BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 6341–6351 6341
DOI: 10.1021/bi100660r

Binding Characteristics of Small Molecules That Mimic Nucleocapsid Protein-Induced Maturation of Stem-Loop 1 of HIV-1 RNA†

Janet Chung,‡ Nikolai B. Ulyanov,§ Christophe Guilbert,§ Anwer Mujeeb,§,∥ and Thomas L. James*,§

‡*Graduate Group in Chemistry and Chemical Biology and* §*Department of Pharmaceutical Chemistry, University of California, San Francisco, 600 16th Street, San Francisco, California 94158-2517* ∥ *Current address: California HIV/AIDS Research Program (CHRP), University of California, Office of the President, 300 Lakeside Dr., Sixth Floor, Oakland, CA 94612-3550.*

Received April 28, 2010; Revised Manuscript Received June 12, 2010

ABSTRACT: As a retrovirus, the human immunodeficiency virus (HIV-1) packages two copies of the RNA genome as a dimer in the infectious virion. Dimerization is initiated at the dimer initiation site (DIS) which encompasses stem-loop 1 (SL1) in the 5′-UTR of the genome. Study of genomic dimerization has been facilitated by the discovery that short RNA fragments containing SL1 can dimerize spontaneously without any protein factors. On the basis of the palindromic nature of SL1, a kissing loop model has been proposed. First, a metastable kissing dimer is formed via standard Watson−Crick base pairs and then converted into a more stable extended dimer by the viral nucleocapsid protein (NCp7). This dimer maturation in vitro is believed to mimic initial steps in the RNA maturation in vivo, which is correlated with viral infectivity. We previously discovered a small molecule activator, Lys-Ala-7-amido-4-methylcoumarin (KA-AMC), which facilitates dimer maturation in vitro, and determined aspects of its structure−activity relationship. In this report, we present measurements of the binding affinity of the activators and characterization of their interactions with the SL1 RNA. Guanidinium groups and increasing positive charge on the side chain enhance affinity and activity, but features in the aromatic ring at least partially decouple affinity from activity. Although KA-AMC can bind to multiple structural motifs, the NMR study showed KA-AMC preferentially binds to unique structural motifs, such as the palindromic loop and the G-rich internal loop in the SL1 RNA. NCp7 binds to SL1 only 1 order of magnitude more tightly than the best small molecule ligand tested. This study provides guidelines for the design of superior small molecules that bind to the SL1 RNA that have the potential of being developed as an antiviral by interfering with SL1−NCp7 interaction at the packaging and/or maturation stages.

As with other retroviruses, human immunodeficiency virus type 1 (HIV-1)[1] packages two homologous copies of its RNA genome into the virion. These RNA molecules form a noncovalent dimer in the 5′-untranslated region (5′-UTR) of the genome (*1*). The dimeric nature of the genome is important in many steps of the replication cycle, including reverse transcription, recombination, and packaging, and, consequently, for infectivity (reviewed in ref *2*); it also enhances the capacity for genetic variability in the virus (*3*). Molecular understanding of the genomic dimerization has been facilitated by the discovery of a short synthetic RNA fragment corresponding to the dimer initiation site (DIS) in the 5′-UTR that can dimerize spontaneously in vitro (*4*−*8*). DIS is located in stem-loop 1 (SL1) of the HIV-1 5′-UTR; it is the minimal RNA domain required to initiate dimerization in vitro (*5*, *7*) and in vivo (*9*). SL1 (Figure 1a) contains a six-nucleotide (nt) palindromic sequence in its apex, which is usually flanked by two purines at the 5′-end and one purine at the 3′-end, especially in the A and B subtypes of HIV-1. An 11 bp stem of SL1 contains a 1−3 G-rich internal loop (GRIL) G·AGG (*5*, *7*). Although there are 64 possibilities for a 6 nt autocomplementary sequence, only two sequences are common, GCGCGC (e.g., in the Lai isolate) and GUGCAC (e.g., in the Mal isolate) (*10*), highlighting the importance of these particular palindromic sequences; furthermore, the palindrome is always six nucleotides. In addition, the flanking purines are also strongly conserved, and mutations or deletions in this region strongly impair dimerization and consequently infectivity (*4*, *11*). Although the exact sequence of the stem does not appear to be as important, maintaining base pairing is essential (*4*), and the sequences of the upper 7 bp portion of the stem, CUUGCUG·CAGCAAG or CUUGCUG·CGGCAAG, also show strong conservation (*12*). Finally, the GRIL motif G·AGG and the lower 4 bp portion of the stem, CUCG·CGAC, are the most conserved features of SL1 among HIV-1 isolates (*12*).

On the basis of the palindromic nature of the 9 nt loop, a kissing loop model of dimerization has been proposed (*5*, *7*) (Figure 1b). The two RNA molecules first establish six Watson−Crick base pairs within the palindrome forming a metastable kissing dimer (KD). The resulting KD is stable at physiological

†This work was supported by Grant AI046967 from the National Institutes of Health to T.L.J.

*To whom correspondence should be addressed. Telephone: (415) 476-1569. Fax: (415) 502-8298. E-mail: james@picasso.ucsf.edu.

[1]Abbreviations: HIV-1, human immunodeficiency virus type 1; SL1, stem-loop 1 RNA; NCp7, proteolytic nucleocapsid protein product of the Gag protein; KA-AMC, Lys-Ala-7-amido-4-methylcoumarin; R-AMC, Arg-7-amido-4-methylcoumarin; RR-AMC, Arg-Arg-7-amido-4-methylcoumarin; R-MHQ, Arg-7-amido-4-methyl-2-hydroxyquinoline; 5′-UTR, 5′-untranslated region; DIS, dimer initiation site; H-bond, hydrogen bond; KD, kissing dimer; ED, extended dimer; SAR, structure−activity relationship; NMR, nuclear magnetic resonance; SL1-fl, full-length SL1 RNA; PAGE, polyacrylamide gel electrophoresis; UV, ultraviolet; TBE, Tris-borate-EDTA buffer; TBM, Tris-borate-$MgCl_2$ buffer; K_d, dissociation constant; PDB, Protein Data Bank; 1D, one-dimensional; 2D, two-dimensional; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; T_1, spin−lattice relaxation time.

Published on Web 06/21/2010

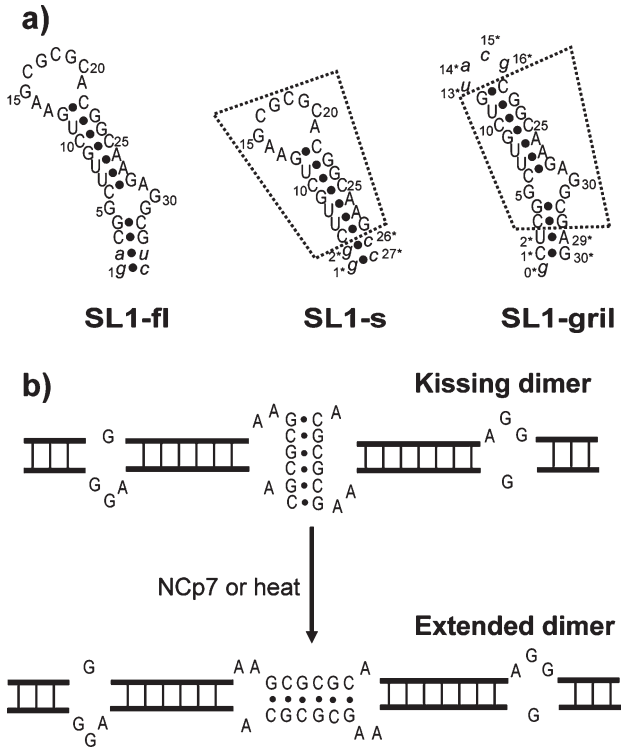

FIGURE 1: SL1 constructs and the kissing loop model of dimer maturation. (a) SL1 of the HIV-1 packaging signal and its fragments used in this study. SL1-fl RNA corresponds to the full-length 35 nt SL1 RNA with two terminal base pairs (shown in lowercase) reversed in order to optimize the in vitro transcription yield. SL1-s RNA includes the palindromic loop and the upper stem of SL1 but not the G-rich internal loop; two G·C base pairs were added to the bottom of the lower stem to increase its stability. SL1-gril RNA consists of the stems and the G-rich internal loop of SL1 with an apical UACG loop added to reduce dimerization during folding and a dangling G residue added at the 5′-end to increase the transcription yield. Dashed boxes enclose wild-type residues from SL1; non-wild-type residues are shown in lowercase and indicated with an asterisk. Two additional constructs were used in this study. SL1-yf consisted of the sequence of SL1-fl with the three flanking purines (adenines) being replaced with pyrimidines (uracils), and SL1-yb consisted of SL1-fl with the GRIL purines all replaced with pyrimidines (two cytosines and two uracils). (b) Two-step model of SL1 dimerization. A metastable kissing dimer (top) is formed using the self-complementary sequence in the apical loops of SL1 monomers, and it can be converted into a more stable extended dimer (bottom) by heat or viral NCp7 protein.

temperature (37 °C), but it is converted into a more stable extended dimer (ED) upon incubation at 55 °C (*8*, *13*) or incubation at room temperature with the viral nucleocapsid protein (NCp7) (*14*), a proteolytic fragment of the Gag polyprotein. This in vitro NCp7-induced dimer conversion apparently mimics part of the maturation process that takes place in vivo, induced by the viral protease cleaving Gag, and is highly correlated with infectivity. Two HIV-1 genomic RNA molecules are selected in the cytoplasm for packaging into the budding virion (*15*), and then the RNA dimer undergoes a series of transformations during maturation inside the newly budded viral particles (*16*). RNA dimers extracted from immature (protease-deficient) viral particles are less thermostable than those from the mature virions (*17*). It is likely that general chaperone properties of NCp7 (*18*, *19*) contribute to the dimer conversion; however, the detailed molecular mechanism of the conversion remains unclear.

