer rarely occurs in ribosomal RNA. Evidence also favors it as a metal ion binding site that assists in the self-cleavage of the group I intron of *Thermus thermophila* (*68*, *70*, *73*). The Tinoco laboratory has demonstrated the capacity of each of the G·U tandems to bind metal ions in the major groove in solution, using cobalt hexammine, Mn^{2+}, and Mg^{2+} as the probes (*67*, *68*). These studies concluded that ion binding has little effect, if any, on the structure of the G·U tandems. Type II tandems always exhibit strong stacking of the six-membered rings of G and U. The stability of type II mismatches depends on the sequence context, with the symmetric arrangement, 5′-C-G-U-G-3′, being particularly unstable (*65*, *74*).

Motif II tandems are characterized by a large helix twist angle, Ω, between the G·U pairs, flanked by smaller twists (*23*). We observe mean values of $\Omega = 42°$ between G3·

Table 2: Distances between Donor and Acceptor Atoms in the G·U Tandem

donor	acceptor	distance[a]
U4,O2	G13,N1	**3.0**
U4,O2	G13,N2	**3.3**[b]
G13,O6	U4,N3	**3.3**[b]
U14,O2	G3,N1	**3.0**
U14,O2	G3,N2	**3.3**[b]
G3,O6	U14,N3	3.4[b]

[a] Standard deviation <0.25 Å unless noted; distances <3.3 Å are in bold. [b] Standard deviation >0.25 Å.

U14|U4·G13, 19° for U4·G13|G5:C12, and 25° for G2:C15|G3·U14. The average of these is 29°, close to the 30−33° of canonical A- or A′-RNA (*75*). Thus, these four base pairs could fit in an A-family stem without large distortion. The tandem G·U formation is located toward the lower left of Figure 4, showing it in juxtaposition to its neighbors; Figure 6 presents detailed views. The alternating small and large twists cause the G2|G3|U4 and G5|G13|U14|C15 stacking patterns discussed earlier. At one end of the tandem (top in Figure 4) there is a discontinuity in the stacking between U4 and neighboring G5, with the latter involved in a cross-strand stack with G13. At the other end of the tandem, there is weak, same-strand stacking of G2|G3 and U14|C15.

Several possibilities for G·U pairing have been presented which can allow these pairs to fit in a normal RNA or DNA helix, albeit with local distortions in the backbone conformation. Tandem G·U pairs may include standard wobble pairing: G,O6···U,H3 and G,H1···U,O2 (*52*, *67*, *68*, *70*, *71*). In low-stability sequences (*74*), there is a bifurcated structure where U,O2 is shared between G,H2 and G,H1, but the distance between G,O6 and U,H3 is too long to make a hydrogen bond (*65*). The wobble and bifurcated G·U pairs are quite similar in appearance with *cis* glycosyl bonds (both N-C1′ vectors pointing toward the minor groove). Allain and Varani have described a *trans* bifurcated G·U pair at the base of the UUCG tetraloop, where the U,O4 carbonyl oxygen H-bonds to the G-amino and -imino hydrogens (*76*).

Details of the G·U tandem in the SL4 16mer are displayed in Figure 6. Strong same-strand stacking of the six-membered rings is observed. A combination of wobble and bifurcated hydrogen bonding is present in nearly all of the refined structures. U,O2 (rendered as a ball in the stacked view) shares one lone pair of electrons with the G-amino-H and another with the G-imino-H, as in the bifurcated structure (*65*). In common with wobble structures, G,O6···U,H3 H-bonds are present. The G3·U14 pair is buckled slightly (Figure 6, left), perhaps to accommodate the pronounced buckle of the adjacent G2·C15 pair in the structure.

Table 2 summarizes the average heavy atom donor−acceptor distances for the G·U hydrogen bonds. Although there are three H-bonds in most of the structures, all have suboptimal angles (average ∼150°), and the outer two are rather long. Thus, each of these hydrogen bonds must contribute less to stability than those in normal Watson−Crick pairs.

