



PubMed
Central

HHMI
HOWARD HUGHES MEDICAL INSTITUTE

AUTHOR MANUSCRIPT
Accepted for publication in a peer-reviewed journal

Published as: *Biochemistry*. 2008 December 9; 47(49): 13064–13073.

Potential Intra- and Intermolecular Interactions involving the Unique-5' Region of the HIV-1 5'-UTR

Shardell Spriggs⁺, Lianko Garyu⁺, Ryan Connor, and Michael F. Summers*

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

Abstract

The 5'-untranslated region (5'-UTR) of the human immunodeficiency virus type-1 (HIV-1) genome regulates multiple RNA-dependent functions during viral replication and has been proposed to adopt multiple secondary structures. Recent phylogenetic studies identified base pair complementarity between residues of the unique 5' element and those near the *gag* start codon (*gag*^{AUG}) that is conserved among evolutionarily distant retroviruses, suggesting a potential long-range RNA-RNA interaction. However, nucleotide accessibility studies led to conflicting conclusions about the presence of such interactions in virions and in infected cells. Here we show that an 11-nucleotide oligo-RNA spanning residues 105-115 of the 5'-UTR (U5) readily binds to oligoribonucleotides containing the *gag* start codon (AUG), disrupting a pre-existing stem loop and forming a heteroduplex stabilized by 11 Watson Crick base pairs ($K_d = 0.47 \pm 0.16 \mu\text{M}$). Addition of the HIV-1 nucleocapsid protein (NC), the transacting viral factor required for genome packaging, disrupts the heteroduplex by binding tightly to U5 ($K_d = 122 \pm 10 \text{ nM}$). The structure of the NC:U5 complex, determined by NMR, exhibits features similar to those observed in NC complexes with HIV-1 stem loop RNAs, including the insertion of guanosine nucleobases to hydrophobic clefts on the surface of the zinc fingers and a 3'-to-5' orientation of the RNA relative to protein. Our findings indicate that the previously proposed long-range U5-*gag*^{AUG} interaction is feasible, and suggest a potential NC-dependent mechanism for modulating the structure of the 5'-UTR.

Keywords

Human Immunodeficiency Virus Type-1 (HIV-1); ribonucleic acid (RNA); psi-site; nucleocapsid protein (NC); nuclear magnetic resonance (NMR)

The 5'-untranslated region (5'-UTR) of the retroviral genome plays a number of important roles during viral replication, including splicing, genome packaging, reverse transcription, and translation (see (1-5) and references there-in). During virus assembly, elements within the 5'-UTR of the full-length genome interact with the nucleocapsid (NC) domains of an unknown number of retroviral Gag polyproteins, resulting in genome incorporation into assembling virions. All retroviruses package two copies of their genomes, and elements that promote dimerization are also located within the 5'-UTR (1-5). The major splice donor site is also typically located within the 5'-UTR, providing a potential mechanism for the specific packaging of the full length (versus spliced) viral RNAs (6, 7). Much of the information about the role of the 5'-UTR in genome packaging has been obtained from studies of the

*To whom correspondence should be addressed. Phone: (410)-455-2527; FAX: (410)-455-1174; Web: www.hhmi.umbc.edu; Email: summers@hhmi.umbc.edu.

⁺These authors contributed equally to this work.
AUTHOR EMAIL: summers@hhmi.umbc.edu

Moloney murine leukemia virus (MLV), an evolutionarily older retrovirus with a simple genome that does not require extensive splicing or encode for accessory proteins. Recent studies suggest that MLV genome packaging may be regulated by a dimerization-dependent RNA conformational switch mechanism, in which elements that bind NC with high affinity are sequestered by base pairing in the monomeric RNA (8) and become exposed to bind NC and promote packaging upon dimerization (9).

RNA conformational switch mechanisms may also regulate functions of the human immunodeficiency virus (HIV) 5'-UTR (1, 10-18). Combinations of nucleotide accessibility mapping experiments, phylogenetic analyses and free energy calculations indicate that the HIV-1 5'-UTR adopts a “branched multiple hairpin (BMH)” conformation that consists of a series of hairpin structures connected by relatively short unstructured linkers (7, 19-27). Residues 1-57, which form the TAR stem loop, promote transcriptional activation by binding to the transcription activator protein, TAT. Although TAR does not participate directly in packaging, a terminal hairpin seems to be required (28). Residues 58-104 form the “poly-A” stem loop, which functions in translation initiation and may form long-range base pairs with residues of *gag* (Figure 1). Residues 116-236 form a branched loop structure that serves as the binding site for the tRNA(Lys) primer. The tRNA binding element is followed by a dimer-promoting stem loop (DIS), a relatively short stem loop (SL2) containing the major splice donor (SD) site, and the so-called Υ stem loop (or SL3) that is important for efficient genome packaging. In the absence of Mg²⁺ ions and NC protein, the HIV-1 5'-UTR appears to adopt a significantly different “long distance interaction (LDI)” structure (17), and interconversion between the BMH and LDI conformers has been proposed to serve as a switch that regulates dimerization, splicing, and possibly other RNA-dependent activities (10-13, 15-17).

Multiple models have also been proposed for residues that overlap with the *gag* start codon (*gag*^{AUG}). Residues 339-352 were originally proposed to form a hairpin structure (7, 20, 25, 29) (historically called SL4; see Figure 1) containing a GAGA tetraloop (a member of the GNRA family of tetraloops; N = G, C, A, or U; R = G or A), and mutagenesis experiments indicate that nucleotides that promote base-pairing in the proposed stem loop are critical for efficient genome packaging (7). However, SL4 oligoribonucleotides do not bind NC with high affinity (30), and since GNRA tetraloops are known to participate in long-range RNA-RNA interactions (31, 32), it was suggested that the SL4 hairpin structure might play a role in stabilizing a particular RNA conformation via GNRA—minor groove interactions (30). More recent phylogenetic studies that included evolutionarily distant retroviruses identified base pair complementarity between residues near the *gag* start codon and an upstream element that links the poly-A and PBS stem loops (U5) (12). This led to the suggestion that residues of *gag*^{AUG} may not form the predicted SL4 stem loop, but might instead participate in long-range U5-*gag*^{AUG} base-pairing, Figure 1 (12, 33). However, chemical accessibility experiments conducted with infected cells and isolated virions indicated that several residues of *gag*^{AUG} (C110 and C111) and U5 (A334 and A336) are strongly modified by alkylating agents and thus not likely to form the proposed long-range interactions at these stages of viral replication (26). More recent secondary structure predictions based on chemical accessibility of ribose groups suggested that long-range base-pairing between AUG and upstream segments in both U5 and the PBS stem loop does occur in infected cells, isolated virions, and *in vitro* transcribed viral RNA (27). The reason for these discrepant results and conclusions has yet to be explained.

Previous NMR studies confirmed that oligoribonucleotides containing the sequence of SL4 adopt the predicted stem loop structure stabilized by a GAGA tetraloop (30, 34), and it was not obvious to us if U5 would be capable of disrupting this structure and forming the proposed long-range base pairs. We therefore conducted a series of biochemical and

structural studies with oligoribonucleotides corresponding to residues 105-115 of the 5'-UTR (U5) and 334-352 of *gag*^{AUG} (AUG). Polyacrylamide gel-shift and NMR results confirm that U5 and AUG readily form a stable heteroduplex, as originally proposed (12, 33). Interestingly, addition of the NC protein disrupts the U5:AUG structure by binding with high affinity to U5. Our findings indicate that the proposed U5:*gag*^{AUG} interactions are feasible, identify a potential high affinity NC binding site within the 5'-UTR, and suggest a new role for NC in modulating the structure of the 5'-UTR.

MATERIALS AND METHODS

RNA preparation and purification

Oligoribonucleotides listed in Table 1 (an 11 nucleotide U5 construct and U5 mutants) and a 19-nucleotide fragment of *gag*^{AUG} (AUG¹⁹) were obtained from Dharmacon Research Inc. with 2'-o-bis(acetoxyethoxy)-methyl (ACE) protecting groups. The 2' ACE groups were removed as described by the manufacturer's protocol and the RNA was purified by ethanol precipitation. The 29 nucleotide *gag*^{AUG} construct (AUG) was generated by T7 RNA polymerase-dependent *in vitro* transcription using a synthetically prepared DNA template containing the T7 promoter sequence. The transcribed RNA was precipitated with ethanol (-20 °C, 12 hrs), lyophilized, and purified by preparative-scale denaturing polyacrylamide gel electrophoresis. After isolation by electroelution (Whatman Elutrap electroelution system), the RNA was concentrated using a Centricon YM-3 ultrafiltration device (Millipore, Bedford, MA) and lyophilized. Samples enriched in ¹⁵N and ¹³C were prepared similarly using isotopically labeled nucleoside triphosphates (Spectra Stable Isotopes).