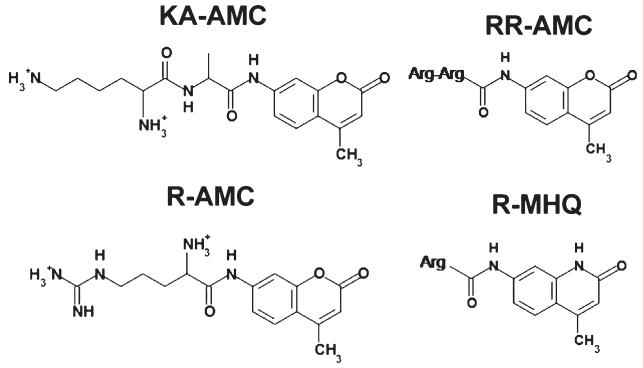

FIGURE 2: Small molecules used for RNA binding: Lys-Ala-7-amido-4-methylcoumarin (KA-AMC), Arg-7-amido-4-methylcoumarin (R-AMC), Arg-Arg-7-amido-4-methylcoumarin (RR-AMC), and Arg-4-methyl-2-hydroxyquinoline (R-MHQ).

In addition to the RNA dimerization, SL1 is also involved in the packaging of the genomic RNA into virions. SL1 and three other stem-loop structures, SL2, SL3, and SL4, comprise a so-called packaging Ψ site; SL1, SL3, SL4, and possibly SL2 together with the NCp7 domain within the Gag polyprotein are required for the packaging (*9*, *20*). NCp7, either as a Gag domain or as a mature protein, binds each of the SL1−SL4 structures in vitro (*21*−*23*). Within SL1, there are two independent determinants specifically recognized by NCp7 in vitro: palindromic apical region with flanking purines and the G-rich internal loop; the apparent dissociation constants for the mature NCp7 binding to the two determinants are 100 and 140 nM, respectively (*23*).

During our experimental screening that aimed to find a small molecule to inhibit the NCp7-mediated SL1 dimer maturation that could serve as a novel HIV-1 therapeutic lead, we unexpectedly found a small molecule activator, Lys-Ala-7-amido-4-methylcoumarin [KA-AMC (Figure 2)], that facilitated the in vitro dimer maturation instead with an effective concentration EC_{50} of 570 μM (*24*). One of the motivations of this study was to investigate whether the activation effectivity of the SL1 maturation by small molecules is correlated with their binding affinity. We used NMR to probe structural features of KA-AMC binding to SL1 RNA and took advantage of fluorescent properties of coumarin to determine the apparent K_d values of the binding of KA-AMC and several related molecules to SL1. One of these molecules, Arg-Arg-7-amido-4-methylcoumarin (RR-AMC), binds SL1 with submicromolar affinity, less than 1 order of magnitude less tightly than NCp7. This study provides insights for discovering and optimizing small molecules inhibiting NCp7−SL1 interaction and thus targeting the HIV-1 RNA dimer maturation and/or HIV-1 RNA packaging.

MATERIALS AND METHODS

RNA Synthesis. All RNA oligonucleotides were synthesized by in vitro transcription (*25*) from a synthetic DNA template (Integrated DNA Technology, Inc.) with T7 RNA polymerase prepared in-house (*26*). The oligonucleotides were then purified by 20% denaturing polyacrylamide gel electrophoresis (PAGE) followed by electroelution, dialyzed in sterile water, and stored at −20 °C. All RNAs were quantified using ultraviolet (UV) spectrophotometry with an extinction coefficient estimated with an online calculator to account for sequence (Ambion, Inc.).

RNA Sample Preparation and Dimerization Assay. A 3 μM aliquot of SL1-fl RNA or its mutants or a 4 μM aliquot of SL1-s RNA (see Figure 1a for nomenclature and sequences of

SL1 constructs and the Supporting Information for sequences of SL1 mutants) in sterile water was heated to 90 °C for 5 min and then snap-cooled on ice for 5 min. An equal volume of prechilled 2× dimerization buffer [20 mM potassium phosphate (pH 7.0), 100 mM NaCl, and 0.2 mM $MgCl_2$] was added, and the sample was incubated on ice for an additional 30 min. Finally, the sample was equilibrated at room temperature for 2 h to fold into the KD conformation. The same RNA constructs were also folded into the ED conformation by being heated to 90 °C followed by slow cooling to room temperature. SL1-yf consists of the sequence of SL1-fl with the three flanking adenines replaced with pyrimidines (uracils), and SL1-yb consists of SL1-fl with the GRIL purines all replaced with pyrimidines (C-UCU). SL1-yf and SL1-yb RNAs were folded into the KD conformation similarly, except for the 3 min heating and snap-cooling periods. The native PAGE assay was used to distinguish KD and ED on the basis of their Mg^{2+} dependence (*27*−*29*). The metastable KD remains a dimer during PAGE only in the presence of Mg^{2+} ions. It dissociates and migrates as a monomer when PAGE is performed in 1× Tris-borate-EDTA (TBE) buffer [89 mM Tris, 89 mM borate (pH 8.4), and 2 mM EDTA], with EDTA as a Mg^{2+} chelator. On the other hand, KD runs as a dimer in the presence of Mg^{2+} ions, such as in the 1× Tris-borate-$MgCl_2$ (TBM) buffer [89 mM Tris, 89 mM borate (pH 8.4), and 3 mM $MgCl_2$]. The more stable ED does not need Mg^{2+} ions to stay as a dimer in the gel and runs as a dimer band under either electrophoresis condition. By performing native PAGE in 1× TBE buffer, we can distinguish these two dimer forms and check the folding state of the RNA sample.

SL1-gril RNA in water was lyophilized and dissolved in 1× dimerization buffer [10 mM potassium phosphate (pH 7.0), 50 mM NaCl, and 0.1 mM $MgCl_2$] at a concentration of ~10 μM and then folded by being heated to 90 °C for 5 min followed by being snap-cooled on ice for 30 min. The folding state of the sample was then checked by native PAGE. After folding, RNA samples were concentrated to a desired concentration with a centrifugal device (Amicon, 3000 MWCO) for fluorescent emission quenching experiments or for NMR studies.

To study the specificity of small molecules binding, we also used HIV-1 SL2 and SL3 and a 18 nt duplex RNA (DUP RNA), consisting of residues 11−17 of the wild-type SL1 sequence preceded by two guanine residues, with the remaining nine residues being self-complementary to the first nine (GGUGAAGCGCG-CUUCACC) to be fully palindromic (see the Supporting Information). The SL2, SL3, and DUP RNA samples in water were lyophilized and dissolved in 1× dimerization buffer to a final concentration of 300−400 μM and folded by being heated to 90 °C and slowly cooled to room temperature. SL2 and SL3 RNA constructs were shown to fold as a monomer previously (*30*, *31*).

Native Gel Electrophoresis. Native PAGE experiments were conducted using a 10% bis:acrylamide (19:1) ratio of matrix at 25 °C in TBE buffer. A native dye, bromophenol blue, containing 10% glycerol was used in each case. Gels were run until the bromophenol blue ran to the bottom at 150 V. Gels were then stained with ethidium bromide and visualized with UV light.

Apparent Dissociation Constant Measurement Using Fluorescence Emission Quenching of Coumarin and Hydroxyquinoline Derivatives. The following small molecules were used for the fluorescence quenching experiments (Figure 2): Lys-Ala-7-amido-4-methylcoumarin (KA-AMC), Arg-7-amido-4-methylcoumarin (R-AMC), Arg-Arg-7-amido-4-methylcoumarin (RR-AMC), and Arg-4-methyl-2-hydroxyquinoline (R-MHQ). KA-AMC and R-AMC were purchased from Sigma; RR-AMC was purchased from Bachem, and R-MHQ was synthesized (*24*). A 300 μL aliquot of a small molecule at a concentration of 1 μM in 1× dimerization buffer was placed in a cylindrical microfluorescence cell (Horiba). The initial emission fluorescence of the coumarin derivatives and R-MHQ was measured at 390 and 364 nm, respectively, with a FluoroMax3 fluorescence spectrometer (Horiba) using an excitation wavelength of 323 nm. Aliquots of RNA ranging from 300 to 400 μM were added to the sample for the "reverse" titration. The samples were mixed briefly, and the fluorescence was measured. The fluorescence quenching data were fitted to the following equation (*32*) by varying parameters θ_b and K_d with Microsoft Excel:

$$\theta = 1 - (1 - \theta_b)\left[R_T + L_T + K_d - \sqrt{(R_T + L_T + K_d)^2 - 4R_T L_T}\right]/(2L_T) \quad (1)$$

where θ ($= F/F_0$) is the ratio of the observed fluorescence intensity F to the initial intensity F_0, θ_b ($= F_b/F_0$) is a similar ratio at the saturating level of binding, R_T and L_T are the total concentrations (bound plus unbound) of RNA and ligand, respectively, and K_d is the dissociation constant.

