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†

Janet Chung, Anwer Mujeeb,‡ Yongying Jiang, Christophe Guilbert, Mrunal Pendke,§ Yanfen Wu,|| and Thomas L. James*

Department of Pharmaceutical Chemistry, University of California, 600 16th Street, San Francisco, California 94158-2517

Received February 7, 2008; Revised Manuscript Received May 26, 2008

ABSTRACT: The type 1 human immunodeficiency virus (HIV-1), like all retroviruses, contains two copies of the RNA genome as a dimer. A dimer initially forms via a self-complementary sequence in the dimer initiation site (DIS) of the genomic RNA, but that dimer is converted to a mature dimer in a process generally promoted by the viral nucleocapsid (NC) protein. Formation of the mature dimer is correlated with infectivity. Study of genomic dimerization has been facilitated by discovery of short RNA transcripts containing the DIS stem−loop 1 (SL1), which can dimerize spontaneously without any protein factors in vitro as well as via the NC protein. On the basis of the palindromic nature of the apical loop of SL1, a kissing loop model has been proposed. First, a metastable kissing dimer is formed via a loop−loop interaction and then converted into a more stable extended dimer by the NC protein. This dimerization process in vitro is believed to mimic the in vivo RNA maturation. During experimental screening of potential inhibitors, we discovered a small molecule, Lys-Ala-7-amido-4-methylcoumarin (KA-AMC), which facilitates the in vitro conversion from kissing dimer to extended dimer. Here we report the structure−activity relationship for KA-AMC for promoting dimer maturation. Guanidino groups and increasing positive charge on the side chain enhance activity. For activity, the charged side chain is preferred on the benzene ring, and O_1 in the coumarin scaffold is essential. NMR studies show that the coumarin derivatives stack with aromatic bases of the RNA. The coumarin derivatives may aid in the investigation of some aspects of dimer maturation and serve as a scaffold for design of maturation inhibitors or of activators of premature maturation, either of which can lead to a potential HIV therapeutic.

All retroviruses, including the type 1 human immunodeficiency virus (HIV-1),[1] contain a genomic dimer that consists of two homologous copies of RNA (*1–4*). The dimeric nature of the RNA is important in viral replication steps such as recombination, packaging, and reverse transcription (reviewed in ref *5*). Molecular understanding of the genomic RNA−RNA interaction has been facilitated by the use of short, synthetic RNA fragments, corresponding to the dimer initiation site (DIS) in the 5′-untranslated region (5′UTR) of the HIV-1 RNA genome, which dimerize spontaneously in vitro in the absence of any proteins (*6–8*). This region of the 5′UTR contains four stem−loop structures. The first stem−loop 1 transcript (SL1) in the packaging sequence corresponds to the dimer initiation site, the minimal domain required for dimerization (*9*). It is also involved in packaging of the RNA genome into new infectious virions (*10, 11*). SL1 has been shown to fold into an 11 bp stem with a four-base internal loop and a nine-nucleotide loop including the hexanucleotide palindrome at the apex (Figure 1a) (*12–14*). Although there are 64 possible autocomplementary hexanucleotide sequences, in nearly all major HIV-1 clades only two hexanucleotide sequences, GCGCGC and GUGCAC, are found (*15, 16*), and while the exact sequence is not critical, maintaining upper and lower stems and the existence of the G-rich internal loop are essential for the HIV replication cycle (*17*). According to the kissing loop model (Figure 1b), the genomic RNA first interacts to form a metastable kissing dimer via standard Watson−Crick base pairing through the palindromic loop of SL1 while in the host cell. The in vitro kissing dimer is stable at physiological temperature (37 °C) (*18*), but it is converted into a more thermostable extended dimer conformation by either the viral nucleocapsid protein (NCp7), a domain of the Gag protein

† This work was supported by Grant AI46967 from the National Institutes of Health to T.L.J.

* To whom correspondence should be addressed. Telephone: (415) 476-1916. Fax: (415) 502-8298. E-mail: james@picasso.ucsf.edu.

‡ Current address: California HIV/AIDS Research Program, University of California, Oakland, CA 94612.

§ Current address: Indian Institute of Chemical Technology, Hyderabad, India.

|| Current address: Department of Medicinal Chemistry, Peking University, Beijing, China.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; NC, nucleocapsid moiety of the Gag protein or proteolytic nucleocapsid protein product; NCp7, proteolytic nucleocapsid protein product of the Gag protein; DIS, dimer initiation site; KD, kissing dimer; ED, extended dimer; KA-AMC, Lys-Ala-7-amido-4-methylcoumarin; 5′UTR, 5′-untranslated region; PAGE, polyacrylamide gel electrophoresis; TBM, Tris-borate-$MgCl_2$ buffer; TBE, Tris-borate-EDTA buffer; SL1, stem−loop 1 transcript; SL1-wt, wild-type sequence of SL1; SAR, structure−activity relationship; EC_{50}, effective concentration at 50% activation; LC−MS, liquid chromatography−mass spectrometry; PDB, Protein Data Bank.

10.1021/bi800230m CCC: $40.75
Published on Web 07/11/2008

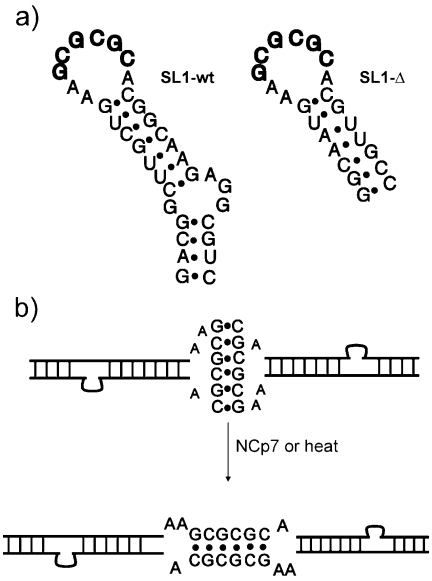


FIGURE 1: Secondary structure of SL1 and the kissing loop model of HIV-1 genomic dimerization. (a) Nucleotide sequences of SL1-wt and SL1-Δ constructs. SL1-wt is a 35-nucleotide construct with the sequence of the HIV-1 Lai isolate but with the bottom two base pairs reversed to produce a higher transcription yield. The highly conserved six-nucleotide palindrome is shown in bold. SL1-Δ is a truncated version of SL1-wt containing only the apex loop and the upper stem used for NMR structure determination of the SL1 kissing dimer (PDB entry 1BAU). (b) Kissing loop model of HIV-1 genomic dimerization. The genomic RNA forms a kissing dimer (top) utilizing the self-complementary hexameric palindrome (bold). This kissing dimer is stable at physiological temperature but can be converted into the more stable extended dimer (bottom) with heat or viral NCp7.