NC preparation and purification

The recombinant 55 residue HIV-1 NC protein was expressed from pRD2 that contains the NC coding region of NC from HIV-1_{NL4-3} subcloned into pET3a (Novagen, WI). pRD2 was transformed into E. coli strain BL21(DE3)pLysE and the overexpressed protein was purified under non-denaturing conditions as described (35), except that a Sephadex-20 column was used for size exclusion.

Isothermal Titration Calorimetry

Dissociation constants for U5:AUG complex formation and for HIV-1 NC binding to U5 and U5 mutants were determined by standard isothermal titration calorimetry (ITC) methods using a VP-Isothermal titration calorimeter (MicroCal Corp., Northampton, MA). RNA and protein concentrations were determined using UV-Vis absorption measurements. The RNA and NC samples containing buffer (10mM Tris-HCl (pH 6.5), 0.1 mM ZnCl₂, 0.1 mM βME) with varying NaCl concentrations (100 mM to 400 mM NaCl) were degassed for 14 minutes prior to titration. The U5:AUG titrations were conducted in buffer containing 150 mM NaCl and 1.0 mM MgCl₂. Exothermic heats of reaction (μcal/s) were measured at 30°C for 28 injections of NC protein (75 - 100 μM) into 1.41 ml of RNA (5 μM). The heats of dilution were obtained by titrating the identical protein sample into a cell containing sample buffer and subtracted from the raw data prior to analysis. Heats of dilution were typically 0.015 μcal/s and heats of binding were 0.3 μcal/s, giving an apparent signal to noise ratio of 20. The syringe mixing speed was 310 rpm as recommended by the vendor. Binding curves were analyzed by non-linear least-squares fitting of the baseline-corrected ITC data to a single binding site model (ITC Origin program, V2.8; MicroCal, Northampton. MA). The dissociation constants (K_d) were reported as the mean ± standard deviation from three independent titration experiments, unless otherwise noted in Table 1.

Native polyacrylamide gel electrophoresis

U5, AUG¹⁹, and AUG RNA samples were prepared at 0.3 mM stock solution in H₂O. For heterodimer formation, equal amounts of U5 and AUG (or AUG¹⁹) were mixed to give a final concentration of 0.3 mM of total RNA. The samples were then incubated at 95°C for 5 minutes and 55°C for 30 minutes (in RNase zapped PCR tubes). These reaction mixtures were then used for NC titrations. Here, 0 µl, 0.25 µl or 0.5 µl NC (0.6 µM) was added to give 1:0, 1:0.5 and 1:1 RNA:NC ratios. Glycerol (50% aqueous; 2 µl) was then added, and the samples (containing 2.5 µg of total RNA) were loaded onto 20% native polyacrylamide gels at 4°C in 0.5x Tris-borate buffer (44.5 mM Tris base, 44.5 mM boric acid). After electrophoresis, the gels were stained with Stains-all (Sigma), which stains negatively charged molecules, and photographed with a gel imaging system (Eastman Kodak Company).

NMR experiments

NMR spectra of exchangeable imino protons were obtained at a sample temperature of 10 °C (~ 1.3 mM RNA, 150 mM NaCl, 10 mM Tris-HCl, pH 7.0). 2D NOESY (36) data were obtained with a mixing time of 200 ms. Samples of the NC-U5 and NC-U5^{GGU} complexes (see Table 1 for definitions) for NMR studies were prepared by titrating small amounts of concentrated NC protein (10 mM d-Tris-HCl, pH 6.5, 10 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM βME) directly into U5 or U5^{GGU} RNA and monitoring the NMR signals of the RNA. NMR data were collected with Bruker AVANCE 800 MHz and DMX 600 MHz spectrometers, processed with NMRPipe/NMRDraw (37) and analyzed with NMRView (38). 2D NOESY and 2D TOCSY (39, 40) NMR data for structural studies were obtained at 15 °C. Proton resonance assignments for the free U5^{GGU} RNA and the NC:U5^{GGU} complex were obtained by standard sequential assignment methods (41).

Structure Calculations

All structure calculations were performed with a 160 processor Apple Xserve cluster equipped with a low-latency Miracom switch. Initial structures were generated with CYANA (version 2.1) using ¹H-¹H distance restraints of 2.7, 3.3 and 5.0 Å corresponding to experimentally observed NOEs of strong, medium, and weak intensity, respectively. Pseudoatom corrections were added for methyl groups and diastereotopically equivalent geminal, methyl, and aromatic protons, and Zn-Cys and Zn-His bonding restraints were employed as described (35).

The 20 CYANA structures with lowest penalties were subjected to additional restrained simulated annealing and energy minimization with AMBER (version 9) (42). One set of structures was calculated with the generalized Born (GB) solvent continuum (43) using the ff99 force field (which was modified to allow zinc coordination to Cys and His residues), a non-bonded cut-off of 12 Å, and no periodic boundary condition. Structures were minimized at 0 K, heated to 298K (8 ps duration), slowly re-cooled to 0 K (52 ps duration), and subjected to 10,000 steps of minimization. A second set of structures was calculated from the 20 original CYANA structures using explicit water conditions, which allowed us to compare predicted intermolecular electrostatic contributions in structures generated using both approaches (44). Each CYANA structure was solvated within an 8 Å octahedral water shell (TIP3P water model (45)) and subjected to 10,000 steps of minimization (cubic spline switch function for direct sum Coulomb interactions; ff99 force field; 10 Å non-bonded cut-off). The structures were then heated to 1000 K (8 ps), slowly cooled to 0 K (52 ps duration), and subjected to 10,000 steps of minimization. Superposition statistics were calculated with Theseus (46).

RESULTS

U5 and AUG oligoRNAs form a stable heteroduplex

Previous NMR studies (30) revealed that oligoribonucleotides containing the sequence of SL4 form a stem loop structure consistent with the predicted secondary structure shown in Figure 1C. The hairpin structure is stabilized by a GAGA tetraloop, a member of the ubiquitous GNRA family, and by both canonical G-C and non-canonical G-U base pairing in the stem. NMR studies confirmed that the AUG oligoribonucleotide shown in Figure 2A also forms the predicted hairpin, and that the isolated U5 oligonucleotide is unstructured (as predicted by MFOLD calculations; data not shown). Addition of U5 to AUG leads to a distinct, stoichiometric band shift in native PAGE data, Figure 2B, indicating that U5 and AUG form a stable heterodimer at 1:1 ratios. No additional bands were observed at U5:AUG heterodimer ratios above 1:1. Isothermal titration calorimetry experiments confirmed that U5 binds AUG stoichiometrically and with high affinity ($K_d = 0.47 \pm 0.16 \mu\text{M}$), Figure 2C. 2D nuclear Overhauser effect (NOESY) spectra obtained for the U5:AUG heterodimer exhibited internucleotide imino-to-imino proton interactions consistent with the predicted base pairing shown in Figure 1B, including both the canonical and non-canonical base pairs.

NC binds to U5 and AUG, disrupting the NC:AUG duplex

The NC domain of the HIV-1 Gag protein is responsible for recognizing and packaging the retroviral genome, and packaging determinants on the RNA reside primarily within the 5'-UTR. Although the HIV-1 NC protein is capable of binding with high affinity to isolated SL2 and SL3 stem loops *in vitro*, most or all of the 5'-UTR is required to efficiently package heterologous RNAs into particles (47). Information regarding the number of NC proteins required for packaging and the number of high affinity NC binding sites within the HIV-1 5'-UTR is currently lacking, but it seems likely that efficient packaging requires 5'-UTR interactions with multiple NC domains. Thirteen independent regions within the 5'-UTR exhibit chemical reactivity properties that are dependent on NC (27). We previously showed that NC binds with modest affinity to the isolated SL4 stem loop ($K_d = 47 \pm 14 \mu\text{M}$) (30). To determine if NC might interact with either U5 or the U5:AUG duplex, we conducted native PAGE NC titration experiments using a 19-nucleotide AUG construct (AUG¹⁹). As shown in Figure 3, addition of NC to U5 leads to a complete band shift at 1:1 NC:U5 stoichiometry, indicating that NC binds stoichiometrically to U5. By comparison, a faint band corresponding to free AUG¹⁹ remains at a NC:AUG¹⁹ ratio of 1:1, consistent with earlier ITC studies that showed only modest affinity for the AUG stem loop (30). Interestingly, addition of NC to the U5:AUG¹⁹ duplex resulted in the disappearance of the U5:AUG¹⁹ band and the appearance of weak bands corresponding to the NC:U5 and NC:AUG¹⁹ complexes. The intensity of the NC:U5 band is weaker than that of the NC:AUG¹⁹ band, possibly due to differences in association/dissociation kinetics of the NC:U5 and NC:AUG¹⁹ complexes.