NMR Experiments for Monitoring Binding of KA-AMC to SL1 RNA. All NMR experiments were performed on a Varian INOVA 600 MHz NMR spectrometer. Spectra in D_2O were recorded at either 25 or 30 °C, and spectra in a 90% H_2O/10% D_2O mixture were acquired at 10 °C. Samples of SL1-fl in either KD or LD forms, SL1-s in KD form, and SL1-gril had concentrations ranging from 250 to 400 μM in RNA strands and varying concentrations of KA-AMC in 1× dimerization buffer. In the case of H_2O samples, water suppression was achieved using a symmetrically shifted shaped pulse (*33*) for the acquisition of two-dimensional (2D) NOESY spectra and using the "watergate" method (*34*) for the acquisition of one-dimensional (1D) spectra. Spectra for samples in D_2O were recorded without any suppression of the residual HDO signal. Changes in RNA proton chemical shifts upon addition of varying concentrations of KA-AMC were monitored in 1D spectra (for imino protons), 1D traces of T_1 inversion recovery experiment (for selective detection of slowly relaxing adenine H2 protons), and 2D NOESY and TOCSY spectra (for other aromatic and sugar protons). Partial proton chemical shift assignments of SL1-fl, SL1-s, and SL1-gril alone and in complexes with KA-AMC were achieved on the basis of homonuclear 2D NOESY (recorded with mixing times varying from 50 to 600 ms) and 2D TOCSY spectra with the help of published assignments for SL1-fl in the LD form (*35*). 1D data were processed and analyzed with vnmr (Varian); 2D data were processed with NMRpipe (*36*) and analyzed with SPARKY (*37*).

Molecular Modeling of a KA-AMC−*SL1 RNA Complex*. Molecular docking of KA-AMC to SL1 RNA was performed using MORDOR (*38*). For docking to the flanking A's site in the KD form, the receptor structure was based on the NMR structure of the KD SL1 RNA lacking the G-rich internal loop [PDB entry 2D19 (*39*)]. The palindromic sequence GCGC-GC was substituted for GUGCAC in 2D19, and the structure was energy-minimized with miniCarlo (*40*) before it was passed to MORDOR. For docking to the GRIL motif, the receptor structure was based on the NMR structure of the full-length SL1 RNA in the ED form [PDB entry 2GM0 (*35*)]. The sequence of the lower stem CUCG·CGAG was substituted for GACG·CGUC in PDB entry 2GM0, and the structure was energy-

minimized. All calculations were done separately for models 1, 2, and 7 of the 2GM0 NMR ensemble, which represent distinct GRIL conformations.

The receptor and ligand preparation and docking procedures with the flexible receptor and ligand were similar to those previously described (*38*, *41*), except that the RNA receptor was allowed more flexibility. For this purpose, distance restraints between heavy atoms compatible with an envelope of A-RNA conformations were imposed on RNA stems on both sides of the binding pocket (ca. 7.4 restraints per residue with an average flat-well width of 0.5 Å), while the binding pocket itself was unrestrained. Compared to using atomic root-mean-square deviation restraints imposed on the overall RNA conformation (*38*), this method allows hinge motions around the binding pocket and a greater flexibility inside the pocket. In addition, the generalized Born method (*42*) was used to account for solvation during some of the docking runs. Between 5000 and 10000 binding poses were generated for each RNA receptor and docking run as described previously (*38*). The docked RNA−KA-AMC complexes were scored according to their interaction energy consisting of van der Waals, electrostatic, and generalized Born solvation energy terms. Approximately 1% of the best scoring poses were examined manually using the ViewDock option of UCSF Chimera (*43*) and then further assessed on the basis of qualitative agreement with NMR observations.

RESULTS

Design of RNA Constructs. SL1 folds into a stem-loop structure with an apical 9 nt loop and a stem containing a 1−3 internal loop (*5*, *7*). Both the apical loop and the internal loop are strongly conserved among HIV-1 isolates; they bind NCp7 in vitro, and mutations in these regions impair viral infectivity. Full-length SL1 [SL1-fl (Figure 1a)] is the same 35 nt RNA construct that we have used previously (*12*, *24*, *35*). It has the sequence corresponding to the wild-type Lai HIV-1 isolate, except for the inversions in the two terminal base pairs, which enabled optimization of the in vitro transcription yield. We have also designed two truncated constructs, each containing one of the important features of SL1, namely, SL1-s ("stable" SL1) and SL1-gril (monomerized SL1 with the GRIL motif) (Figure 1a). The SL1-s construct lacks GRIL but contains the palindromic apical loop and the upper stem of the full-length construct with two additional GC base pairs at the terminus to increase stem stability. Increasing stem stability reduces the likelihood of KD-to-ED dimer conversion during experiments and simplifies NMR data analysis significantly.

SL1-gril contains the stem-G-rich internal loop-stem motif of wild-type SL1 with the addition of an apical UACG tetraloop. The tetraloop increases the construct stability, reduces the level of dimerization during folding, and furthermore provides a unique signature in NMR (*44*). In addition, we used the SL1-yf and SL1-yb mutants of SL1-fl (*12*). In SL1-yf (see the Supporting Information), the adenines flanking the palindrome, i.e., A13, A14, and A21, were substituted with uracils. In SL1-yb, the 1−3 internal loop G-AGG was substituted with an all-pyrimidine 1−3 internal loop C-UCU. Finally, to study the specificity of the small molecules binding, some experiments were conducted for the HIV-1 SL2 and SL3 stem-loops and for an 18 bp duplex RNA. The self-complementary RNA sequence used to form the duplex, GGUGAA-**GCGCGC**UUCACC, has the SL1 palindromic hexanucleotide in the middle (shown in bold). The wild-type sequence GGCGACU**GGUG**AGUA**C**GCC was used for the SL2 construct (the tetraloop and the A-bulge are shown in bold). Three base pairs were added to the SL3 sequence ggaCUAGC**GGAG**GCUAGucc to optimize the yield of in vitro transcription (the tetraloop is shown in bold, and the non-wild-type residues are shown in lowercase).

Table 1: Apparent Dissociation Constants K_d (micromolar) for the Binding of Small Molecules to SL1-Related RNA Constructs[a]

RNA	KA-AMC	R-AMC	RR-AMC	R-MHQ
SL1-fl KD	19.3 (4.2)	5.2 (0.3)	0.5 (0.1)	3.5 (0.3)
SL1-s KD	50.0 (3.7)	12.6 (0.1)	1.4 (0.2)	5.9 (1.0)
SL1-gril	20.9 (0.2)	12.2 (0.6)	0.9	7.1 (0.8)
duplex RNA	122.1	82.2	16.5	98.5

[a]Apparent dissociations constants were determined on the basis of fluorescence quenching of small molecules upon RNA titration as explained in Materials and Methods. Standard deviations are shown in parentheses, determined in two to four independent experiments. The data for binding to duplex RNA and binding of RR-AMC to SL1-gril are preliminary, determined on the basis of a single experiment.

RNA constructs SL1-fl and SL1-s were folded as desired, in the KD or ED form, and constructs SL1-yf and SL1-yb were folded in the KD form, which was then verified using the magnesium-dependent native PAGE assay as described in Materials and Methods (data not shown). Native PAGE was also used to verify that SL1-gril, SL2, and SL3 were folded as monomers (data not shown).

Binding Affinity of Small Molecules for SL1 RNA by Fluorescence Emission Quenching. We used the fluorescence properties of the small molecules shown in Figure 2 to measure their binding affinity for SL1 RNA and its mutants. The fluorescence emission intensity is quenched when these small molecules bind to RNA. The apparent dissociation constants K_d were determined by fitting the fluorescence intensity binding curves to eq 1, which assumes a single binding site on the receptor, hence the apparent character of the K_d values. For monomeric constructs SL1-gril, SL2, and SL3, the total RNA concentration in eq 1 represents the concentration of the binding sites, and the resulting K_d values characterize the affinity for this site. In the case of dimeric RNA constructs, SL1-fl, SL1-s, SL1-yf, and SL1-yb, each symmetric dimer carries at least two equivalent binding sites. Because the RNA concentration was expressed in terms of RNA strands (which is again equivalent to the binding site concentration), the resulting K_d values characterize the affinity for each binding site, assuming that the two symmetric binding sites do not interact with each other. There are, however, two pairs of distinct binding sites on the full-length SL1-fl RNA construct, one associated with the palindromic region of SL1 and another associated with the GRIL motif (see also NMR chemical shift mapping below). The palindromic and GRIL binding sites are dissociated from each other in the SL1-s and SL1-gril constructs. If all four (two pairs) binding sites had equal affinity and were noninteracting, then the apparent K_d values measured for SL1-fl must be one-half of those for SL1-s and SL1-gril, which is indeed approximately the case for R-AMC, RR-AMC, and R-MHQ (Table 1). However, in the case of KA-AMC, the GRIL motif appears to be the dominant binding site with the apparent affinity for SL1-gril being more than twice as strong as the affinity for SL1-s (Table 1).

Matters are further complicated by the fact that each of the four small molecules studied here binds to the duplex RNA, albeit with lower affinities (Table 1). This was expected, of course, because of the negative charges on the RNA backbone and positive charges on small molecules; this further highlights the apparent character of the measured K_d values. We must note, however, that this is also the case for NCp7, which also binds

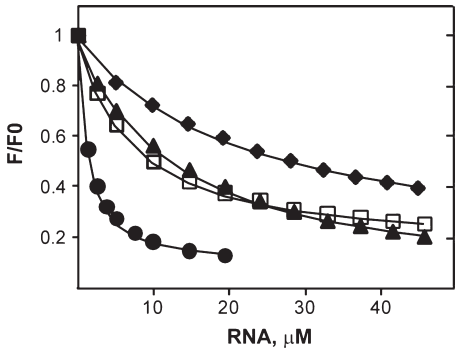

FIGURE 3: Representative fluorescence quenching curves for small molecules binding to SL1-gril RNA: KA-AMC (◆), R-AMC (▲), R-MHQ (□), and RR-AMC (●). The lines show the results of the fitting to eq 1.