Our AMBER refinement protocol includes tethering hexahydrated sodium ions to the phosphates to simulate solvent dielectric effects in a computationally efficient manner (*33*, *34*). The AMBER force field includes a linearly increasing penalty for large violations of restraints. This prevents one or two large violations from dominating the potential energy. An interesting result of this is that the sodium ions sometimes "lose" the phosphates to which they are tethered and migrate to regions of higher electronegative density. This occurred for 13 of the 15 structures shown in Figure 6. The hydrated ions were within H-bonding distance to all four carbonyl oxygens in the major groove, G3/13,O6 and U4/14,O4, in about half of the structures. In most of the other cases, the ions were localized in the plane of the guanine rings, near locations that would enable them to H-bond to G,N7 and G,O6. The studies with cobalt hexaammine placed the ion such that the cobalt lies about 1.5 Å further from the O6 and O4 sites than the sodium in the present report, consistent with the larger size of the liganded cobalt.

Wobble pairs can be generated from our structure by displacing each uracil base ∼1 Å toward the major groove (toward the ions in the figure). The bifurcated pattern can be generated by pivoting each base stack about its U,O2 to open the tandem toward the major groove. This displaces G,O6 from U,H3 by an additional ∼1 Å and disrupts the O···H−N bond. We tested the likelihood of these modifications by introducing additional restraints and performing 10 ps AMBER MD simulations on the 15 final DYANA structures (see Experimental Procedures). The AMBER energies increased slightly but not by a statistically significant amount, so this is not a sufficient basis to choose between the alternatives.

It may be difficult to resolve issues regarding these small displacements given the current state of the art in NMR and X-ray diffraction technology. It is possible that the disagreements result from the variety of data collection and refinement protocols used in each of our laboratories. It is also likely that the details of this type II G·U tandem vary with its sequence context. All of the structures in this comparison have the G·U tandem embedded between G·C pairs. The Colmenarejo et al. construct (*68*) had a high stability 5′-G-G-U-C-3′ symmetric sequence, while the Chen et al. study (*65*) involved the low-stability C-G-U-G symmetric tandem sequence. Our 5′-G-G-U-G-3′ sequence is not symmetric and mixes the neighbor properties of the low- and high-stability tandems; it is possible the H-bonding properties should also be mixed. Due to the low stability of the SL4 16mer, it may be among the low-stability sequences for type II G·U tandems. It is worth noting that a G·U tandem at the center of a symmetric duplex loses important NOE contacts due to magnetic equivalence. For instance, the counterparts of {G3,1-G13,1} and {U4,3-U14,3} are not observable in a symmetric duplex, the intraresidual {G3,x-G3,y} are not distinguishable from the corresponding{G13,x-G13,y}, and likewise for the {U4,x-U4,y}/{U14,x-U14,y} (here x and y denote the various base and sugar protons within a single residue). The natural lack of equivalence in hairpin constructs generally gives rise to many more usable NOEs than in symmetric duplexes. Finally, the differences in base-pairing details between the structures being compared are at about the level one might expect from thermal fluctuations. It is reasonable that each of the three alternatives represents a substantially populated metastable state. Perhaps type II G·U tandems introduce extra flexibility to helical segments, a property that is selected against in the relatively rigid scaffolding of ribosomal RNA.

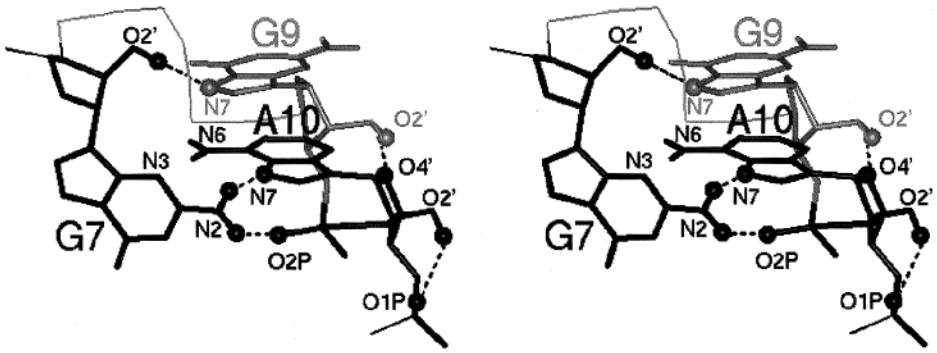

FIGURE 7: Hydrogen-bonding network in the SL4 tetraloop. Dashed lines indicate H-bonds detected by InsightII in several of the 15 structures; donor and acceptor atoms are rendered as balls. This view of G7, G9, and A10 from beneath was created by a nearly 30° rotation about *x* from Figure 4 (top). G9 is shown in lighter outline. Average rmsd = 0.5 Å from the reference structure for residues 7, 9, and 10. Not all heavy atoms are shown.