NOESY NMR signals observed for the Watson Crick imino protons of the U5:AUG heteroduplex disappeared upon addition of NC (data not shown) under physiological-like conditions ([NaCl] = 150 mM; [MgCl₂] = 1 mM), indicating that NC disrupts the U5:AUG base pairing. These data collectively indicate that NC disrupts the U5:AUG complex by binding individually to both U5 and AUG.

To determine the affinity of NC for the U5 RNA, isothermal titration calorimetry experiments were performed. Under conditions of low ionic strength (at which several previous NC-RNA binding studies were reported; 10 mM NaCl, 10 mM Tris, 0.1 mM ZnCl₂, 0.1 mM βME, pH 6.5) isotherms indicative of very tight binding (approaching the limits of reliable K_d determination) were obtained. We therefore conducted NC titration ITC

experiments over a range of ionic strengths ($[NaCl] = 100\text{ mM}, 200\text{ mM}, 300\text{ mM}$ and 400 mM NaCl), Figure 3B. All of the ITC experiments gave binding isotherms that fit best to a single binding site model (see Table 1). The extrapolated dissociation constant of 2.24 nM at 10 mM NaCl is considerably lower than those observed previously for SL2 and SL3 hairpin RNAs under similar low ionic strength conditions ($K_d = 110 \pm 50\text{ nM}$ and $170 \pm 65\text{ nM}$ for SL2 and SL3, respectively).

Tight NC binding is dependent on U5 guanosines

Previous structural studies of NC complexes with SL2 and SL3 stem loop RNAs indicated that the zinc knuckles of NC interact specifically with exposed guanosines. Since the U5 construct employed in the current studies contains three guanosines, we hypothesized that NC binding might occur via multiple modes. It was alternatively possible that more than one NC molecule might bind to U5, although the PAGE and ITC data described above indicated the presence of only a single high affinity NC binding site. To determine the potential roles of the three guanosines in NC binding, we engineered a series of single, double, and triple G to U mutations into U5, Table 1. ITC experiments with these mutants revealed that the construct in which the 3'-G was mutated to U ($G112U, U5^{GGU}$) binds NC with affinity that is actually slightly greater than that measured for the native construct under similar conditions ($81 \pm 8\text{ nM}; [NaCl] = 100\text{ mM}$), whereas the other single mutants ($G108U$ and $G106U$; called $U5^{GUG}$ and $U5^{UGG}$, respectively) bind with reduced affinities ($\sim 250\text{ nM}$), Table 1. Constructs containing two G to U mutations bind with significantly reduced affinities ($540\text{-}800\text{ nM}$), and binding could not be detected by ITC for the construct that lacks guanosines ($U5^{UUU}$). The reduced affinity of the native U5 construct may be due to the presence of a G112-dependent RNA secondary structure that inhibits NC binding, and additional studies to test this hypothesis are underway. Interestingly, G112 is well removed from G106 and G108, and the spacing of G106 and G108 is identical to the spacing of tetraloop guanosines that interact with the NC zinc knuckles in the NC:SL2 and NC:SL3 complexes (35, 48).

1D 1H NMR spectra obtained for native U5 upon titration with NC exhibited broad signals with multiple features, consistent with conformational heterogeneity. Although we have yet to unambiguously identify the nature of the heterogeneity, the most likely cause is multiple binding modes in which the two zinc knuckles interact with different combinations of the three guanosines. However, high quality NMR spectra consistent with a single binding mode were obtained for the $U5^{GGU}$ construct, which was shown by ITC to bind NC with very high affinity. Structural studies were therefore conducted with this construct.

Structure of the NC:U5 GGU complex

1H NMR signals for NC:U5 GGU were assigned by conventional 2D methods, and portions of the 2D NOESY spectra that exhibit intermolecular NOEs involving the aromatic protons are shown in Figure 4. For example, moderate-intensity NOEs were observed between the C109-H6 proton and the side chain methyl protons of Ile24. G108-H8 exhibited NOEs with Phe 16-H $^{\delta,e}$, Ala25-C $^{\beta}$ H₃, and Ile24-C $^{\gamma,\delta}$ H₃ of the N-terminal zinc knuckle. Strong NOEs were also observed between the Phe16 aromatic protons and U107-H6, and between C109-H6 and the methyl groups of Ile24. The G106-H8 and -H1' protons exhibited NOEs with the Trp37 aromatic protons of the C-terminal zinc knuckle. This pattern of NOE cross peaks is consistent with an NC binding mode that involves the interactions of U107, G108 and C109 with the N-terminal zinc knuckle and U105, G106 and U107 with the C-terminal knuckle.

The 3D structure of the NC:U5 GGU complex was determined by a two-step approach, in which initial models generated with CYANA using a simple van der Waals force field were subjected to annealing and minimization with AMBER using a more sophisticated force

field (ff99, with generalized Born implicit solvent simulation (43)). The NMR data and structure statistics are summarized in Table 2. Best-fit superpositions of the backbone atoms of the N- and C-terminal zinc knuckles for the 20 conformers with lowest target function are shown in Figure 5. Each zinc knuckle domain has a well-defined backbone conformation with a fold that is consistent with structures determined previously for free and RNA-bound forms of HIV-1 NC. In contrast, the structure of the linker that connects the two zinc knuckles (Arg32, Lys33, Lys34) is not well defined because rapid chemical exchange precluded detection of the backbone amide protons for these residues. It is likely that the heterogeneity observed in the models reflects true conformational heterogeneity and dynamics in solution for this short stretch of basic residues. Backbone NH signals were not detected for most residues of the N- and C-terminal tails (residues Met1-Lys14 and Thr50-Asn55, respectively), and these residues did not give rise to intermolecular or long-range NOEs and were not restrained during structure refinement. The zinc knuckles pack against each other via interactions between Phe16 and Trp37 of the N- and C-terminal zinc knuckles, respectively. Similar inter-knuckle packing was observed in the NC-SL2 and -SL3 complexes (35, 48).

Nucleotides that exhibit NOEs with NC are reasonably well-defined in the ensemble, particularly G106, U107 and G108 (Figure 5). In contrast, nucleotides C110-U115 exhibited only weak sequential intramolecular NOEs and no intermolecular NOEs, and were thus not restrained during the structure calculations. U105 and G106 pack against the C-terminal knuckle, U107 interacts with the linker and makes contacts with both knuckles, and G108 and C109 pack against the N-terminal zinc knuckle, Figure 6a. Overall, the structure has a clam-like appearance, with the two knuckles acting as the shells of the clam that clamp together over U105, G106, U107, and G108, Figure 6b. The nucleobases of G106 and G108 pack tightly into the hydrophobic “guanosine binding pockets” of the C- and N-terminal zinc knuckles, respectively. The bases of these residues are poised to form a network of hydrogen bonds with backbone atoms of the protein, including G106-O6 to Trp37-NH and Met46-NH; G106-N1H and -N2-H21 to Gly35-O; G108-O6 to Phe16-NH and Ala25-NH; G108-N1H and -N2-H21 to Lys14-O (Figure 6c,d). These zinc knuckle-guanosine interactions are very similar to those observed previously in NC-RNA complexes (35, 48, 49). In addition, the relative orientation of the NC and U5 strands (N-terminal zinc knuckle binding to the 3'-guanosine) is the same as that observed in the NC-SL2 and -SL3 NMR structures.

DISCUSSION

Several structural models have been proposed for the HIV-1 5'-UTR based on combinations of chemical and enzymatic probing, phylogenetic analyses, and free energy calculations (7, 19-27). Although some features are common among most models, such as the base pairing in the TAR, PolyA, DIS, SL2 and SL3 stem loops, other features are significantly different. In particular, most (7, 20, 25, 29), but not all (19), of the early studies suggested that AUG forms a stem loop, and NMR studies confirmed that isolated AUG RNAs readily form this predicted structure (30). More recent phylogenetic studies suggested an alternate structure for AUG, in which nucleotides A334-G344 form long-range base pairs with residues U105-U115, respectively of U5 (12). Base-pair complementarity was observed for a number of different retroviruses, suggesting that U5:AUG base pairing may be evolutionarily conserved (33). However, these potential interactions were not supported by *in vivo* chemical accessibility mapping experiments conducted by Ehresmann and co-workers using infected cells and isolated virions (26). In these studies, several nucleic acid bases predicted to participate in U5:AUG interactions were readily modified by chemical probes, indicating that they are exposed and not involved in stable base pairing. However, more recent mutagenesis studies suggested that these long range interactions may be required for replication (50). More recent chemical modification experiments conducted using probes

HHMI Author Manuscript HHMI Author Manuscript HHMI Author Manuscript

that target the 2'-hydroxyl group of conformationally flexible, non-base paired residues (SHAPE) suggested that residues of AUG do, in fact, base pair with residues of U5 (27) (although the proposed base pairing pattern differed somewhat from earlier predictions (12, 33)). The SHAPE method is purportedly more sensitive to unstructured elements than traditional chemical probing techniques, and it is therefore unclear why the residues of U5 and AUG were reactive to traditional chemical probes but poorly reactive to the SHAPE probes, both of which were performed in cells and in isolated virions under native-like conditions. Interestingly, treatment of virions with agents that eject zinc from NC (resulting in protein unfolding and reduced RNA affinity) led to a decrease in the sensitivity of U5 residues to SHAPE agents, leading to the suggestion that NC has a destabilizing effect on the U5-AUG duplex.