Table 2: Specificity of Small Molecules Binding to SL1-Related RNA Constructs[a]

RNA	KA-AMC	R-AMC	RR-AMC	R-MHQ
SL1-fl KD	6.3	15.8	33.0	28.1
SL1-s KD	2.4	6.5	11.8	16.7
SL1-gril	5.8	6.7	18.3	13.9

[a]Defined as the ratio of the apparent K_d for the binding to duplex RNA and apparent K_d for the binding to the SL1-related RNA construct (Table 1).

nonspecifically RNA and DNA stems (see, e.g., refs *19* and *45*) in addition to specific binding to a number of structural RNA motifs; the measured affinities of NCp7 for such motifs must therefore also be considered as apparent (*23*). This parallel is important, as it allows a direct comparison of apparent affinities of the small molecules measured here and apparent NCp7 affinities (*23*).

Figure 3 shows the results of fitting of the fluorescence binding curves measured in representative experiments for SL1-gril. Despite the fact that the small molecules bind to the stems in addition to the G-rich internal loop, the curves can be fit reasonably well to eq 1, probably because the continuing ligand binding at higher RNA concentrations is absorbed by the apparent value of $\theta_b = F_b/F_0$.

Each of the four small molecules binds SL1-related RNA constructs in the low micromolar or submicromolar range (Table 1), with the apparent dissociation constants decreasing in the following order: KA-AMC > R-AMC > R-MHQ > RR-AMC for each RNA. The arginine guanidinium group makes a ligand a tighter binder compared to the localized amino group on lysine, and increasing the ligand positive charge also improves the affinity. This trend parallels the order of effectivity of KA-AMC, R-AMC, and RR-AMC in promoting the KD-to-ED conversion, but this order is reversed for R-AMC and R-MHQ (*24*). Substituting lysine for arginine and increasing the positive charge also lead to tighter nonspecific binding to duplex RNA (ranging from 122 μM for KA-AMC to 16 μM for RR-AMC), which is expected due to the negatively charged RNA backbone. What was not expected is that the specificity of binding also increases in approximately the same order [defined as the ratio of the apparent K_d values for the binding to the duplex RNA compared to the binding site of interest, either palindromic loop or GRIL (Table 2)]. For the two tighter ligands, R-MHQ and RR-AMC, the specificity is a factor of ~10. At the same time, the apparent binding affinity of the tightest ligand, RR-AMC, approaches that of NCp7 binding to the same RNA sequences (*23*), with a factor of 14 difference for the palindromic loop (SL1-s) and a factor of 6 for the G-rich internal loop (SL1-gril). RR-AMC and KA-AMC bind somewhat better (by factors of 1.6 and 2.4, respectively) to GRIL than to the palindromic loop, while R-AMC and R-MHQ bind both sites equally within experimental error (Table 1).

Table 3: Apparent Dissociation Constants K_d (micromolar) for the Binding of KA-AMC to Various RNA Constructs[a]

SL1-fl ED	SL1-s ED	SL1-yf KD	SL1-yb KD	SL2	SL3
22.6 (2.5)	47.6 (2.1)	22.9 (1.1)	26.3 (1.9)	47.5 (2.4)	65.1 (0.4)

[a]Apparent dissociation constants were determined on the basis of fluorescence quenching of small molecules upon RNA titration as explained in Materials and Methods. Standard deviations are shown in parentheses, determined in two to four independent experiments.

KA-AMC binds the KD and ED isomers with similar affinities, with apparent K_d values of 19−23 and 48−50 μM for SL1-fl and SL1-s, respectively (compare Tables 1 and 3). Under the experimental conditions used in this study, the four small molecules do not convert KD into ED; we verified this for RR-AMC (data not shown), which was shown to be the most efficient conversion activator of the four ligands (*24*). Indeed, such a conversion was not expected, because the KD-to-ED conversion requires much higher, submillimolar concentrations of the small molecules (*24*).

In addition to SL1 binding and nonspecific binding to the RNA duplex, KA-AMC also binds to SL2 and SL3 HIV-1 stem-loop constructs (Table 3) with an apparent affinity similar to that of the palindromic loop (SL1-s). Interestingly, NCp7 has similar binding properties; in fact, it binds SL2 and SL3 more tightly (K_d values of 23 and 28 nM, respectively) than it binds the palindromic loop and GRIL (K_d values of 100 and 140 nM, respectively) (*23*). KA-AMC also binds the KD forms of a number of SL1 mutants, including the ones with the flanking adenines of the palindromic loop (SL1-yf) or purines of the GRIL motif (SL1-yb) changed to pyrimidines, with an apparent affinity similar to that for SL1-fl (Table 3). While the affinities of NCp7 for these mutants have not been measured, NCp7 is likely to bind them as well, because it promotes their conversion from KD to ED form as readily as it does for SL1-fl (*12*).

Chemical Shifts and Line Width of KA-AMC Protons upon Binding to RNA. Proton NMR resonances of KA-AMC undergo significant shifts and line broadening upon binding to SL1-fl KD (*24*). Similar effects are also observed for the KA-AMC binding to SL1-s KD (Supporting Information) and SL1-gril (Figure 4). Resonances of protons on the aromatic ring of KA-AMC experience very strong upfield shifts, indicative of ring stacking with bases on RNA, with the effect being even more pronounced for binding to SL1-gril than to SL1-s KD (Supporting Information). Resonances of the alanine side chain protons display a modest upfield shift, and lysine protons display a modest downfield shift upon binding. The line broadening of KA-AMC protons suggests an exchange process at the intermediate NMR time scale, either between bound and unbound states or between different binding modes. The line broadening is especially severe for the ring protons of KA-AMC, which is correlated with the large changes in chemical shifts for these protons. The observed moderate shift of the resonances of the ring KA-AMC protons back in the downfield direction with increasing concentrations of the added ligand (Figure 4d−f) is most likely due to the simultaneous increase in

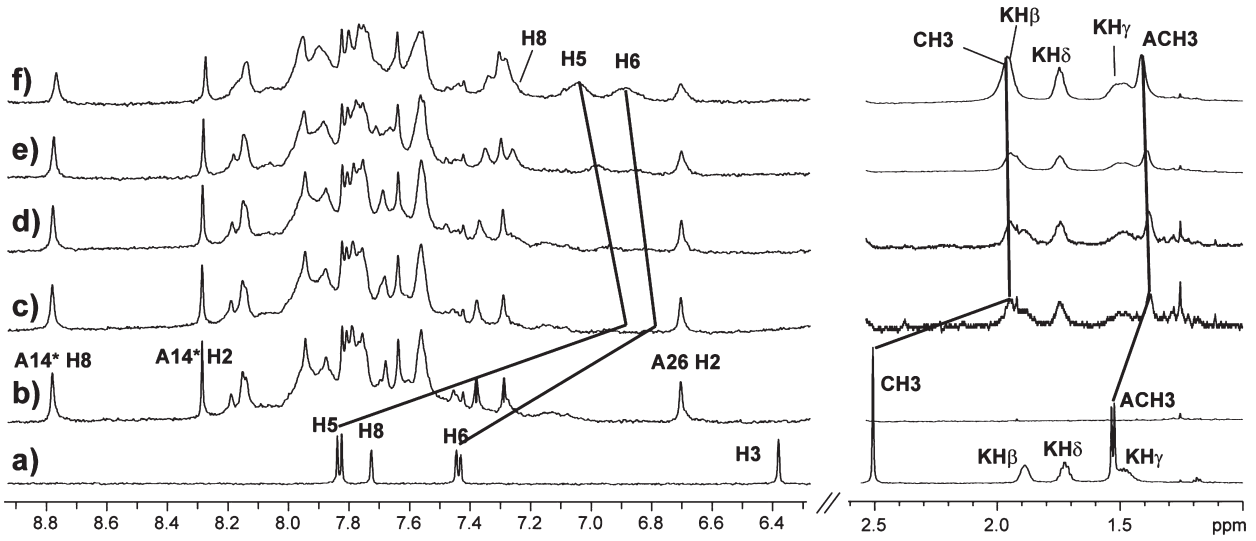

FIGURE 4: 1D proton NMR spectra at 25 °C of KA-AMC (0.5 mM) and SL1-gril RNA (280 μM). (a) KA-AMC alone, (b) SL1-gril alone, and the following SL1-gril−KA-AMC complexes: (c) 1:0.25, (d) 1:0.5, (e) 1:1, and (f) 1:2. The vertical scale in the traces in the right panel of parts c and d is increased compared with those of the rest of the traces.

the fraction of unbound ligand. At the same time, there is no detectable ligand line sharpening at higher ligand concentrations, which would be expected if the line broadening was entirely due to intermediate exchange between bound and free states. Further, the RNA proton signals continue to change with the continuing titration of the ligand (Figure 4). In summary, these results strongly suggest the existence of exchange between multiple binding modes.

