Stem of SL1 RNA in HIV-1: Structure and Nucleocapsid Protein Binding for a 1 × 3 Internal Loop[†,‡]

YiQiong Yuan, Deborah J. Kerwood, Andrew C. Paoletti, Michael F. Shubsda,[§] and Philip N. Borer*

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

Received January 15, 2003; Revised Manuscript Received March 12, 2003

ABSTRACT: The 5′-leader of HIV-1 RNA controls many viral functions. Nucleocapsid (NC) domains of gag-precursor proteins select genomic RNA for packaging by binding several sites in the leader. One is likely to be a stem defect in SL1 that can adopt either a 1 × 3 internal loop, SL1i (including G247, A271, G272, G273) or a 1 × 1 internal loop (G247 × G273) near a two-base bulge (A269−G270). It is likely that these two conformations are both present and exchange readily. A 23mer RNA construct described here models SL1i and cannot slip into the alternate form. It forms a 1:1 complex with NCp7, which interacts most strongly at G247 and G272 (K_d = 140 nM). This demonstrates that a linear G-X-G sequence is unnecessary for high-affinity binding. The NMR-based structure shows an easily broken G247:A271 base pair. G247 stacks on both of its immediate neighbors and A271 on its 5′-neighbor; G272 and G273 are partially ordered. A bend in the helix axis between the SL1 stems on either side of the internal loop is probable. An important step in maturation of the virus is the transition from an apical loop−loop interaction to a dimer involving intermolecular interactions along the full length of SL1. A bend in the stem may be important in relieving strain and ensuring that the strands do not become entangled during the transition. A stem defect with special affinity for NCp7 may accelerate the rate of the dimer transformation. This complex could become an important target for anti-HIV drug development, where a drug could exert its action near a high-energy intermediate on the pathway for maturation of the dimer.

The 5′-leader sequence of HIV-1 has been the subject of intensive study to understand its importance in different stages of the viral life cycle, including dimerization, encapsidation, and reverse transcription (*1*). The major packaging domain (nt ∼240−360) within this sequence is proposed to have four stem loops termed SL1 to SL4[1], and these four loops are highly conserved (*2*, *3*). Among the four loops, SL3 provides the majority of specificity for packaging the genome into virions by its interaction with the 55-residue nucleocapsid domain (NC) of the gag and gag-pol precursor proteins (*2*−*9*). However, NC also has binding affinity to the other three stem loops (*2*, *10*). It is likely that optimal packaging efficiency is rendered through multiple RNA−protein interactions.

SL2, SL3, and SL4 all have an apical tetraloop. It has been shown that the strongest affinity for NC occurs when two G residues are in a GNG sequence in the tetraloops (*10*, *11*). Each G residue interacts with a zinc finger from NC (*12*−*14*). The weaker affinity of SL4 for NC is attributed to the fact that the loop residues of SL4 are ordered in a manner similar to other GNRA tetraloops (*15*), and the G residues are not as readily accessible for NC binding as those in SL2 and SL3 (*16*, *17*).

SL1 has both an apical 9mer loop and a defect in its stem. Both of these regions have unpaired G residues, which we have shown to bind significantly to the mature nucleocapsid protein, NCp7 (*10*). The stem defect is predicted to have two possible secondary structures: an *i*(1 × 3) internal loop or an *i*(1 × 1) − *b*(*2*) form with a 1 × 1 internal loop and a two-base bulge (Figure 1a,b). Free energy predictions indicate that both defects are similarly stable. Here, we constructed various RNAs to model the former structure, as well as alterations within the loop to study their importance in binding to NCp7. The results reveal that two G residues on opposite sides of the loop, G5 and G17 (corresponding to G247 and G272),[1] are most important for tight binding.

[†] Supported in part by NIH Grant GM32691, Grant 02740-30-RGT from the American Foundation for AIDS Research, and Syracuse University.

* Corresponding author. E-mail: pnborer@syr.edu. Tel: (315) 443-5925. Fax: (315) 443-4070.

[‡] An ensemble of 23 structures has been deposited with the Protein Data Bank, Accession Number 1OSW.

[§] Current address: State University of New York, Upstate Medical University, Department of Biochemistry and Molecular Biology, Syracuse, NY 13210.

[1] Abbreviations: HIV-1 sequences are counted from the CAP base as number 1, corresponding to the convention adopted by Coffin et al. (*1*); to obtain the numbering used in some references, one must add 453−456 to account for the proviral U3 stretch that is not present in genomic or mRNA. NCp7, the 55 amino acid nucleocapsid protein of HIV-1, Human Immunodeficiency Virus; SL1, SL2, SL3, and SL4, short stem-loop segments of RNA from the 5′-leader region of HIV-1; DIS, the self-complementary dimer initiation site in the apical loop of SL1; NC, nucleocapsid. 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}). 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. Other abbreviations: NOE, nuclear Overhauser effect; COSY, correlation spectroscopy; HMQC, heteronuclear multiple quantum coherence; rmsd, root-mean-square deviation.

10.1021/bi034084a CCC: $25.00
Published on Web 04/10/2003

While two G residues are required for high affinity binding, a linear GNG sequence is not necessary.

SL1 also plays the key role in dimerizing the viral genome (*18*−*20*). The apical SL1 loop contains a self-complementary dimer initiation site, DIS. However, recent mutation and deletion studies indicated that the four-purine stem defect, the subject of the present work, is also required for dimerizing genomic RNA (*21*). While the apical loops and stem regions of SL2, SL3, and SL4 are highly conserved, the SL1 stem defect is almost completely conserved (Table 1) (*21*−*23*). This suggests that the defect does indeed have important roles in the viral life cycle.

High-resolution structures have been published for RNA constructs to model SL1 (*23*−*31*), SL2 (*12*, *32*), SL3 (*13*, *33*, *34*), and SL4 (*16*, *17*). The published studies on SL1 have focused mainly on the apical DIS loop. Recently Greatorex et al. reported the structure of the defect with two variations in its four-purine sequence (*23*). By introducing the double variation, the internal loop was stabilized, but the structure of the *i*(1 × 3) loop does not correspond to the highly conserved loop sequence. It also does not correspond to the structure we report here for the *i*(1 × 3) loop that has the conserved sequence.

The present 3-D structure was determined for a 23mer RNA construct, SL1i (Figure 1c), which has alterations to the stems on both sides of the loop to prevent it from slipping from the *i*(1 × 3) loop into the alternative stem defect. We find evidence for a local G5:A16 base pair and partial ordering of unpaired G17-G18 in the loop. The loop has considerable flexibility, and the stem appears to make an overall bend around the defect. Both of these factors may have important influences in the biological function of the stem defect. In addition, we report affinities of NCp7 for the 23mer and variants to define the most important interaction sites in the loop and discuss models for generating the mature extended dimer from the loop−loop complex.

EXPERIMENTAL PROCEDURES

RNA Synthesis and Purification. The 23mer SL1i RNA constructs were purchased from Dharmacon (Lafayette, CO); all were synthesized using the 2′-ACE protecting strategy (*35*). NMR samples were prepared from 1 μmol scale syntheses and purified as described previously (*16*), with minor modifications to optimize the HPLC gradient and desalting protocol. RNA samples for the protein binding assay were prepared on the 50 nmol scale and purified using procedures in ref *10*. Samples were disaggregated by heating to 85−95 °C for a few minutes and then cooling quickly to 0 °C prior to NMR, UV, and protein binding assays.

Protein Binding Assays. Aliquots of ∼100 μM RNA solution were added to a 1 mL aliquot of 0.3 μM NCp7 (for NCp7 preparation, see refs *10* and *11*) in the fluorescence buffer (0.2 M NaCl, 5 mM sodium phosphate, pH 7.0, 0.1 mM $ZnCl_2$, and 0.01% poly(ethylene glycol)). After mixing, the fluorescence was measured at 350 nm with a Hitachi F-4010 fluorescence spectrophotometer using a 290 nm excitation wavelength, a 5 nm excitation band-pass, and either a 3 or 5 nm emission band-pass. The dissociation constant K_d was calculated based on the assumption that the RNA−NCp7 complex adopted a 1:1 stoichiometry. The description of the model is detailed previously (*10*, *11*). Fits to all of the titration curves were excellent, suggesting that no alternative stoichiometry is required to explain complex formation.