The present studies confirm that isolated U5 is capable of binding with high affinity to AUG-containing oligonucleotides, disrupting the pre-existing GNRA tetraloop structure. The U5:AUG hetero-dimer that forms is stabilized by 11 base pairs, consistent with the original phylogenetic analysis (12). Of course, our studies do not confirm the presence of these interactions *in vivo*, nor do they rule out other potential long-range base pairing possibilities. However, they clearly demonstrate that the proposed long-range interactions are feasible, a fact that was not obvious to us *a priori* due to the inherent stability of the AUG stem loop (30). Our findings also indicate that NC is capable of disrupting the U5:AUG duplex by binding to U5. Although U5 contains three guanosines that serve structural roles in the U5:AUG complex by forming G-C and G*U base pairs, only two (G106 and G108) are required for high affinity binding to NC. Tight NC binding is mediated by the zinc knuckles of NC, which clamp together over U105, G106, U107, and G108.

It is noteworthy that residues within the U5 element exhibited reduced SHAPE reactivity upon treatment of virions with an agent that ejects zinc from the NC zinc knuckles (27). This led to the suggestion that the U5-AUG duplex, which was predicted to exist in virions on the basis of SHAPE chemistry (27), is destabilized to some extent by NC (27), and that zinc ejection leads to stabilization of the duplex. We propose that the U5-AUG interactions do not exist in virions, at least for a significant subset of particles, and that U5 instead associates with NC. The elevated SHAPE reactivity of U5 residues (27) and the significant reactivity observed by traditional chemical mapping for residues C110, C11, A330, A332 and A334 (26) appear to be consistent with this model. The increase in protection of the U5 ribose groups that occurs upon treatment with zinc ejectors would therefore be attributed to the formation of U5:AUG base pairing, rather than conversion of a less stable duplex to a more stable duplex. The proposed long range U5:AUG interactions could play a role in an earlier replication event that occurs prior to RNA association with Gag or NC.

Studies of 5'-UTR fragments have provided a wealth of information regarding RNA and protein-RNA structures that are likely to play important roles in genome packaging, reverse transcription, and other RNA-dependent events that are required for replication. The biological relevance of these relatively small structures can be difficult to independently verify, in part because retroviruses often use multiple, redundant mechanisms that can confound interpretation of mutagenesis experiments. For example, it is likely that the NC domains of multiple Gag proteins participate in genome recognition and packaging (each virion contains several thousand copies of Gag), and mutagenesis experiments involving one potential Gag binding site may not lead to a significant change in phenotype (7, 25, 51). Atomic-level 3D structural studies of the intact 5'-UTR are clearly warranted. In this regard, we have initiated NMR studies of RNAs containing the intact HIV-1 5'-UTR, in which the residues of *gag*^{AUG} have been isotopically-labeled by segmental synthesis. Analysis of the NMR spectra obtained for these RNAs should be possible using the behavior and spectra of the isolated AUG and U5:AUG fragments as a guide.

Acknowledgments

We are grateful to Kun Lu, Rob Edwards, Yu Chen (HHMI, UMBC) for advice and technical assistance.

SUPPORT: Support from the NIH (5R01GM042561) is gratefully acknowledged. S.S. was a recipient of a Ruth L. Kirschstein NRSA grant (1 F31 GM070097). S. S. and L. G. were supported by an NIGMS initiative for minority student development grant R25-GM55036.

Abbreviations

2D	two dimensional
3D	three dimensional
A	adenosine
BME	β -mercaptoethanol
C	cytosine
CA	capsid protein
G	guanosine
HIV-1	human immunodeficiency virus Type-1
HMQC	heteronuclear multiple quantum coherence
ITC	isothermal titration calorimetry
HIV-1	human Immunodeficiency virus type-1
NC	nucleocapsid protein
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RNA	ribonucleic acid
U	uracil

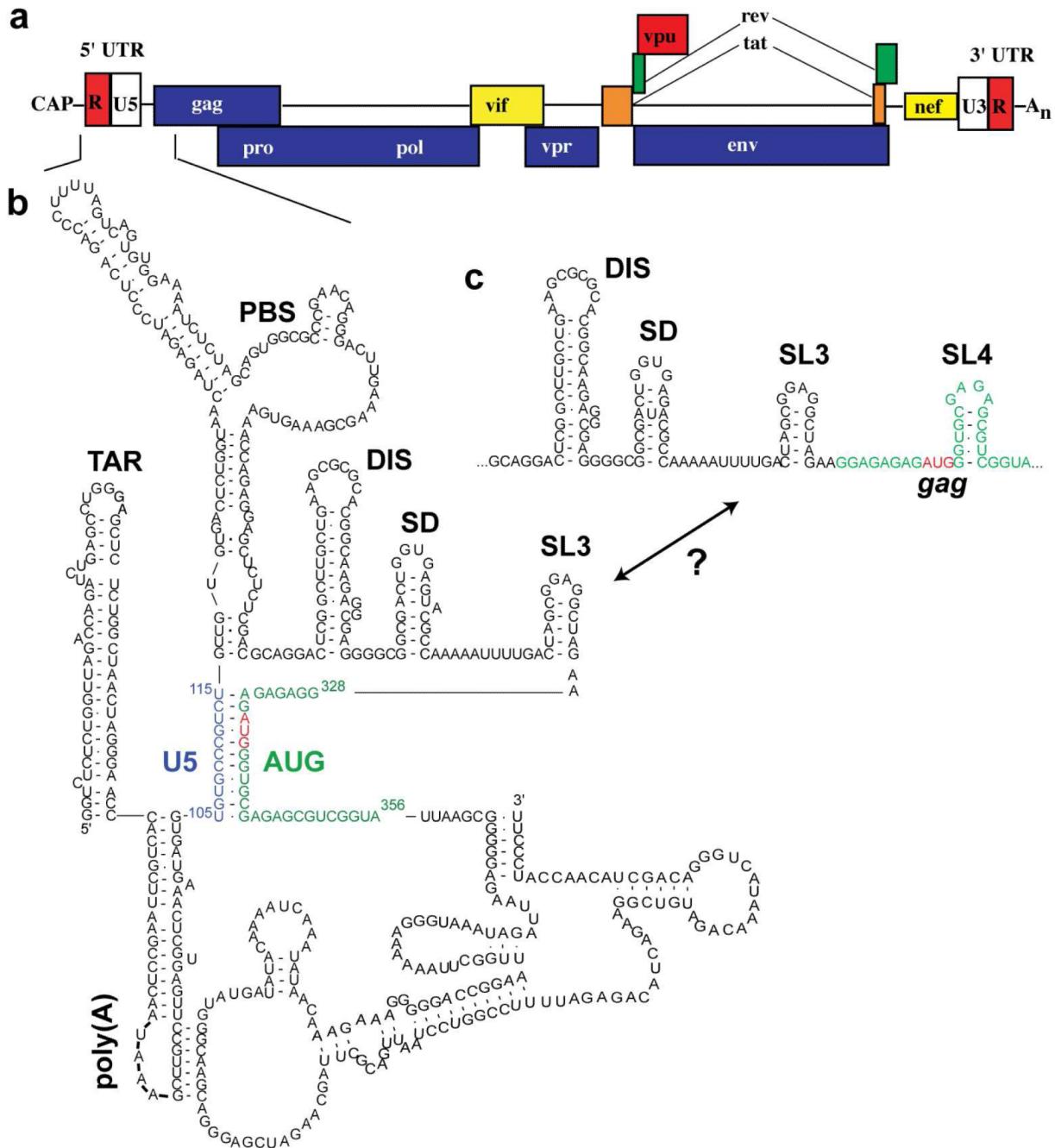
References

1. Berkhout, B. *Prog. Nucl. Acid Res. and Mol. Biol.* Academic Press, Inc.; 1996. Structure and function of the human immunodeficiency virus leader RNA; p. 1-34.
2. Berkowitz R, Fisher J, Goff SP. RNA packaging. *Curr. Top. Microbiol. Immun.* 1996; 214:177–218.
3. Coffin, JM.; Hughes, SH.; Varmus, HE. *Retroviruses*. Cold Spring Harbor Laboratory Press; Plainview, N.Y.: 1997.
4. Swanstrom, R.; Wills, JW. Synthesis, assembly and processing of viral proteins. In: Coffin, JM.; Hughes, SH.; Varmus, HE., editors. *Retroviruses*. Cold Spring Harbor Press; Plainview, N.Y.: 1997. p. 263-334.
5. D'Souza V, Summers MF. How retroviruses select their genomes. *Nature Reviews Microbiology*. 2005; 3:643–655.
6. Mann R, Baltimore D. Varying the position of a retrovirus packaging sequence results in the encapsidation of both unspliced and spliced RNA. *J. Virol.* 1985; 54:401–407. [PubMed: 3989912]
7. McBride MS, Panganiban AT. The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* 1996; 70:2963–2973. [PubMed: 8627772]