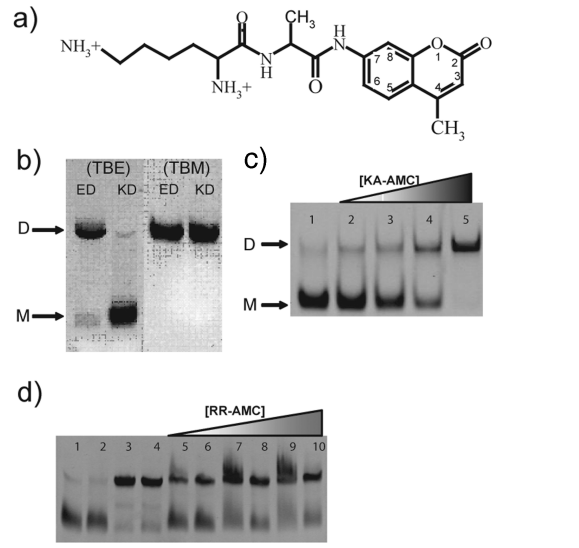


FIGURE 2: Native gel shift assay. (a) Chemical structure of Lys-Ala-7-amido-4-methyl-coumarin (KA-AMC), the first compound discovered to facilitate the dimer maturation of a 35-nucleotide synthetic RNA fragment representing the dimer initiation site of the HIV-1 genome in vitro. (b) The metastable kissing dimer (KD) dissociates when no Mg^{2+} is present. Therefore, it runs as a dimer (marked by D) in a TBM gel, but it dissociates into monomers (marked by M) in a TBE gel. The more stable extended dimer (ED) does not require Mg^{2+} and runs as a dimer in either gel. (c) Native TBE gel showing KA-AMC mediates SL1 dimer maturation. The kissing dimer is stable at room temperature (lane 1, marked by M). KA-AMC facilitates SL1 dimer maturation at room temperature which is evident with an elevated proportion of extended dimer (marked by D) with an increased concentration of KA-AMC (lanes 2−5, with 125, 250, 500, and 1250 μM KA-AMC, respectively). (d) The dimer band present in the gel is due to the extended dimer and not a stabilized kissing dimer−small molecule complex. The dimer band in the sample persists even after washing with excess buffer. There are five pairs of samples in the TBE gel (lanes 1 and 2, 3 and 4, 5 and 6, 7 and 8, and 9 and 10), and only the even-numbered sample of the pair has been washed with excess buffer. From left to right, the five pairs correspond to KD, ED, and KD with 37.5, 75, and 150 μM RR-AMC, respectively.

(*19*), or heat (55 °C) (*20, 21*). This nucleocapsid-induced dimer conversion observed in vitro may correspond to a part of the maturation process induced by the protease-cleaved NCp7 in the budding virion in vivo (*22*); this dimer maturation is essential for optimal infectivity (see ref *5* for a review).

We hypothesized that if we could find a small molecule that would bind to the kissing dimer, in effect forming a termolecular complex (one ligand binding to the kissing dimer), it might inhibit NCp7-mediated maturation and abrogate the infectivity of the viral particles. Using the high-resolution NMR structure of the SL1 kissing dimer (*23*), we used computational screening to generate a list of potential small molecule inhibitors targeting the SL1 kissing dimer. During experimental screening of some of the compounds, we found that one small molecule, Lys-Ala-7-amido-4-methylcoumarin [KA-AMC, **1** (Figure 2a)], activated maturation instead. That is, KA-AMC induced maturation of a short SL1 transcript at room temperature in vitro. Therefore, KA-AMC may serve as a simplified system for studying nucleocapsid-mediated maturation, and it may serve as a starting point for structure-based design of inhibitors that block this process or activators that prematurely cause mature dimer formation while still in the cytoplasm rather than the budding virion.

The unexpected discovery of an activator gives rise to some questions: among them, what features of the activator promote maturation and whether there are other small molecules, exogenous or endogenous, that might similarly potentiate retroviral activity. We decided to investigate first features of the activator, leading to this report on the structure−activity relationship (SAR) of KA-AMC. Such a study enables us to identify the types of compounds that might facilitate the activation, but perhaps more importantly, it was our intent to optimize the activator for subsequent studies to test activity in vivo and to investigate how these coumarin derivatives potentiate maturation.

MATERIALS AND METHODS

RNA Synthesis. The SL1 RNA oligonucleotide with the wild-type sequence (SL1-wt) (Figure 1a) was synthesized by in vitro runoff transcription (*24*) with T7 RNA polymerase, prepared in house (*25*), from a synthetic DNA template (Integrated DNA Technology, Inc.) and purified with 20% denaturing polyacrylamide gel electrophoresis (PAGE), followed by electroelution and dialysis in sterile water. The dialyzed RNA was quantified using ultraviolet (UV) spectrophotometry and stored at −20 °C.

RNA Sample Preparation and Dimerization Assay. The assay for monitoring dimer maturation has been previously described (*26*). For each assay, 5 μM SL1-wt RNA was heated to 90 °C for 3 min, followed by snap-cooling on ice for 3 min. An equal volume of prechilled 2× dimerization

buffer [20 mM potassium phosphate (pH 7), 100 mM NaCl, and 0.2 mM $MgCl_2$] was added, and the sample was incubated on ice for an additional 30 min. As a final step, the RNA sample was equilibrated for 2 h at room temperature to fold into the kissing dimer (KD) or at 55 °C for the extended dimer. These dimers can be easily distinguished by 10% native PAGE in 1× TBE buffer [89 mM Tris, 89 mM borate (pH 8.4), and 2 mM EDTA] or 1× TBM buffer [89 mM Tris, 89 mM borate (pH 8.4), and 3 mM $MgCl_2$] (Figure 2b) on the basis of the difference in Mg^{2+} dependence that has been reported by several groups (*8, 20, 27*). The metastable kissing dimer requires Mg^{2+} to remain as a dimer during electrophoresis and dissociates into monomers as the EDTA chelates the Mg^{2+} in TBE buffer. Therefore, the kissing dimer exhibits a monomer band in TBE gel electrophoresis and a dimer band in TBM gel electrophoresis. In contrast, the more stable extended dimer has no Mg^{2+} dependence, exhibiting a dimer band in both TBE and TBM gel electrophoresis.