RNA Chemical Shift Mapping of Interactions of KA-AMC with SL1 Constructs. To identify the KA-AMC binding sites and characterize ligand−RNA interactions, we monitored changes in proton NMR chemical shifts of SL1-fl KD and ED, SL1-s KD, and SL1-gril during the KA-AMC titration. Chemical shifts of unbound SL1-fl ED have been assigned previously (*35*), and chemical shifts of other unbound RNAs were partially assigned using homonuclear NMR techniques. As expected, chemical shifts of residues of the palindromic loop and flanking adenines were similar in SL1-fl KD and SL1-s KD, and chemical shifts or residues in or near the GRIL motif were similar in SL1-fl KD and SL1-fl ED (data not shown). However, GRIL residues had somewhat different chemical shifts in the SL1-gril and SL1-fl constructs; G4 and G5 fingerprint cross-peaks could not be assigned in SL1-gril using homonuclear NMR due to the spectral overlap (data not shown). These differences are likely to reflect somewhat different structures and/or dynamics at this site and must be due to the different sequence of the lower 4 bp stem in these RNA constructs (Figure 1a): the sequence is wild-type with the dangling G at the 5′-end in SL1-gril, and it has a blunt end with two reversed base pairs in SL1-fl. The solution conformation of this site is known to be sensitive to relatively minor changes in the sequence (see ref *35*).

We followed the changes in proton chemical shifts of RNA constructs upon addition of KA-AMC using well-resolved regions of various NMR spectra: imino protons in 1D spectra in water, pyrimidine H5−H6 cross-peaks, and occasional H2′−H1′ cross-peaks (for flexible residues and residues in the C2′-endo conformation) in 2D TOCSY spectra. The adenine H2 resonances were monitored using the T_1 inversion recovery 1D traces with an appropriate delay after the inversion 180° pulse, which allowed for the recovery of most aromatic protons but kept the H2 resonances negative due to a longer T_1 relaxation time for these protons. In addition, we followed the H1′−H6/H8 fingerprint cross-peaks in 2D NOESY spectra for SL1-s and SL1-gril. Some of the NMR spectra are shown in the Supporting Information, and the results are summarized in Figure 5. Figure 5 shows maximum absolute values of changes in chemical shifts among protons in each RNA residue upon KA-AMC titration to SL1-fl KD, SL1-s KD, and SL1-gril.

It is clear that most RNA residues are affected at least to some extent by addition of KA-AMC, even if it is only line broadening for some residues. Part of this may relate to the capability of KA-AMC to bind even to the RNA duplex (Table 1). In particular, the RNA termini are affected by ligand binding (see, e.g., G1 H2′−H1′ cross-peak in Figure 6a), consistent with the coumarin ring stacking to the stem end, but the interiors of RNA stems are affected as well, especially at higher ligand ratios (Figure 5a). Nevertheless, the maximum chemical shift changes occur for the residues at or near two distinct sites: the palindromic loop with flanking adenines (SL1-fl KD and ED and SL1-s KD) and the GRIL motif (SL1-fl KD and ED and SL1-gril), consistent with the higher apparent affinity of KA-AMC for these sites than for duplex RNA (Table 1). Indeed, the small shifts and broadening seen for residues in the interior of the stems may arise from the ligand binding to these two preferred sites. For the palindromic loop site in the KD form, SL1-fl KD and SL1-s KD, we observed KA-AMC-induced perturbation of the imino signals of U11, G23, G12, and G17 and of overlapping imino signals in the palindrome GCGCGC itself, H5−H6 TOCSY cross-peaks of C16, C20, and C22, H1′-H2′ TOCSY cross-peaks of G12 and A14, and A13 H8 proton and H2 protons in some of the flanking adenines resolved in the inversion recovery experiment (Supporting Information). Additionally, we compared 2D NOESY spectra of the free RNA versus that of a 2:2.25 RNA−KA-AMC complex for SL1-s KD and observed a loss of intraresidue H1′−H6/H8 NOE cross-peaks in residues G12, A13, G15, A21, and C22 in the complex and a loss of sequential H1′−H6/H8 NOE cross-peaks for the U11−G12, G12−A13, A13−A14, G15−C16, C20−A21, and A21−C22 connectivities. All these changes are centered at the region of the flanking adenines next to the palindromic loop. Loss of the

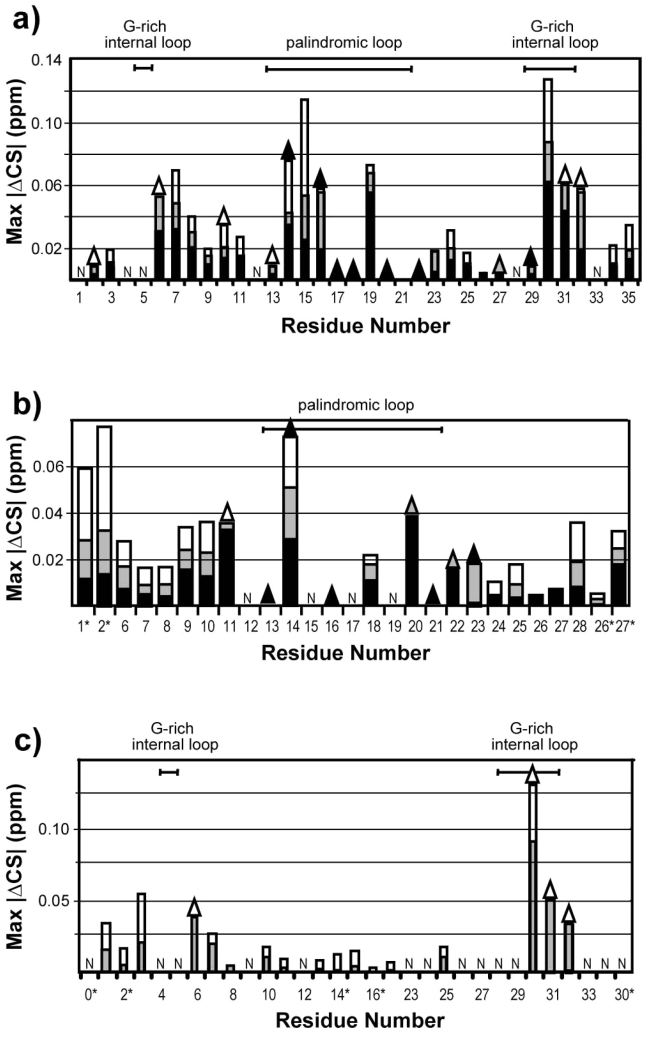

FIGURE 5: Changes in chemical shifts of SL1 constructs upon binding of KA-AMC. The titration data for SL1-fl KD (a), SL1-s KD (b), and SL1-gril (c) are presented as the absolute values of the maximum chemical shift differences (Max|ΔCS|). Data from the following NMR experiments were included in this index: 1D spectra in water, TOCSY and T_1 inversion recovery spectra in D_2O at 25 °C for SL1-fl; 1D spectra in water at 10 °C, TOCSY and T_1 inversion recovery spectra in D_2O at 30 °C for SL1-s; TOCSY in D_2O at 25 °C for SL1-gril. Water 1D spectra provide chemical shifts of imino protons H1 and H3; inversion recovery experiments provide adenine H2 chemical shifts, and TOCSY experiments provide chemical shifts of H5 and H6 protons in pyrimidines and H1′ and H2′ protons for riboses in the C2′-endo conformation. The SL1-fl and SL1-s data for the 2:1, 2:2, and 2:4 RNA−ligand complexes (using RNA strand concentrations) are shown as black, gray, and white bars, respectively. The SL1-gril data for the 1:1 and 1:2 RNA−ligand complexes are shown as gray and white bars, respectively. Black, gray, and white triangles indicate residues undergoing resonance broadening observed starting with the 2:1, 2:2, and 2:4 complexes, respectively. N indicates that information about a residue is not available for comparison. RNA regions corresponding to the palindromic loop and the GRIL motif are indicated above the graph. Residues are numbered according to Figure 1a.

sequential H1′−H6/H8 NOE cross-peaks may reflect changes in RNA structure upon KA-AMC binding. However, intraresidue H1′−H6/H8 NOEs are expected for any RNA conformation, and their loss can result from only exchange broadening of proton resonances. We also observed broadening of many other NMR signals (shown with triangles in Figure 5a,b), such as imino protons of G12, U11, and G23, overlapping imino signals in the palindrome, H5−H6 TOCSY cross-peaks in C16, C20, and C22, H1′−H2′ TOCSY cross-peak in G12, adenine H2 protons in the

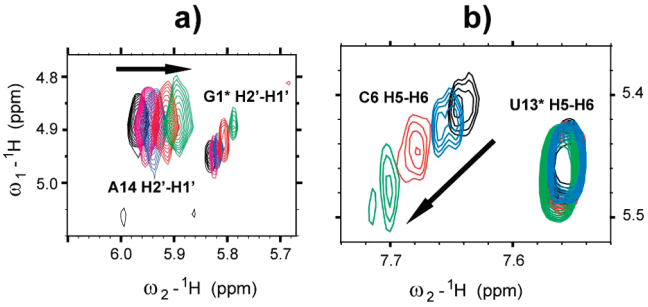

FIGURE 6: Examples of NMR line broadening for RNA protons upon addition of KA-AMC showing superposition of TOCSY spectra for various RNA:ligand ratios. (a) SL1-s KD complexes with KA-AMC. The RNA strand concentration was 400 μM. The spectrum of RNA alone is colored black, and spectra for the 2:0.5, 2:1, 2:2, and 2:4 RNA strand:ligand ratios are colored magenta, blue, red, and green, respectively. (b) SL1-gril with an RNA concentration of 280 μM. The spectrum of RNA alone is colored black, and spectra for the 1:0.5, 1:1, and 1:2 RNA:ligand ratios are colored blue, red, and green, respectively. Arrows show the direction of chemical shift changes for the A14 H2′−H1′ (a) and C6 H5−H6 (b) cross-peaks.

flanking A's region, and A13 H8. Interestingly, KA-AMC binding to another dimeric form, SL1-fl ED, caused a distinctly different pattern of chemical shift perturbations around the flanking A's site (data not shown), suggesting a distinct binding mode. Although line broadening was also observed for many residues in SL1-fl ED upon ligand binding, the imino protons in the palindromic GC pairs remained sharp (data not shown).