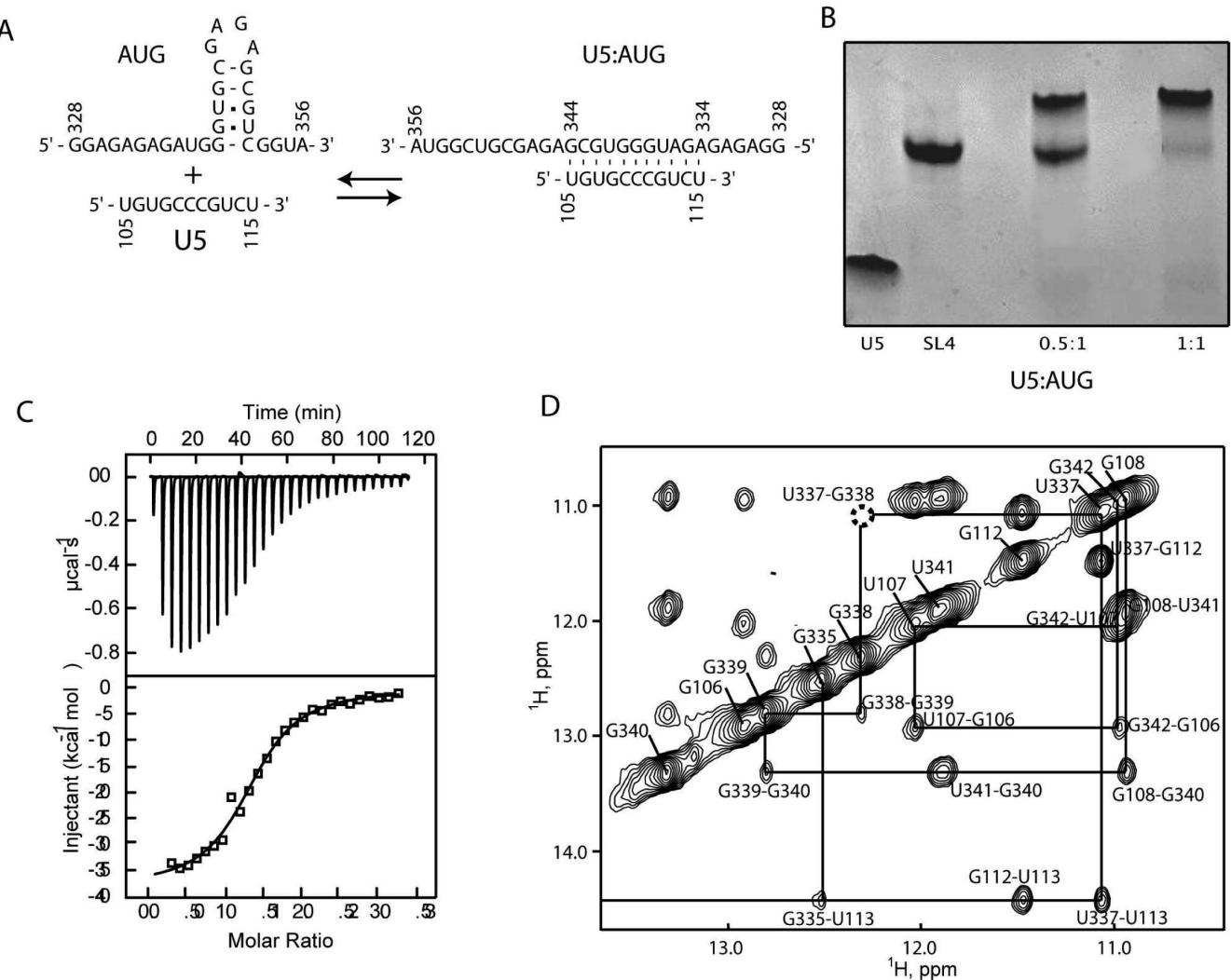
8. Tounekti N, Mougel M, Roy C, Marquet R, Darlix J-L, Paoletti J, Ehresmann B, Ehresmann C. Effect of dimerization on the conformation of the encapsidation psi domain of Moloney murine leukemia virus RNA. *J. Mol. Biol.* 1992; 223:205–220. [PubMed: 1731069]
9. D'Souza V, Summers MF. Structural basis for packaging the dimeric genome of Moloney Murine Leukaemia Virus. *Nature*. 2004; 431:586–590. [PubMed: 15457265]
10. Abbink TEM, Berkhou B. RNA structure modulates splicing efficiency at the human Immunodeficiency Virus Type 1 major splice donor. *J. Virol.* 2008; 82:3090–3098. [PubMed: 18160437]
11. Ooms M, Huthoff H, Russell R, Liang C, Berkhou B. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* 2004; 78:10814–10819. [PubMed: 15367648]
12. Abbink TEM, Berkhou B. A novel long distance base-pairing interaction in Human Immunodeficiency Virus Type 1 RNA occludes the Gag start codon. *J. Biol. Chem.* 2003; 278:11601–11611. [PubMed: 12458192]
13. Huthoff H, Berkhou B. Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry*. 2002; 41:10439–10445. [PubMed: 12173930]
14. Dirac AMG, Huthoff H, Kjems J, Berkhou B. Regulated HIV-2 RNA dimerization by means of alternative RNA conformations. *Nucl. Acids Res.* 2002; 30:2647–2655. [PubMed: 12060681]
15. Berkhou B, Ooms M, Beerens N, Huthoff H, Southern E, Verhoef K. *In vitro* evidence that the untranslated leader of the HIV-1 genome is an RNA checkpoint that regulates multiple functions through conformational changes. *J. Biol. Chem.* 2002; 277:19967–19975. [PubMed: 11896057]
16. Huthoff H, Berkhou B. Two alternating structures of the HIV-1 leader RNA. *RNA*. 2001; 7:143–157. [PubMed: 11214176]
17. Berkhou B, Van Wamel JLB. The leader of the HIV-1 RNA genome forms a compactly folded tertiary structure. *RNA*. 2000; 6:282–295. [PubMed: 10688366]
18. Klasens BI, Thiesen M, Virtanen A, Berkhou B. The ability of the HIV-1 AAUAAA signal to bind polyadenylation factors is controlled by local RNA structure. *Nucleic Acids Res.* 1999; 27:446–454. [PubMed: 9862964]
19. Harrison GP, Lever AML. The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* 1992; 66:4144–4153. [PubMed: 1602537]
20. Baudin F, Marquet R, Isel C, Darlix J-L, Ehresmann B, Ehresmann C. Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol.* 1993; 229:382–397. [PubMed: 8429553]
21. Hayashi T, Shioda T, Iwakura Y, Shibuta H. RNA packaging signal of human immunodeficiency virus type 1. *Virology*. 1992; 188:590–599. [PubMed: 1585635]
22. Hayashi T, Ueno Y, Okamoto T. Elucidation of a conserved RNA stem-loop structure in the packaging signal of human immunodeficiency virus type 1. *FEBS*. 1993; 327:213–218.
23. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 1994; 91:4945–4949. [PubMed: 8197162]
24. Clever J, Sassetti C, Parslow TG. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of Human Immunodeficiency Virus Type 1. *J. Virol.* 1995; 69:2101–2109. [PubMed: 7884856]
25. McBride MS, Panganiban AT. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation *in vivo*. *J. Virol.* 1997; 71:2050–2058. [PubMed: 9032337]
26. Paillart JC, Dettenhofer M, Yu X-F, Ehresmann C, Ehresmann B, Marquet R. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J. Biol. Chem.* 2004; 279:48397–48403. [PubMed: 15355993]
27. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*. 2008; 6:883–899.