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

Molecular Docking of a Small Molecule to the Kissing Dimer Structure. The ~200000 compounds in the Available Chemicals Directory (ACD), 1998 version (MDL Information Systems), were scanned for their van der Waals (VDW) and electrostatic (ES) complementarity to the surface created by formation of the kissing loop using the NMR structure of SL1-Δ in the KD form (Figure 1a) [PDB entry 1BAU (*23*)] first using the virtual screening program DOCK 4.0 (*28*) and subsequently with the virtual screening program ICM (*29*). DOCK is fast for matching compounds with different conformations to a specific site that is held rigid on the basis of intermolecular VDW and ES interactions; however, because the scoring function at the time of screening did not account for other important interactions such as hydrogen bonds, it could not predict true binding affinity. We used DOCK as a selection filter to pick the top 50000 potential compounds for further calculation using ICM. ICM could better predict binding mode and affinity, because it is possible to incorporate some receptor flexibility and include additional hydrogen bond interaction and solvation terms in the scoring function (*30*). However, because it is a more elaborate procedure, the time required to dock each compound is longer. When docking was completed, we examined the top-ranking 1000 compounds manually and created a list of 75 compounds to test experimentally on the basis of their binding pose, likely bioavailability, diversity in the scaffold, availability for purchase, and price.

Construction of the SAR Library. Chemicals (see Table 1) were purchased as indicated: **1**, **1a**, **1b**, **3**, and **6** (Sigma), **5** (Bachem), and **2**, **4**, and **10** (Chem Impex); those found using the AMC base as a substructure in the ACD were 7-aminocoumarin **7a** (Toronto Research Chemicals), 6-aminocoumarin **8a** (VWR), 3-aminocoumarin **9a**, and 7-amino-4-methylquinolin-2-one **11a** (Sigma). Preparative TLC plates were purchased from Analtech (Silica Gel GF, 1000 μm). Flash column chromatography was performed using silica gel (Merck, 230−400 mesh).

Compounds **7**, **8**, and **11** were synthesized by coupling Boc-Arg(Boc)$_2$-OH with the corresponding amine using the Boc anhydride method according to Scheme 1. Generally, (Boc)-Arg-(Boc)$_2$-OH (Bachem) was activated by reacting it with 1.1 equiv of Boc anhydride (Aldrich) in pyridine under argon for 1 h at room temperature. The resulting mixed anhydride was then reacted with 1 equiv of amine for 24−72 h at room temperature. The reaction mixture was extracted with ethyl acetate (EtOAc). The extract was washed with 5% citric acid, saturated $NaHCO_3$, and saturated NaCl and dried over Na_2SO_4. Solvent was removed by evaporation, and the residue was purified by preparative thin-layer chromatography (10:1 $CHCl_3$/MeOH) to give compounds **7b**, **8b**, and **11b**. Boc groups in compounds **7b**, **8b**, and **11b** were removed by 30% trifluoroacetic acid (TFA) in dichloromethane (DCM) at room temperature for 18 h. After evaporation of solvents, the residue was dried under vacuum to give target compounds **7**, **8**, and **11**.

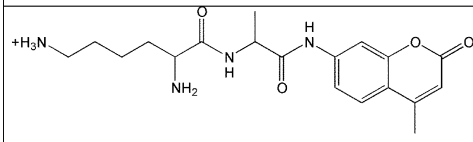
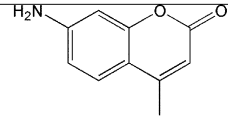
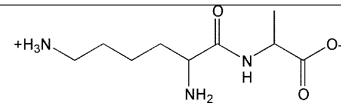
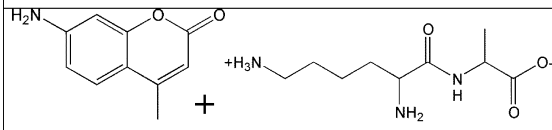
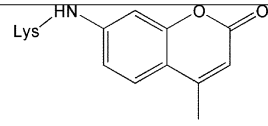
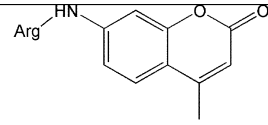
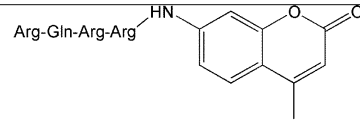
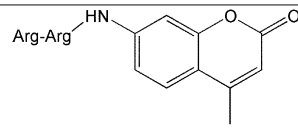
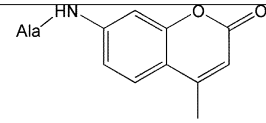
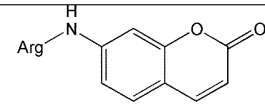
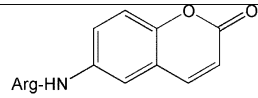
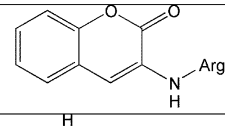
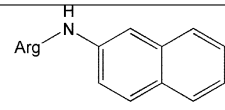
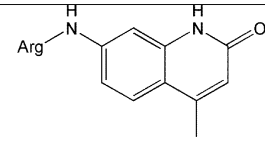
Compound **9** was synthesized by coupling Boc-Arg(Boc)$_2$-OH with the corresponding amine using phosphoryl chloride ($POCl_3$). Under argon, (Boc)-Arg-(Boc)$_2$-OH and 1 equiv of **9a** were mixed together in pyridine and then chilled to −15 °C. To the solution was added 2 equiv of $POCl_3$ (Sigma), and the reaction was continued for 1 h at −15 °C. The reaction was quenched by adding H_2O and the mixture extracted with EtOAc. The organic phase was washed with 5% citric acid, saturated $NaHCO_3$, and saturated NaCl and dried over Na_2SO_4. Solvent was removed by evaporation, and the residue was purified by flash column chromatography (10:3 hexane/EtOAc) to give compound **9b**. The Boc groups in **9b** were removed as before to give target compound **9**.

