

# The three-way junction structure of the HIV-1 PBS-segment binds host enzyme important for viral infectivity

Zhenwei Song<sup>1</sup>, Thomas Gremminger<sup>1</sup>, Gatikrushna Singh<sup>2</sup>, Yi Cheng<sup>1,3,4</sup>, Jun Li<sup>1,3,4</sup>, Liming Qiu<sup>1,3,4,5</sup>, Juan Ji<sup>1</sup>, Margaret J. Lange<sup>6</sup>, Xiaobing Zuo<sup>7</sup>, Shi-Jie Chen<sup>1,3,4</sup>, Xiaoqin Zou<sup>1,3,4,5</sup>, Kathleen Boris-Lawrie<sup>2,\*</sup> and Xiao Heng<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry, University of Missouri, Columbia, MO, 65211, USA, <sup>2</sup>Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA, <sup>3</sup>Department of Physics and Astronomy, University of Missouri, Columbia, MO 65211, USA, <sup>4</sup>Institute for Data Science and Informatics, University of Missouri, Columbia, MO 65211, USA, <sup>5</sup>Dalton Cardiovascular Research Center, University Missouri, Columbia, MO 65211, USA, <sup>6</sup>Department of Molecular Microbiology and Immunology, University of Missouri, Columbia, MO 65211, USA and <sup>7</sup>X-Ray Science Division, Argonne National Laboratory, Lemont, IL, 60439, USA

Received March 18, 2021; Revised April 15, 2021; Editorial Decision April 16, 2021; Accepted April 19, 2021

## ABSTRACT

HIV-1 reverse transcription initiates at the primer binding site (PBS) in the viral genomic RNA (gRNA). Although the structure of the PBS-segment undergoes substantial rearrangement upon tRNA<sup>Lys3</sup> annealing, the proper folding of the PBS-segment during gRNA packaging is important as it ensures loading of beneficial host factors. DHX9/RNA helicase A (RHA) is recruited to gRNA to enhance the processivity of reverse transcriptase. Because the molecular details of the interactions have yet to be defined, we solved the solution structure of the PBS-segment preferentially bound by RHA. Evidence is provided that PBS-segment adopts a previously undefined adenosine-rich three-way junction structure encompassing the primer activation stem (PAS), tRNA-like element (TLE) and tRNA annealing arm. Disruption of the PBS-segment three-way junction structure diminished reverse transcription products and led to reduced viral infectivity. Because of the existence of the tRNA annealing arm, the TLE and PAS form a bent helical structure that undergoes shape-dependent recognition by RHA double-stranded RNA binding domain 1 (dsRBD1). Mutagenesis and phylogenetic analyses provide evidence for conservation of the PBS-segment three-way junction structure that is preferentially bound by RHA in support of efficient reverse transcription, the hallmark step of HIV-1 replication.

## INTRODUCTION

The HIV-1 5' untranslated region (5'UTR) consists of a series of structural elements that coordinate various viral replication events. The hallmark event of reverse transcription initiates on the 5'UTR by annealing of tRNA<sup>Lys3</sup> with the complementary primer binding site (PBS) (1). Annealing does not occur spontaneously at physiological temperature, as both the PBS-containing segment (PBS-segment) of viral RNA and tRNA<sup>Lys3</sup> are highly structured. Complete annealing requires mature nucleocapsid (NC) protein, which is produced after maturation of the virion proteins (2–5).

Several discrete regions in the PBS-segment are known to interact with viral and host factors to promote reverse transcription (6–14). The primer activation signal (PAS) has been reported to promote tRNA priming and enhance reverse transcription efficiency (6–9), but it is under debate whether PAS directly interacts with tRNA<sup>Lys3</sup> during reverse transcription initiation (15). Lysyl-tRNA synthetase (LysRS) has been reported to facilitate the annealing step by placing tRNA<sup>Lys3</sup> onto the PBS-segment. The process is mediated by the interactions between the anticodon binding domain of LysRS and the tRNA-like element (TLE) within the PBS-segment (10), which mimics the tRNA<sup>Lys3</sup> anticodon for LysRS binding and increases the efficiency of tRNA<sup>Lys3</sup> annealing (11–14).

DHX9/RNA helicase A (RHA) co-assembles with the HIV-1 genomic RNA (gRNA) and bolsters virion infectivity (16,17). RHA belongs to the DExH-box superfamily and unwinds RNA duplex in the 3'- to 5'- direction. It has two double-stranded RNA binding domains (dsRBDs) at the

\*To whom correspondence should be addressed. Tel: +1 573 882 3953; Email: [hengx@missouri.edu](mailto:hengx@missouri.edu)  
Correspondence may also be addressed to Kathleen Boris-Lawrie. Tel: +1 612 625 2700; Email: [kbl@umn.edu](mailto:kbl@umn.edu)

N-terminus, followed by two RecA-like domains in the core helicase region, and an RG-rich domain on the C-terminus. Both dsRBDs are required for RHA to interact with the 5'UTR (16–18). A point mutation that disrupted the structure of the PBS-segment was deficient in interacting with the N-terminus domain of RHA (dsRBD1+dsRBD2) and resulted in reduced infectivity (19). RHA does not re-arrange the annealed RNA complex for reverse transcription initiation, but enhances processivity of reverse transcriptase (RT) during the elongation phase by unwinding the RNA secondary and/or tertiary structures (20).

Hence, prior to tRNA<sup>Lys3</sup> annealing, the PBS-segment RNA serves as a scaffold to interact with viral and host factors, including aforementioned NC, tRNA<sup>Lys3</sup>, LysRS, RHA and possibly other yet to be characterized factors (6–14). These interactions may be mechanistically linked to achieve appropriate primer annealing for the temporal control of reverse transcription initiation. Chemical probing, enzymatic probing, computational modeling, as well as recent small angle X-ray scattering (SAXS) data supported formation of the TLE stem loop and PAS stem in the PBS-segment, connected by a long flexible loop that encompasses the tRNA<sup>Lys3</sup> annealing site (21–25). However, these structural features leave gaps in knowledge of structural basis for the recruitment of some factors, including RHA. Lack of high-resolution structural information of the PBS-segment has hampered the structure-function investigation of the RNA: protein interactions necessary for efficient reverse transcription initiation.

Here, we report the solution structure of the HIV-1 PBS-segment RNA and its crucial role in RNA helicase interaction necessary for efficient reverse transcription. The results characterize a previously unidentified three-way junction structure that is composed of the TLE, tRNA annealing arm and PAS. Phylogenetic analysis of HIV-1 clinical sequences documents that the secondary structure was highly conserved among various HIV-1 strains, and computational mutagenesis scanning analysis shows that the base pairings in the bottom of the TLE stem are necessary to maintain the three-way junction structure. SAXS data of PBS-segment mutant RNAs demonstrate that disruption of base pairings at the bottom of TLE stem altered RNA folding, and reduced infectivity of progeny viruses attributable to deficient recruitment of RHA to the PBS-segment, as would be expected based on previous results (19). Combining electrophoretic mobility shift assay (EMSA), isothermal titration calorimetry (ITC), NMR and advanced molecular docking, we show that the three-way junction structure of the PBS-segment is specifically recognized by RHA. RHA's dsRBD1 and dsRBD2-Core domains are likely to manifest the selective shape recognition mechanism. In summary, our data demonstrated that the three-way junction structure of the PBS-segment structure is conserved to support infectivity of progeny viruses.

## MATERIALS AND METHODS

### Plasmids

The PBS-segment, 5'UTR, 5'UTR<sup>ΔTAR-PolyA</sup> and 5'UTR<sup>ΔPBS</sup> plasmids used for RNA *in vitro* transcription were constructed as previously described (19). The

PBS-segment mutant template plasmids for RNA transcription were generated by site-directed mutagenesis. The vector pNL4-3-CMV-EGFP used for infectivity measurement had deletion of *vif*, *vpr*, *vpu*, *nef* and *env* open reading frames (ORF), and contained a GFP ORF driven by a CMV promoter (26). The PBS mutations in pNL4-3-CMV-EGFP for infectivity assays were generated by multiple DNA fragment assembly using primers listed in Supplementary Tables S1 and S2. (GenBuilder Cloning Kit, GenScript). All of the sequences were confirmed by Sanger sequencing (DNA Core, University of Missouri).

### RNA *in vitro* transcription

All of the RNA constructs, including 5'UTR, 5'UTR<sup>ΔTAR-PolyA</sup>, 5'UTR<sup>ΔPBS</sup>, TAR-PolyA, PBS-segment and mutants, TLE, PAS stem, hairpin-control and tRNA<sup>Lys3</sup>, were made by *in vitro* T7 transcription reactions. The DNA templates of 5'UTR, 5'UTR<sup>ΔTAR-PolyA</sup>, 5'UTR<sup>ΔPBS</sup>, TAR-PolyA, PBS-segment and mutants were made by PCR amplifying corresponding region of the plasmids with a 5'-Top17 sequence (5'-GAAATTAATACGACTCACTATA-3'), the five extra base pairs at the 5'-end were added to promote T7 binding and enhance RNA transcription yield). The template DNAs of TLE, PAS stem, hairpin-control and tRNA<sup>Lys3</sup> were custom synthesized (IDT). The nucleotide specific <sup>2</sup>H labeled samples, including A<sup>2R</sup>GU<sup>R</sup>-, A<sup>2R</sup>C<sup>R</sup>U-, A<sup>2R</sup>C<sup>R</sup>U<sup>R</sup>-, A<sup>2R</sup>G<sup>R</sup>-, AC- and AG-PBS-segment (nomenclature here: the letters denote the nucleosides containing <sup>1</sup>H which is visible by NMR, R = ribose, A<sup>2R</sup> means H2 and ribose hydrogens are protonated; A<sup>2R</sup>GU<sup>R</sup>-PBS-segment means the H2 and ribose of adenosine, ribose of uracil and guanosine are protonated, and cytidine is fully deuterated) were prepared by incorporating corresponding deuterated and protonated rNTPs in T7 transcriptions as previously described (19,27). Fully deuterated rNTPs and H5,H6-deuterated CTP and UTP were purchased from Silantes (Silantes GmbH, Munich) and Cambridge Isotope Laboratories (CIL, Andover, MA). H8-deuterated ATP and GTP were prepared in the lab (19). <sup>13</sup>C/<sup>15</sup>N labeled RNAs were synthesized by incorporating <sup>13</sup>C/<sup>15</sup>N labeled rNTPs (CIL, Andover, MA) in T7 transcription.

### RNA refolding

The PBS-segment RNA used for biophysical studies was refolded by preparing the RNA in 10 mM Tris-HCl, pH 7.5, incubating at 95°C for 3 min, snap cooling on ice, and then mixing with salts to reach final salt concentrations of binding buffer A (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM KCl and 1 mM MgCl<sub>2</sub>). The RNA samples were then incubated at 37°C for 30 min prior to EMSA and ITC experiments.