RNA Secondary Structure Prediction. Structures were predicted using Mfold (*36*, *37*) and energy parameters from the Turner lab (*38*). The presence of alternative folds within 10% of the lowest free energy in the predictions were considered likely to exist in appreciable population in solution.

UV Melting Curves. UV melting experiments were performed on a Beckman DU 640 spectrophotometer. The SL1i 23mer was melted in the NMR buffer (see next section) with a heating rate of 1 °C/min. The melting temperature was calculated using the first derivative of the melting profile.

NMR Data Collection. The NMR sample in D_2O was prepared from the purified SL1i powder, which was reconstituted in 99.9% D_2O and dried using a SpeedVac. After repeating this procedure two times, the sample was dissolved in 700 μL of NMR buffer [25mM NaCl, 5 mM $Na_{1.4}H_{1.6}PO_4$, 0.1 mM EDTA, 0.01% NaN_3 in D_2O (99.996%), measured pH = 6.3]. An additional sample was prepared in the same NMR buffer but in 90% H_2O/10% D_2O adjusted to pH 5.5. Both samples had a strand concentration of about 1 mM. 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. Chemical shifts of protons were reported relative to DSS.

Exchangeable Proton Spectra. The 1-D spectra in water were acquired with a 3−9−19 water peak suppression sequence (*39*) with pulsed field gradients applied along the *Z* direction. The sweep width was 10 500 Hz. 1-D spectra were acquired at temperatures ranging from −9 to 45 °C with a 5 °C temperature step from 0 to 45 °C. For studies below 0 °C, the sample was centrifuged to remove all suspended matter and decanted carefully to a clean NMR tube. The temperature was slowly decreased at a rate of 1 °C per hour from 0 to −9 °C. 2-D NOE experiments were acquired at −6 and 5 °C with a mixing time of 200 ms, 4*k** × 512*k** hypercomplex points, 128 scans per t1 increment, and a 3 s relaxation delay time.

Nonexchangeable Proton Spectra. A number of experiments were done for assignment purposes. These include 2-D NOE at 5, 18, 30, and 40 °C. These were acquired with 4*k** × 512*k** hypercomplex points, 64 scans per t1 increment, 3 s relaxation delay, and 100 or 300 ms mixing times. In addition to the 2-D NOE, natural abundance ^{1}H-^{13}C HSQC, TOCSY, and ^{1}H-^{31}P COSY were also performed as previously described (*16*). The 2-D NOE experiment for quantitative analysis was performed at 30 °C, mixing time of 300 ms, 8 s relaxation delay time with 4*k** × 512*k** points, 64 scans per t1 increment. The resulting spectrum was processed using XWIN NMR (Brüker) and deconvolved as previously described (*16*) using NMRZ and TRIAD software (Tripos, St. Louis, MO).

NMR-Based Restraints. Cross-peaks from the 2-D NOE spectrum at 30 °C and 300 ms mixing time were integrated using NMRZ software and then categorized as previously described (*16*). The intensities from cross-peaks with moderate to no overlap were used as input for MARDIGRAS to calculate target distances corrected for spin diffusion (*40*). These target distances were used in DYANA (*41*) and AMBER (*42*) calculations with the target range based on

the amount of overlap and strength of the cross-peak: no overlap, strong to medium intensity, range was set to 0.3 Å; slight to moderate overlap, strong to medium intensity, range set to 0.5 Å; and moderate overlap or weak peak intensity, range set to 1 Å. For regions of extensive overlap, cross-peaks were grouped into strong (1.8−3 Å), medium (2−4 Å), weak (2.5−5 Å), and very weak (4−6 Å). If such grouping was not possible, a range of 1.8−5 Å was included. A lower limit of 4.5 Å was set if no cross-peak was seen for a certain proton−proton pair. The H5′1 and H5′2 were not stereospecifically assigned. If {*n*,8-*n*,5′1} and {*n*,8-*n*,5′2} had nearly the same intensity, the same distance range was assigned from the base proton to H5′1 and H5′2 (in rare cases this was also applied to other protons having cross-peaks to *n*,H5′1 and *n*,H5′2, where *n* is the same for both). Distance ranges based on the exchangeable protons were given a broad range of 1.8−6 Å, except for intense cross-peaks within a base pair, which were assigned a range of 1.8−4 Å.

Hydrogen bond restraints were applied to the seven base pairs of the stem in which the corresponding imino proton signals were clearly visible in the 1-D H_2O spectrum. In both the DYANA and the AMBER calculations, the distances used were 2.82 ± 0.1 Å for AN1 to UN3, 2.95 ± 0.1 Å for AN6 to UO4 and GN1 to CN3, 2.91 ± 0.1 Å for GO6 to CN4, and 2.86 ± 0.1 Å for GN2 and CO2. A hydrogen bond angle range of 170−190° was included in the AMBER calculations.

A number of dihedral angle restraints were included in both the DYANA and the AMBER calculations. These include α and ζ (excluding the trans conformation if the phosphorus chemical shift was within the main signal envelope) (*43*), *β* and *ϵ* (based on the 3-bond H−P coupling constants in the H−P COSY), and *δ* (based on H−H coupling constants). The procedures used for defining the restraints have been described earlier (*16*, *44*). A list of the distance and angle restraints are provided in the Supporting Information.

Structure Determination. The NMR-based restraints were used as input to the torsion-angle dynamics program, DYANA (*41*); this produced a number of structures (typically 400) ranked in order of their target function values. The DYANA calculations were initiated when about 2/3 of the final number of restraints was known and repeated several times with increasing numbers of restraints. This process enables one to test assignments based on preliminary structures and to accumulate absent NOE restraints that have a meaningful impact on the refinement; absent NOEs were included with lower-bound restraints of 4.5 Å. Fifty of the structures from DYANA that had the lowest target values were further refined in calculations using AMBER 5.0 (*42*). Specifics of both calculations have been described previously (*16*) with the exception that explicit solvent was included in the current AMBER calculations. First, sodium atoms were placed around the phosphates, and then water molecules were added up to a distance of 5 Å from the RNA fragment. The assembly was subjected to 1000 steps of conjugate-gradient energy minimization and then a 10 ps molecular dynamics trajectory at 300 K. The water molecules were removed for the final minimization, which consisted of another 1000 steps minimization. All restraints were kept constant throughout the calculations, with the distance and hydrogen-bond restraints having force constants four times that of the dihedral angle restraints. An ensemble of 24 final structures were chosen to represent the SL1i 23mer; these had the lowest AMBER energies, did not have tangled strands, and had only minor restraint violations.

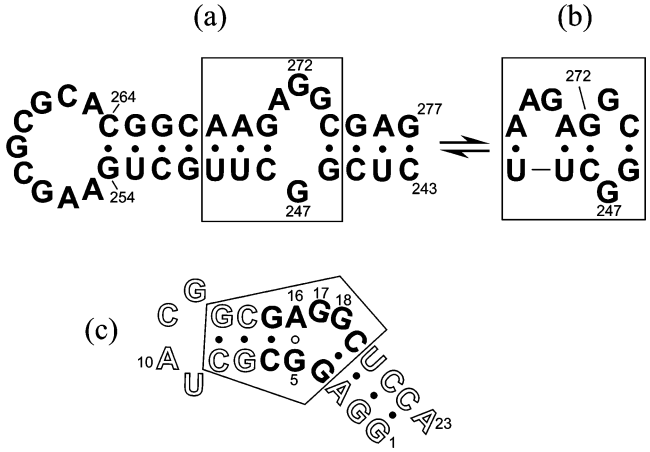

FIGURE 1: (a) Consensus sequence for SL1 in the HIV-1 major packaging domain, noting G247 and G272, which are the two most important loci for interaction with the nucleocapsid protein (see text). The boxed sites are nearly invariant (see Table 1) and can exist in an alternate secondary structure (b) with nearly equivalent stability. A 23mer construct in (c) models the 1 × 3 internal loop in panel a, holding the closing base pairs constant; the non-native sequence is shown in outline font. A G5:A16 base pair is observed (G247:A273), and the stems to the right and left of the internal loop are kinked (see Figure 4).

The structures were displayed using InsightII (Accelerys, San Diego, CA). All structures were superimposed to a reference structure that had the lowest average rmsd to the other structures based on the heavy atoms of the base and sugar rings for residues 1−22. InsightII was used to identify hydrogen bonds in each structure using the default requirements: (1) donor-H−acceptor angle greater than 120°, (2) donor−acceptor distance less that 3.0 Å, and (*3*) H−acceptor less than 2.0 Å. Dihedral angles for each structure were determined using the CARNAL module of AMBER.