The identity and purity of all intermediates and final products were confirmed by ^{1}H NMR (600 MHz Varian) and LC−MS (ESI+, Waters).

Small Molecule-Mediated Maturation of the SL1 RNA Kissing Dimer. For each compound in the SAR study, various concentrations of the compound were added to 2.5 μM kissing dimer and incubated at room temperature for 2 h. In the meantime, two aliquots of the same kissing dimer were incubated at 25 or 55 °C to fold into the extended dimer to serve as internal controls of 0 or 100% conversion, respectively. Immediately after the incubation period, loading dye containing bromophenol blue was added to each sample which was run on a 10% native TBE gel at 150 V until the dye reached the bottom. The bands were visualized after staining with ethidium bromide under UV light. To demonstrate that the dimer band does indeed come from ED and not from KD stabilized by the ligand, some samples were removed prior to gel electrophoresis, and the activator was washed out by being diluted with excess dimerization buffer five times, transferred to a centrifugal filter device (Amicon, 3000 MWCO), and centrifuged at 3000 rpm for 45 min. All AMC derivatives in our studies have a unique absorbance at 325 nm; the A_{325} of the filtrate was monitored after each wash to detect any remaining activator present in the retentate.

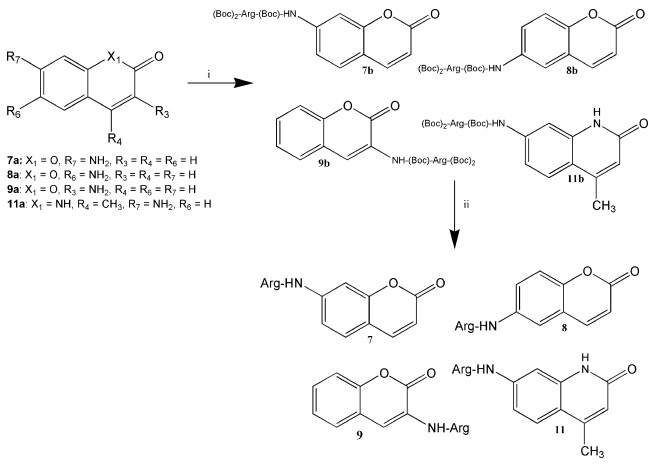
SAR Evaluation. The native gel shift assay of dimer maturation for each compound was performed in duplicate. The intensities of the monomer (i.e., kissing dimer) and dimer (i.e., extended dimer) bands were quantified with the

Table 1: Dimer Maturation Activity of Small Molecules[a]

	#	Name	EC_{50}, μM
	**1**	Lys-Ala-7-amido-4-methylcoumarin	540-600
	**1a**	7-amino-4-methyl-coumarin	ND[b]
	**1b**	Lys-Ala dipeptide	ND
		1:1 mixture of **1a** and **1b**	ND
	**2**	Lys-7-amido-4-methyl-coumarin	560-610
	**3**	Arg-7-amido-4-methyl-coumarin	310-320
	**4**	Arg-Gln-Arg-Arg-7-amido-4-methyl-coumarin	50-53
	**5**	Arg-Arg-7-amido-4-methyl-coumarin	100-110
	**6**	Ala-7-amido-4-methyl-coumarin	1600-1700
	**7**	Arg-7-amido-coumarin	500-540
	**8**	Arg-6-amido-coumarin	610-640
	**9**	Arg-3-amido-coumarin	1400-1500
	**10**	Arg-7-amido-β-naphthalene	720-760
	**11**	Arg-7-amido-4-methyl-2-hydroxyquinoline	790-940

[a] Effective concentration of KA-AMC derivatives to promote 50% dimer maturation (EC_{50}) of the SL1-wt kissing dimer as determined by the native gel shift assay. Determined with 2.5 μM SL1-wt kissing dimer and incubation at room temperature for 2 h. [b] ND, not detected.

Scheme 1: Synthesis of Arginine-Conjugated Aminocoumarins and 7-Amino-4-methylquinoline-2-one[a]



[a] Reagents and conditions: (i) for compounds **7**, **8**, **11**, Boc-Arg(Boc)$_2$-OH, Boc$_2$O, pyridine, 1−3 days, room temperature; for compound **9**, Boc-Arg(Boc)$_2$-OH, POCl$_3$, pyridine, 1 h, −15 °C; (ii) 30% TFA, DCM, 18 h, room temperature.

Quantity-One software supplied with Gel-Doc gel imager (Bio-Rad Laboratories). The percentage conversion of all samples were normalized on the basis of the amount of residual of extended dimer and kissing dimer, present in the kissing and extended dimer controls, respectively. The fraction of ED was calculated by taking the ratio of pixels after the correction of the dimer band to those of both bands. The averaged fraction ED values were fitted with eq 1 using KaleidaGraph to calculate the effective concentration at 50% activation (EC_{50}) for each compound:

$$\text{fraction of ED} = \frac{1}{1 + 10^{n[\log(EC_{50}) - \log[L]]}} \quad (1)$$

where n is the Hill slope, EC_{50} the effective concentration (molar) at 50% activation, and [L] the activator concentration (molar). EC_{50} was used to compare activities of compounds in the studies.

The standard deviation s was calculated via eq 2 using the STDEV function in Microsoft Excel:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad (2)$$

where x is the sample measurement, $\bar{x}$ the average measurement, and n the number of sample measurements (e.g., $n = 2$ for duplicates).

The standard error was calculated with eq 3 and plotted as the y-error bars in Figures S1−S4.

$$s_e = \frac{s}{\sqrt{n}} \quad (3)$$

where s_e is the standard error, s the standard deviation, and n the number of sample measurements.