### Protein purification

The plasmid for recombinant RHA expression was a kind gift from Dr William Clay Brown (University of Michigan). The recombinant RHA contained N-terminal His<sub>6</sub>- and Mocr-tags and was expressed in insect cells. Purification of RHA was performed as previously described (20)

and the protein was stored in storage buffer (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM β-mercaptoethanol and 10% glycerol). The recombinant N-terminally His<sub>6</sub>-tagged dsRBD1 was expressed in *Escherichia coli* BL21(DE3)-pLysS cells (Invitrogen) treated with 0.3 mM IPTG for 4 h at 37°C. To express <sup>15</sup>N labeled dsRBD1, the bacteria were grown in the LeMaster and Richards minimal medium with <sup>15</sup>NH<sub>4</sub>Cl (0.5 g/l) as the sole nitrogen source. The harvested cells were resuspended in lysis buffer (40 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 7.5) with lysozyme, sonicated and centrifuged. The supernatant was applied to a Cobalt column (HisPur Cobalt Resin, Thermo Scientific), and the protein was eluted with lysis buffer containing additional 150 mM imidazole. To remove the His<sub>6</sub>-tag, the protein was digested overnight at 4°C with TEV protease (1 mg protease per 20 mg of His<sub>6</sub>-dsRBD1) and dialyzed in low-salt buffer (30 mM Tris, 100 mM NaCl, 2 mM EDTA, 5 mM β-mercaptoethanol). The protein was then subjected to ion exchange and size exclusion chromatography, and stored in storage buffer at -80°C. The plasmid for recombinant NC (HIV-1 NL4-3 strain) expression was a kind gift from Dr Michael F. Summers (University of Maryland, Baltimore County). The NC protein was expressed and purified as described elsewhere (28). The pRT-Dual plasmid (29) for recombinant RT expression (HIV-1 HXB2 strain) was a kind gift from Dr. Donald Burke (University of Missouri, Columbia). The RT protein was expressed and purified as described previously (30).

### Cells and viruses

293FT cells (Invitrogen) were propagated in Dulbecco's modified Eagle's medium (DMEM, Sigma) supplemented with 10% fetal bovine serum (FBS). Wildtype and mutant viruses were produced by transfection as previously described (26). Briefly, 293FT cells were plated in 10% FBS DMEM at density of  $5 \times 10^5$  cells per well in six-well plates overnight and then co-transfected with 500 ng of pNL4-3-CMV-EGFP or mutants together with 100ng pMD-G (AIDS Reagent Program), which encodes the vesicular stomatitis virus glycoprotein, VSV-G, for pseudotyping. The mixtures were added dropwise to the cells and incubated for 6 hrs in 5% CO<sub>2</sub> at 37 °C, rinsed, and cultured in 2 ml of 10% FBS DMEM. The supernatant was harvested 48 h post-transfection by centrifugation at 1000 rpm for 10 min at 4°C, and the vector virus preparations passed through 0.45 μm filters. The transfected cells were collected and fixed with 4% of paraformaldehyde, and transfection efficiency was monitored by detection of EGFP on an Accuri C6 Flow Cytometer (BD Biosciences, San Jose, CA, USA). The Gag p24 was quantified by ELISA using pre-coated MicroFluere plates (Optofluidic Bioassay, MI).

### Infectivity assay

Equivalent amounts of VSV-G-pseudotyped vectors viruses (250 ng of Gag p24) with or without mutations in the PBS-segment were used to transduce  $2 \times 10^5$  of TZM-bl cells in 12-well plates. The infected cells were collected and fixed with 4% paraformaldehyde 24 hrs post-transduction, and the percentage of cells expressing EGFP was measured using the BD Accuri C6 Flow Cytometer (BD Biosciences,

San Jose, CA, USA). The infectivity was calculated as the percentage of EGFP-expressing cells in mutant virus infected cells compared to that of the wild type (WT).

### Quantification of reverse transcription products in infected cells

MT4 lymphocytes ( $5 \times 10^5$  cells in 0.5 ml of RPMI medium) were seeded in 12-well plates and incubated with equivalent cell-free medium containing vector viruses normalized to 200 ng of Gag p24 and spinoculated for 2 h, washed twice in RPMI medium and cultured for 4 h. The inactivated virus control sample was heated to 100°C for 5 min, cooled to room temperature, centrifuged at 12 000 rpm for 1 min, and treated with DNase I for 1 h. The infected cells were collected and washed twice with 1× phosphate buffered saline and DNA was extracted in the QIAamp DNA blood mini kit (Qiagen). Cellular DNA (600 ng) was evaluated by qPCR with specific primer sets that detect early and late products of HIV-1 reverse transcription (Supplementary Table S3) (31). The abundance of copies was determined relative to standard curves generated on gene-specific PCR amplicons compared to mock and heat-inactivated virus infection.

### In vitro tRNA annealing, primer extension and tRNA placement assays

The annealing reaction was initiated by mixing 20 pmol PBS-segment, 20 pmol tRNA<sup>Lys3</sup>, and 600 pmol NC (~6 nucleotides per NC molecule) in 20 μl of buffer B (10 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl and 5 mM MgCl<sub>2</sub>). The mixture was incubated at 37°C for 30 min, and the reaction was terminated by incubating with 0.5% SDS and 0.2 mg/ml Proteinase K at 37°C for 30 min. The RNA was purified by phenol-chloroform extraction and electrophoresed on 10% polyacrylamide gels.

Primer extension assays were performed with 5'-Cy3-labeled tRNA<sup>Lys3</sup> annealed on PBS-segment and mutants in the presence of NC. To label tRNA<sup>Lys3</sup> on the 5'-terminus, GMP-primed tRNA<sup>Lys3</sup> was synthesized by mixing GMP with rNTPs in T7 transcription, and the RNA was fluorophore labeled with a Cy3-NHS ester (Lumiprobe) (20,32). The primer extension reaction was initiated by mixing 0.33 mM dNTP with a solution containing 0.1 μM template and 0.4 μM RT, incubated in 37°C for 0, 0.5, 1.0, 1.5, 2.0, 3.0 and 4.0 min, and quenched by 25 mM EDTA. The proteins were removed from the reactions by phenol-chloroform extraction, and the primer extension products were resolved on 10% denaturing polyacrylamide gels.

To check the tRNA placement on the WT and mutant viral RNAs, 200 μl of WT and mutant pseudotyped viruses were harvest from transfected 293FT cells and treated with Dnase I and DpnI at 37°C for 90 min. The viral RNA was extracted with TRIzol (Invitrogen). To check if the viral RNA was pre-annealed with tRNA<sup>Lys3</sup> to prime reverse transcription, 2 ng of extracted RNA was mixed with dNTP (10 mM each), 5× First-Strand Buffer (Invitrogen), 0.1 M DTT and RNaseOUT™ (40 units/μl), and the mixture was incubated at 42°C for 2 min. Primer extension was initiated by adding 1 μl (200 units) of SuperScript™ II RT (Invitrogen), and the reaction was incubated

at 42°C for 50 min. To quantify the (-)cdDNA products, 2  $\mu$ l of cdDNA generated from the previous step was mixed with the primer/probe set targeting the RU5 region (Supplementary Table S3) for qPCR (Luna<sup>®</sup> Universal Probe qPCR Master Mix, NEB). To quantify the total gRNA, the primer/probe set (Supplementary Table S3) targeting gag region was used in RT-qPCR (Luna<sup>®</sup> Universal Probe One-Step RT-qPCR, NEB). To detect if plasmid DNA were carried over through RNA extraction, the same amount of RNA (0.2 ng) and primer/probe set targeting the gag region were used in qPCR. Both qPCR and RT-qPCR were performed on CFX96 Real-Time PCR Detection System (Bio-Rad).

### NMR spectroscopy

The PBS-segment and its control fragment samples were prepared (200–300  $\mu$ M) and pre-incubated in 10 mM Tris-d<sub>11</sub>, pH 7.5 and 1 mM MgCl<sub>2</sub> at 37°C for 30 min. Two-dimensional <sup>1</sup>H–<sup>1</sup>H NOESY and <sup>1</sup>H–<sup>1</sup>H TOCSY data for control RNA fragments, and <sup>1</sup>H–<sup>1</sup>H NOESY data for fully protonated and site-specific deuterated PBS-segment RNA and in complex with dsRBD1 were collected in D<sub>2</sub>O (99.96%, CIL) at 308 K. <sup>1</sup>H–<sup>13</sup>C HMQC data were collected for <sup>13</sup>C/<sup>15</sup>N labeled control RNA fragments. One- and two-dimensional imino proton spectra were collected for the PBS-segment RNA and control RNAs (200–300  $\mu$ M) prepared in buffer containing 10% D<sub>2</sub>O + 90% H<sub>2</sub>O. HN-COSY data were collected for <sup>13</sup>C/<sup>15</sup>N labeled TLE stem loop RNA. <sup>1</sup>H–<sup>15</sup>N TROSY data were collected for 200  $\mu$ M dsRBD1 titrated with 0, 54, 108, 136 and 160  $\mu$ M of the PBS-segment RNA in binding buffer B (10 mM Tris-d<sub>11</sub>, pH 7.5, 500 mM KCl and 1 mM MgCl<sub>2</sub>) in 10% D<sub>2</sub>O and 90% H<sub>2</sub>O at 318 K. All the NMR data were collected on a Bruker Avance III 800 MHz spectrometer equipped with TCI cryoprobe (NMR Core, University of Missouri). The NMR data were processed by NMRPipe (33) and analyzed by NMRViewJ (34).

### SAXS data collection and processing

SAXS data of the PBS-segment RNA were collected at Sector 18 of the Advanced Photon Source, Argonne National Laboratory. In-line data were collected by eluting 9 mg/ml PBS-segment on a Superdex-200-Increase SEC column (GE) at 0.75 ml/min flowing rate. SAXS data of the mutant RNA samples were collected at beamline 12-ID-B of Advanced Photon Source at Argonne National Laboratory, using an in-line AKTA micro FPLC setup with a Superdex 75 Increase 5/150 GL size exclusion column. The wavelength,  $\lambda$ , of X-ray radiation was set to 0.9322 Å. Scattered X-ray intensities were measured using a Pilatus 2M detector. The sample-to-detector distance was set such that the detecting range of momentum transfer  $q$  [ $= 4\pi \sin \theta / \lambda$ , where  $2\theta$  is the scattering angle] was 0.005–0.85 Å<sup>-1</sup>. The sample passed through the FPLC column and was loaded to a flow cell for SAXS measurements. The flow cell is a cylindrical quartz capillary 1.5 mm in diameter and 10  $\mu$ m wall thickness. The exposure time was set to 1 s to reduce radiation damage and data were collected at every other second. The 2D scattering images were converted to 1D SAXS

( $I(q)$  versus  $q$ ) curves through azimuthally averaging after solid angle correction and then normalizing with the intensity of the transmitted X-ray beam flux, using the beamline software. The SAXS data were then processed using ATSAS suite 2.8 and 3.0 (35). The radius of gyration ( $R_g$ ) was determined by the Guinier plot in its linearity region and pairwise distribution was plotted using GNOM (36). The *ab initio* models were generated by averaging 10 independent DAMMIF runs on DAMAVER (37).