RESULTS

SL1i Construction. The wild type SL1 stem defect can have two alternative forms (Figure 1), the asymmetric internal loop (Figure 1a) and the internal loop/bulge form (Figure 1b). According to free energy calculations, these should differ in free energy by only 0.1 kcal/mol (*38*). *Mfold* (*36*) was used to evaluate candidate sequences, concluding that the 23mer SL1i construct (Figure 1c) prevents the *i*(1 × 3) loop from slipping into the alternate form. It is not possible to preserve the base sequence needed to study the *i*(1 × 1) − *b*(2) form; any sequence with AGAGG opposite GCU will slip between the two alternatives. The SL1i construct has several important features: (1) the four loop bases and the two closing Watson−Crick base pairs match the wild-type (indicated in solid font in Figure 1c), (2) the encompassing base pairs are altered to prevent slippage and ensure stable stems on either side of the loop (the 5′-terminal sequence was also chosen for high yield from T7 RNA polymerase preparation in case labeling was needed; however, natural abundance NMR spectra are sufficient to determine the RNA structure), (3) it has a stable apical UACG tetraloop to ensure proper folding (*45*), (4) it has a

Table 1: Sequence Variation in SL1 from HIV-1[a]

stem-5′		DIS loop		stem-3′	
243	0.3%	253	0.3%	266	0.3%
244	0.5%	254	0.5%	267	0.0%
245	0.3%	255	1.9%	**268**	**0.3%**
246	**0.3%**	256	18%	**269**	**1.2%**
247	**0.0%**	257	0.5%	**270**	**0.3%**
248	**0.2%**	258	51%	**271**	**0.2%**
249	**0.3%**	259	0.0%	**272**	**0.2%**
250	**0.0%**	260	0.3%	**273**	**0.2%**
251	0.0%	261	47%	**274**	**0.0%**
252	0.2%	262	0.2%	275	0.3%
		263	28%	276	0.0%
		264	1.1%	277	1.2%
		265	35%		

[a] Mean variation from the consensus (*22*) (average of 627 sequences per site from Genbank). Boldface type indicates bases that are boxed in Figure 1.

Table 2: Affinities of SL1i Constructs for NCp7

247	271−273	K_d (nM)[a]	RA[b]
G	AGG	140	20%
G	AGA	120	23%
G	AAG	270	10%
A	AGG	290	10%
A	AAA	1300	2%
G(wt)[c]	AGG(wt)[c]	210	13%

[a] Dissociation constant measured in 0.2 M NaCl, 25 °C. [b] Relative affinity as compared to SL3 (K_d = 28 nM) (*10*). [c] For the 29mer: GGUCGGCUUGCUACGGCAAGAGGCGACCA. Underlined bases correspond to the boxed regions in Figure 1a,b.

3′-dangling A to stabilize the terminal base pair (*38*, *46*), (5) it also has an alternating G-C-G near the tetraloop to lessen overlap in the NMR spectra. Predictions from *Mfold* indicated that only one stable form exists for the 23mer SL1i construct. The UV melting curve has a single cooperative transition with T_m = 67 °C.

NCp7 Binding Constants. Table 2 shows the RNA constructs used in the NCp7 binding assay and their dissociation constants, K_d, measured by the Trp-37 fluorescence assay. All of the 23mer RNA constructs (the first five in the table) kept the same secondary structure according to *Mfold*. The last construct in Table 2 has a stem corresponding to the wild-type sequence, so it is capable of adopting the two interconverting forms, $i(1 \times 3)$ and $i(1 \times 1) - b$(*2*).

Assignment of Exchangeable Proton Resonances. Figure 2 shows imino proton spectra for the SL1i 23mer at several temperatures. Assignments were obtained from a series of 1-D experiments and 2-D NOE experiments in H_2O. There are 12 G,1 and U,3 protons in SL1i; all but G5,1, G17,1, and G18,1 were identified through correlating cross-peaks in the 2-D NOE spectra. The G5,1, G17,1, and G18,1 in the internal loop appear flexible and in fast exchange with the solvent above 20 °C; they probably are the broad peaks around 11 ppm and the shoulder on the G2,1 peak. No unambiguous cross-peaks could be found to establish their assignments. The dangling A stabilizes the end base pair G1:C22, so G1,1 is resistant to exchange and disappears at about 45 °C. G4,1 is fairly broad at all temperatures, probably because the closing base pair for the internal loop is subject to faster exchange. Residues G2, U20, and G7 are in the stems and have sharp imino proton peaks at the higher temperatures. G12,1 is unusually upfield and

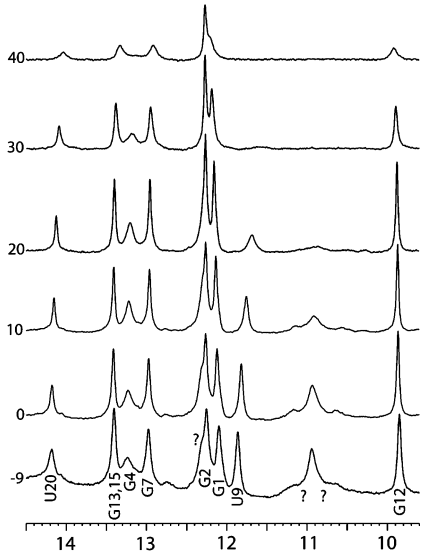

FIGURE 2: Imino proton spectra of the 23mer RNA construct as a function of temperature (°C) at 500 MHz, pH = 5.5. Resonances for G5, G17, G18 are probably located at "?".

sharp, which is a characteristic of the UACG tetraloop (*45*), while U9,3 is broad and disappears quickly as the temperature increases. G13,1 and G15,1 overlap at all temperatures, and at least one of them is quite resistant to exchange broadening.

Nonexchangeable Proton and Phosphorus Assignments. Nonexchangeable protons and phosphorus were assigned following the methods previously described (*16*, *44*), mainly based on classical NOE-walks between sequential residues. 1- and 2-D NOE spectra in D_2O at different temperatures and TOCSY, ^{1}H-^{31}P COSY, as well as ^{1}H-^{13}C HSQC spectra were used for the assignments. The peaks in the 2-D NOE spectrum were sharpened by (−6) Hz deconvolution, which preserves the relationship of peak integrals (*47*, *48*). Figure 3 shows the 1′ to base NOE walk (H1′−H8/H6 connectivities). The NOE walk went from G1 through A23 except for missing peaks in the apical loop at {A10,1′−C11,6}, {C11,1′−G12,8}, and {G12,1′−G13,8}. The very high intensity at {G12,1′−G12,8} indicates that G12 exists mainly in syn rotamers about the glycosyl bond. The H5−H6 cross-peaks have high intensities and splitting patterns ($^3J \sim 8$ Hz) in 2-D NOE and TOCSY spectra. The {*n*5-*n*6}, {*n*,8-(*n* + 1),5}, and several base to base NOEs helped to confirm the assignments. The unusual upfield shift of G13,1′ was confirmed by ^{1}H-^{13}C HSQC. The characteristics of peaks in 1-D spectra also provide support for the assignments, including sharp peaks for H2, doublets for H6, as well as in ascertaining the number of protons in overlapped peak areas by careful integration. Some overlaps were also resolved in 2-D NOE taken at different temperatures.

The assignments of sugar protons (2′,3′,4′,5′,5") were extended from the base (*2*, *5*, *6*, *8*) and 1′ protons using other regions of the 2-D NOE, TOCSY, and ^{1}H-^{31}P COSY spectra. It was not possible to assign all of the H4′; however, their assignments are not necessary for structure determination. H5′ and H5″ were not stereospecifically distinguished.