NMR Experiments. A 10 μM SL1-wt RNA solution was folded into the kissing dimer as described above, except all solutions were deuterated. Various concentrations of compound (final concentrations of 50, 100, and 200 μM) were added to the RNA solution. The NMR spectra were acquired with a Varian Inova 600 MHz NMR spectrometer at 25 °C. Saturation transfer difference (STD) NMR spectra were acquired by internal subtraction via phase cycling (*31*). On-resonance radiation was set to 5.5 ppm. Presaturation of RNA resonances was achieved by an appropriate number of band-selective G4 Gaussian cascade pulses to give a saturation time of 2 s. Assignments of proton resonances of AMC and KA-AMC alone in solution were based on chemical shifts, J coupling patterns, and integrated signal intensities. Assignments of KA-AMC protons in RNA complexes additionally relied on two-dimensional TOCSY NMR spectra (not shown) aided by incremental titration of the RNA with the ligand.

RESULTS

Identification of a Small Molecule Activator of HIV-1 Dimer Maturation. Molecular docking against the Available Chemicals Directory (ACD) gave a ranked list of small organic compounds predicted to bind to the high-resolution structure of the HIV-1 kissing dimer complex. From the top 1000 ranked compounds, we picked an initial list of 75 compounds to test experimentally. Criteria for the selection included molecular mass (generally <500 Da), solubility in neutral aqueous buffers, and price. A native gel shift assay based on Mg^{2+} ion dependence of the dimers was utilized for screening. The semistable kissing dimer requires Mg^{2+} ions to remain as a dimer during electrophoresis, while the more stable extended dimer does not. Therefore, the kissing dimer runs as a monomer band, while the extended dimer runs as a dimer band in the absence of Mg^{2+}, such as in a native TBE gel (Figure 2b), and can be easily distinguished. A few inhibitors were identified (data not shown), but more interesting was the appearance of a small molecule activator, KA-AMC, that facilitates dimer maturation (Figure 2c). The kissing dimer, shown as a faster migrating monomer band in the TBE gel, is stable at room temperature (Figure 2c, lane 1), and to refold into the extended dimer, it requires a RNA chaperone, such as NCp7, or heat to overcome the activation energy barrier. KA-AMC facilitates the dimer maturation much in the manner of a RNA chaperone, evident by the increasing amount of extended dimer, shown as the slower migrating dimer band in TBE gel, correlated with increasing KA-AMC concentration (Figure 2c, lanes 2−4). The dimer band present in the TBE gel is not due to a stabilized kissing dimer−small molecule complex but originates from conversion to the extended dimer: these dimer bands in the TBE gel persist even after the activator has been washed out (Figure 2d, lanes 6, 8, and 10).

The following sections explore in detail the impact on activation of dimer maturation with variations of the originally identified KA-AMC structure. Derivatives were compared on the basis of the effective concentration at 50% activation (EC_{50}), namely the concentration of the small molecule activator required for 50% dimer maturation.

Structure−Activity Relationship for the Small Molecule Activation of Dimer Maturation. (i) Lys-Ala and AMC Must Be Covalently Linked for Activation. KA-AMC (**1** in Table 1) can be divided into two segments: the bicyclic aromatic AMC base (**1a**) and the positively charged KA dipeptide (**1b**). We observe that **1a** must be covalently linked to **1b** for activity. A 1:1 mixture of each segment or each segment alone shows no activity (Table 1). It is evident that both segments behave synergistically to initiate binding and activation.

(ii) The Guanidino Group on Arg Is Preferred over the Localized Amino Group in Lys. The originally discovered activator (**1**) has the KA dipeptide with a charge of +1 on the 7 position of AMC. We have tested other AMC derivatives with the 7 substituent having a +1 charge, such as Lys-AMC (**2**) and Arg-AMC (**3**). First, we found **1** and **2** have similar EC_{50} values [540−600 μM (Table 1 and Figure S1)], suggesting the alanine residue in the KA dipeptide is dispensable and the length of the carbon chain does not affect activity. Therefore, it seems only the positive charge is required for activity. Second, we investigated the chemical nature of the positive charge and found the guanidino group in Arg is preferred over the localized amino group in Lys by ∼2-fold [EC_{50} ∼ 300 and 540−600 μM, respectively (Table 1 and Figure S1)], although both carry a +1 charge. This is possible because the guanidino group can form more hydrogen bonds with the RNA than the amino group.

(iii) The Positive Charge on the Substituent at the Seventh Position of the AMC Base Enhances Activity. Next, we varied the net charge of the substituent at the 7 position and observed a positive correlation between the number of positive charges and activity. The +3 charged Arg-Gln-Arg-Arg-AMC (**4**) is more active than the +2 charged Arg-Arg-AMC (**5**), which is in turn more active than the +1 charged Arg-AMC (**3**) [EC_{50} = 50, 100, and ∼300 μM, respectively (Table 1 and Figure S2)]. In contrast, an uncharged AMC derivative, such as Ala-AMC (**6**), has an at least 5-fold lower activity [EC_{50} = 1600−1700 μM (Table 1 and Figure S2)] compared to those with a net positive charge. Note, however, that we found spermidine, a compound with a high positive charge, to inhibit the conversion from kissing dimer to extended dimer (data not shown), indicating that positive charge alone cannot account for the activities of these small molecule activators.

(iv) Substitution on the Benzene Ring in the Coumarin Scaffold Is Preferred. With the results listed in Table 1, we can compare Arg-amidocoumarin analogues to examine whether activation is dependent on the position of the charged substituent. For better activity, Arg is preferred on the benzene ring, as in the case of Arg-7-amidocoumarin (**7**) or Arg-6-amidocoumarin (**8**) [EC_{50} ∼ 500 and 600 μM, respectively (Table 1 and Figure S3)], rather than the α-pyrone ring, as in the case of Arg-3-amidocoumarin (**9**) [EC_{50} = 1400−1500 μM (Table 1 and Figure S3)]. The analogue with Arg on the α-pyrone ring has a 2−3-fold decrease in activity, implying that the activator inserts into the RNA in a specific orientation.