### Structure calculation

The PBS-segment structure was initially calculated using NMR-derived restraints by CYANA (38). Standard torsion angle restraints were used for regions of A-helical geometry in PAS stem and TLE stem loop. Standard hydrogen bonding restraints for Watson–Crick pairs in PAS stem (nt 123–131 and 217–225), TLE stem loop (nt 135–150, 158–167 and 171–177), and tRNA annealing arm (nt 187–189 and 195–197), and phosphate distances in A-form RNA helical structures in these regions were employed to maintain the major groove width in an A-form helical structure (Table 1) (38,39). The theoretical solution scattering curves of the lowest structures were calculated and compared with the experimental SAXS data by CRY SOL v2.8 (40). A total of 10 structures of lowest energy and lowest  $\chi^2$  values were selected for further molecular dynamics (MD) refinement. For each of the 10 structures, 1 ns MD simulation was performed in 1 mol/L NaCl solution under 300 K and 1 bar with NMR restraints using Amber software package (41), and the structure of lowest  $\chi^2$  value ranging from 1.8 to 2.5 in the trajectory was extracted for further refinement. In the third step, Xplor-NIH software (42) was used to further refine the structures to reduce  $\chi^2$  values. For each of the 10 structures, 100 simulated annealing MD simulations were performed with NMR and SAXS restraints. After Xplor-NIH refinement, the  $\chi^2$  values of the 1000 (10  $\times$  100) structures were reduced to 1.35–1.5. The 10 structures of the lowest energies were deposited to the RCSB protein databank (PDB: 7lva).

### Computational mutagenesis scanning

Computational mutation scanning was performed using the Vfold2D RNA structure folding model (43–46) to predict the RNA structure with single nucleotide variation at each nucleotide position. Structure prediction of a total of 309 mutants containing single point mutation in the PBS-segment was carried out. Models of lowest energy were analyzed and grouped based on the structural changes.

### EMSA

EMSA was carried out by preparing mixture solutions of 0.5  $\mu$ M RNA (5'UTR, 5'UTR <sup>$\Delta$ TAR-PolyA</sup> or 5'UTR <sup>$\Delta$ PBS</sup>) and recombinant RHA at concentrations of 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4  $\mu$ M in binding buffer A. To compare RHA binding to PBS-segment and TAR-PolyA, 0.5  $\mu$ M RNA was incubated with 0, 1, 2, 3 and 4  $\mu$ M of RHA. The mixtures were incubated at 37°C for 10 min and resolved on 1.5% agarose gels stained with 0.5  $\mu$ g/ml ethidium bromide. The gels were imaged using Gel Doc XR+ (Bio-Rad).

**Table 1.** NMR Restraints and statistics of calculated structures

Cyana <sup>a</sup>	
NOE-derived restraints	504
Intra-residue	184
Sequential	306
Long range $ i - j  > 1$	14
H-bond restraints	170
NOE restraints/residue	4.89
Torsion angles	434
Target function ( $\text{\AA}^2$ )	$0.2 \pm 0.008$
Upper distance viol. ( $\text{\AA}^2$ )	$0.0114 \pm 0.0002$
Lower distance viol. ( $\text{\AA}^2$ )	$0.0026 \pm 0.0003$
Sum VDW viol. ( $\text{\AA}^2$ )	$1.25 \pm 0.85$
RMSD ( $\text{\AA}$ )	$2.9 \pm 1.6$
SAXS	
$\chi^2$	$1.4 \pm 0.02$
MolProbity analysis <sup>b</sup>	
Clashscore	0, 100th percentile (0)
Probably wrong sugar pucker (%)	1 (0.97)
Bad backbone conformation	3 (12.6)
Bad bonds (%)	0 (0)
Bad angles (%)	2 (0.05)

<sup>a</sup>Statistics of 10 lowest energy models.

<sup>b</sup>The 10 structures refined by MD simulations were evaluated using the MolProbity Server (90,91).

EMSA experiments were repeated three times and a representative gel is shown.

## ITC

The ITC experiments were carried out by titrating 610  $\mu\text{M}$  of dsRBD1 into 17  $\mu\text{M}$  PBS-segment, and titrating 1.15 mM of dsRBD1 into 18  $\mu\text{M}$  hairpin-control RNA at 30°C on a VP-ITC (MicroCal, GE Healthcare). Heat of dilution titrations were performed by titrating the dsRBD1 protein into a matching buffer in the same experimental settings. Both protein and RNA samples were prepared in binding buffer B. The baseline was corrected by subtracting the heat of dilution, and the data were fitted using ‘one-site’ non-linear least square regression for hairpin-control and ‘two-site’ non-linear least square regression for PBS-segment. The statistics of thermodynamic parameters ( $N$ ,  $K_d$ ,  $\Delta H$ ,  $\Delta S$  and  $\Delta G$ ) were summarized from three experiments (Supplementary Table S5).

## Molecular docking

Molecular docking of dsRBD1 onto PBS-segment was guided by NMR experimental data. Using the crystal structure of human RHA dsRBD1 (PDB: 3vyy) and 1040 PBS-segment RNA structures that had been relaxed in MD simulations, a free docking by our in-house MDockPP program (47) was carried out to generate a total of 56 160 000 potential complex structures. The modeled complex structures were subsequently sorted by the protein–RNA interaction-specific scoring function ITCscorePR in a descending order based on their energy scores (48). Then, the NMR-derived constraints related to residues of the PBS segment and the dsRBD1 were incorporated to filter out compliant candidates as described below. On the RNA side, restraints impose a maximum distance of 7  $\text{\AA}$  between both H1' and C1

of G129 in PBS-segment and dsRBD1 because of the significantly shifted H1' of G129 upon dsRBD1 titration, and a minimum distance of 5  $\text{\AA}$  between atom and C2 of A220 of PBS and dsRBD1 because A220-H2 remains unshifted in the NMR titration spectra. Additionally, restraints requesting U141 and U172 of PBS-segment be within 7  $\text{\AA}$  of dsRBD1 were also applied. On the protein side, restraints were generated from the  $^1\text{H}$ – $^{15}\text{N}$  TROSY titration data, which demand the following residues, F7, G13, K14, K16, M17, T18, Y21, K29, K54, K55 and Q58, be within 7  $\text{\AA}$  of PBS-segment. These residues exhibited significant chemical shift perturbation upon PBS-segment titration ( $\Delta\delta > 0.05$  ppm). Distance restraints were also set for residue Y43 that should not be within 5  $\text{\AA}$  distance of RNA because no chemical shift perturbation was observed in the TROSY titration. Any qualifying candidate must satisfy all the specified restraints, resulting in a total of 69 modeled complex structures for further evaluation with experimental data. The best model with lowest energy was further optimized with UCSF Chimera (49).

Using Modeller (50), the structure of the dsRBD2-Core (RHA without dsRBD1 and RGG domains, 149–1150) was constructed by homology modeling based on the template structure of Drosophila MLE helicase (PDB: 5aor) (51), an orthologous protein sharing 51% sequence identity with the human RHA. Molecular docking was performed using our in-house docking program, MDockPP (47). A 6° angular interval was used to sample the relative orientations between the dsRBD2-Core and PBS-segment RNA. The resulting 54,000 potential binding poses were then filtered with three constraints: (i) The  $\alpha 1$  helix of the dsRBD2 interacts with the RNA, (ii) Region 3 residues (K235, K236) interact with the RNA and (iii) The dsRBD2-Core does not have any steric clash with the dsRBD1 domain when both bind to the RNA. For constraints (i) and (ii), a 7  $\text{\AA}$  distance cutoff was used to define the proximity to the RNA. Namely, the constraint (i) was met if at least one atom in a residue was within the cutoff distance to the RNA for every residue in the  $\alpha 1$  helix. The constraint (ii) regarding K235 or K236 was likewise determined. The constraint (iii) was based on the fact that both dsRBD1 and dsRBD2-Core bind to the RNA simultaneously in physiological condition. The assessment of this constraint was based on the buried solvent accessible surface area (SASA) between the dsRBD1 and the dsRBD2-Core. The SASA was calculated using NACCESS (52). The buried SASA equals  $(SASA_{dsRBD1} + SASA_{dsRBD2-Core} - SASA_{dsRBD1+dsRBD2-Core})/2$ , where  $SASA_{dsRBD1}$  and  $SASA_{dsRBD2-Core}$  stand for the SASA of dsRBD1 alone and the SASA of the dsRBD2-Core alone, respectively.  $SASA_{dsRBD1+dsRBD2-Core}$  represents the SASA of the dsRBD1 and the dsRBD2-Core when they were viewed as a single structure after the dsRBD1: RNA structure and the dsRBD2-Core:RNA structure were superimposed on the PBS-segment RNA. Due to the inaccuracy in numerical calculations and the specific algorithm of the NACCESS, it is possible to obtain a non-zero, albeit small, SASA value for two molecules that are close but not touching each other. To counter this problem, a threshold value of 30  $\text{\AA}^2$  was used to determine if two molecules have atomic clashes. The dsRBD1 and

the dsRBD2-Core were deemed to have no atomic clash if the calculated buried SASA was below this threshold value.