GAGA Tetraloop. GNRA tetraloops are recognized as important structural elements of ribosomal and other RNAs (*25*−*29*, *69*, *77*−*86*). Pardi and co-workers first elucidated their structures, including the GAGA, GAAA, and GCAA sequences (*24*, *87*, *88*). The GNRA loop is also a metal ion binding site (*85*, *89*). Schmitz et al. recently refined structures with Mg^{2+} bound to a GGAA tetraloop (*84*, *90*). Mg^{2+} ion binding does not appear to exert a large effect on the loop structure.

These loops allow a tight hairpin turn in the RNA chain by means of a sheared G·A base pair and other hydrogen bonds involving the residues corresponding to G7, G9, and A10 in the SL4 16mer. Stacking interactions also stabilize the loop, especially A8|G9|A10 (see Figure 4, top). These are the primary distinguishing features of GNRA loops and are also present in the structures from the Pardi laboratory (*24*). Their GAGA loop corresponds exactly to our sequence, C6 through G11 (our sequence numbers are those in 1ZIG + 2, where 1ZIG is the Protein Data Bank accession number). The internal agreement among the loop structures in Figure 4 is excellent (average rmsd = 0.6 Å from the reference structure for residues 6−11). There is a strong visual similarity of the structures to 1ZIG, as well. However, least-squares superpositions have rmsd ≅ 3.8 Å, implying there are differences in details of the structural solutions.

Chemical shift is probably the most sensitive NMR parameter to changes in local conformation. The chemical shifts for the loop residues are nearly the same as those reported earlier for 1ZIG (*24*). The average magnitude of the deviations is only 0.03 ppm for all of these residues, with the largest (∼0.1 ppm) associated with C6 (H4′, H5′) and G7 (H4′) for which our assignments are ambiguous. The close agreement indicates that the conformations are identical or very nearly so.

Details of the G7·A10 pair in the SL4 16mer are shown in Figure 7. The G·A pair has a strong propeller twist, whereas it is twisted, but more nearly coplanar in 1ZIG. A second difference between our structures and 1ZIG concerns the location of A8. This residue occupies the "N" in GNRA; any base can occur at this position, and it can reside in many orientations and still stack upon G9. Thus, differences in its location are of little consequence. The final difference is that our structures show stacking of G7|G11, while others suggest that the stacking at the base of the loop is G7|C6 and A10|G11 (*24*, *86*). None of these stacking overlaps is particularly strong, so these also are not major differences.

Table 3: Distances between Donor and Acceptor Atoms in the GAGA Tetraloop

		mean distance[a]			
donor	acceptor	SL4	GAGA	GCAA	GAAA
G7,O2′	G9,N7[c]	**3.0**	**2.9**	**2.9**	**3.0**[b]
G7,N2	A10,OP	**3.0**	**2.5**	**2.4**	**2.7**
A10,O2′	G11,OP	**3.0**	4.1[b]	4.8	4.3
G7,N2	A10,N7	**3.1**	**2.5**	**2.6**	**3.2**[b]
G9,O2′	A10,O4′	**3.2**	**2.6**	**3.3**	4.0[b]
G7,N1	A10,OP	3.9[b]	3.5	**3.2**	3.9[b]
G7,O2′	G9,O6[d]	4.4[b]	**2.7**	**2.4**	3.6[b]
A10,N6	G7,N3	5.0[b]	**3.2**[b]	4.1[b]	4.3[b]
A10,N6	G7,O2′	6.4[b]	**2.7**[b]	3.7	**3.2**[b]

[a] Standard deviation <0.25 Å unless noted; distances <3.3 Å are in bold. [b] Standard deviation >0.25 Å. [c] A9,N7 for GCAA and GAAA. [d] A9,N6 for GCAA and GAAA.

We see evidence for intermediate exchange broadening at G7,8, so this base is more mobile than its neighbors.