(v) A Methyl Group and O1 of AMC Are Important for Activity. In investigating the important features of the AMC scaffold, we removed the methyl group at the 4 position and substituted other bicyclic aromatic scaffolds with different hydrogen bonding capacities. First, removal of the methyl group decreases the activity ∼2-fold if one compares the EC_{50} values of **3** and **7** [∼300 and ∼500 μM, respectively (Table 1 and Figure S4)]. Second, the hydrogen bond acceptor O_1 is important for activity. When we compare the activity of **7** and Arg-7-amido-β-naphthalene (**10**) [EC_{50} ∼ 500 and 740 μM, respectively (Table 1 and Figure S4)], it is not surprising to observe the decrease in activity because the β-naphthalene ring does not have the ability to form hydrogen bonds that might be important for RNA−ligand interactions. Although both **10** and Arg-7-amido-4-methyl-2-hydroxyquinoline (**11**) have similar EC_{50} values [∼740 and ∼850 μM, respectively (Table 1 and Figure S4)], the ring system of **11** is less active than **10** when one takes into the account the 1.4-fold decrease in activity caused by deletion of the methyl group. This implies O_1 in the coumarin ring is an important hydrogen acceptor, and it is probably involved in RNA−ligand interaction for binding, activation, or both; substitution of O_1 with a hydrogen bond donor or even elimination of the possibility of H-bonding at this position results in a substantial decrease in activity.

NMR Study of the Activator−RNA Interaction. The saturation transfer difference NMR experiment (*31*) was used to identify the binding epitope on the ligand for its interaction with SL1-wt KD. For STD NMR, a selective pulse is used to irradiate a narrow spectral region (typically between 5 and 6 ppm) of RNA signals, where no ligand proton resonances are present. Due to efficient spin diffusion in macromolecules, all RNA proton resonances can be saturated in a few seconds. When a ligand binds specifically to the RNA, magnetization is transferred to the ligand protons that interact with the RNA proportional to the proximity to any saturated RNA protons. Due to rapid exchange between the bound and free ligands in solution, the degree to which the ligand protons are saturated will still be manifest in the average signal from the ligand. In the STD spectrum, resulting from subtracting spectra obtained with and without the selective pulse, only those ligand protons that interact with the RNA exhibit signals with signal intensity corresponding to the proximity to the RNA.

Both KA-AMC and AMC bind to SL1-wt KD (panels a and b of Figure 3, respectively), which is evident in the significant line broadening (shown in spectra ii) and STD effects (spectra iii) in the presence of RNA. Clearly, KA-AMC is affected much more than AMC, however. It is interesting to note that protons on the benzene ring of the AMC scaffold interact with RNA for both KA-AMC and AMC. For KA-AMC, aromatic protons and the methyl group on the AMC scaffold exhibit a significant upfield chemical shift change, undoubtedly arising from ring currents due to stacking with the aromatic bases of the RNA. Resonances of Ala H_α and Lys H_ε also manifest STD effects. Unfortunately, we cannot distinguish whether the signal at 1.5 ppm in the STD spectrum belongs to the methyl group of Ala or H_γ of Lys due to line broadening and similar chemical shifts.

DISCUSSION

The study reported here was motivated by the surprising discovery of a small molecule, KA-AMC, that activates RNA dimer maturation. As that activity is roughly 10^4 times weaker than that of the nucleocapsid protein, it was deemed important to optimize the activity and concomitantly to discern the features of the small molecule that contributed to the activity. We hypothesize that KA-AMC utilizes its positively charged KA dipeptide moiety to form not only nonspecific electrostatic interactions with backbone phosphoryl groups on the RNA but also hydrogen bonds, while the AMC moiety stacks with the RNA aromatic bases. This hypothesis is supported by the correlation that increasing positive charge enhances activation. A more positively charged derivative has a stronger electrostatic interaction and consequently binds to the RNA more tightly, allowing a

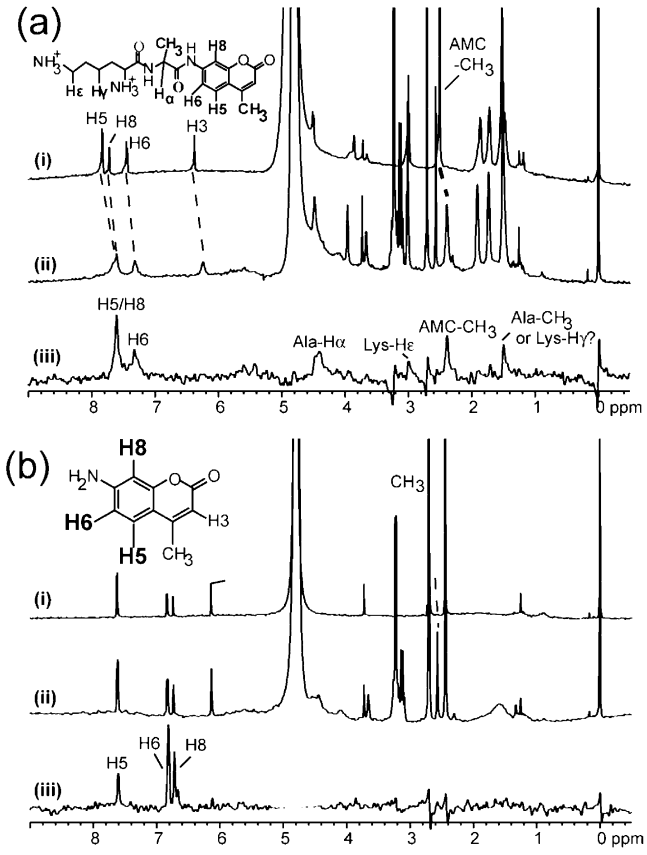


FIGURE 3: NMR spectra of KA-AMC and AMC. For all spectra, the compound is at a concentration of 200 μM and the SL1-wt KD is at a concentration of 5 μM. (a) (i) Reference spectrum of KA-AMC alone. (ii) Reference spectrum of KA-AMC with the SL1-wt KD. All of the aromatic protons and the methyl group of the AMC base have an upfield shift (shown with dashed lines) in addition to line broadening. (iii) STD NMR spectrum of KA-AMC with the SL1-wt KD. The RNA is irradiated at 5.5 ppm. The ligand protons that manifest a STD effect (shown in bold in the chemical structure) include all aromatic protons and the methyl group on the AMC base and some aliphatic protons in the Lys-Ala side chain. (b) (i) Reference spectrum of AMC alone. (ii) Reference spectrum of AMC with the SL1-wt KD. All ligand protons are broadened slightly due to binding. Only the methyl group shows a significant upfield shift when binding to the RNA. (iii) STD NMR spectrum of AMC in the presence of the SL1-wt KD. The RNA is irradiated at 5.5 ppm. Only protons on the benzene ring of the AMC show a STD effect (shown in bold in the chemical structure).