28. Das AT, Harwig A, Vrolijk MM, Berkout B. The TAR hairpin of human immunodeficiency virus Type 1 can be deleted when not required for Tat-mediated activation of transcription. *J. Virol.* 2007; 81:7742–7748. [PubMed: 17494072]
29. Clever JL, Parslow TG. Mutant Human Immunodeficiency Virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* 1997; 71:3407–3414. [PubMed: 9094610]
30. Amarasinghe GK, Zhou J, Miskimon M, Chancellor KJ, McDonald JA, Matthews AG, Miller RA, Rouse MD, Summers MF. Stem-loop SL4 of the HIV-1 Ψ -RNA packaging signal exhibits weak affinity for the nucleocapsid protein. Structural studies and implications for genome recognition. *J. Mol. Biol.* 2001; 314:961–969. [PubMed: 11743714]
31. Woese CR, Winder S, Gutell RR. Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proc. Natl. Acad. Sci. USA.* 1990; 87:8467–8471. [PubMed: 2236056]
32. Michel F, Westhof E. Modeling of the three-dimensional architecture of Group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 1990; 216:585–610. [PubMed: 2258934]
33. Damgaard CK, Andersen ES, Knudsen B, Gorodkin J, Kjems J. RNA interactions in the 5' region of the HIV-1 genome. *J. Mol. Biol.* 2004; 336:369–379. [PubMed: 14757051]
34. Heus HA, Pardi A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science.* 1991; 253:191–194. [PubMed: 1712983]
35. De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ -RNA recognition element. *Science.* 1998; 279:384–388. [PubMed: 9430589]
36. Kumar A, Ernst RR, Wuthrich K. A Two-Dimensional Nuclear Overhauser Enhancement (2nd Noe) Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules. *Biochemical and Biophysical Research Communications.* 1980; 95:1–6. [PubMed: 7417242]
37. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 1995; 6:277–293. [PubMed: 8520220]
38. Johnson BA, Blevins RA. NMRview: a Computer Program for the Visualization and Analysis of NMR Data. *J. Biomol. NMR.* 1994; 4:603–614. [PubMed: 22911360]
39. Braunschweiler L, Ernst RR. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. *J. Magn. Reson.* 1983; 53:521–528.
40. Bax A, Davis DG. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.* 1985; 65:355–360.
41. Wüthrich, K. *NMR of Proteins and Nucleic Acids.* John Wiley & Sons; New York: 1986.
42. Case DA, Cheatham TEI, Darden T, Gohlke H, Luo R, Merz KMJ, Onufriev A, Simmerling C, Wang B, Woods R. The Amber biomolecular simulation programs. *J. Computat. Chem.* 2005; 26:1668–1688.
43. Mongan J, Simmerling C, McCammon JA, Case DA, Onufriev A. Generalized born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* 2007; 3:156–169. [PubMed: 21072141]
44. Okur A, Wickstrom L, Layten M, Geney R, Song K, Hornak V, Simmerling C. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *J. Chem. Theory Comput.* 2006; 2:420–433.
45. Jorgensen WL, Chandrasekhar J, Madura J, Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983; 109:926–935.
46. Theobald DL, Wuttke DS. Accurate structural correlations from maximum likelihood superpositions. *PLOS Computational Biology* in press. 2008
47. Berkowitz RD, Hammarskjold M-L, Helga-Maria C, Rekosh D, Goff SP. 5' Regions of HIV-1 RNAs are not sufficient for encapsidation: Implications for the HIV-1 packaging signal. *Virology.* 1995; 212:718–723. [PubMed: 7571442]
48. Amarasinghe GK, De Guzman RN, Turner RB, Chancellor K, Wu Z-R, Summers MF. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ -RNA packaging signal. *J. Mol. Biol.* 2000; 301:491–511. [PubMed: 10926523]

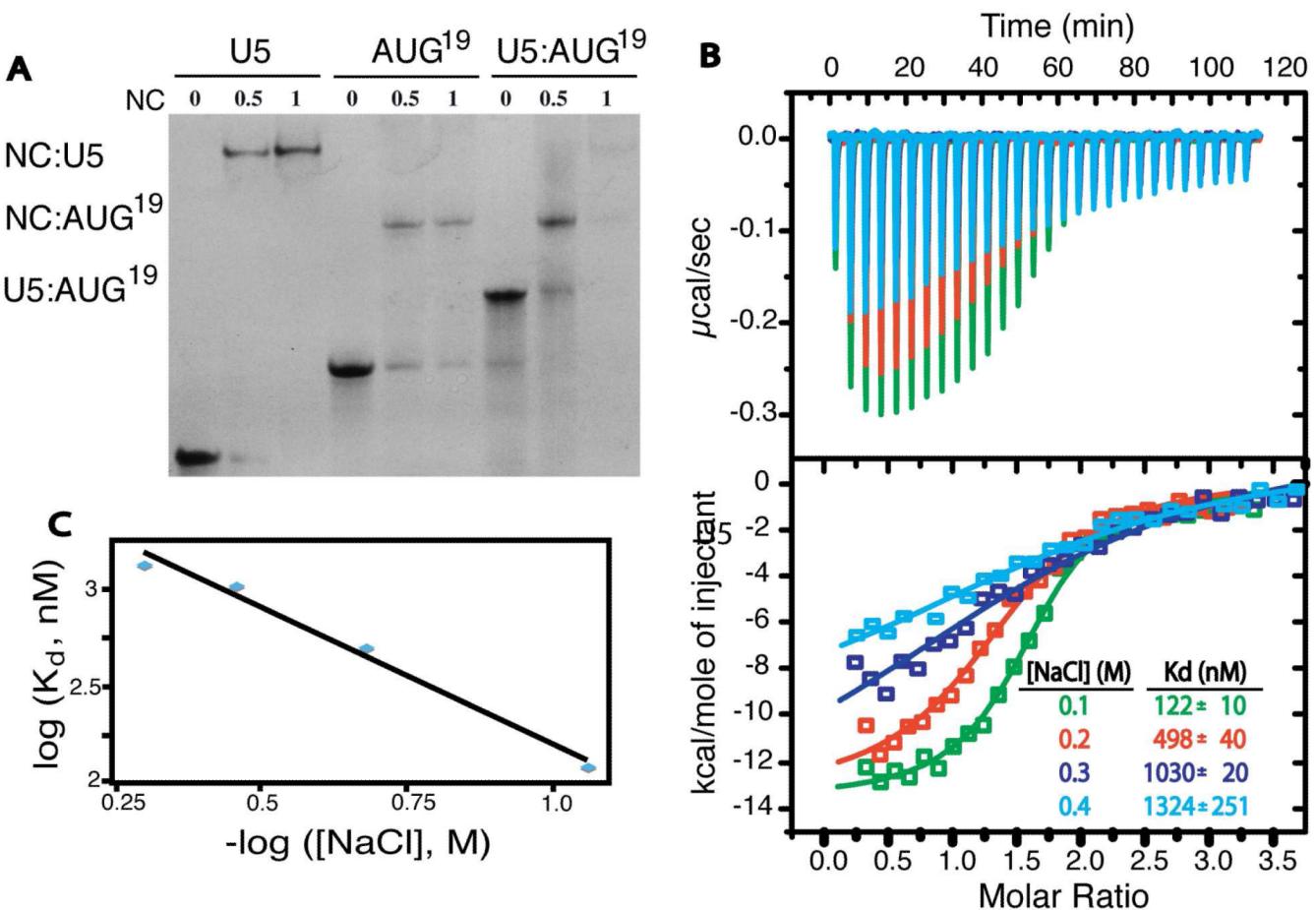
49. South TL, Summers MF. Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* 1993; 2:3–19. [PubMed: 8443588]
50. Nikolaitchik O, Rhodes TD, Ott D, Hu W-S. Effects of mutations in the Human Immunodeficiency Virus Type 1 gag gene on RNA packaging and recombination. *J. Virol.* 2006; 80:4691–4697. [PubMed: 16641262]
51. McBride MS, Schwartz MD, Panganiban AT. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J. Virol.* 1997; 71:4544–4554. [PubMed: 9151848]