Another contrast between the Jucker and Pardi structures (*24*) and ours involves the network of hydrogen bonds within the tetraloop. Figure 7 and Table 3 present the details. We observe four H-bonds in nearly all of our structures, corresponding to the first four entries in the table; the fifth is present in about half of the structures in the ensemble. Four of the five, G7,O2′-H···G9,N7, G7,N2-H···A10,OP (*pro-R* phosphate oxygen), G7,N2-H···A10,N7, and G9,O2′-H···A10,O4′, are also short in 1ZIG [the underlined H-bonded pairs have been described (*24*)]. An interesting addition from our structure is A10,O2-H···G11,OP (*pro-S* oxygen). This H-bond, seen at the lower right of Figure 7, sends the backbone chain (thin line) toward the left in the figure. This turn allows G7|G11 stacking and forces G11,-H1′ to lie directly beneath A10, resulting in this proton's large upfield shift.

We see no evidence for the other, previously suggested, H-bonds listed at the bottom of Table 3. Prominent among these is A10,N6-H···G7,N3, although it is still possible for it to exist as a water-mediated H-bond (*24*). Introducing a restraint to hold the A10,N6-G7,N3 distance to 3.0 Å increases the AMBER energies by a statistically insignificant amount, so this provides no basis to eliminate the possibility that the atoms are H-bonded. The lower four entries in the table are already too long for a normal H-bond in at least one of the Jucker and Pardi GNRA structures. In this domain rich in donors and acceptors, H-bonds may simply rearrange in response to changes in the sequence in the stem or loop

(*24*). This possibility is supported by thermodynamic studies with modified nucleotides that prevent certain of the H-bonds from existing (*91*). Most substitutions resulted in only very small changes in stability. The exception is that loss of G7,-N2-H···A10,N7 clearly destabilizes the loop (*24*).

Structural Refinement Protocols. We have noted differences between our results and those of others for the G·U tandem (*65*, *68*) and the GAGA tetraloop (*24*, *86*). The present work has a higher density of restraints than these studies, more than double in some cases. Many of our distance bounds have been corrected for spin diffusion, providing more restrictive limits. However, flexibility also causes asymmetric averaging of NOEs, weighting them toward distances that are shorter than the time average. Therefore, we used larger target ranges for RANDMARDI distances than in previous structures determined in our laboratory (*11*, *34*). We repeated the DYANA calculations, setting all lower distance ranges to 1.8 Å; no significant changes were observed in the overall structure or the local structures for the tetraloop, G·U mismatches, or the four terminal residues. An interesting issue is whether the variation among the structures may arise from the different refinement protocols and programs used in each laboratory; in addition, the quoted studies incorporate distances involving exchangeable protons and restraints on the sugar conformation in different ways. Also, NMR technology keeps improving so older reports may need revision.

There are important issues that confront NMR spectroscopists in producing and comparing structural models for nucleic acids. These include representing molecular averages using structural ensembles, least-squares superposition of flexible and underdetermined conformations, the density of NMR and holonomic (covalent bonds, angles, ring closure) constraints, and the effects of different structural refinement protocols. We will now briefly consider the latter and postpone a more comprehensive discussion (Kerwood and Borer, in preparation).

We have mentioned previously that the overall rmsd = 0.6 Å for the heavy atoms of residues 6−11, using a reference structure for the least-squares superposition; let us call this reference structure, R1. However, 1ZIG fits R1 with rmsd = 3.8 Å for the same residues. When we used the Jucker and Pardi structural restraints for residues 5−12 in our DYANA-AMBER protocol to produce 20 structures and identified a new reference structure, R2, the rmsd = 0.9 Å to the other 19. These structures are very similar to those reported in Figures 4 and 7, with rmsd = 1.2 Å for R2 to R1 and 3.8 Å for 1ZIG to R2. Thus, the major difference between our tetraloop structures and 1ZIG can be attributed to the refinement protocol.

Further work is required to ascertain the effects of differing protocols on NMR-based RNA structures. Three items may be worth mentioning here. First, we reported above the use of restraints to force a hydrogen bond in the loop or destroy others in the G·U tandem. The result was a statistically insignificant change in the AMBER energies. This suggests that the surface of NMR restraint and molecular potential energies is rather flat near the structure that fits the restraints. Thus, details of RNA structures at the 1−2 Å level can be strongly influenced by the details of the refinement protocol. Second, there is no universal consensus on introducing nonexperimental constraints to enforce A-like stems and flat base pairs. This is especially a problem when unusual stems and loops are under consideration. Finally, we emphasize that most of the variance in atomic positions is associated with the backbone atoms, which are the most dynamic and ill-determined features of RNA structures. The focus belongs upon the base and sugar rings, whose positions are defined by the greatest density of NMR and holonomic constraints. With this focus, all of the GNRA loop and tandem G·U structures are more similar than different.