KA-AMC-induced perturbations in and near the GRIL site (in SL1-fl KD and ED, and in SL1-gril) include the H1′−H2′ cross-peaks in G30 and G31 and the H5−H6 cross-peaks in C3, C6, and C32. NMR signals of the tetraloop UACG in SL1-gril were not affected significantly during titration. In addition, we observed significant broadening of NMR signals for residues near the GRIL motif upon addition of KA-AMC, including the H1′−H2′ TOCSY cross-peaks in G30 and G31, H5−H6 TOCSY cross-peaks in C6 and C32, and many intra- and inter-residue H1′−H6/H8 cross-peaks in A29, G30, G31, and C32 (shown with triangles in Figure 5c). NMR line broadening at both the palindromic loop and GRIL sites suggests exchange processes at an intermediate NMR time scale, which could be associated either with bound and unbound conformations or with multiple binding modes of KA-AMC (see Discussion, below).

Docking. No reliable intermolecular NOE cross-peaks were observed in any of the KA-AMC complexes with SL1 RNA constructs due to severe line broadening for RNA and especially ligand signals. This made impossible any NMR-based structure determination. Therefore, we conducted several computational docking experiments to explore possible poses for KA-AMC binding to the flanking adenines and GRIL sites using the flexible docking program MORDOR (*38*). MORDOR permits both RNA and ligand flexibility, so they can bind to each other via induced fit while predicting complex structures that are energetically feasible. In most high-scoring (low-energy) poses for binding to the SL1-s receptor, the coumarin ring of KA-AMC forms stacking interactions with RNA bases, consistent with the strong upfield shift of the ring protons of the ligand as well as base protons. A number of stacking arrangements have resulted from the docking; the most prevalent motifs for binding to the flanking adenine region are shown in panels a and b of Figure 7. In both cases, A13 is displaced from its position in the free RNA (*39*) with the coumarin ring. In the first motif (Figure 7a), the coumarin

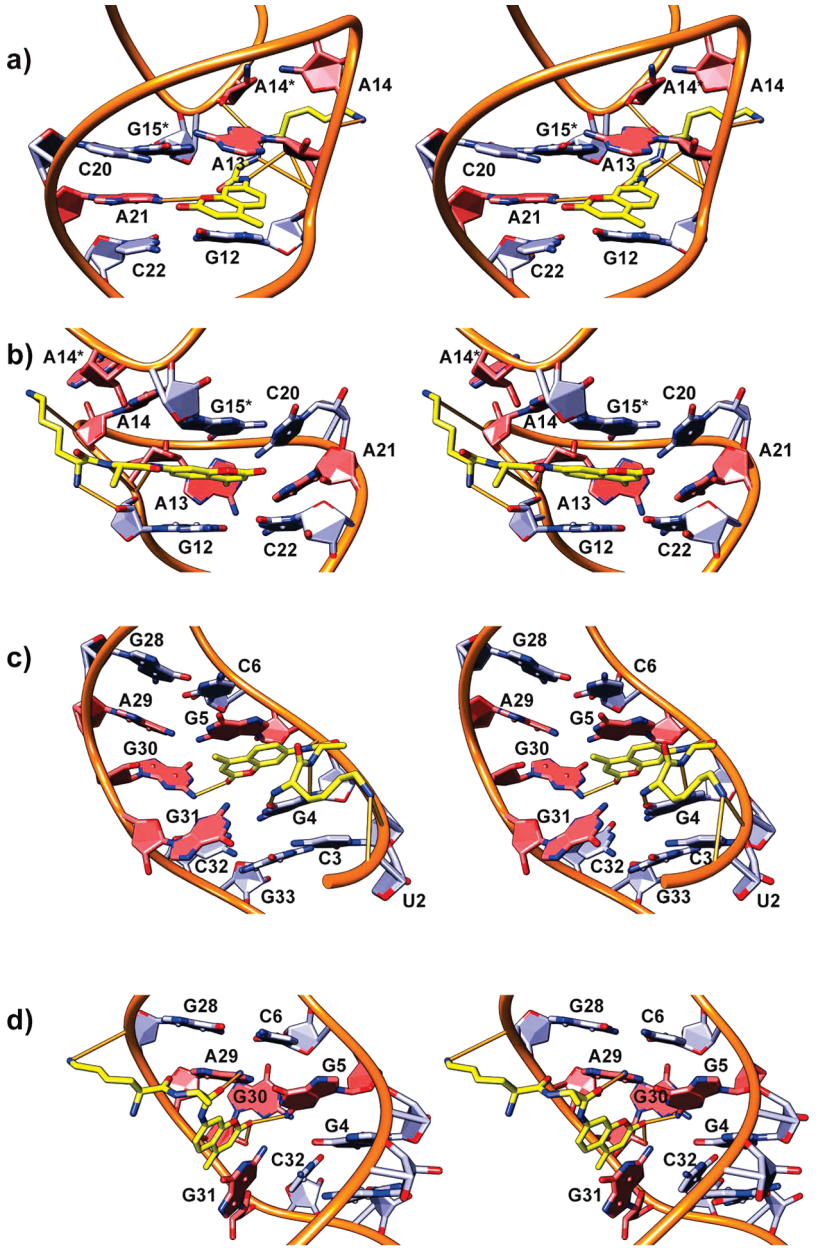

FIGURE 7: Stereoviews of binding poses of KA-AMC at the flanking adenines (a and b) and GRIL (c and d) binding sites calculated with MORDOR (*38*). RNA is shown schematically using a ribbon representation of the backbone and filled representation for bases and sugars with UCSF Chimera (*50*). Adenines flanking the palindromic apical sequence (a and b) and purines of the GRIL motif are colored light red. Carbon atoms of KA-AMC are colored yellow. Hydrogen bonds between KA-AMC and RNA are represented as thin orange lines.

ring is stacked directly under A13, and the coumarin ring oxygen is forming an H-bond with the amino group of A21. The alanine residue of KA-AMC is located above G12, which can explain moderate upfield shifts of its proton resonances (see above). The amide and guanidinium groups of KA-AMC are making H-bonds with the phosphate and sugar groups on both strands of SL1-s. In the second motif (Figure 7b), the coumarin ring is stacked between G12 of the one strand and G15 of the other strand of SL1-s, and the amide and guanidinium protons are making H-bonds with the phosphate and sugar groups from the lower half of SL1-s. It is interesting that RNA residues that are stacked directly against or hydrogen bonded with the coumarin ring in at least one of the binding poses, i.e., G12, A13, G15, and A21, are not observed in NMR spectra of the SL1-s complex with KA-AMC (Figure 5b), presumably due to a severe line broadening. While the absence of NMR signals cannot be considered as a proof of binding poses presented in panels a and b of Figure 7, it is certainly consistent with the existence of multiple modes of binding, which are exchanging on the intermediate time scale. Neighboring residues, such as A14, C20, and C22, are still observed in NMR spectra (Supporting Information), but they typically experience chemical shift perturbations and moderate line broadening upon KA-AMC binding (Figures 5b and 6a), probably caused by the propagation of a structural perturbation due to ligand binding.

Although the GRIL motif is structurally very different from the adenines flanking the palindromic kissing loop, KA-AMC binding poses calculated with MORDOR share common features for the two binding sites. First, there are multiple low-energy binding poses for the KA-AMC docked with the GRIL motif. Second, the coumarin ring is stacked against RNA bases, and the amide and guanidinium groups of KA-AMC are making H-bonds with sugars and phosphates of RNA in high-scoring poses for SL1-gril.

Two typical motifs are shown in panels c and d of Figure 7. In the first pose (Figure 7c), the coumarin ring is stacked between G4 and G5 while the carbonyl oxygen of coumarin is making an H-bond with the amino group of G30, and the 7-amido group and the lysine guanidinium group are making H-bonds with RNA phosphate oxygens. At the same time, the amide protons of alanine and lysine form H-bonds with the Hoogsteen edge of G4. In the second pose (Figure 7d), the coumarin ring is stacked directly above G31 and partially (at an angle) below A29. The coumarin carbonyl oxygen is making bifurcating H-bonds with the amino group of G5 and the ribose hydroxyl group of G30. In addition, the alanine amide oxygen is making an H-bond with the amino group of A29, and the lysine guanidinium group is H-bonding with a phosphate oxygen. Comparison of the predicted binding poses with NMR data is complicated by the lack of assignments for residues G4 and G5. Nevertheless, the maximum chemical shift and line width perturbations are observed for the purines of the GRIL motif (A29, G30, and G31) and for the neighboring residues (C6 and A27) (Figures 5c and 6b). An interesting spectral feature of the binding of KA-AMC to SL1-gril is a significant downfield shift of H1′ and H2′ resonances in G30 and a simultaneous upfield shift of H1′ and H2′ in G31 (Supporting Information). On the basis of ring current effect considerations, it seems that the G31 ribose must be directly above or below the coumarin ring while the ribose of G30 should be in the same plane with the coumarin ring in some of the binding poses. The ribose hydroxyl group of G30 is hydrogen-bonded to the coumarin carbonyl oxygen in the pose shown in Figure 7d; this indeed brings the G30 H2′ proton approximately into the plane of the coumarin ring. However, these binding poses do not explain the upfield shifts of the G31 ribose. Another possible explanation may involve changes in the conformational equilibrium of the GRIL motif, which displayed conformational heterogeneity even without a ligand (*35*) and remained flexible upon KA-AMC binding.