greater fraction of the AMC moiety to be bound and activating; as a consequence, the overall concentration of the activator required to refold RNA is lower. While charge on the substituent is a big plus for activation, it is not the sole determinant: the nature of the positive side chain and position on the AMC ring system are equally important. For activity, the guanidino group of Arg is better than the localized amino group, although both have a net +1 charge, implying that the capacity to form hydrogen bonds with RNA also enhances activity. It is well-known that RNA favors binding to the guanidino group of arginyl residues in proteins (*32*). The hypothesis is also consistent with the observation that the KA dipeptide and AMC moieties must be covalently bonded for activity. The positively charged side chain provides a driving force to bring AMC into the proximity of the RNA for further interaction. We note, however, that AMC alone has exceedingly weak activity and the dipeptide alone has no detectable activity. In contrast to some nucleic acid intercalators, which do not have such specificity, AMC apparently stacks with the RNA in a certain orientation inducing specific interactions between the small molecule and RNA. This is shown by the strong requirement of a hydrogen bond acceptor O_1 in AMC and the optimal placement of the positive charge on the benzene ring in the scaffold.

The SAR results are consistent with the NMR results. Figure 3 shows that many KA-AMC proton resonances are broadened and shifted upfield in the presence of the SL1-wt KD. Clearly, it is binding and stacking with aromatic rings on the RNA. The AMC alone does bind very weakly, which is evident in the slight line broadening and at least one signal being shifted upfield. This observation is consistent with the notion that the KA dipeptide and AMC base must be covalently linked for optimal activity.

The STD NMR data (Figure 3) enable us to map the ligand binding epitope. All protons on the benzene moiety of the AMC scaffold are used for binding in KA-AMC and AMC (panels a and b of Figure 3, part iii). This result is consistent with the finding that the optimal placement of the positive charge is on the benzene ring of the AMC scaffold, and placement of the charge on the α-pyrone ring decreases the activity by 2−3-fold. Furthermore, the methyl group of AMC, H_α of Ala, and H_ε of Lys show STD effects in KA-AMC. An additional STD signal at 1.5 ppm also arises from the side chain, but it is not clear whether it belongs to the methyl group of Ala (1.53 ppm in the free form) or H_γ of Lys (1.47 ppm in the free form) of KA-AMC due to line broadening in the presence of RNA. It is interesting to find that the methyl group of AMC exhibits a STD effect in KA-AMC but not in the AMC methyl group, although there is an upfield shift in both cases. The importance of the methyl group is reflected in the SAR studies: removal of the methyl group decreases activity ~2-fold. Lastly, the STD effect of H_α supports the notion that AMC and the KA dipeptide need to be covalently linked together for activity, while the STD effect of H_ε can be a consequence of the terminal amino group interacting with the RNA.

Interestingly, AMC derivatives, such as RR-AMC, can bind to other stem−loop RNA (HIV-1 SL2 and SL3 hairpins) and the RNA duplex in a manner similar to NCp7 binding to other nucleic acids; however, the intermolecular interaction with the different RNA constructs does not appear to be the same for the different constructs, based on our preliminary NMR experiments (data not shown). We are currently investigating the mechanism of action of RR-AMC in dimer maturation and how it recognizes different RNA molecules.

Although KA-AMC does not inhibit dimer maturation, it could potentially lead to development of an antiretroviral agent in one of two ways. The most obvious way is to use our growing knowledge of KA-AMC analogue binding features to design a compound that would bind and inhibit dimer maturation rather than promote dimer maturation. The less obvious, but potentially more effective, way is by inducing early maturation of the genomic dimer prior to budding. Dimerization is a highly regulated process and has been postulated to be regulated by a riboswitch mechanism (*33*): The HIV-1 leader can adopt two mutually exclusive conformations, branched multiple hairpin (BMH) and long distance interaction (LDI), and the DIS is only exposed in the BMH conformation. This BMH−LDI riboswitch model

has been used to explain regulation of dimerization in vitro (*34*); a change in equilibrium between these conformations affects RNA dimer formation (*35*). Should this riboswitch model be proven, our small molecule activator may potentially be modified to affect the equilibrium between the BMH and LDI conformations and consequently adversely affect dimerization as a novel HIV-1 therapeutic.

It has been shown that both unproteolyzed Gag and mature NCp7 proteins can promote dimer maturation in vitro (*36*), yet the stability of genomic RNA extracted from immature virions differs from that of mature virions (*22*), presumably because the kissing dimer is present in the immature virion. Recently, it has been shown that the HIV-1 virion infectivity factor (Vif) protein inhibits NCp7-induced genomic maturation in vitro and may be a temporal regulator of RNA folding (*37*). Moreover, Vif is a RNA-binding protein that interacts with the genomic RNA to form a 40S mRNA−protein particle (mRNP) in the cytoplasm of the virus-producing cell, thus protecting the RNA from degradation, preventing improper engagement with ribosomes, and maintaining proper folding that is suitable for packaging. In the presence of Gag precursors, the affinity of Vif for RNA decreases. It was then postulated that Vif in the mRNP complex initiates binding of NC domains of the Gag precursors with the RNA. As the RNA binds to NC domains of the Gag precursors, the binding affinity of Vif decreases, and it eventually dissociates from the RNA. This displacement process may guide the proper folding and condensation of viral RNA during packaging (*38*). Disrupting the proper folding of the RNA by inducing early genomic maturation during packaging could be detrimental and a novel strategy for combating HIV. In addition, there have been no studies on the effects of early genomic maturation and its impact, so KA-AMC analogues could serve as a useful chemical biology tool.