The splitting of H1′ for A10, C11, G17, and G18 in 2-D NOE, along with the appearance of their strong {1′-2′} cross-peaks in TOCSY, indicate that their sugars are similar to

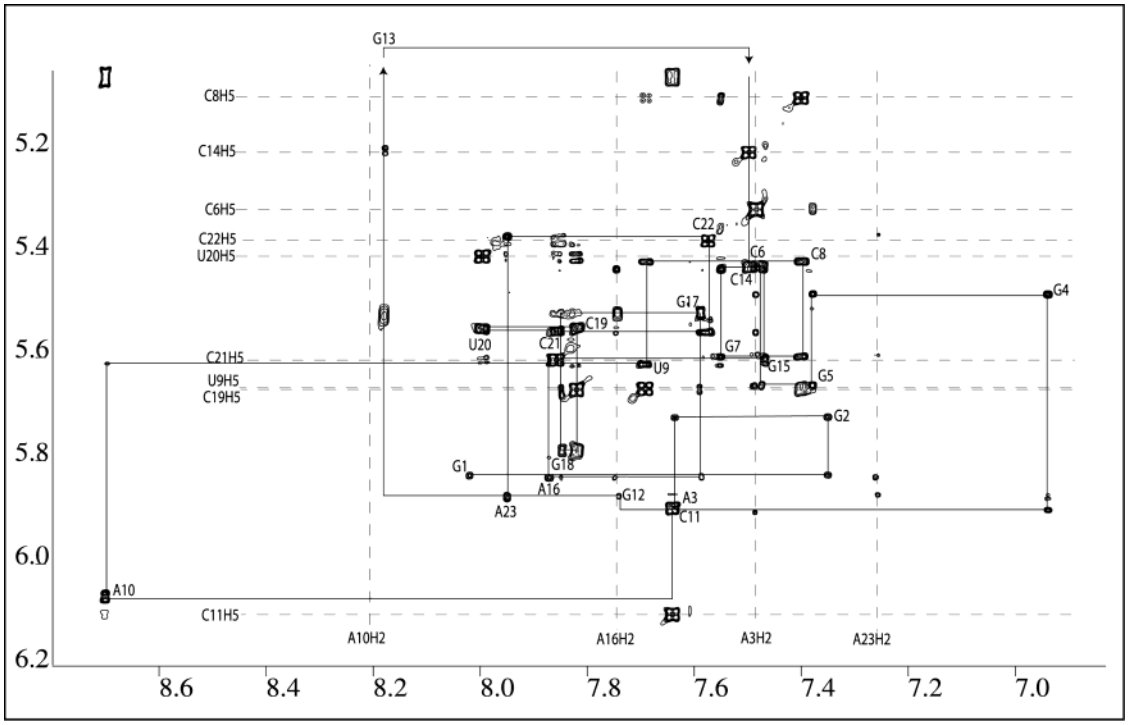

FIGURE 3: Base-1′ proton assignments in the 2-D NOE spectrum of the SL1i 23mer RNA (30 °C, 600 MHz spectrum after deconvolution by −6 Hz). The H6−H5 cross-peaks are doublets with ~8 Hz couplings in each dimension.

those in B-DNA structures (S-pucker). G5, G15, and A16 have weak {1′-2′} cross-peaks in TOCSY and less distinct double peaks in the 2-D NOE; therefore, they have mixed N/S sugar puckers. Cross-peaks involving the H8 and H1′ of G17 and G18 show broadness in 2-D NOE spectra, implying that these residues are relatively flexible.

RNA Structure. The structures are displayed in Figures 4−8, and their details are described in the Discussion. Table 3 shows the numbers in each category of restraints and the statistics for violations. A total of 502 restraints were used in the refinement (mean = 22 per residue), with the number applied to a given residue varying from 24−58 (distribution shown in Supporting Information).

Modeling. The seven base-pair stem nearest the 9mer DIS loop was modeled using an A-helical 7mer duplex, (CGCGCUG):(CGGCGCG). The terminal (CGC):(GCG) segment was superimposed on (6−8):(13−15) from the reference NMR structure of the SL1i 23mer using the sugar ring atoms as the basis for superposition (rmsd = 2.3 Å). A similar procedure was used for the 4 base-pair stem (1−4):(19−22). The CURVES program (*49*) was used to determine the local helix axes for the 7mer and 4mer duplexes.

Table 3: Restraint Statistics for the SL1i 23mer Refinement

NMR-Based Restraints	
residue *i* to *i*	171
residue *i* to *j*, *j* ≠ *i*	208
subtotal	379
H-bond	40
dihedrals	83
total restraints	502
Average Violations per Structure	
distances	0.06 Å
>(C−C)[a]	0
>(C−C)/2	2
>(C−C)/4	12
dihedrals	4°
>60°	0
>30°	2
>15°	7
AMBER Statistics (±SD)	
E_{Total} (kcal/mol)	−2000 ± 100
$E_{Constraint}$ (kcal/mol)	210 ± 10

[a] Standard C−C bond length = 1.537 Å (*52*).

DISCUSSION

Sequence Analysis. The helix defect in the stem of SL1 is important for full-length genome packaging (*21*). It is also critical to form the stable, extended dimer linkage found in mature virus particles (*28*, *29*), and mutational studies have linked this region with viral infectivity (*50*). The stem of SL1 is strongly conserved, with usually less than 1% variation at sites corresponding to the helix imperfection (see Table 1 and Figure 1). Variations below 0.5% are at about the level one might expect from sequencing and database errors. By contrast, the apical DIS loop exhibits a high degree of variation, together with strong covariation at 258 and 261 that preserves base pairing in the dimer. The low degree of variation in the stem defect region is probably correlated with essential biological function. Unfortunately for phylogenetic analysis, this also means that covariation is not observed that could be used to argue for one secondary structure over another. G247 (G5 in the 23mer construct) is absolutely conserved, and A271, G272, and G273 (A16, G17, G18) each vary only once. The most variable position is A269, which changes eight times to G in the database. This would replace the U249:A269 pair with U249:G269 in Figure 1a and would replace A269 with G269 in the two-base bulge of Figure 1b. Neither change would be expected to create a large difference in the relative stability of the two alternative stem defects.

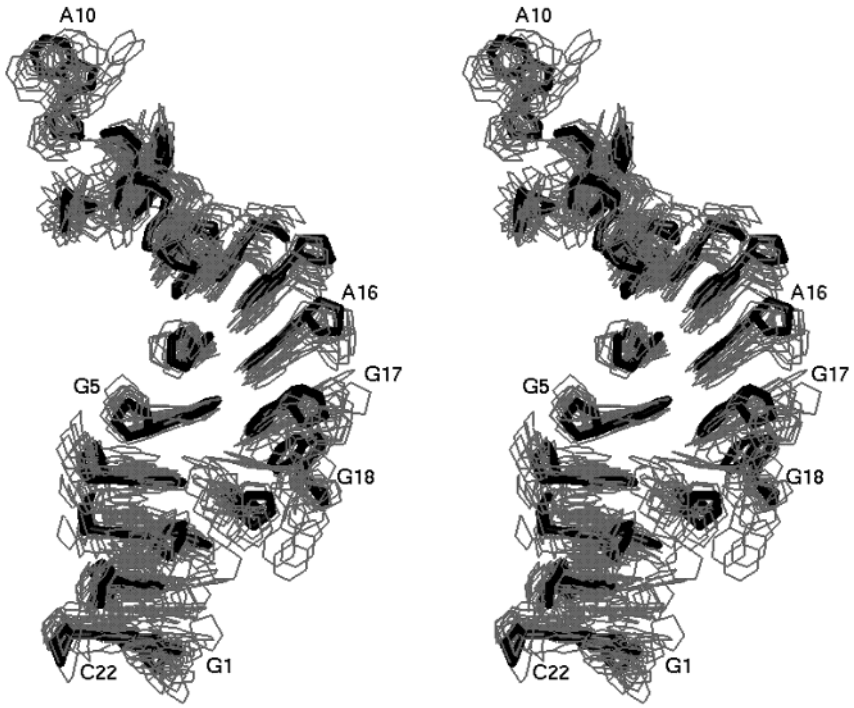

FIGURE 4: Superposition of 24 refined structures of the SL1i 23mer. The reference structure with the lowest rmsd to the others is shown in heavy black lines (relaxed stereoview).