## RESULTS

### The PBS-segment RNA adopts a three-way junction structure

To investigate the PBS-segment structure, an RNA of nucleotides (nts) 125–223 was synthesized with two terminal G-C pairs for efficient transcription (Figure 1A). NMR assignments were facilitated by referencing the PBS-segment RNA spectra with that of control RNA fragments, including TLE stem loop (nt 135–177) and PAS stem (nt 125–223 with a GAGA tetraloop connecting residue U131 and G217) (Supplementary Figure S1D). Attempts to make a control RNA fragment for the tRNA annealing arm were not successful. We tried various 5'- and 3'- end residues but none of the RNA fragments gave rise to similar NMR cross peak patterns in the  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum as that of the PBS-segment RNA. Imino proton resonances of the PBS-segment were very broad (Supplementary Figure S1B, C). The base pairs in the PAS stem and TLE stem control RNAs were confirmed by referencing the imino proton spectrum with that of the control RNA fragment spectra (Supplementary Figure S1E, F). Additional imino proton resonances were observed and assigned to G190, G195, U200 and G206, suggesting base pairings in the tRNA annealing arm (Supplementary Figure S1C). Assignments of the non-exchangeable protons of residues in these control RNA samples were used as references for the assignment of the entire PBS-segment. To overcome the spectral overlap and enhance spectral resolution, nucleotide-specific  $^2\text{H}$ -labeled samples (Materials and Methods) were prepared for NMR data collection and assignments of non-exchangeable protons. This strategy enabled us to assign residues in and near the tRNA annealing arm that are not covered by the control RNA fragments. In addition to imino proton resonances, we also relied on the sequential and long-range NOEs that involve adenosine H2 to obtain secondary structure information within the PBS-segment (53) (Figure 1A). Comparison of the non-exchangeable proton assignments with database predictions also indicate the canonical structures in PAS and TLE region, and non-canonical structures in the junction and tRNA annealing arm are formed (Supplementary Figure S3) (54).

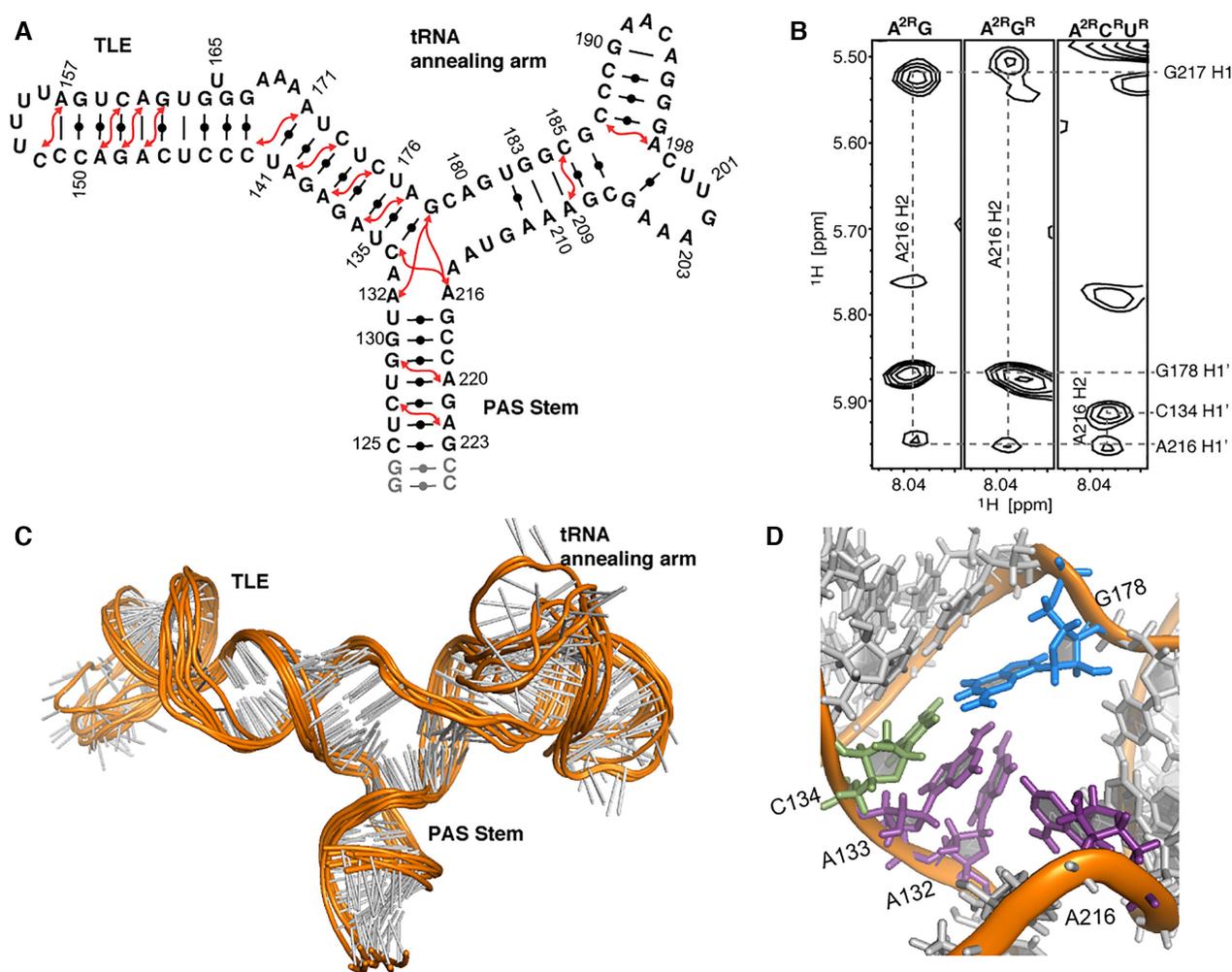
In agreement with previous structural models, our NMR data confirm the formation of PAS stem and the TLE stem in the PBS-segment (21–25). The assignment of residues in the tRNA annealing arm indicate that these residues are not completely flexible and unstructured, but instead forming some non-canonical structures. While no imino proton resonances were observed for residues U182–C185 and G208–G212, sequential stacking of these residues are indicated by the NOESY walk (Supplementary Figure S2). Long-distance NOEs were detected and assigned to the residues in the tRNA annealing arm (nt 179–214), including A209.H2–C185.H1' (Supplementary Figure S2), and A198.H2–187.H1', demonstrating base pairing in the tRNA annealing arm. However, some cross-peaks, such as G183.H8–U182.H1' and A209.H2–C185.H1', are very

weak (Supplementary Figure S2), suggesting structural dynamics of these residues in the tRNA annealing arm. They are likely in rapid exchange between a stacked/paired position and unstructured positions. The junction that converges PAS stem, TLE hairpin and tRNA annealing arm is adenosine-rich (A132, A133 and A216), and NOE cross peaks between A216-H2 and the H1's of G178, G217 and C134 were detected in spectra with different  $^2\text{H}$  labeling strategies (Figure 1B). In summary, our NMR data reveal that the PBS-segment RNA adopts a three-way structure with an adenosine-rich junction (Figure 1C and D).

SAXS data were collected to obtain restraints of the overall shape of the PBS-segment RNA. Gel-filtration analysis of the PBS-segment RNA showed a minor peak before the main elution peak, indicating that a small portion of the sample may be aggregated (Figure 2A). The aggregated RNAs are not likely to give rise to measurable NMR signals due to their large molecular sizes and low population (<10%), but could introduce large errors when measuring the RNA dimensions by SAXS. To obtain better homogeneity of the RNA sample, in-line SAXS data were collected by eluting the PBS-segment RNA on a Superdex-200-Increase SEC column (GE), and SAXS data collected for samples eluted in the main peak were used for structural analysis. The averaged scattering data, Kratky Plot, Guinier plot and the distribution of the interatomic distance  $P(r)$  are shown in Figure 2B–E, respectively. *Ab initio* models generated from the SAXS data recapitulated a three-way junction structure highly consistent with our NMR model (Figure 2F). Efforts were made to truncate one of the arms in the RNA sample to help unambiguously identify the TLE and the tRNA annealing arm in the SAXS *ab initio* model. However, NMR analysis of the truncated PBS-segment revealed that the secondary structure was altered due to the truncation. SAXS profiles of the NMR-derived structures were back calculated and compared with the experimental SAXS data. The structures with lowest energy and  $\chi^2$  values were selected for further energy optimization by MD simulation (Figure 2G). The statistics of calculated structures are summarized in Table 1 and Supplementary Table S4.

### The base pairs within the three-way junction structure are phylogenetically conserved

To determine if the three-way junction structure of the PBS-segment is phylogenetic conserved, analysis of HIV-1 sequences deposited in the HIV database (<http://www.hiv.lanl.gov/>) was performed. A total of 1,908 PBS-segment containing sequences (C125–G223) were aligned using ClustalX2 (55). There are 421 sequences in subtypes A, G and some circulating recombinant forms containing a 23-nt insertion downstream of the PBS (56,57). For the purposes of analysis, this group of sequences were excluded as they may result in an alternative folding of the PBS-segment (30,58–60). Analysis of the remaining 1487 HIV-1 sequences from various subtypes revealed that majority of the nucleotides in the PBS-segment are highly conserved, underlining the importance of PBS-segment in virus replication. The nearly 100% conservation of the 18-nt tRNA annealing residues (U182–C199) are expected, given necessity to maintain complementary to the 3'-end