Finally, there are many coumarin-based anti-HIV inhibitors targeted at reverse transcription, integration, and proteolysis [recently reviewed (*39*)] due to its diverse pharmacological and biochemical properties, depending on substitution patterns. It is possible to chemically modify KA-AMC further to improve affinity and efficacy for promoting early dimer maturation or as a scaffold for structure-based design of dimer maturation inhibitors targeted at the SL1 RNA−NCp7 interaction. In essence, it might be possible to develop a compound that could be an effective drug with multiple mechanisms, although it would not be the first multiple-mechanism drug.

ACKNOWLEDGMENT

We thank Dr. Nikolai B. Ulyanov for helpful discussions and advice and Dr. Ken Lind for some preliminary docking calculations.

SUPPORTING INFORMATION AVAILABLE

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

REFERENCES

1. Kung, H. J., Hu, S., Bender, W., Bailey, J. M., Davidson, N., Nicolson, M. O., and McAllister, R. M. (1976) RD-114, baboon, and woolly monkey viral RNA's compared in size and structure. *Cell 7*, 609–620.
2. Bender, W., Chien, Y. H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B., and Davidson, N. (1978) High-molecular-weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J. Virol. 25*, 888–896.
3. Bender, W., and Davidson, N. (1976) Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell 7*, 595–607.
4. Murti, K. G., Bondurant, M., and Tereba, A. (1981) Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. *J. Virol. 37*, 411–419.
5. 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.
6. Roy, C., Tounekti, N., Mougel, M., Darlix, J. L., Paoletti, C., Ehresmann, C., Ehresmann, B., and Paoletti, J. (1990) An analytical study of the dimerization of in vitro generated RNA of Moloney murine leukemia virus MoMuLV. *Nucleic Acids Res. 18*, 7287–7292.
7. Darlix, J. L., Gabus, C., Nugeyre, M. T., Clavel, F., and Barre-Sinoussi, F. (1990) Cis elements and trans-acting factors involved in the RNA dimerization of the human immunodeficiency virus HIV-1. *J. Mol. Biol. 216*, 689–699.
8. 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.
9. 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.
10. 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.
11. Laughrea, M., Jetté, L., Mak, J., Kleiman, L., Liang, C., and Wainberg, M. A. (1997) Mutations in the kissing-loop hairpin of human immunodeficiency virus type 1 reduce viral infectivity as well as genomic RNA packaging and dimerization. *J. Virol. 71*, 3397–3406.
12. Harrison, G. P., and Lever, A. M. (1992) The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol. 66*, 4144–4153.
13. Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B., and Ehresmann, C. (1993) Functional sites in the 5′ region of human immunodeficiency virus type 1 RNA form defined structural domains. *J. Mol. Biol. 229*, 382–397.
14. Clever, J., Sassetti, C., and Parslow, T. G. (1995) RNA Secondary Structure and Binding Sites For Gag Gene Products in the 5′ Packaging Signal of Human Immunodeficiency Virus Type 1. *J. Virol. 69*, 2101–2109.
15. 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.
16. Leitner, T., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J. W., Wolinsky, S., and Korber, B., Eds. (2005) *HIV Sequence Compendium*, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
17. 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.
18. Paillart, J. C., Berthoux, L., Ottmann, M., Darlix, J. L., Marquet, R., Ehresmann, B., and Ehresmann, C. (1996) A dual role of the putative RNA dimerization initiation site of human immunodeficiency virus type 1 in genomic RNA packaging and proviral DNA synthesis. *J. Virol. 70*, 8348–8354.
19. 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.
20. 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.
21. Huthoff, H., and Berkhout, B. (2002) Multiple secondary structure rearrangements during HIV-1 RNA dimerization. *Biochemistry 41*, 10439–10445.

22. 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.
23. Mujeeb, A., Clever, J. L., Billeci, T. M., James, T. L., and Parslow, T. G. (1998) Structure of the Dimer Initiation Complex of the HIV-1 Genomic RNA. *Nat. Struct. Biol. 5*, 432–436.
24. 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.
25. 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.
26. 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 stem-loop of HIV-1: Sequence effects and mechanism of RNA refolding. *Nucleic Acids Res. 35*, 2026–2034.
27. 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.
28. Kick, E. K., Roe, D. C., Skillman, A. G., Liu, G., Ewing, T. J. A., Sun, Y., Kuntz, I. D., and Ellman, J. A. (1997) Structure-Based Design and Combinatorial Chemistry Yield Low Nanomolar Inhibitors of Cathepsin D. *Chem. Biol. 4*, 297–307.
29. Abagyan, R., and Totrov, M. (1994) Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations For Peptides and Proteins. *J. Mol. Biol. 235*, 983–1002.
30. Lind, K. E., Du, Z., Fujinaga, K., Peterlin, B. M., and James, T. L. (2002) Structure-Based Computational Database Screening, In vitro Assay, and NMR Assessment of Compounds that Target TAR RNA. *Chem. Biol. 9*, 185–193.
31. Mayer, M., and James, T. L. (2002) Detecting ligand binding to a small RNA target via saturation transfer difference NMR experiments in D_2O and H_2O. *J. Am. Chem. Soc. 124*, 13376–13377.
32. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) Arginine-Mediated RNA Recognition: The Arginine Fork. *Science 252*, 1167–1171.
33. Ooms, M., Huthoff, H., Russell, R., Liang, C., and Berkhout, B. (2004) A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol. 78*, 10814–10819.
34. Huthoff, H., and Berkhout, B. (2001) Two alternative structures of the HIV-1 leader RNA. *RNA 7*, 143–157.
35. Abbink, T. E., Ooms, M., Haasnoot, P. C., and Berkhout, B. (2005) The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry 44*, 9058–9066.
36. 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.
37. Henriet, S., Sinck, L., Bec, G., Gorelick, R. J., Marquet, R., and Paillart, J. C. (2007) Vif is a RNA chaperone that could temporally regulate RNA dimerization and the early steps of HIV-1 reverse transcription. *Nucleic Acids Res. 35*, 5141–5153.
38. Zhang, H., Pomerantz, R. J., Dornadula, G., and Sun, Y. (2000) Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. *J. Virol. 74*, 8252–8261.
39. Kostova, I., Raleva, S., Genova, P., and Argirova, R. (2006) Structure-Activity Relationships of Synthetic Coumarins as HIV-1 Inhibitors. *Bioinorg. Chem. Appl.,* 68274.

BI800230M