**Figure 1.**

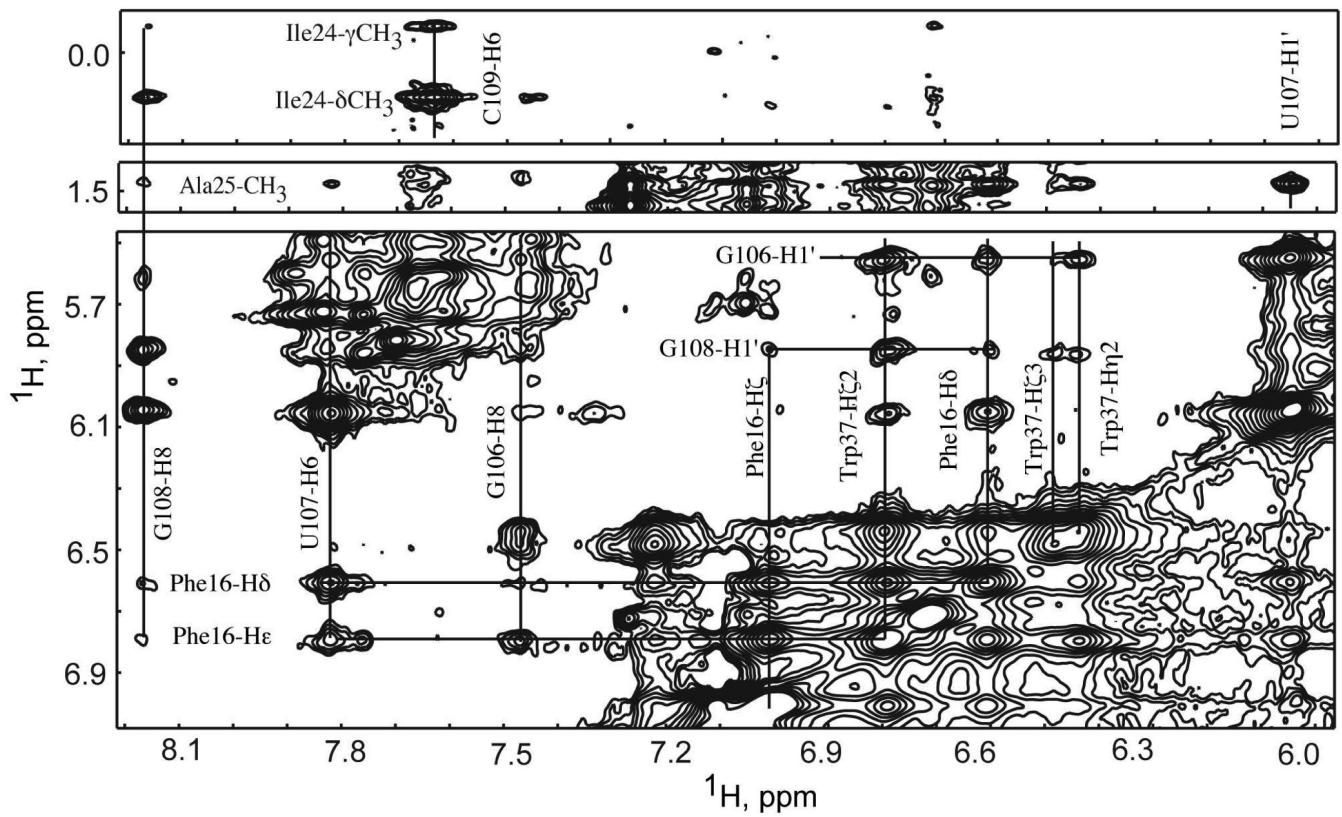
(a) HIV-1 genome showing the location of the 5'- Untranslated Region (UTR). (b) One of the predicted secondary structures of the HIV-1 5'- UTR showing proposed U5-gag^{AUG} long-range interactions (12, 33). The U5 sequence is shown in blue and the AUG sequence is shown in green, with the gag start codon in red. (c) Alternate secondary structure showing the hairpin structure of AUG (also called SL4) that was predicted on the basis of chemical modification and mutagenesis experiments (7, 20, 25, 29).

**Figure 2.**

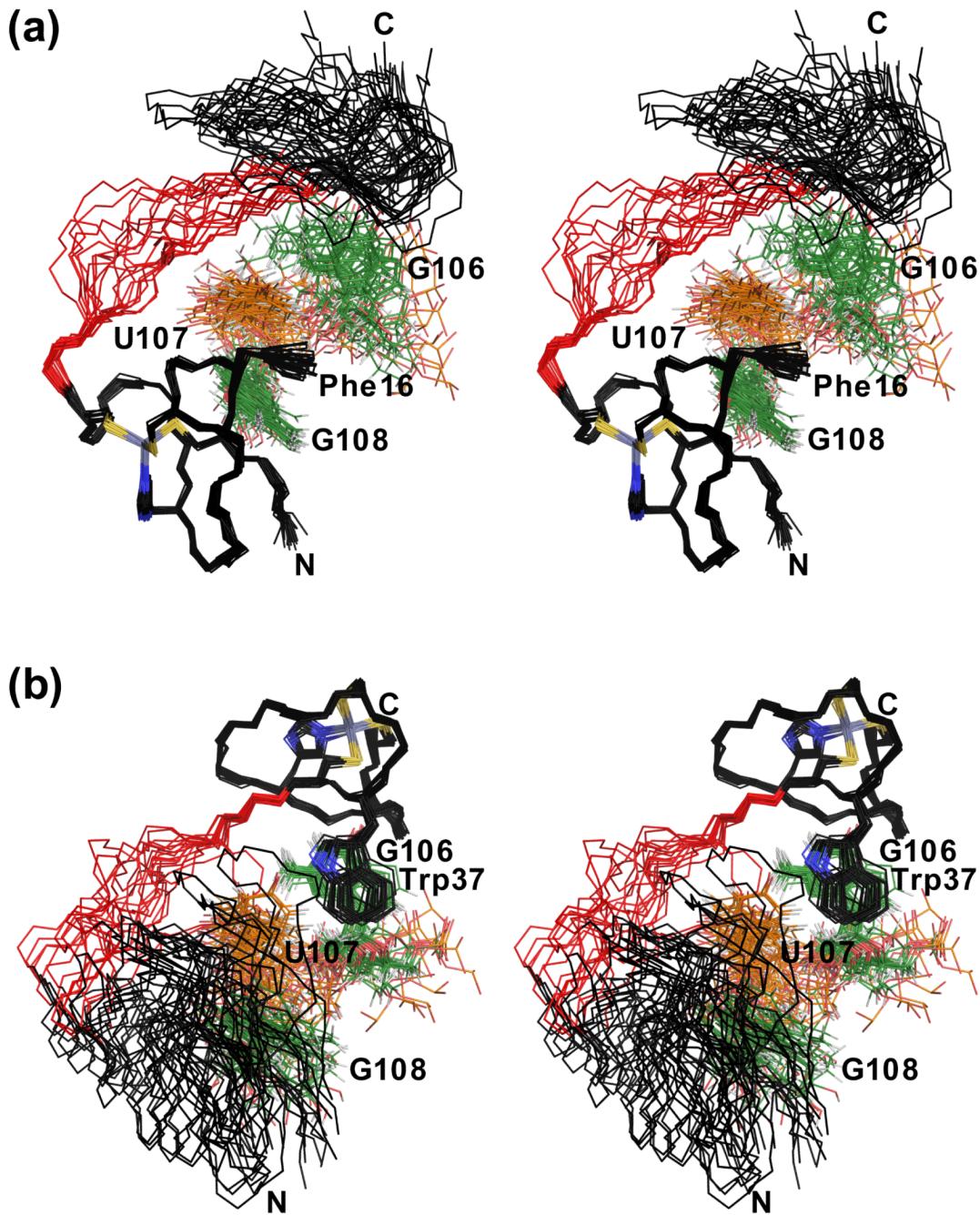
U5 and AUG form a stable duplex. **(a)** U5 and AUG RNA constructs used for this study. **(b)** Native PAGE results for AUG RNA upon titration with U5 RNA showing that U5 and AUG form a stable heterodimer at 1:1 U5:AUG molar ratios. No additional bands were observed at U5:AUG ratios greater than 1:1. **(c)** Representative ITC data obtained upon titration of AUG with U5 ($K_d = 0.47 \pm 0.16 \mu\text{M}$). **(d)** Downfield region of the 2D ^1H - ^1H NOESY spectrum obtained for the U5:AUG duplex showing sequential connectivities between base-paired Watson-Crick imino protons. A weak cross peak between the U337 and G338 imino protons is observed at lower contour levels (dashed circle).