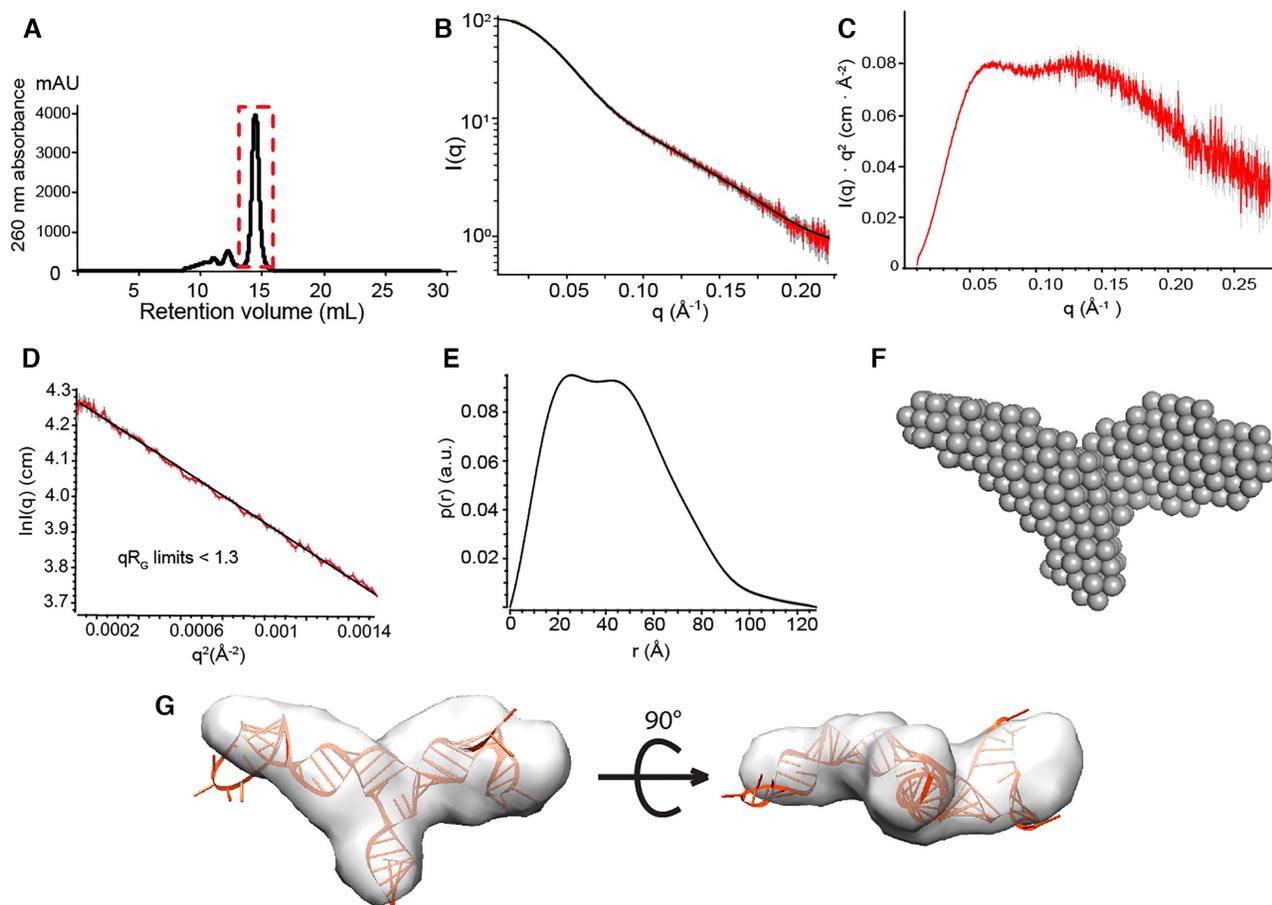


**Figure 1.** PBS-segment adopts a three-way junction structure. (A) Secondary structure of the PBS-segment supported by NMR assignment is shown. Red arrows denote the inter-helical NOEs involving adenosine H2 protons. (B) Portions of the 2D NOESY spectra with NOE connectivities supporting the three-way junction are shown. From left to right, spectra of  $A^{2R}G$ -,  $A^{2R}G^R$ - and  $A^{2R}C^R U^R$ -labeled PBS-segment. (See the nomenclature in the Materials and Methods). (C) NMR-derived secondary structure model of the PBS-segment. (D) A zoomed-in view of the adenosine-rich three-way junction in the PBS-segment.

of human tRNA<sup>Lys3</sup> primer for reverse transcription initiation. The PAS stem and the lower stem of TLE are structurally conserved, with base-pair substitution of U128–A220 by a U–G pair in the PAS stem, and conversion of G137–C175 to a U–A pair (Figure 3A). Maintaining the base pairing in PAS is essential for the dimeric conformation of 5'UTR and efficient reverse transcription (61–63), whereas the structural conservation of the lower stem of TLE remained unclear. We therefore hypothesized that the structural integrity of TLE lower stem is necessary to maintain the overall three-way junction structure of the PBS-segment.

To address this hypothesis, computational mutation scanning was performed to scan the nucleotides in the PBS-segment and predict their impact on the RNA secondary structure. We applied the Vfold2D RNA structure folding model (14, 15) to compute the structures for a given RNA sequence. For the WT PBS-segment RNA (HIV-1 NL4–3), the Vfold2D-predicted structure agrees with the experimen-

tal result. We performed a total of  $3 \times 10^3$  ( $n = 309$ ) single nucleotide mutations, to exhaustively predict structures for each mutant, and learned that different mutations can cause different structural changes. Critical nucleotides whose mutation cause disruption of the folding were identified. Three major types of structural changes were observed: (i) in the central three-way junction region, (ii) the TLE stem loop and (iii) the tRNA-annealing stem. Some of mutant structures maintain the WT three-way junction global fold while some form an alternative elongated stem-loop structure. We found that the disruption of the central three-way junction can result in folding changes in both the TLE stem loop and the tRNA annealing arm. The mutations that lead to the disruption of the central three-way junction are distributed mostly in the lower stem of TLE (U135–A140 and U172–G178). The result shows the TLE stem is stable as long as the lower stem is maintained. Mutations in the upper stem of TLE only cause local structural variations, but do not change the overall three-junction structure. Similarly, sin-



**Figure 2.** SAXS data of the PBS-segment support a three-way junction structure. (A) SEC profile of the PBS-segment RNA is shown. The SAXS diffraction data of red dash boxed region were used for structural analysis of the PBS-segment. (B–E) Averaging scattering profile (B), Kratky plot (C), Guinier plot (D) and the pair distance distribution function (E) for the scattering curve are shown. (F) An *ab initio* model of the PBS-segment generated with DAMMIF is shown. (G) Different views of the *ab initio* model superimposed with a representative structure of PBS-segment calculated by NMR-derived restraints and refined with the SAXS data.

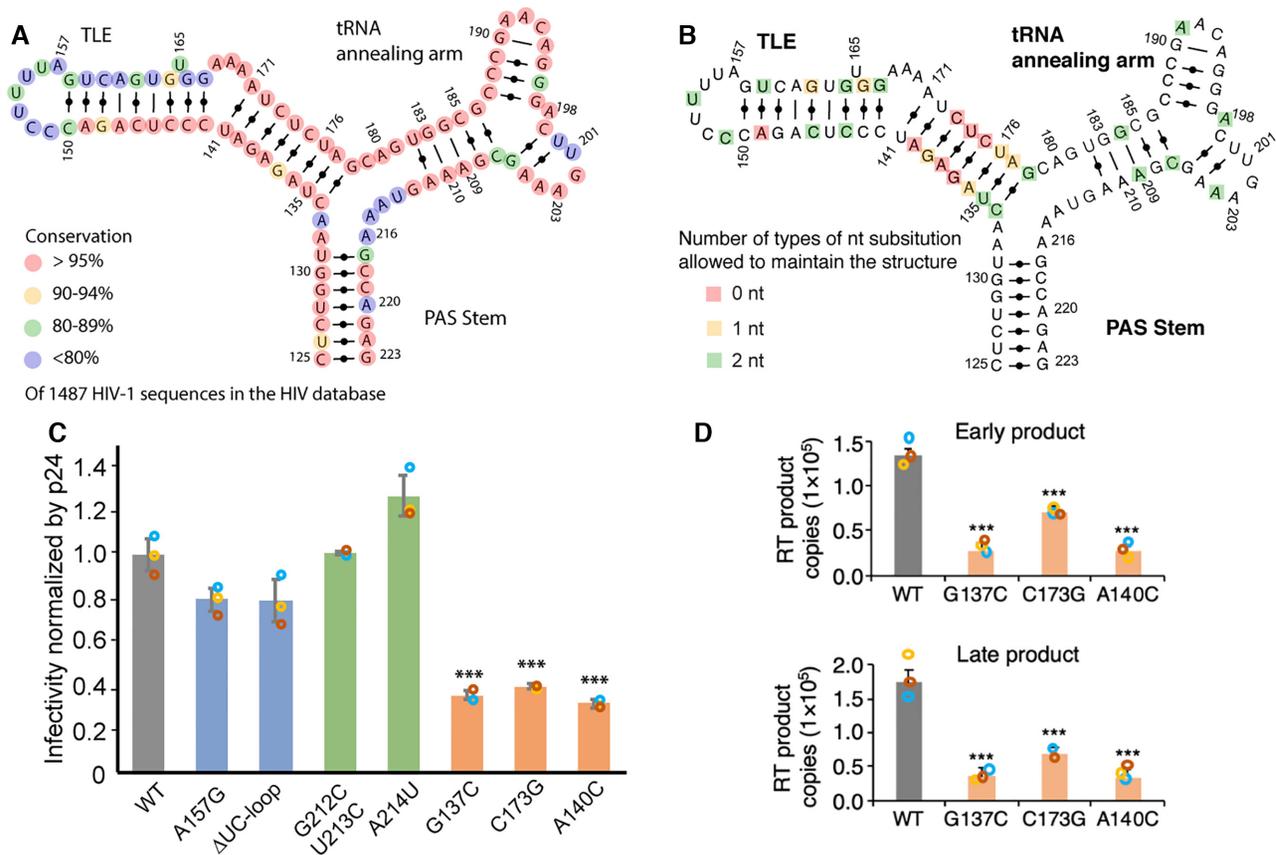
gle point mutations in PAS and the tRNA annealing arm lead to only local changes without disrupting the three-way structure of the PBS-segment. In summary, the predicted low-tolerance nucleotides reside in the lower stem of TLE, while majority of the nucleotides outside of this region can be mutated (Figure 3B). These results indicate that the lower stem of TLE is important to maintain the three-way junction of the PBS-segment structure. Other regions are sensitive to mutations, but single nucleotide mutations only cause local structural changes.

### The three-way junction structure is important for virion infectivity

Based on the computational mutation analysis, a series of mutations were designed to (i) disrupt the TLE loop, A157G and  $\Delta$ UC-loop (C151CUUUUA157 was substituted by CGAGAG), (ii) perturb tRNA annealing arm, G212C/U213C and A214U and (iii) disrupt the three-way junction structure, G137C, A140C and C173G. To avoid possible disruption of tRNA annealing and reverse transcription initiation, none of the mu-

tation sites selected are predicted to be involved in intermolecular interactions with tRNA<sup>Lys3</sup> based on previously reported tRNA<sup>Lys3</sup>:PBS complex models (7–9,57,62,64–66). Vfold secondary structure prediction (43) suggests that the mutant RNAs in groups 1 and 2 adopt the three-way junction structure, whereas group 3 mutants, G137C, A140C and C173G, mainly adopt a long stem loop structure without a three-way junction feature (Supplementary Figure S4). Consistently, SAXS data reveal that these three-way junction disrupting mutants have a longer paired distance ( $D_{max}$ ), and their shapes are different compared with the WT PBS-segment (Supplementary Figure S5).