Affinity for NCp7. SL1i models of the *i*(G × AGG) loop were constructed, as well as A × AAA, G × AGA, G × AAG, and A × AGG loops to compare their binding affinities to NCp7 (see Figure 1c and Table 2). In the absence of Mg^{2+}, we have found that it is essential to measure affinities at or above 0.2 M salt to suppress nonspecific binding of RNA to NCp7 (ref *10* and A. C. Paoletti and P. N. Borer, unpublished); this is also near to physiological ionic strength (*51*). Under these conditions, the K_d for *i*(A × AAA) is 1300 nM, which is only 2% of the binding affinity of SL3 to NCp7 (*10*). This demonstrates that Gs in the internal loop are important for binding. K_d for the G × AGG and G × AGA loops are 140 and 120 nM, respectively, indicating that G5 and G17 are required for tightest binding with NCp7 among the SL1i constructs; this G_2 locus has bases located across from each other in the stem instead of in a continuous GNG sequence as in SL3 and SL2. The loops that have two Gs in other positions have a 50% reduction in binding affinity, G × AAG and A × AGG. This is similar to the situation in SL3 loop variants, where two G residues confer some stability to complexes with NCp7, even when they do not contain the optimal G-N-G sequence (*11*). NCp7 binding to a construct that models the wild-type SL1 stem was also studied (the apical 9-residue loop was replaced by UACG). This stem is capable of adopting the two interconverting forms. Its affinity (apparent K_d = 210 nM) is about 70% of the *i*(G × AGG) construct. It is possible that this binding constant represents an average weighted according to the populations of the two species and their intrinsic affinities for NCp7. If K_d = 140 nM represents the affinity for the *i*(1 × 3) form, the *i*(1 × 1) − *b*(*2*) form must also exist in an appreciable amount and must bind less tightly to NCp7.

Choice of an Appropriate RNA Construct. Thermodynamic analysis suggested that the native SL1 stem defect should exist in two conformations in solution, differing by about 0.1 kcal/mol in free energy. Therefore, it is likely that these conformations will exchange rapidly on the NMR time scale, making it impossible to determine a structure for the native stem defect. Greatorex et al. also recognized this issue and introduced a double variation from *i*(G × AGG) to *i*(G × GGA) to stabilize the internal loop against conformational exchange (*23*). Its *i*(***G*** × G***GA***) structure has interesting noncanonical hydrogen bonding involving the italicized residues. However, this sequence never occurs in the HIV-1 database, and this noncanonical structure cannot exist in the native sequence. Therefore, it is unlikely to be relevant to the structure and function of the stem defect in SL1. By contrast, we stabilized the wild-type internal loop and closing base pairs by altering the surrounding sequence. This makes only the assumption that the neighbors distant from the internal loop are not essential to maintaining the *i*(G × AGG) structure. The NCp7 binding studies indicate that our 23mer SL1i model for the *i*(1 × 3) stem defect in HIV-1 SL1 has substantial affinity for the nucleocapsid. The most important binding sites for NC are G5 and G17, which correspond to G247 and G272 in the HIV-1 genome.

Overall Structure. An ensemble of 24 refined structures for the SL1i 23mer is presented in Figure 4. The average rmsd to the reference structure is 2.2 Å for the entire molecule, which reflects local disorder and differences in the degree to which the structures bend.

The level of restraint violations (Table 3) is similar to the SL3 RNA structure, with a flexible GGAG tetraloop, which we have published previously (*33*). Restraint violations are compared with a typical C−C bond length (1.537 Å) (*52*) and rotamer interval (60°); conformational uncertainties smaller than these rarely detract from the overall similarity of structures, and they are readily understood by scientists who are not NMR experts. Our use of MARDIGRAS-derived distances (about 40% of the total number of distance restraints) gives much tighter ranges than is commonly used

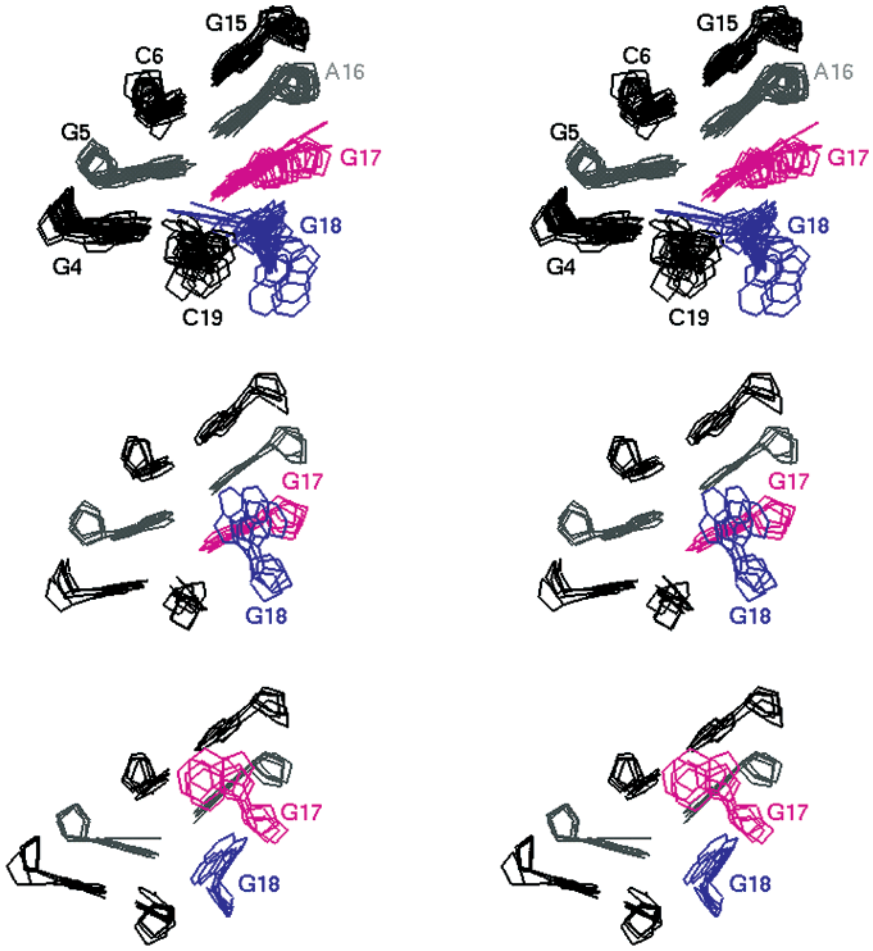

FIGURE 5: Three families of structures for the internal loop; all have a G5−A16 base pair, G4|G5|C6 and G15|A16 stacking, and a discontinuity in stacking between G18 and C19. Top: Cluster 1 has A16|G17 stacking and G18 in the deep groove of the helix. Middle: Cluster 2 still has A16|G17 stacking but G18 in the shallow groove. Bottom: Cluster 3 has A16|G18 stacking and G17 in the shallow groove.

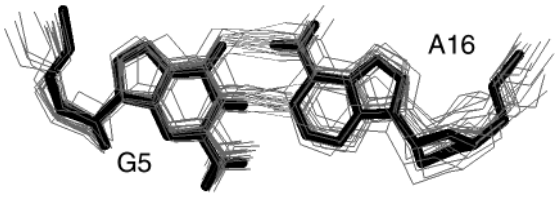

FIGURE 6: G5−A16 pair present in most of the refined structures. The reference structure is shown in heavy black lines. Light gray lines between the bases are H-bonds distinguished by parameters programmed into InsightII (see Experimental Procedures).

in structures derived from ^{13}C,^{15}N-labeled RNA, where 2/3 of the NOEs may effectively give only an upper bound restraint of 6 Å, with the lower bound enforced by van der Waals restrictions. With many distance ranges less than 1 Å, we therefore observe somewhat larger violations. The three largest violations of restraints are associated with the flexible residues in the apical UACG tetraloop, A10 and C11 (average violation 0.6−1.1 Å). This might be expected as distances derived from NOEs are weighted toward the closest approach of atoms rather than the ensemble average. The other substantial violations (all <1/2 C−C) are associated with G4, G17, and G18 at the stem defect. As with the apical loop, the NOE-derived distances are shorter than is typical in the refined structures. This suggests that flexibility is the origin of NOEs that are larger than those consistent with an ensemble of relaxed structures. As these distance violations occur in the internal loop, it is necessary to treat the degree of overall bending of the stem with caution. The dihedral angle restraints are relatively uncertain, so large force constants for these restraints were not applied; a few violations on the order of a rotamer jump are observed, although the average is less than one per structure.

The UACG tetraloop is well-defined except for A10 and C11. All of the characteristic cross-peaks are found in 2-D NOE, ^{1}H-^{31}P COSY, and ^{1}H-^{13}C HSQC spectra, and their chemical shifts are in the same range as those published (*45*, *53*−*55*). Residues A10 and C11 include S-puckered sugars and high-anti glycosyl torsions, and G12 has a syn glycosyl torsion (see Table 4). We also observe the sheared U9−G12 base pairs in our structures as described previously (*43*).