DISCUSSION

Previously, we reported small molecules that facilitate HIV-1 SL1 RNA dimer maturation in vitro and performed an SAR study to identify features important for its activity (*24*); however, it remained unclear whether the enhancement in activity correlates with RNA binding affinity or the specificity of binding. We utilized the intrinsic fluorescence emission of coumarin and hydroxyquinoline derivatives and their quenching by RNA to estimate the binding affinity of these molecules for the SL1 RNA. We also probed structural features of KA-AMC−SL1 RNA interactions via NMR.

We observed our activator molecules bind to many structural motifs with varying affinities, in a manner similar to that of the viral NCp7 protein. Although the total charge of a small molecule is an important determinant for binding affinity (compare R-AMC with RR-AMC in Table 1), a particular pattern of donor and acceptor atoms is also important (compare the binding affinity of KA-AMC and R-AMC in Table 1). Although increasing positive charge facilitates both specific and unspecific binding to RNA, we found that the four small molecules studied have a binding preference for unique structural motifs [16-fold preference for SL1-fl KD over a duplex RNA (apparently stacking with a terminal base pair) for R-AMC and 33-fold for RR-AMC (Table 2)]. KA-AMC exhibited preferential binding to the GRIL motif of SL1 (Table 1) compared to other stem-loops in the HIV-1 ψ locus, SL2 and SL3 (Table 3). Furthermore, the coumarin derivatives appear to have some preference for GRIL (in SL1-gril RNA) over the flanking A's site (SL1-s), while R-MHQ binds slightly better to the flanking A's (Table 1). Finally, there is no complete correlation between the SL1 maturation activation in vitro and SL1 binding. While the levels of both maturation activation and binding increase in the order KA-AMC < R-AMC < RR-AMC, this order is reversed for R-AMC and R-MHQ activation versus binding. The SAR study highlighted the importance of the coumarin ring oxygen as the H-bond acceptor for activity. When this oxygen in R-AMC is substituted with an H-bond donor NH group in R-MHQ, the activity decreases 3-fold (*24*), while the apparent affinity is higher for R-MHQ both for binding to the full-length SL1-fl KD and for binding to the individual sites of SL1 (Table 1). This shows that the difference in the maturation activity does not solely depend on the apparent binding affinity.

It is significant that we discovered a strong SL1 RNA binder, RR-AMC, with an apparent binding constant in the high nanomolar range. SL1 ligands may potentially have an antiviral effect by competing with the SL1−NCp7 interaction and thus either inhibiting RNA packaging or affecting RNA maturation. RR-AMC binds the flanking A's site and GRIL 14- and 6-fold less tightly, respectively, than NCp7. At the same time, RR-AMC binds these two sites 12- and 18-fold more tightly, respectively, than it binds duplex RNA. Most known clinically useful RNA ligands bind RNA with micromolar affinity and have a molecular mass of > 500 Da (*46*, *47*). RR-AMC has a molecular mass of 488 Da with a total charge of +3; it has the potential to be further optimized into an even tighter and more specific SL1 binder, although gaining specificity is expected to be a major difficulty.

We probed structural features of the KA-AMC−SL1 RNA interactions using NMR. In a manner consistent with binding constant measurement by fluorescence, we observed that KA-AMC binds to many different RNA motifs; however, it binds preferentially to the flanking adenines and GRIL in the SL1-s and SL1-gril constructs, respectively. Upon binding, ligand and RNA residues in the binding site undergo conformational exchange in the intermediate NMR time range, indicated by the severe line broadening. Line broadening due to possible aggregation effects was excluded for the experimental conditions utilized by using NMR to determine diffusion coefficients for the RNA signals (data not shown). Broadening of NMR signals may be caused by chemical exchange either between bound and free RNA conformations or between multiple bound conformations (*48*). One can distinguish these possibilities by titrating a great excess of ligand into the receptor. If exchange between free and bound species is the cause of broadening, adding ligand in large excess over that required to saturate the binding site should result in line narrowing for the RNA signals to a value reflecting the molecular tumbling of the RNA−ligand complex since the exchange term contributing to line broadening becomes vanishingly smaller. On the other hand, if the ligand has multiple binding modes that are comparable in energy, there would always be heterogeneity in RNA conformations, so the line shape would remain broad for RNA signals even in a great excess of ligand. Our NMR data show no sharpening of any signals at higher KA-AMC ratios (see A14 H8 and A26 H2 peaks in Figure 4 and the Supporting Information), implying that KA-AMC has multiple binding modes. On the other hand, these arguments are strictly speaking applicable to the situation with a single binding site. For example, assuming a single binding site for the GRIL motif with a K_d of 21 μM, 76% of RNA and ligand is expected to be in the bound state for the 1:1 RNA−ligand complex (Figure 4e). For the 1:2 RNA−ligand complex (Figure 4f), 93% of RNA but only

47% of ligand are expected to be bound. Then, we would expect significant line sharpening and significant recovery of the chemical shift perturbation for the ligand signals, which clearly occurs for the ligand residues. However, there is no evident narrowing of signals from the RNA. The situation is complicated by a continuing nonspecific ligand binding to other RNA residues, consistent with its ability to bind RNA duplex (Table 1). Therefore, we cannot completely exclude a contribution of intermediate exchange between bound and unbound states to the observed RNA line broadening. Nevertheless, the existence of multiple binding poses in each specific site, GRIL and the flanking adenines of the KD form, is consistent with our computational docking results (Figure 7). We could not determine the solution structures of the complexes because of the absence of observable intermolecular NOE cross-peaks, also likely because of the conformational averaging. However, many features of predicted binding poses are in qualitative agreement with chemical shift perturbations. Furthermore, preliminary studies using molecular dynamics simulations of selected binding poses with explicit water using AMBER 9 (*49*) show that many of these features remain stable during the simulation time. For example, for the pose shown in Figure 7a, the specific pattern of H-bonds has changed (the H-bond between the coumarin ring oxygen and the A21 amino group broke, and the one between the coumarin carbonyl oxygen and C20 amino group formed instead), but the coumarin ring remained stacked under A13 during 1 ns of the simulation (data not shown). The existence of multiple binding poses is likely to contribute entropically to the observed apparent ligand affinities.

Several coumarin and hydroxyquinoline derivatives conjugated with a short peptide can promote the conversion of the SL1 RNA from the KD to ED conformation (*24*), thus mimicking both the in vivo (*17*) and in vitro (*14*) NCp7 behavior. The similarities between NCp7 and small molecules studied here go further. NCp7 binds with high affinity to various structural motifs of HIV-1 RNA, such as the palindromic region and GRIL on the SL1 hairpin and G-rich apical loops of SL2 and SL3 hairpins with apparent K_d values of 23−140 nM (*23*). KA-AMC binds the palindromic region and the GRIL motif of SL1 with apparent K_d values of 50 and 21 μM, respectively (Table 1); it also binds SL2 and SL3 (Table 3). Further, NCp7 also binds fairly nonspecifically to double-stranded nucleic acids (*19*). KA-AMC may bind the stem regions of SL1 as well, based on small proton chemical shift perturbations (Figure 5), and it binds a duplex RNA sample, although less tightly than the two specific sites (Tables 1 and 2). Here the similarities end. It appears that approximately stoichiometric amounts of NCp7 (two NCp7 molecules and two RNA strands) are required for the full KD-to-ED conversion of SL1wt (*12*). Assuming the apparent K_d of 100 nM (*23*), ~82% of SL1-fl RNA is bound by NCp7 under the experimental conditions used, and most of the RNA was converted to the ED conformation (*12*). With an apparent K_d of 19 μM (Table 1), 88% of SL1-fl KD would be bound by KA-AMC at the 100-fold ligand excess, yet most of the RNA would not have converted to ED under these conditions, because the EC_{50} value, the effective concentration of KA-AMC for achieving 50% conversion, is 540−600 μM (*24*). Clearly, NCp7 not only binds better but also, independent of that, is a better refolding chaperone. We plan to use structural insights obtained in this study to further optimize small molecules binding SL1 to inhibit the NCp7−SL1 interaction and thus target the HIV-1 RNA dimer maturation and/or HIV-1 RNA packaging.

ACKNOWLEDGMENT

We thank Abram Calderon for assistance in fluorescence quenching experiments and Tuck C. Wong for help with diffusion coefficient measurements.