We next examined the virion infectivity of these mutants in single-round infectivity assays using an NL4–3-derived reporter vector virus. The *env* gene of the reporter virus was replaced by an EGFP open reading frame and thus EGFP was produced when the cells become transduced (26). TZM-bl cells were incubated with vector virus samples containing 100 ng of Gag p24 and harvested 24 h later. The percentage of cells producing EGFP were quantified by flow cytometry to compare infectivity. Mutations altering the TLE



**Figure 3.** The three-way junction structure of the PBS-segment is conserved for infectivity. (A) Phylogenetic analysis of 1487 clinical sequences spanning the PBS-segment is shown. The conservation at each nucleotide position is color labeled. (B) Mutational scanning was performed to substitute each one nucleotide and predict the secondary structure of each mutant. The red, yellow, and green squares highlight the residues can be substituted by 0, 1 and 2 nucleotides in order to maintain the three-way junction structure. Unlabeled residues do not contribute to the three-way junction structure in a sequence specific manner. The low stem of TLE is shown in blue dashed box. (C) Mutations that disrupt the three-way structure of the PBS-segment reduced the virus infectivity. Infectivity was normalized by the p24 protein level. Gray bar, WT; blue bars, mutations in the TLE loop; green bars, mutations in the tRNA annealing arm; orange bar, mutations in the TLE stem. (D) Mutations that disrupt the three-way junction structure of the PBS-segment diminished HIV early reverse-transcription intermediates in target cells. Cyan, red, and yellow circles represent data from three independent experiments. Statistically significant differences were measured by Student's *t* test: \*\*\*  $P \leq 0.001$ .

loop (A157G and  $\Delta$ UC-loop) resulted in modest reduction of the single-round infectivity (Figure 3C), consistent with the previously reported role of TLE mimicking tRNA<sup>Lys3</sup> to facilitate tRNA<sup>Lys3</sup> loading for reverse transcription initiation (10). Mutations in the tRNA annealing arm showed none (G212C/U213C), or slightly positive impact (A214U) on viral infectivity (Figure 3C). These mutations are predicted to maintain the three-way junction structure (Supplementary Figure S4). On the other hand, an approximate 60% infectivity reduction was observed for the three-way junction disrupted mutants (G137C, A140C and C173G) (Figure 3C).

Since the PBS-segment is the reverse transcription initiation site for (-)cDNA synthesis, we then examined the reverse transcription activity of these mutants in lymphocytes. MT4 lymphocytes ( $5 \times 10^5$  cells in 0.5 ml of RPMI medium) were transduced with the pseudotyped virions in cell-free medium containing 200 ng of Gag p24 by spinoculation. Six hrs post-infection, cellular DNA from the transduced MT4 cells was isolated and analyzed by qPCR using primer

pairs to amplify the early and late products of reverse transcription (31). The results demonstrated that the amount of early RT product, (-)ssDNA, in MT4 cells was significantly diminished for the three-way junction disrupted mutants (G137C, A140C and C173G) ( $P < 0.001$ ) (Figure 3D). The synthesis of late RT products was similar, indicating the significant reduction in early RT activity was carried forward. These results indicate that the three-way junction disruption decreased virion infectivity by reducing the early reverse transcription activity.

To test whether the point mutations in PBS-segment affect tRNA<sup>Lys3</sup> annealing or RT loading, we examined tRNA<sup>Lys3</sup> annealing by EMSA, compared the *in vitro* primer extension efficiencies, and quantified the tRNA placement on viral RNA in virions. The WT and mutant RNAs were annealed with tRNA<sup>Lys3</sup> in the presence of NC. None of the mutations disrupted formation of an RNA:RNA duplex, and the migration rates of the mutant complexes on a polyacrylamide gel were similar to the WT (Supplementary Figure S6A). *In vitro* primer extension assays

were carried out by incubating the annealed RNA template with RT/dNTP, and monitoring the synthesis of (–)ssDNA. No difference in the primer extension efficiency between the WT and mutants was observed (Figure S6B–E). We then investigated tRNA<sup>Lys3</sup> placement on WT and mutant RNAs using RNA extracted from viral particles produced from transfected 293FT cells. The RNA was mixed with an RNase H activity reduced RT (SuperScript II, Invitrogen) and dNTPs. If the mutations in the PBS-segment do not affect tRNA<sup>Lys3</sup> placement on gRNA, the co-purified tRNA<sup>Lys3</sup> will serve as a primer to generate (–)cDNA when supplemented with RT (67). To normalize the (–)cDNA generated in WT and mutant viral RNAs, the input gRNAs were measured by RT-qPCR with primers/probe targeting the gag region (Supplementary Figure S6F). All of the mutant viral RNA successfully generated (–)cDNA at a similar level as the WT (Supplementary Figure S6G). The amount of proviral DNA in each sample was quantified by qPCR to ensure minimal DNA contamination (<1% RNA copy numbers). Collectively, our data show that the three-way junction disrupting mutations did not affect tRNA<sup>Lys3</sup> annealing and reverse transcription under *in vitro* conditions. Their negative impact on viral replication is likely to happen prior to tRNA annealing.

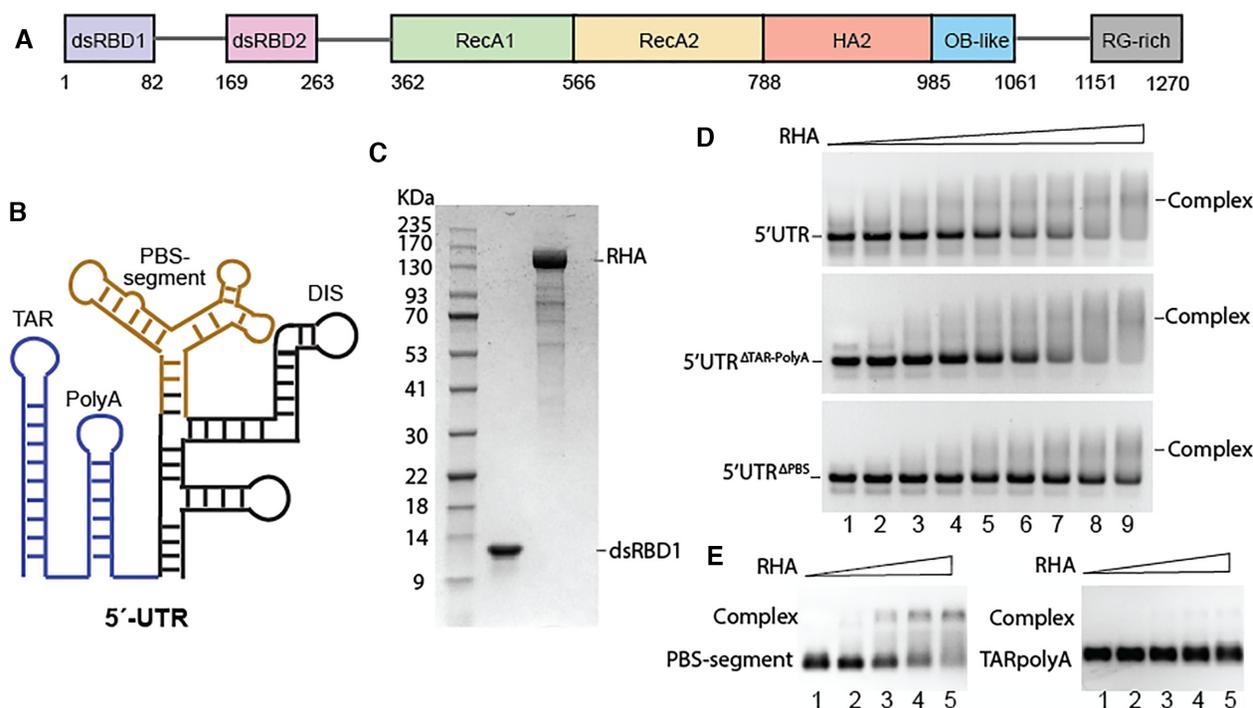
### RHA preferentially binds to the PBS-segment of the HIV-1 5'UTR

We have previously reported that HIV-1 recruits host RHA during virus assembly to serve as a processivity enhancement factor for RT (20). The recruitment is mediated by protein: RNA interactions that involve the PBS-segment of HIV-1 5'UTR (19). Therefore, the reduced infectivity and decreased reverse transcription products of the mutant viruses (Figure 3C and D) are likely due to the failure of RHA loading onto the PBS-segment. Our previous biophysical assays show that the N-terminal domain of RHA (dsRBD1 + dsRBD2) preferentially bind to the PBS-segment RNA (19). To confirm that PBS-segment is the binding target of RHA, recombinant full-length RHA protein (Figure 4A) expressed in a baculovirus system was purified (Figure 4C). The recombinant RHA was titrated to a series of HIV-1 RNA constructs, including the 5'UTR RNA (nt 1–344), 5'UTR with TAR and PolyA truncated (5'UTR<sup>ΔTAR-PolyA</sup>) and with PBS-segment deleted (5'UTR<sup>ΔPBS</sup>). These 5'UTR constructs have the 3'-half of the AUG hairpin truncated to favor RNA dimerization (Figure 4B) (53,63,68). As expected, the EMSA results show that the relative affinity of RHA for 5'UTR<sup>ΔTAR-PolyA</sup> is similar to 5'-UTR, but much weaker for 5'UTR<sup>ΔPBS</sup> (Figure 4D), as at RHA: RNA ratios of 7:1 and 8:1 (lanes 8 and 9) almost no free RNA bands were visible in the 5'UTR and 5'UTR<sup>ΔTAR-PolyA</sup> samples, but significant amount of free RNA bands were observed in the 5'UTR<sup>ΔPBS</sup> samples. When comparing the RHA binding affinity for smaller RNA fragments, TAR-PolyA (nt 1–104, 33 kDa) and PBS-segment (33 kDa), RHA preference towards the PBS-segment was observed (Figure 4E). Together, these data demonstrate that RHA preferentially binds to PBS-segment within the 5'-UTR.

### The binding interface between dsRBD1 and PBS-segment was mapped by NMR

To map the RNA-binding residues on dsRBD1, <sup>15</sup>N-labeled dsRBD1 was titrated with PBS-segment in a series of <sup>1</sup>H-<sup>15</sup>N-TROSY NMR experiments. Due to the severe precipitation issues when mixing PBS-segment RNA with dsRBD1 at NMR concentrations (200 μM), the titration was carried out in a high salt buffer and at high temperature (500 mM KCl and 1 mM MgCl<sub>2</sub>, 318K) to reduce non-specific electrostatic interactions between protein and RNA. As a result, the affinity between PBS-segment and dsRBD1 was weakened and thus fast exchange regime was observed (Figure 5A). Chemical shift perturbations (CSPs) were observed for a group of residues in α1 helix (region 1), the loop connecting β1–β2 (region 2) and α2 helix (region 3), suggesting these residues are within or close to the binding interface (Figure 5D and Supplementary Figure S7). These residues are consistent with the RNA binding interface reported in the crystal structure of RHA dsRBD1 in complex with a non-specific target (GC)<sub>10</sub> RNA (PDB: 3vyy) (69). To identify the RNA residues within or close to the dsRBD1 binding interface, 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra of the PBS-segment in complex with dsRBD1 was collected. In addition to fully protonated RNA, site-specific deuterated samples, including AG-PBS-segment and A<sup>2R</sup>C<sup>R</sup>U<sup>R</sup>-PBS-segment RNAs, were used to simplify the spectra for unambiguous assignment. G129–H1' was shifted upon dsRBD1 titration (Figure 5B), as we previously reported (19). Chemical shifts of residues in the lower TLE stem region also exhibited moderate shifts upon dsRBD1 binding, including A140, U141, C142 and U172 (Figure 5C). The protein and RNA residues identified within and close to the binding interface by NMR titrations are labeled in Figure 5D. Both the TLE and the PAS stems are involved in dsRBD1 binding, and the distance/dimension matches the RNA binding interface in dsRBD1 (Figure 5D).