Relationship of the SL4 Structure to HIV-1 Packaging. Recent work in our laboratory has used the quenching of tryptophan fluorescence to measure K_d (∼400 nM in 0.2 M NaCl) for NCp7 and the SL4 16mer; this is about 10% of the affinity we determined for the SL3 20mer (Shubsda, Paoletti, Hudson, and Borer, unpublished). Clever et al. (*2*) measured similar affinity for a 34mer SL4 binding to the NC precursor, p15, fused at its C-terminus to bacterial glutathione *S*-transferase (∼200 nM at ∼0.1 M ionic strength). It is likely that SL3 is the primary recognition site in the 5′-leader (*92*) but that NC domains from several *gag* precursors combine to bind other loops to distinguish full-length HIV-1 RNA from spliced subsets and cellular mRNA.

Our results suggest a relationship of the structure of SL4 to its functional role in the major packaging domain of HIV-1. None of the likely sites for interaction with the nucleocapsid domain play a role in intraloop H-bonds. This includes donor and acceptor sites on G7, A8, and G9 that comprise a favored G-N-G site for interaction with NC (*10*, *12*). Furthermore, only rather weak H-bonds need to be disrupted for interaction to occur at G9-A10-G11 or G3-U4-G5. We have demonstrated the low stability of the stem, so the energetic cost will be low for one or more NC domains to bind these bases. SL4 may also be a critical element in establishing a 3D fold for the 5′-leader of HIV-1 RNA. It has been shown that GNRA tetraloops are favored sites for loop−receptor interactions (*25*−*30*).

NOTE ADDED IN PROOF

The laboratory of Summers has studied the interaction of NCp7 with 14mer and 22mer SL4 constructs (*93*). Low enthalpies of binding were observed by isothermal titration calorimetry, and no single complex was observable by NMR at ∼1 mM RNA and NCp7.

ACKNOWLEDGMENT

We thank the SUNY College of Environmental Science and Forestry, where some spectra were collected at their 600 MHz NMR facility. Sean Studer and Lucia Pappalardo conducted studies on the feasibility of the 16mer construct as a structural target. We thank Profs. Michael Summers, University of Maryland, Baltimore County, and James Dabrowiak, Syracuse University, for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Two figures illustrating regions from the 2D NOE spectrum and two tables giving the ^{1}H and ^{31}P assignments and the distance and angle restraints. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997) *Retroviruses*, Cold Spring Harbor Laboratory Press, Plainview, NY.