SUPPORTING INFORMATION AVAILABLE

Additional experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hoglund, S., Ohagen, A., Goncalves, J., Panganiban, A. T., and Gabuzda, D. (1997) Ultrastructure of HIV-1 genomic RNA. *Virology 233*, 271–279.
2. Paillart, J. C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004) Dimerization of retroviral RNA genomes: An inseparable pair. *Nat. Rev. Microbiol. 2*, 461–472.
3. Hu, W. S., and Temin, H. M. (1990) Genetic consequences of packaging two RNA genomes in one retroviral particle: Pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U.S.A. 87*, 1556–1560.
4. Clever, J. L., Wong, M. L., and Parslow, T. G. (1996) Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J. Virol. 70*, 5902–5908.
5. Skripkin, E., Paillart, J. C., Marquet, R., Ehresmann, B., and Ehresmann, C. (1994) Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc. Natl. Acad. Sci. U.S.A. 91*, 4945–4949.
6. Paillart, J. C., Marquet, R., Skripkin, E., Ehresmann, B., and Ehresmann, C. (1994) Mutational Analysis of the Bipartite Dimer Linkage Structure of Human Immunodeficiency Virus Type 1 Genomic Rna. *J. Biol. Chem. 269*, 27486–27493.
7. Laughrea, M., and Jette, L. (1994) A 19-Nucleotide Sequence Upstream of the 5′ Major Splice Donor Is Part of the Dimerization Domain of Human Immunodeficiency Virus 1 Genomic RNA. *Biochemistry 33*, 13464–13474.
8. Muriaux, D., Fosse, P., and Paoletti, J. (1996) A kissing complex together with a stable dimer is involved in the HIV-1Lai RNA dimerization process in vitro. *Biochemistry 35*, 5075–5082.
9. Sakuragi, J., Sakuragi, S., and Shioda, T. (2007) Minimal region sufficient for genome dimerization in the human immunodeficiency virus type 1 virion and its potential roles in the early stages of viral replication. *J. Virol. 81*, 7985–7992.
10. Lodmell, J. S., Ehresmann, C., Ehresmann, B., and Marquet, R. (2000) Convergence of natural and artificial evolution on an RNA loop-loop interaction: The HIV-1 dimerization initiation site. *RNA 6*, 1267–1276.
11. Paillart, J. C., Westhof, E., Ehresmann, C., Ehresmann, B., and Marquet, R. (1997) Non-canonical interactions in a kissing loop complex: The dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol. 270*, 36–49.
12. Mujeeb, A., Ulyanov, N. B., Georgantis, S., Smirnov, I., Chung, J., Parslow, T. G., and James, T. L. (2007) Nucleocapsid protein-mediated maturation of dimer initiation complex of full-length SL1 stemloop of HIV-1: Sequence effects and mechanism of RNA refolding. *Nucleic Acids Res. 35*, 2026–2034.
13. Laughrea, M., and Jette, L. (1996) Kissing-loop model of HIV-1 genome dimerization: HIV-1 RNAs can assume alternative dimeric forms, and all sequences upstream or downstream of hairpin 248−271 are dispensable for dimer formation. *Biochemistry 35*, 1589–1598.
14. Muriaux, D., De Rocquigny, H., Roques, B. P., and Paoletti, J. (1996) NCp7 activates HIV-1Lai RNA dimerization by converting a transient loop-loop complex into a stable dimer. *J. Biol. Chem. 271*, 33686–33692.
15. Moore, M. D., Nikolaitchik, O. A., Chen, J., Hammarskjold, M. L., Rekosh, D., and Hu, W. S. (2009) Probing the HIV-1 genomic RNA trafficking pathway and dimerization by genetic recombination and single virion analyses. *PLoS Pathog. 5*, e1000627.
16. Song, R., Kafaie, J., Yang, L., and Laughrea, M. (2007) HIV-1 viral RNA is selected in the form of monomers that dimerize in a three-step protease-dependent process: The DIS of stem-loop 1 initiates viral RNA dimerization. *J. Mol. Biol. 371*, 1084–1098.
17. Fu, W., Gorelick, R. J., and Rein, A. (1994) Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions. *J. Virol. 68*, 5013–5018.

18. Feng, Y. X., Campbell, S., Harvin, D., Ehresmann, B., Ehresmann, C., and Rein, A. (1999) The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: Possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J. Virol. 73*, 4251–4256.
19. Urbaneja, M. A., Wu, M., Casas-Finet, J. R., and Karpel, R. L. (2002) HIV-1 nucleocapsid protein as a nucleic acid chaperone: Spectroscopic study of its helix-destabilizing properties, structural binding specificity, and annealing activity. *J. Mol. Biol. 318*, 749–764.
20. Marquet, R., Paillart, J. C., Skripkin, E., Ehresmann, C., and Ehresmann, B. (1994) Dimerization of human immunodeficiency virus type 1 RNA involves sequences located upstream of the splice donor site. *Nucleic Acids Res. 22*, 145–151.
21. Berkhout, B. (1996) Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucleic Acid Res. Mol. Biol. 54*, 1–34.
22. Clever, J. L., and Parslow, T. G. (1997) Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol. 71*, 3407–3414.
23. Shubsda, M. F., Paoletti, A. C., Hudson, B. S., and Borer, P. N. (2002) Affinities of packaging domain loops in HIV-1 RNA for the nucleocapsid protein. *Biochemistry 41*, 5276–5282.
24. Chung, J., Mujeeb, A., Jiang, Y., Guilbert, C., Pendke, M., Wu, Y., and James, T. L. (2008) A small molecule, Lys-Ala-7-amido-4-methylcoumarin, facilitates RNA dimer maturation of a stem-loop 1 transcript in vitro: Structure-activity relationship of the activator. *Biochemistry 47*, 8148–8156.
25. Milligan, J. F., and Uhlenbeck, O. C. (1989) Synthesis of small RNAs using T7 RNA polymerase. In Methods in Enzymology (Dahlberg, J. E., and Abelson, J. N., Eds.) pp 51−62, Academic Press, New York.
26. Grodberg, J., and Dunn, J. J. (1988) *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol. 170*, 1245–1253.
27. Marquet, R., Baudin, F., Gabus, C., Darlix, J. L., Mougel, M., Ehresmann, C., and Ehresmann, B. (1991) Dimerization of human immunodeficiency virus (type 1) RNA: Stimulation by cations and possible mechanism. *Nucleic Acids Res. 19*, 2349–2357.
28. Laughrea, M., and Jette, L. (1996) HIV-1 genome dimerization: Formation kinetics and thermal stability of dimeric HIV-1Lai RNAs are not improved by the 1−232 and 296−790 regions flanking the kissing-loop domain. *Biochemistry 35*, 9366–9374.
29. Takahashi, K. I., Baba, S., Chattopadhyay, P., Koyanagi, Y., Yamamoto, N., Takaku, H., and Kawai, G. (2000) Structural requirement for the two-step dimerization of human immunodeficiency virus type 1 genome. *RNA 6*, 96–102.
30. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) NMR structure of stem-loop SL2 of the HIV-1 ψ RNA packaging signal reveals a novel A-U-A base-triple platform. *J. Mol. Biol. 299*, 145–156.
31. Pappalardo, L., Kerwood, D. J., Pelczer, I., and Borer, P. N. (1998) Three-dimensional folding of an RNA hairpin required for packaging HIV-1. *J. Mol. Biol. 282*, 801–818.
32. Elalaoui, A., Divita, G., Maury, G., Imbach, J. L., and Goody, R. S. (1994) Intrinsic tryptophan fluorescence of bovine liver adenosine kinase, characterization of ligand binding sites and conformational changes. *Eur. J. Biochem. 221*, 839–846.
33. Smallcombe, S. H. (1993) Solvent suppression with symmetrically-shifted pulses. *J. Am. Chem. Soc. 115*, 4776–4785.
34. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR 2*, 661–665.
35. Ulyanov, N. B., Mujeeb, A., Du, Z., Tonelli, M., Parslow, T. G., and James, T. L. (2006) NMR structure of the full-length linear dimer of stem-loop-1 RNA in the HIV-1 dimer initiation site. *J. Biol. Chem. 281*, 16168–16177.
36. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J. Biomol. NMR 6*, 277–293.
37. Goddard, T. D., and Kneller, D. G. (1998) SPARKY, version 3.0, University of California, San Francisco.
38. Guilbert, C., and James, T. L. (2008) Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model. 48*, 1257–1268.
39. Baba, S., Takahashi, K., Noguchi, S., Takaku, H., Koyanagi, Y., Yamamoto, N., and Kawai, G. (2005) Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J. Biochem. 138*, 583–592.
40. Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A., and Jernigan, R. L. (1991) Static and Statistical Bending of DNA Evaluated by Monte Carlo Simulations. *Proc. Natl. Acad. Sci. U.S.A. 88*, 7046–7050.
41. Gómez Pinto, I., Guilbert, C., Ulyanov, N. B., Stearns, J., and James, T. L. (2008) Discovery of Ligands for a Novel Target, the Human Telomerase RNA, Based on Flexible-Target Virtual Screening and NMR. *J. Med. Chem. 51*, 7205–7215.
42. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical Treatment of Solvation For Molecular Mechanics and Dynamics. *J. Am. Chem. Soc. 112*, 6127–6129.
43. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.
44. Comolli, L. R., Ulyanov, N. B., Soto, A. M., Marky, L. A., James, T. L., and Gmeiner, W. H. (2002) NMR structure of the 3′ stem-loop from human U4 snRNA. *Nucleic Acids Res. 30*, 4371–4379.
45. Stewart-Maynard, K. M., Cruceanu, M., Wang, F., Vo, M. N., Gorelick, R. J., Williams, M. C., Rouzina, I., and Musier-Forsyth, K. (2008) Retroviral nucleocapsid proteins display nonequivalent levels of nucleic acid chaperone activity. *J. Virol. 82*, 10129–10142.
46. Ryu, D. H., and Rando, R. R. (2001) Aminoglycoside binding to human and bacterial A-site rRNA decoding region constructs. *Bioorg. Med. Chem. 9*, 2601–2608.
47. Aboul-ela, F. (2010) Strategies for the design of RNA-binding small molecules. *Future Med. Chem. 2*, 93–119.
48. Reibarkh, M., Malia, T. J., Hopkins, B. T., and Wagner, G. (2006) Identification of individual protein-ligand NOEs in the limit of intermediate exchange. *J. Biomol. NMR 36*, 1–11.
49. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2006) AMBER 9, University of California, San Francisco.
50. Couch, G. S., Hendrix, D. K., and Ferrin, T. E. (2006) Nucleic acid visualization with UCSF Chimera. *Nucleic Acids Res. 34*, e29.