With the guidance of the experimental data, a data-driven molecular docking study was exploited to model the structure of the PBS-segment: dsRBD1 complex. The crystal structure of RHA dsRBD1 (PDB: 3vyy) was used to dock the PBS-segment structure using an in-house program MDockPP (47). Protein residues that are in the RNA binding interface of the reported crystal structure (PDB: 3vyy) and showed large CSP (Δδ > 0.05 ppm), including F7, G13, K14, K16, M17, T18, Y21, K29, K54, K55 and Q58, were selected to generate distance restraints (Supplementary Figure S7). The RNA residues (G129, A140, U141, C142 and U172) were also defined to be close to dsRBD1. The distance between protein and RNA was set to 7 Å because this is the shortest distance restraint that resulted in docking models without steric clashes. The lowest energy model (Figure 6B) shows that region 1 (α1 helix) of dsRBD1 interacts with the minor groove of the bottom stem of TLE, region 2 is in close proximity to the PAS stem, and K54 and K55 in region 3 are near RNA residues in or adjacent to the three-way junction (A132, A133, and U176). The TLE: regions 1 and PAS: region 2 interactions are supported by both protein and RNA NMR data (Figure 5). The NMR signals for the junction residues A132, A133, and U176 were broadened and became undetectable upon dsRBD1 titra-



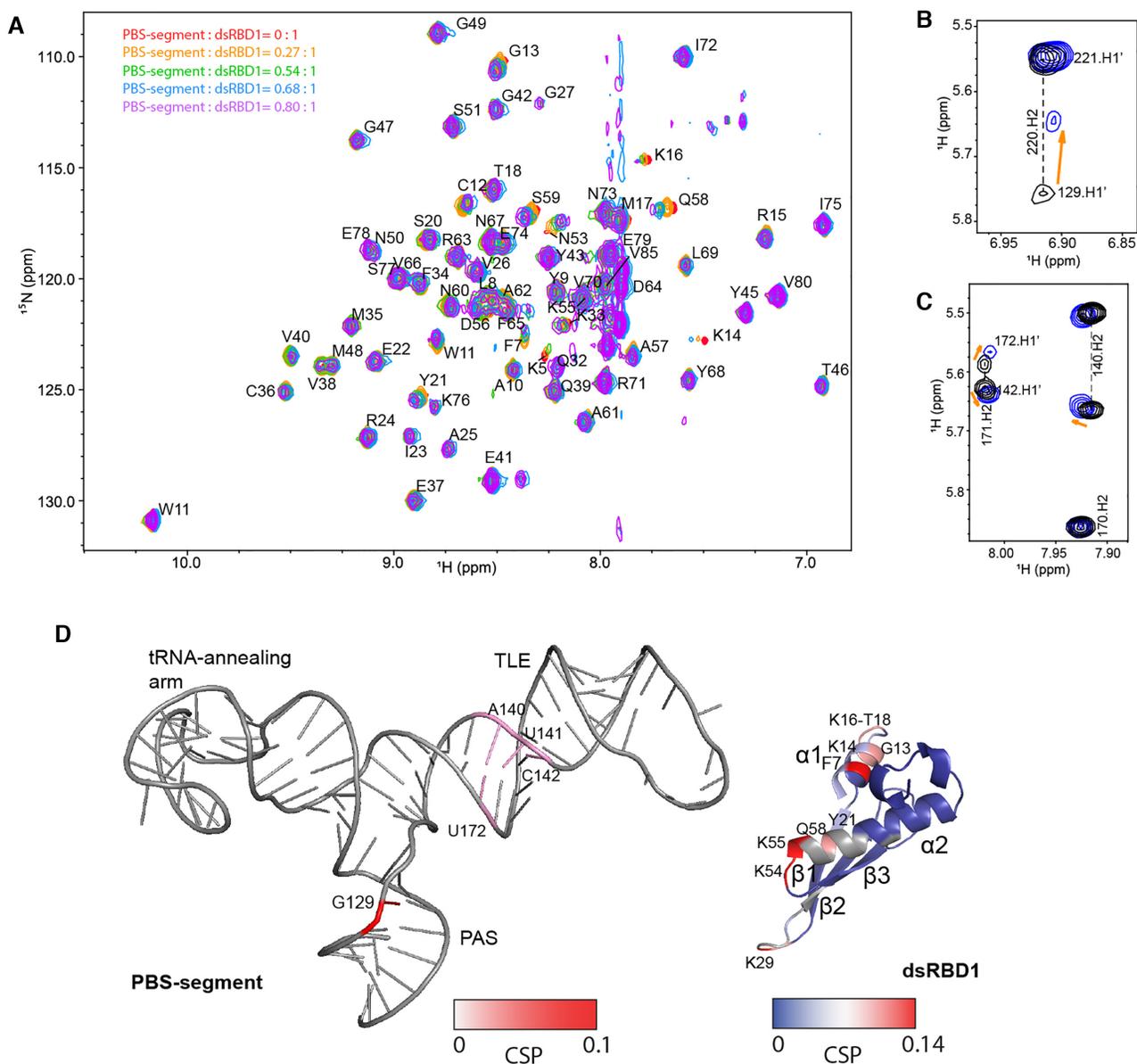
**Figure 4.** RHA preferentially binds to the PBS-segment region of the 5'UTR. (A) The domain organization of RHA is shown. (B) The secondary structural model of the 5'UTR is shown. TAR-PolyA, blue; PBS-segment, brown. (C) A representative SDS-PAGE shows the molecular mass markers (left), recombinant dsRBD1 purified from *E. coli* (middle), and recombinant RHA purified from insect cells (right). (D) EMSA of recombinant RHA binding to 5'UTR, 5'UTR<sup>ΔTAR-PolyA</sup> and 5'UTR<sup>ΔPBS</sup> (from top to bottom). The RNA samples were pre-dimerized, and the final mixture contained 0.5 μM of RNA and various concentrations of RHA (from lane 1 to 9, RHA concentrations were 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 μM). The unbound RNA and complex were separated in a 1.5% native agarose gel. The representative gel of three independent experiments is shown. (E) EMSA of recombinant RHA binding to PBS-segment and TAR-PolyA. The final mixtures contained 0.5 μM of RNA and 0, 1, 2, 3 and 4 μM of RHA (from lane 1 to 5). The representative gel of three independent experiments is shown.

tion in 2D NOESY spectra, which could be caused by protein: RNA interactions. Overall, the docking model is consistent with the NMR data, as majority of the residues near or within the protein: RNA interface exhibited CSP in the NMR titrations (Supplementary Figure S7).

The docking model suggests a shape recognition of the PBS-segment by dsRBD1. The tRNA-annealing arm introduces a ~108° bend of the PAS away from co-axial stacking on TLE stem, resulting in ~15 Å departure of the G129 from the straight A-form helical RNA structure model. In the crystal structure of dsRBD1: (GC)<sub>10</sub> complex, in order to have all the three regions to interact with RNA, region 1 is slightly distant from the RNA so that region 2 could be close. In the case of the PBS-segment, due to the angle between the TLE lower stem and PAS, dsRBD1 is closer to RNA with regions 1–3 all in contact with RNA. This leads to formation of additional hydrogen bonds between dsRBD1 and PBS-segment as compared to a straight (GC)<sub>10</sub> RNA shown in the crystal structure (Figure 6A–B).

To validate the shape-dependent recognition predicted by the docking model, ITC experiments were performed to compare the titration isotherms of dsRBD1 into PBS-segment and a hairpin-control RNA with the tRNA annealing arm truncated. The hairpin-control RNA contains base pairings in the PAS stem (nt 125–131 and 217–223) and TLE lower stem (nt 134–141 and 171–178), and is enclosed by a CG pair and a UUCG tetraloop (Figure 6C).

The isothermal data of titrating dsRBD1 into the hairpin-control RNA were fit using one-site non-linear least square regression. Data fitting shows that it can interact with multiple dsRBD1 molecules ( $N = 5.2 \pm 0.2$ ) at  $K_d = 18.5 \pm 1.1 \mu\text{M}$ , indicative of non-specific electrostatic interactions (Figure 6D). The ITC data of dsRBD1: PBS-segment titration were fit using two-sites binding mode, which shows that PBS-segment contains one high-affinity dsRBD1 binding site ( $K_{d1} = 1.17 \pm 0.23 \mu\text{M}$ ) and a second set of weak binding sites ( $K_{d2} = 45 \pm 8.1 \mu\text{M}$ ) (Figure 6E). The affinity of the second set of binding is comparable to the dsRBD1:hairpin-control interaction, suggesting that after the high-affinity binding site on PBS-segment is occupied, dsRBD1 starts to non-specifically interact with the RNA. It also explains that the <sup>1</sup>H-<sup>15</sup>N TROSY NMR titrations were collected in high salt buffer and at high temperature to eliminate non-specific dsRBD1: RNA interactions (Figure 5A). Assuming no cooperativity among these dsRBD1 non-specific binding sites, the averaged thermodynamic contribution of each dsRBD1 binding site on the hairpin-control were calculated (Supplementary Table S5). Both the PBS-segment: dsRBD1 and hairpin-control: dsRBD1 interactions are enthalpy ( $\Delta H$ )-driven, but the enthalpy change associated with PBS-segment binding is significantly greater than hairpin-control binding ( $-7845 \pm 123 \text{ cal/mol per site}$  for the high-affinity site in the PBS-segment versus  $-1762 \pm 37 \text{ cal/mol per site}$  in the hairpin-control) (Supplementary



**Figure 5.** The dsRBD1:PBS-segment binding interface was mapped by NMR. (A)  $^1\text{H}$ - $^{15}\text{N}$  TROSY spectra were collected for dsRBD1 titrated with various molar equivalents of PBS-segment RNA. The RNA: protein ratios are listed on the top right corner. (B, C) Portions of the 2D NOESY spectra of the PBS-segment RNA (black) and in complex with one molar equivalent of dsRBD1 protein (blue) are shown. (D) Chemical shift perturbations are mapped on the structure of PBS-segment and dsRBD1.

Table S5). For a low affinity titration system, even though the binding event is near but not completely saturated at the end of titrations,  $K_d$  can be accurately determined but  $\Delta H$  can be underestimated (70). The second set of binding sites in PBS-segment is not saturated because the heat in the final injections are not near zero, and thus the  $\Delta H$  for these binding sites are less negative than the dsRBD1:hairpin-control interaction (Supplementary Table S5). In summary, the ITC data are in agreement with the docking model that more hydrogen bonds are formed when dsRBD1 binds to a specific target RNA (PBS-segment) than a non-specific target ((GC)<sub>10</sub> or hairpin-control) (Figure 6A, B). Recognition of PBS-segment by dsRBD1 is RNA shape-dependent.