The local fits in the stems are excellent, with rmsd <0.6 Å for G1−G4:C19−C22 (lower stem in Figure 4), and C6−C8:G13−G15 (upper stem). The structures and the pattern of NOEs indicate that stems on both sides of the internal loop are base paired and stacked in a manner similar to a typical right-handed RNA helix.

Details of the Internal Loop. The internal loop residues are numbered near the center of Figure 4, where it can be seen that G17 and G18 are somewhat disordered, while the other loop bases are well-defined. G5 stacks on its neighbors, G4 and C6, and A16 stacks on the upper stem as confirmed by NOEs from C6 to A16,2. There are sequential NOEs throughout both sides of the loop but none that relate G17 and G18 to residues across the loop. Normal Watson−Crick

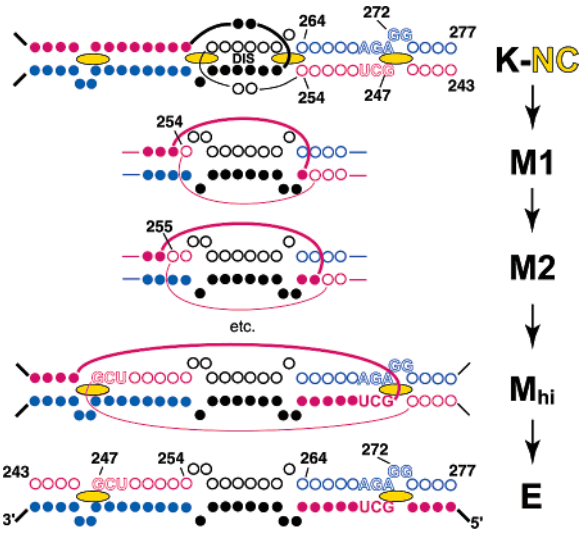

FIGURE 7: Possible mechanism to form the extended dimer, E, in SL1 of HIV-1. The kissing loop duplex, K, is shown at the top, complexed to several NCp7 proteins (ovals). Both chains are identical in the complex, with chain A represented as filled circles and chain B as open circles. The 5′-ends are shown in red and the 3′-ends in blue. In this model the intermolecular stem propagates outward from the $(DIS)_2$ helix through branch-migration intermediates, M1, M2, ... (only the central region is shown). The branches pass through M_{hi}, a high-energy intermediate, and finally to E. Propagation is assisted by the mature NCp7 protein, which may bind strongly to the branch points.

Table 4: Unusual Residue Conformations in the Refined Structures

residue[a]	δ-Tors[b]	sugar[c]	base[d]
G5	0	N	
C6	0	n/S	
A10	S	N/s	high-anti
C11	S	n/S	anti/high-anti
G12	N	N	syn
G13	0	N/s	
C14	N	N/s	
G15	0	N/s	
A16	0	N/s	
G17	S	N/S	anti/syn
G18	S	n/S	anti/high-anti
C19	0	N	
A23	0	N/s	

[a] Residues not denoted in the table have standard N-anti conformations in nearly all of the structures. See Supporting Information for restraint list and full distributions for δ and χ. [b] Input δ-torsion angle restraint favors: N, N-conformers; S, S-conformers, 0, neither. [c] Output sugar puckers: N = almost all N, S = almost all S, N/s = mostly N, n/S = mostly S, and N/S = almost equal. [d] Output orientation about the glycosyl bond (no torsional restraint in the input).

base pairs surround the internal loop (G4:C19 and C6:G15) as confirmed by {G4,1−C19,4} and {C6,4−G15,1} NOEs. All 24 of the structures in Figure 4 have at least one donor−acceptor pair within H-bonding distance to form a noncanonical G5:A16 pair in the loop (more details given below). The network of NOEs relating the surrounding G:C pairs to G5, A16, G17, and G18 force a bend around G17−G18 (see also the discussion of Figure 8, below). It should be cautioned that the overall details of bending are not sensitive functions of the current NMR observables, and that there is evidence for rapid conformational exchange. The complex nature of the bending equilibrium may be difficult to estimate with confidence even upon measurement of residual dipolar coupling constants in partially oriented samples (*56*−*60*).

Three major clusters of structures for the internal loop can be distinguished according to the orientation of G17 and G18 (see Figure 5). These loop residues are unpaired in all of the structures. It can be seen that the loop-closing base pairs, G4:C19 and C6:G15, occupy similar positions in all three panels of the figure, which all have a G5:A16 pair and G15|A16 stacking and no stacking of loop bases on C19. Thus, the degree of bending of the stem appears roughly the same for each cluster, and it is very likely that the average bend is similar in direction to that in Figure 4. Cluster 1 (top) has A16|G17 stacking and G18 unstacked in the deep groove of the helix. G18 can occupy many positions and accounts for most of the rmsd (1.7 Å to the reference structure). Cluster 2 again has A16|G17 stacking, but now G18 lies in the shallow groove (rmsd = 0.8 Å). Cluster 3 (bottom) has A16|G18 stacking, and G17 unstacked in the shallow groove (rmsd = 0.8 Å). The G17 and G18 {*i*,1′−*i*,8} cross-peaks are strong, and the cross-peaks to their neighbors are weak, indicating that they have substantial populations of syn glycosyl rotamers. Furthermore, their $^3J_{H1'-H2'}$ values are large, so they must have significant populations of S-puckered sugars. These points are borne out in the structures, which have a wide distribution of χ and δ for these residues (Table 4 and Supporting Information). The relative broadness of the imino, base, and 1′ signals of G5, A16, G17, and G18 further indicates that the internal loop is quite flexible. Given that twice as many structures occur in Cluster 1 as in the others combined, and that the peaks are broad, it would be difficult to obtain clear NMR evidence for the existence of the minor forms. However, we note that there are several broad peaks near 11 ppm that could indicate multiple positions for G17 and G18 (see Figure 2, low temperatures). Multiple loop conformations may be important in making the G_2 locus more accessible to NC—the distance between G5 and G17 is about 9 Å in Cluster 3, nearly the same as between G318 and G320 in the SL3−NCp7 complex (*13*).

There is strong evidence for a base pair between G5 and A16, although neither G5N1−A16N1 nor G5O6−A16N6 distance restraints were used to force the pair. The appearance of this noncanonical base pair is driven only by the network of restraints between the neighboring inter- and intranucleotide restraints. We predicted this and several other G:A pairs in our proposed secondary structure for the HIV-1 major packaging domain (*33*). Among the 24 lowest-energy structures, 18 form two H-bonds between G5N1−A16N1 and G5O6−A16N6 (see Figure 6), and the rest form at least one. Elgavish et al. have described this as an imino G:A base pair (*61*); their survey indicated that G:A oppositions at the end of a stem tend to form this base pair when the G is 5′ to the helix. We note that there is no evidence for G5N1H peak in the H_2O spectra (see Figure 2 and associated text, above). This is in accord with other NMR studies of G:A pairs (*23*, *62*−*64*), suggesting that this G-imino exchanges rapidly with solvent. The average G5C1′−A16C1′ distance for the 24 structures is 13 Å, as in similar G:A oppositions collected in a compendium of noncanonical interactions (*65*). It is interesting that the G:A base pair is not strong enough to preclude binding G5 to NCp7. Apparently, the interactions in the loop are easily rearranged in response to NC binding.