2. Clever, J., Sassetti, C., and Parslow, T. G. (1995) *J. Virol. 69*, 2101−2109.
3. McBride, M. S., and Panganiban, A. T. (1996) *J. Virol. 70*, 2963−2973.
4. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) *Struct. Fold. Des. 7*, 1439−1449.
5. Mujeeb, A., Clever, J. L., Billeci, T. M., James, T. L., and Parslow, T. G. (1998) *Nat. Struct. Biol. 5*, 432−436.
6. Mujeeb, A., Parslow, T. G., Zarrinpar, A., Das, C., and James, T. L. (1999) *FEBS Lett. 458*, 387−392.
7. Takahashi, K., Baba, S., Hayashi, Y., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) *J. Biochem. (Tokyo) 127*, 681−686.
8. Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J., and Lancelot, G. (1999) *J. Biomol. Struct. Dyn. 16*, 1145−1157.
9. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) *J. Mol. Biol. 299*, 145−156.
10. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., Chancellor, K. J., Wu, Z. R., and Summers, M. F. (2000) *J. Mol. Biol. 301*, 491−511.
11. Pappalardo, L., Kerwood, D. J., Pelczer, I., and Borer, P. N. (1998) *J. Mol. Biol. 282*, 801−818.
12. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−388.
13. Zeffman, A., Hassard, S., Varani, G., and Lever, A. (2000) *J. Mol. Biol. 297*, 877−893.
14. Hayashi, T., Shioda, T., Iwakura, Y., and Shibuta, H. (1992) *Virology 188*, 590−599.
15. Hayashi, T., Ueno, Y., and Okamoto, T. (1993) *FEBS Lett. 327*, 213−218.
16. Clever, J. L., and Parslow, T. G. (1997) *J. Virol. 71*, 3407−3414.
17. Clever, J. L., Eckstein, D. A., and Parslow, T. G. (1999) *J. Virol. 73*, 101−109.
18. McBride, M. S., and Panganiban, A. T. (1997) *J. Virol. 71*, 2050−2058.
19. McBride, M. S., Schwartz, M. D., and Panganiban, A. T. (1997) *J. Virol. 71*, 4544−4554.
20. South, T. L., and Summers, M. F. (1993) *Protein Sci. 2*, 3−19.
21. Morellet, N., Demene, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournie-Zaluski, M. C., and Roques, B. P. (1998) *J. Mol. Biol. 283*, 419−434.
22. Vogt, V. M., and Simon, M. N. (1999) *J. Virol. 73*, 7050−7055.
23. Masquida, B., and Westhof, E. (2000) *RNA 6*, 9−15.
24. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
25. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature 372*, 111−113.
26. Hedenstierna, K. O., Siefert, J. L., Fox, G. E., and Murgola, E. J. (2000) *Biochimie 82*, 221−227.
27. Jagath, J. R., Matassova, N. B., de Leeuw, E., Warnecke, J. M., Lentzen, G., Rodnina, M. V., Luirink, J., and Wintermeyer, W. (2001) *RNA 7*, 293−301.
28. Abramovitz, D. L., and Pyle, A. M. (1997) *J. Mol. Biol. 266*, 493−506.
29. Jaeger, L., Michel, F., and Westhof, E. (1994) *J. Mol. Biol. 236*, 1271−1276.
30. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1696−1699.
31. Scaringe, S. A. (2000) *Methods Enzymol. 317*, 3−18.
32. Skalicky, J. J., Sukumaran, D. K., Mills, J. L., and Szyperski, T. (2000) *J. Am. Chem. Soc. 122*, 3230−3231.
33. Borer, P. N., Pappalardo, L., Kerwood, D. J., and Pelczer, I. (1997) *Adv. Biophys. Chem.*, 173−216.
34. Kerwood, D. J., and Borer, P. N. (1996) *Magn. Reson. Chem. 34*, S136−S146.
35. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) *J. Biomol. NMR 6*, 390−402.
36. Borgias, B. A., Gochin, M., Kerwood, D. J., and James, T. L. (1990) *Prog. NMR Spectrosc. 22*, 83−100.
37. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−298.
38. Pearlman, D., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1991) *AMBER*, University of California, San Francisco, CA.
39. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
40. Lavery, R., and Sklenar, H. (2001) *Curves v. 5.1: Helical analysis of irregular nucleic acids, http://www.ibpc.fr/UPR9080/Curonline.html*, Laboratoire de Biochimie Théorique URA 77 CNRS, Paris, France.
41. Mathews, D. H., Burkard, M. E., and Turner, D. H. (1999) *RNAstructure v. 3.5, http://128.151.176.70/RNAstructure.html*, University of Rochester, Rochester, NY.
42. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
43. Feigon, J., Leupin, W., Denny, W. A., and Kearns, D. R. (1983) *Biochemistry 22*, 5943−5951.
44. Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., and Reid, B. R. (1983) *J. Mol. Biol. 171*, 319−336.
45. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
46. Jeong, G. W., Borer, P. N., Wang, S. S., and Levy, G. C. (1993) *J. Magn. Reson. A 103*, 123−134.
47. Borer, P. N., and Levy, G. C. (1994) *Methods Enzymol. 239*, 257−288.
48. Weisz, K., Shafer, R. H., Egan, W., and James, T. L. (1994) *Biochemistry 33*, 354−366.
49. Pikkemaat, J. A., and Altona, C. (1996) *Magn. Reson. Chem. 34*, S33−S39.
50. Lankhorst, P. P., Haasnoot, C. A., Erkelens, C., and Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387−1405.
51. Davies, D. B. (1978) *Prog. NMR Spectrosc. 12*, 135−225.
52. Trikha, J., Filman, D. J., and Hogle, J. M. (1999) *Nucleic Acids Res. 27*, 1728−1739.
53. Bubienko, E., Cruz, P., Thomason, J. F., and Borer, P. N. (1983) *Prog. Nucleic Acid Res. Mol. Biol. 30*, 41−90.
54. SantaLucia, J., Kierzek, R., and Turner, D. H. (1990) *Biochemistry 29*, 8813−8819.
55. SantaLucia, J., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
56. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
57. Wu, M., SantaLucia, J., and Turner, D. H. (1997) *Biochemistry 36*, 4449−4460.
58. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) *Biochemistry 38*, 14214−14223.
59. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., and Zuker, M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
60. Walter, A. E., and Turner, D. H. (1994) *Biochemistry 33*, 12715−12719.
61. Walter, A. E., Wu, M., and Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
62. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. H. (1985) *Biochemistry 24*, 4533−4539.
63. Burkard, M. E., Kierzek, R., and Turner, D. H. (1999) *J. Mol. Biol. 290*, 967−982.
64. Gautheret, D., Konings, D., and Gutell, R. R. (1995) *RNA 1*, 807−814.
65. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) *Biochemistry 39*, 8970−8982.
66. McDowell, J. A., He, L., Chen, X., and Turner, D. H. (1997) *Biochemistry 36*, 8030−8038.
67. Kieft, J. S., and Tinoco, I. (1997) *Structure 5*, 713−721.
68. Colmenarejo, G., and Tinoco, I. (1999) *J. Mol. Biol. 290*, 119−135.
69. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
70. Cate, J. H., and Doudna, J. A. (1996) *Structure 4*, 1221−1229.
71. Biswas, R. S. M. (1997) *J. Mol. Biol. 270*, 511−519.