**Figure 3.**

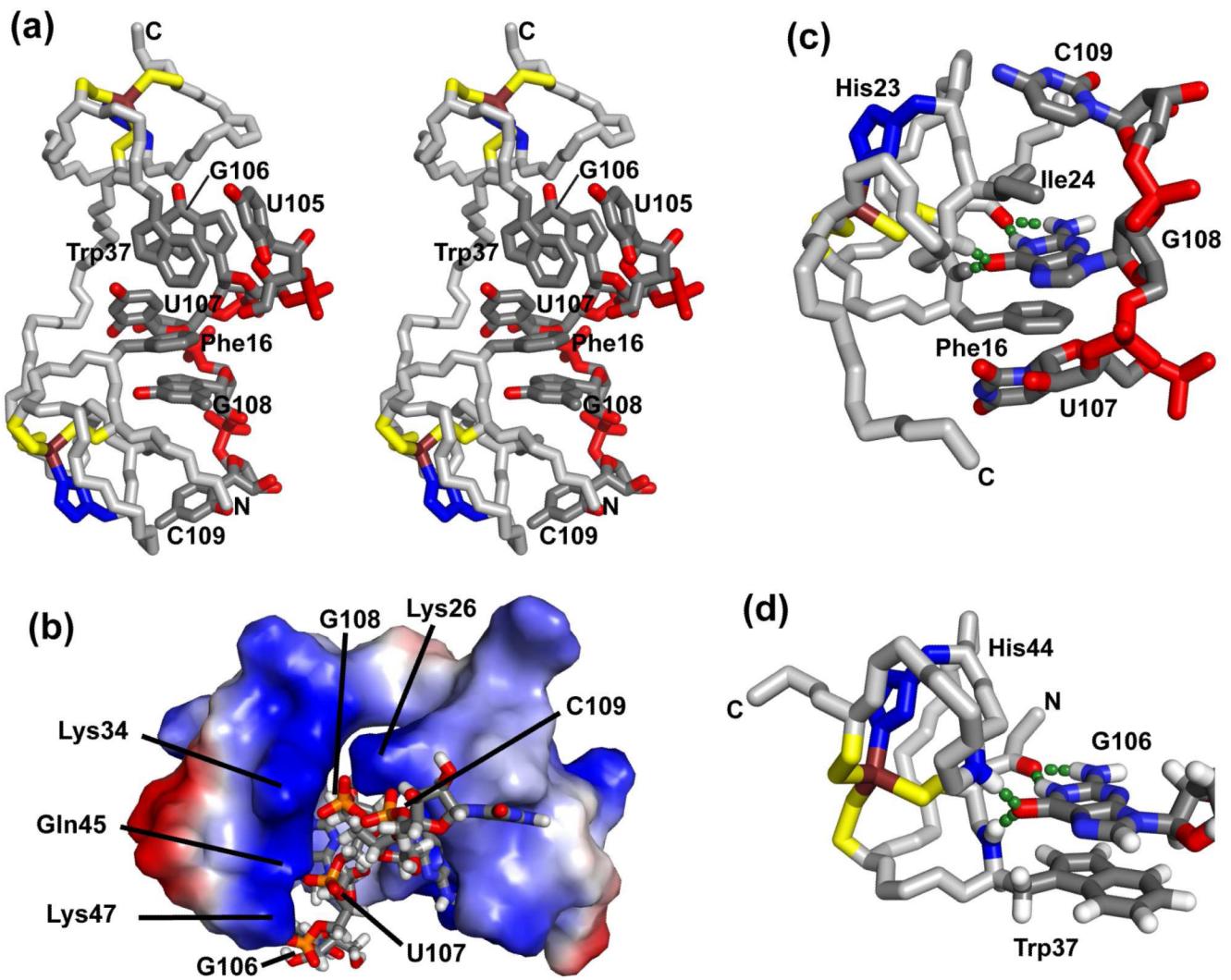
(a) Native PAGE results obtained upon titration of U5 (left), a 19-residue form of AUG that spans residues A334-C352 (AUG¹⁹; center) and U5:AUG¹⁹ (right) with NC (NC/RNA ratios are shown at the top). NC binds to both U5 and AUG¹⁹, disrupting the U5:AUG¹⁹ complex. NC:U5 migrates slowly due to the low overall negative charge. Weak band intensities observed at NC:U5:AUG¹⁹ ratios of 1:1:1 are due to significant smearing that appears to be related to exchange between the different equilibrium species. (b) ITC data obtained upon titration of U5 with NC at varying NaCl concentrations. Upper Panel: the raw data showing the time-dependent heat profiles obtained upon NC injection. Bottom panel: Calculated binding isotherms obtained after subtraction of blank (heat of dilution) data. (c) Plot of the log (K_d) versus -log [NaCl].

**Figure 4.**

Portions of the 800 MHz 2D NOESY data obtained for the unlabeled NC:U5^{GUU} complex. Selected intermolecular cross peaks are shown for the aromatic protons of Phe 16 (N-terminal zinc knuckle) and Trp 37 (C-terminal knuckle) are labeled. Selected intermolecular NOEs involving the Ile24-δCH₃, Ile24-γCH₃ and Ala25-CH₃ protons are shown in the upper panels.

**Figure 5.**

Stereoviews of the 20 lowest-energy structures calculated for the NC:U5^{GGU} complex. **(a)** Best-fit superposition of the backbone atoms of the N-terminal zinc knuckle. The figure displays the backbone atoms of residues Val 13-Cys 49 (black), heavy atoms of the N-terminal zinc knuckle (Cys and His in yellow and blue, respectively), and RNA residues U107 (orange), G106 and G108 (green). The backbone atoms of residues that link the two zinc knuckles are shown in red. **(b)** Same as in (a), except that the backbone atoms of the C-terminal zinc knuckle are superposed.

**Figure 6.**

Structure of the lowest-energy NC:U5^{GGU} structure. **(a)** Stereoview showing the relative positions of the two zinc knuckles and the RNA residues that interact with NC. **(b)** Electrostatic surface potential representation (positive and negative potential shown in blue and red, respectively) of residues Val 13 - Cys 49 in the NC:U5^{GGU} complex. Residues G106-C109 are shown as sticks. Phosphate groups and basic residues of NC that are poised for favorable electrostatic interactions are labeled. **(c, d)** Stick figure representations showing H-bonding interactions between G108 and G106 and the N- and C-terminal zinc knuckles, respectively (Cys = yellow, His = blue, Zn = brown).

Table 1ITC derived dissociation constants for native and mutants forms of U5^a

RNA	[NaCl], M	K _d , nM	N
U5- (UGUGCCCGUCU)	0.1	122 ± 10	1.09 ± 0.01
	0.2	498 ± 40	1.21 ± 0.15
	0.3	1030 ± 20	1.24 ± 0.23
	0.4	1324 ± 251	1.17 ± 0.14
GGU- (UGUGCCCUUCU)	0.1	81 ± 8	1.06 ± 0.09
GUG- (UGUUCCCCGUCU)	0.1	239 ± 20	0.98 ± 0.07
UGG- (UUUGCCCGUCU)	0.1	267 ± 20	0.99 ± 0.08
GUU- (UGUUCCCCUUCU)	0.1	880 ± 91	1.03 ± 0.16
UGU- (UUUGCCCUUCU)	0.1	812 ± 91	1.08 ± 0.33
UUG- (UUUUCCCCGUCU)	0.1	547 ± 55	1.05
UUU- (UUUUCCCCUUCU)	0.1	No detected binding	

^aReported values reported as the mean ± standard deviation from three experiments, except UUU (one experiment).

Table 2

Statistics for the NMR Structure Determination

NMR-derived restraints	
¹ H- ¹ H distance restraints	
Intraresidue	41
Sequential ($ i-j = 1$)	49
Medium range ($ i-j = 2-5$)	43
Long range ($ i-j > 5$)	37
Intermolecular	29
Hydrogen-bond restraints (4 per H-bond)	176
Zinc-ligand restraints	69
Total number of restraints	444
Number of restrained residues	24
NOE restraints per restrained residue	14.9
Total restraints per restrained residue	35.3

CYANA Statistics	Mean ± Std. Dev.	Max ± Std. Dev.
Target Function (Å ²)	0.86 ± 0.25	1.16
Restraint violations		
Upper distance viol. (Å)	0.029 ± 0.002	0.21 ± 0.05
Lower distance viol. (Å)	0.018 ± 0.002	0.11 ± 0.01
Sum of VDW viol. (Å)	3.9 ± 0.6	0.35 ± 0.23
Torsion angle viol. (deg.)	0.146 ± 0.381	0.25 ± 0.65

AMBER Statistics (kCal/mol)	Generalized Born	Explicit Water
Average Amber energy	-5,150.9	-112,990.9
Average Restraint energy	5.0	5.6
Average Distance penalty	5.0	5.6
Average Torsion penalty	0	0
Procheck-NMR Statistics		
Number of structures	20	20
Residues analyzed	13-31, 35-49	13-31, 35-49
Residues in Most Favored Regions	462 (82.5%)	469 (83.8%)
Residues in Additional Allowed Regions	97 (17.3%)	90 (16.1%)
Residues in Generously Allowed Regions	1 (0.2%)	1 (0.2%)
Residues in Disallowed Regions	0 (0.0%)	0 (0.0%)
Convergence Statistics (Pairwise RMSD)		
F1 (residues 13-28; backbone)	0.27	0.32
F2 (residues 34-49; backbone)	0.31	0.34
F1-F2 (residues 13-49; backbone)	2.41	1.77