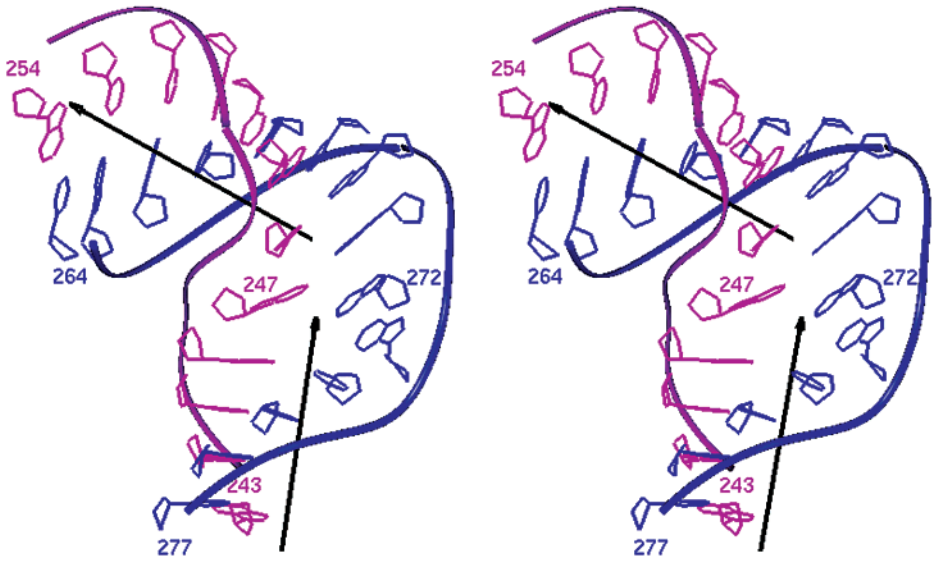

FIGURE 8: UACG-tetraloop has been removed from the SL1i 23mer, and the upper helix was extended to seven base pairs plus G247:A273; this model has the full length of the stem in Figure 1a. Local helix axes show that the stem changes direction around the stem defect as indicated by the local helix axes. The orientation is the same as in Figure 4. Residues 243 and 254 are separated further than 264 and 277. The details of bending in nature may be different from those shown here, as they are not sensitive functions of the NMR observables. (The small discontinuities in the ribbons occur because the upper helix in the SL1i 23mer deviates somewhat from ideal A-form geometry.)

Model for Dimer Initiation and Propagation. Dimer formation is crucial for HIV-1 infectivity and as a reservoir for recombination. The dimerization of HIV-1 genomic RNA begins with an initiation complex between self-complementary regions in the apical DIS loops of SL1 (residues 257−262; see Figure 1a). This metastable kissing dimer (K) then progresses to an extended dimer (E) (*66*, *67*), chaperoned by the mature nucleocapsid protein (*68*−*70*). In vitro experiments have shown that the stem defect is also necessary for this two-step dimerization process in the presence of NCp7 (*29*) and that the stem defect is a site for specific binding to the NC protein (*10*, *71*) (see also Table 2). Shen et al. altered the stem defect in several mutants of genomic RNA and found that nearly all changes reduce the potential for dimerization and cause large reductions in infectivity (*21*).

Figure 7 provides a plausible mechanistic outline for the K → E transformation. Some aspects of this mechanism are similar to one proposed independently by Rist and Marino (*69*), although their work dealt with truncated SL1 constructs that did not involve the stem defect. The $(DIS)_2$ region and symmetry-related stem bases occupy very similar spatial locations in both the K and the E dimers for which structures have been determined by X-ray diffraction (*24*, *25*). The X-ray structures also did not include the stem defect. The two stems in K must open and reanneal in the stems that flank $(DIS)_2$ in E. A likely mechanism for the transformation involves branch migration, as depicted in Figure 7. Such a mechanism would have the minimal number of broken base pairs between intermediate steps in the pathway. The highest energy intermediate is rate limiting in a standard kinetic scheme, and for branch migration this would be similar to M_{hi} in Figure 7, where the annealing is about half complete. The presence of NCp7 bound at the stem defect may assist in the transition of the branch point through this intermediate. There are also geometrical and topological problems that could be encountered in the K → E transition. Figure 7 depicts branch migration led by the strand beginning at 254 (red); the flat diagram makes it appear that single phosphodiester bonds are severely stretched over the long arcs in M_{hi}. In addition, the strands cannot disentangle if they make more than a full turn about the helix axis between fixed ends.

How can the stretching and tangling issues be resolved? Ennifar et al. (*24*) suggested a ribozyme-like action to cut and religate the chains, although to our knowledge no direct evidence has been published to support this possibility. It is also inconsistent with the kinetics of resolution of the K → E transformation in heterodimeric SL1 constructs that do not contain the stem defect (*69*). Bends in the stem at the $(DIS)_2$ segment (*69*) and at the defect may help to alleviate both problems. Rist and Marino propose that the complex bends around $(DIS)_2$, which in Figure 7 would bring the branch points out of the plane of the page toward the viewer. The branch points could then come close to each other. There is no direct evidence for such a bend, but the $i(1 \times 2)$ defects immediately adjacent to $(DIS)_2$ could contribute the necessary flexibility (*69*).

The stem defect may also contribute toward relieving strain in branch migration (or other) intermediates. The upper stem of the NMR structure was extended in Figure 8 to the model the length that would occur in the SL1 stem. The end-to-end distance is longer for the 243−254 segment (red) than for 264−277 (blue) if a bend occurs as predicted by the present NMR structure. (A bend with a similar direction is likely for the alternative $i(1 \times 1) - b(2)$ stem defect.) Of course, such a bend does not imply that a single phosphodiester bond can stretch over a longer distance. Rather, in three-space the tendency of a helix bent in such a fashion is to allow one of the strands to traverse a longer distance than the other. Together with the potential for bending about the $(DIS)_2$ segment (*69*), this may considerably reduce the barrier for K → E.

With regard to tangling, the strands from the two different chains are not entangled in K, so the border of one fixed region is the end of the central $(DIS)_2$ segment. The second fixed region is not so well-defined, but it probably is at

243:277 (*11*, *33*) as shown in Figure 7. The stem would then have 12 base pairs plus two residues in stem defects, longer than a full turn of a regular A-helix. This could present a topological dilemma for separating and reannealing the strands, even if a bend surrounds $(DIS)_2$. Figure 8 shows that the blue segment does not completely encircle the red segment. Thus, it should be relatively easy to open the stems without tangling the strands in the K → E transformation, especially with the elastic stem defect(s) acting to relieve strains. There will only be a tangling problem if the stem continues much beyond 243:277 as in the Lever secondary structure (*23*, *72*) where the second fixed region would be at 236:282.

Figure 7 also incorporates suggestions regarding the role of NCp7 in the K → E transition. It is likely that the K form is incorporated into the budding virus particle, and NCp7 only becomes available later, during maturation (*66*−*68*). A large excess of NCp7 is then produced, which binds at many RNA sites, including the SL1i defect (*10*) and near $(DIS)_2$ (*69*), as indicated in Figure 7. It is possible that NCp7 has a high affinity for the branch point where at least a few nucleotides must be unpaired. It is well-known that NCp7 binds single-stranded RNA most strongly (reviewed in ref *73*). Especially high affinity may occur at the stem defect, where one might expect considerable freedom of orientation for G247, G270, G272, and G273 as the branch junction passes. It is also possible that one zinc finger of an NCp7 can bind a G-base from chain A and another G-base from chain B at the SL1 stem defect.

Conclusion. An RNA construct to model the internal loop form of the SL1 stem defect has the wild-type loop and surrounding base pairs. This loop cannot slip into the alternate form that should coexist in the native SL1 stem. Key guanosyl residues have been identified that interact with the mature nucleocapsid protein in the SL1i construct. The NMR-based structure has a noncanonical G:A base pair in the loop and a bend in the stem segments that surround the loop. A bend in the stem may be important in the transition from an apical loop−loop interaction to a dimer involving intermolecular interactions along the full length of SL1. The affinity of NCp7 for the stem defect may accelerate the rate of the transition. The SL1i−NCp7 interaction could be an important new target for anti-AIDS drug development. Such drugs could affect the dimer transformation at a high-energy intermediate and therefore could strongly attenuate viral reproduction.

ACKNOWLEDGMENT

We are grateful to Dr. Yong Lin for analyzing the frequency of mutations in the 5′-leader and Prof. Bruce Hudson for assistance with the Trp-37 fluorescence assay and helpful discussions. A plasmid vector containing the NCp7 gene was a generous gift from Prof. Michael Summers. We are also grateful for NSF and NIH support for NMR instrumentation and to the SUNY College of Environmental Science and Forestry for access to the 600 MHz NMR instrument.