72. Deng, J., and Sundaralingam, M. (2000) *Nucleic Acids Res. 28*, 4376−4381.
73. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nat. Struct. Biol. 4*, 553−558.
74. He, L., Kierzek, R., SantaLucia, J., Walter, A. E., and Turner, D. H. (1991) *Biochemistry 30*, 11124−11132.
75. Arnott, S., Hukins, D. W., Dover, S. D., Fuller, W., and Hodgson, A. R. (1973) *J. Mol. Biol. 81*, 107−122.
76. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
77. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467−8471.
78. Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994) *Nat. Struct. Biol. 1*, 273−280.
79. Murphy, F. L., and Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.
80. Dieckmann, T., Suzuki, E., Nakamura, G. K., and Feigon, J. (1996) *RNA 2*, 628−640.
81. Jiang, F., Kumar, R. A., Jones, R. A., and Patel, D. J. (1996) *Nature 382*, 183−186.
82. Cai, Z., Gorin, A., Frederick, R., Ye, X., Hu, W., Majumdar, A., Kettani, A., and Patel, D. J. (1998) *Nat. Struct. Biol. 5*, 203−212.
83. Perbandt, M., Nolte, A., Lorenz, S., Bald, R., Betzel, C., and Erdmann, V. A. (1998) *FEBS Lett. 429*, 211−215.
84. Schmitz, U., Behrens, S., Freymann, D. M., Keenan, R. J., Lukavsky, P., Walter, P., and James, T. L. (1999) *RNA 5*, 1419−1429.
85. Schmitz, U., Freymann, D. M., James, T. L., Keenan, R. J., Vinayak, R., and Walter, P. (1996) *RNA 2*, 1213−1227.
86. Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y., and Nishikawa, S. (1993) *Nucleic Acids Res. 21*, 5670−5678.
87. Heus, H. A., and Pardi, A. (1991) *Science 253*, 191−194.
88. Jucker, F. M., and Pardi, A. (1995) *RNA 1*, 219−222.
89. Greenbaum, N. L., Mundoma, C., and Peterman, D. R. (2001) *Biochemistry 40*, 1124−1134.
90. Schmitz, U., James, T. L., Lukavsky, P., and Walter, P. (1999) *Nat. Struct. Biol. 6*, 634−638.
91. SantaLucia, J., Kierzek, R., and Turner, D. H. (1992) *Science 256*, 217−219.
92. Damgaard, C. K., Dyhr-Mikkelsen, H., and Kjems, J. (1998) *Nucleic Acids Res. 26*, 3667−3676.
93. Amarasinghe, G. K., Zhou, J., Miskimon, M., Chancellor, K. J., McDonald, J. A., Matthews, A. G., Miller, R. R., Rouse, M. D., and Summers, M. F. (2001) *J. Mol. Biol.* (in press).

BI0111909