SUPPORTING INFORMATION AVAILABLE

Two tables (chemical shifts and restraints) and three figures (restraints, δ, and χ vs residue number). 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 Lab Press, Plainview, NY.
2. Clever, J., Sassetti, C., and Parslow, T. G. (1995) *J. Virol. 69*, 2101−9.
3. McBride, M. S., and Panganiban, A. T. (1996) *J. Virol. 70*, 2963−73.
4. Hayashi, T., Shioda, T., Iwakura, Y., and Shibuta, H. (1992) *Virology 188*, 590−9.
5. Hayashi, T., Ueno, Y., and Okamoto, T. (1993) *FEBS Lett. 327*, 213−8.
6. Clever, J. L., and Parslow, T. G. (1997) *J. Virol. 71*, 3407−14.
7. Clever, J. L., Eckstein, D. A., and Parslow, T. G. (1999) *J. Virol. 73*, 101−9.
8. McBride, M. S., Schwartz, M. D., and Panganiban, A. T. (1997) *J. Virol. 71*, 4544−54.
9. McBride, M. S., and Panganiban, A. T. (1997) *J. Virol. 71*, 2050−8.
10. Shubsda, M. F., Paoletti, A. C., Hudson, B. S., and Borer, P. N. (2002) *Biochemistry 41*, 5276−82.
11. Paoletti, A. C., Shubsda, M. F., Hudson, B. S., and Borer, P. N. (2002) *Biochemistry 41*, 15423−8.
12. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) *J. Mol. Biol. 299*, 145−56.
13. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) *Science 279*, 384−8.
14. 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*, 419−34.
15. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−80.
16. Kerwood, D. J., Cavaluzzi, M. J., and Borer, P. N. (2001) *Biochemistry 40*, 14518−29.
17. 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. 314*, 961−70.
18. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4945−9.
19. Paillart, J. C., Westhof, E., Ehresmann, C., Ehresmann, B., and Marquet, R. (1997) *J. Mol. Biol. 270*, 36−49.
20. Laughrea, M., and Jette, L. (1994) *Biochemistry 33*, 13464−74.
21. Shen, N., Jette, L., Wainberg, M. A., and Laughrea, M. (2001) *J. Virol. 75*, 10543−9.
22. Lin, Y. Ph.D. Thesis, Syracuse University, Syracuse, NY, 2002.
23. Greatorex, J., Gallego, J., Varani, G., and Lever, A. (2002) *J. Mol. Biol. 322*, 543.
24. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C., and Dumas, P. (2001) *Nat. Struct. Biol. 8*, 1064−8.
25. Ennifar, E., Yusupov, M., Walter, P., Marquet, R., Ehresmann, B., Ehresmann, C., and Dumas, P. (1999) *Structure Fold Des. 7*, 1439−49.
26. Girard, F., Barbault, F., Gouyette, C., Huynh-Dinh, T., Paoletti, J., and Lancelot, G. (1999) *J. Biomol. Struct. Dyn. 16*, 1145−57.
27. Theilleux-Delalande, V., Girard, F., Huynh-Dinh, T., Lancelot, G., and Paoletti, J. (2000) *Eur. J. Biochem. 267*, 2711−9.
28. Takahashi, K., Baba, S., Hayashi, Y., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) *J. Biochem. (Tokyo) 127*, 681−6.
29. Takahashi, K. I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) *RNA 6*, 96−102.
30. Mujeeb, A., Parslow, T. G., Zarrinpar, A., Das, C., and James, T. L. (1999) *FEBS Lett. 458*, 387−92.
31. Mujeeb, A., Clever, J. L., Billeci, T. M., James, T. L., and Parslow, T. G. (1998) *Nat. Struct. Biol. 5*, 432−6.
32. 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.
33. Pappalardo, L., Kerwood, D. J., Pelczer, I., and Borer, P. N. (1998) *J. Mol. Biol. 282*, 801−18.
34. Zeffman, A., Hassard, S., Varani, G., and Lever, A. (2000) *J. Mol. Biol. 297*, 877−93.
35. Scaringe, S. A. (2000) *Methods Enzymol. 317*, 3−18.
36. Zuker, M. (2002) http://bioinfo.math.rpi.edu/~zukerm/.

37. Zuker, M., Mathews, D., and Turner, D. H. (1999) in *NATO ASI Series, RNA Biochemistry and Biotechnology* (Barciszewski, J., and Clark, B. F. C., Eds.) pp 11−43, Kluwer Academic Publishers, New York.
38. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−40.
39. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−5.
40. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) *J. Biomol. NMR 6*, 390−402.
41. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J. Mol. Biol. 273*, 283−98.
42. Case, D. A., Pearlman, D., Caldwell, J. C., Cheatham, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) *AMBER 5*, University of California, San Francisco.
43. Allain, F. H., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−53.
44. Borer, P. N., Pappalardo, L., Kerwood, D. J., and Pelczer, I. (1997) in *Advances in Biophysical Chemistry* (Bush, C. A., Ed.) pp 173−216.
45. Molinaro, M., and Tinoco, I. (1995) *Nucleic Acids Res. 23*, 3056−63.
46. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. H. (1985) *Biochemistry 24*, 4533−9.
47. Jeong, G. W., Borer, P. N., Wang, S. S., and Levy, G. C. (1993) *J. Magn. Reson. A 103*, 123−34.
48. Borer, P. N., and Levy, G. C. (1994) *Methods Enzymol. 239*, 257−88.
49. 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.
50. Harrison, G. P., Miele, G., Hunter, E., and Lever, A. M. L. (1998) *J. Virol. 72*, 5886−96.
51. Kratz, A., and Lewandrowski, K. B. (1998) *N. Engl. J. Med. 339*, 1063−72.
52. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
53. Varani, G., Cheong, C., and Tinoco, I. (1991) *Biochemistry 30*, 3280−9.
54. Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M., and Uesugi, S. (1990) *Nucleic Acids Res. 18*, 3831−9.
55. Abdelkafi, M., Ghomi, M., Turpin, P. Y., Baumruk, V., Herve du Penhoat, C., Lampire, O., Bouchemal-Chibani, N., Goyer, P., Namane, A., Gouyette, C., Huynh-Dinh, T., and Bednarova, L. (1997) *J. Biomol. Struct. Dyn. 14*, 579−93.
56. Bax, A., Kontaxis, G., and Tjandra, N. (2001) *Methods Enzymol. 339*, 127−74.
57. Wu, Z., and Bax, A. (2002) *J. Am. Chem. Soc. 124*, 9672−3.
58. Wu, Z., Tjandra, N., and Bax, A. (2001) *J. Biomol. NMR 19*, 367−70.
59. Sibille, N., Pardi, A., Simorre, J. P., and Blackledge, M. (2001) *J. Am. Chem. Soc. 123*, 12135−46.
60. Hennig, M., Carlomagno, T., and Williamson, J. R. (2001) *J. Am. Chem. Soc. 123*, 3395−6.
61. Elgavish, T., Cannone, J. J., Lee, J. C., Harvey, S. C., and Gutell, R. R. (2001) *J. Mol. Biol. 310*, 735−53.
62. Collier, A. J., Gallego, J., Klinck, R., Cole, P. T., Harris, S. J., Harrison, G. P., Aboul-Ela, F., Varani, G., and Walker, S. (2002) *Nat. Struct. Biol. 9*, 375−80.
63. Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J., and Feigon, J. (1994) *Biochemistry 33*, 5357−66.
64. Peterson, R. D., and Feigon, J. (1996) *J. Mol. Biol. 264*, 863−77.
65. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) *Nucleic Acids Res. 30*, 3497−531.
66. Muriaux, D., De Rocquigny, H., Roques, B. P., and Paoletti, J. (1996) *J. Biol. Chem. 271*, 33686−92.
67. Muriaux, D., Fosse, P., and Paoletti, J. (1996) *Biochemistry 35*, 5075−82.
68. Fu, W., Gorelick, R. J., and Rein, A. (1994) *J. Virol. 68*, 5013−8.
69. Rist, M. J., and Marino, J. P. (2002) *Biochemistry 41*, 14762−70.
70. Darlix, J. L., Cristofari, G., Rau, M., Pechoux, C., Berthoux, L., and Roques, B. (2000) *Adv. Pharmacol. 48*, 345−72.
71. Damgaard, C. K., Dyhr-Mikkelsen, H., and Kjems, J. (1998) *Nucleic Acids Res. 26*, 3667−76.
72. Harrison, G. P., and Lever, A. M. (1992) *J. Virol. 66*, 4144−53.
73. Darlix, J. L., Lapadat-Tapolsky, M., de Rocquigny, H., and Roques, B. P. (1995) *J. Mol. Biol. 254*, 523−37.

BI034084A