### Molecular docking of dsRBD2-Core onto PBS-segment suggests possible specific RHA: RNA interactions

Previous studies suggested that dsRBD2 is indispensable in directing RHA binding to target RNAs (18,69,71). Crystal structure of dsRBD2 in complex with a (GC)<sub>10</sub> RNA reveals that it binds to RNA similarly as dsRBD1, with equivalent  $\alpha 1$  helix (region 1) and loop connecting  $\beta 1$ - $\beta 2$  (region 2) binding to two successive minor grooves and  $\alpha 2$  helix (region 3) binding to the major groove in between (Supplementary Figure S8A) (69). Crystal structure of a homologous protein MLE reveals that its dsRBD2 is tightly associated with the helicase core domain. While its region 1 and region 3 are surface exposed and could interact with an



RNA target, region 2 of dsRBD2 is surrounded by residues in RecA2 domain and is not surface exposed (Supplementary Figure S8B) (51). RHA and MLE share 51% sequence identity and 69% sequence similarity, so it is very likely that dsRBD2 of RHA is also in close contact with its RecA2 domain. Thus, *in vitro* study of dsRBD2 alone in RNA interactions may not represent real interactions with RNA in the context of full-length RHA, as the RecA2 domain of RHA may create spatial hindrance to prevent dsRBD2 region 2 residues from binding to a target RNA. To avoid this problem, we modeled the structure of RHA dsRBD2-Core (Figure 4A, residue 169–1150) using homology modeling based on the MLE structure (PDB: 5aor) (Supplementary Figure S8C), and docked it onto the PBS-segment RNA. The docked models were filtered using criteria that dsRBD2 regions 1 and 3 participate in RNA binding, and the dsRBD2-Core binding site does not overlap with the previously determined dsRBD1 binding site. The lowest energy model shows that the  $\alpha 1$  helix (region 1) and K236 (region 3) of dsRBD2 interact with the three-way junction of the PBS-segment, and RecA2 residues in the Core domain make contacts with the tRNA annealing arm (Supplementary Figure S8D, E). Region 2 of dsRBD2 is not involved in PBS-segment RNA binding, as it is surrounded by residues in RecA2. These results are consistent with a previous report that substitution of a region 2 residue in dsRBD2 did not affect RHA-mediated RISC assembly (69). In summary, the docking model suggests that the dsRBD2-Core could also contribute to the recognition of the three-way junction structure of the PBS-segment.

## DISCUSSION

### The three-way junction structure of the PBS-segment is important for infectivity

Solving the structure of the PBS-segment has been challenging as the tRNA annealing arm is highly metastable thus making crystallization difficult and traditional characterization by NMR impractical. Metastable residues in the tRNA annealing arm were proposed to be unstructured by chemical probing, enzymatic probing, SAXS and computational modeling (21–25). In our study, we established the PBS-segment sequence boundaries encompassing the three-way junction in the HIV-1 RNA encapsidation signal delineated by prior NMR studies, in which residues C125–G130 form base pairs with C218–G223 (53,63). The structure reported for the PBS-segment here should resemble its parent structure in the dimeric 5'UTR prior to reverse transcription initiation that directs viral RNA genome packaging (63,72). By combining site-specific deuteration NMR strategies with SAXS methodology, EMSA with full length RHA and biological validation experiments, we were able to determine one conformation of the PBS-segment important for viral infectivity.

The identified three-way junction structure is reinforced by phylogenetic analysis of HIV sequences in patient samples, which document conserved pairings at the bottom of TLE (Figure 3A). In line with the phylogenetic findings, we show that single point mutations in this region could completely disrupt the three-way junction (Figure 3B). The

mutants adopt relatively extended structures (Supplementary Figures S4–5) that failed to efficiently produce reverse transcription products in infected cells (Figure 3C, D). Our structural study of the PBS-segment implements the current knowledge of the structure of the 5'UTR of the gRNA and provide structural basis for the recruitment of beneficial host factors during genome packaging to bolster virion infectivity.

### The tRNA annealing arm is important for RHA: PBS-segment interaction

Annealing of tRNA<sup>Lys3</sup> onto PBS requires a chaperone, which can be the NC domain of Gag polyprotein or the processed mature NC protein (3,4,73–75). *In vitro* primer extension on RNA extracted from protease-deficient virions exhibited lower nucleotide incorporation rate than WT, but could be rescued by incubating the RNA with mature NC protein, suggesting Gag facilitates a partial annealing of tRNA onto PBS and NC promotes a complete annealing (4). The two-step annealing model is supported by the *in vivo* SHAPE analysis of the HIV-1 viral RNA in WT and protease-deficient virions (5). The matrix (MA) domain was reported to modulate the chaperone activity of Gag, which is negatively regulated by MA: RNA interactions and can be stimulated by MA interacting with inositol phosphate on the plasma membrane (76). Hence tRNA annealing promoted by Gag is likely to happen at the plasma membrane where virus assembly takes place. RHA is predominantly localized in nucleus and shuttles between nucleus and cytoplasm (77). The interactions between RHA and PBS-segment may occur as early as in nucleus. It is possible that RHA is brought by the 5'UTR of gRNA to the virus assembly site, where Gag interacts with inositol phosphate and promotes tRNA annealing. The spatial and temporal control of these molecular interactions need to be investigated in future studies.

The PBS-segment structure is composed of an adenosine-rich three-way junction at the confluence of three subdomains: PAS, TLE and the tRNA annealing arm (Figure 1). It is worth mentioning that a similar SAXS envelope was reported for an almost identical PBS-segment RNA from the NL4-3 isolate (nt 125–223 with three non-native G–C pairs at the 5'/3' termini) (24), but the residues in the tRNA annealing arm were single stranded and unstructured in the reported SAXS-derived model. In our studies, base pairings and base stackings in the tRNA annealing arm were apparent in the NMR data. Sequential NOESY walk of residues in the tRNA annealing arm was observed in the NMR spectra, but the inter-residue NOE peak intensities were weak (Supplementary Figure S2), suggesting that residues U182–C185 and G208–G212 are exchanging between flexible/unpaired and base stacked structures. The structural flexibility is believed to be favored for reverse transcription as the secondary structure must be unwound for tRNA annealing. The tRNA annealing arm in the *ab initio* model generated from SAXS data appeals wider than a typical A-form dsRNA (Figure 2G). Both the NMR and SAXS data are in agreement with an intrinsically dynamic domain in which several residues in tRNA annealing arm remain accessible for chaperone-mediated tRNA annealing.

The importance of the TLE stem loop and the PAS stem facilitating tRNA loading and activating reverse transcription have been previously addressed (7,8,10,62). Unlike PAS and TLE, residues in the tRNA annealing arm are less phylogenetically conserved (Figure 3A). Many mutations in the tRNA annealing arm are not predicted to disrupt the three-way junction structure of the PBS-segment (Figure 3B). Our study suggest that the tRNA annealing arm contributes to the overall shape of the PBS-segment RNA recognized by RHA. The bent TLE-PAS stem is preferred by dsRBD1 over the hairpin-control, a straight dsRNA helical structure which can be considered as TLE stacking on PAS (Figure 6C–E). TAR-PolyA is another example of straight helical structure as PolyA co-axially stacks on TAR (24,60). EMSA data show the affinity of dsRBD1 for TAR-PolyA is much weaker than PBS-segment (Figure 4E). Thus, even though the tRNA annealing arm does not directly interact with dsRBD1, it participates in the shape-dependent dsRBD1 recognition by preventing TLE from co-axially stacking on PAS. The affinity differences indicate two RNA-binding modes of RHA: the non-specific interactions with dsRNA when RHA functions as a helicase, and the specific interactions with PBS-segment recruited by HIV-1 as a beneficial host factor.

#### RHA dsRBD1 shape-dependent recognition contributes to loading RHA to the PBS-segment

DsRBDs exist in many proteins and influence various steps of RNA metabolism. Some of the proteins have more than one dsRBD, and the contribution of each dsRBD to target RNA selection is usually not equal. MLE helicase facilitates the incorporation of the roX lncRNA into the dosage compensation complex via its dsRBDs interacting with the roX lncRNA. Studies revealed that dsRBD2 has 10-fold greater affinity than dsRBD1, and dsRBD2 plays the major role in recognition of the R2H1 and SL7 stem loops of roX lncRNA (78,79). Both the R2H1 and SL7 stem loops contain Watson-Crick and G:U wobble base pairs that form straight A-form helices. The crystal structure of dsRBD1+dsRBD2 in complex with R2H1 shows that MLE dsRBD1 binds to RNA (PDB: 5ztm) in a similar way as RHA dsRBD1 binds to a (GC)<sub>10</sub> dsRNA (PDB: 3vyy). These dsRBD1: dsRNA crystal structures indicate that dsRBD1 binds to straight A-form dsRNA helices without sequence specificity.

Efforts have been made to investigate the sequence and shape recognition of dsRBDs in directing the proteins to their target RNAs. In general, dsRBDs use conserved residues in region 1, 2 and 3 to interact with two consecutive minor grooves in dsRNAs (78,80–85). Sequence specificity is achieved by some species-specific amino acids interacting with non-canonical structures in RNA. For example, dsRBD of Rnt1p, a member of RNase III family of dsRNA endonucleases, binds to the sn47 precursor RNA to direct the endonuclease to its RNA substrates by recognition of the AGNN tetraloop of the target RNAs. NMR studies show that in addition to using regions 1–3 interacting with two consecutive minor grooves in sn47 RNA, S376 and R372 in  $\alpha 1$  form hydrogen bonds with the 2'OH of the two non-conserved 3'-nucleotides of the

tetraloop, and M368 is stacked on the ribose of the adenosine in the tetraloop (86). The interactions between the dsRBM (double-stranded RNA binding motifs, an alternative name of dsRBD) of adenosine deaminases that act on RNA (ADARs) and its target GluR-2 RNA provide another example of specific recognition. ADARs are a group of enzymes that selectively deaminate adenosine. Solution studies of dsRBMs of ADAR2 in complex with GluR-2 RNA show that the dsRBM1 of ADAR2 binds to the top stem of the GluR-2 RNA with specific contacts between  $\alpha 1$  of dsRBM1 and RNA tetraloop. The interactions include M84 making a sequence-specific contact with the A–U pair that is adjacent to the UCCG tetraloop, and E88 forming a hydrogen bond with the amino group of the first C in the tetraloop (87). UCGG belongs to the UNGC tetraloop family and adopts a specific tetraloop fold with extra thermostability (88). Therefore the recognition of GluR-2 by ADAR2 dsRBM1 is sequence-specific and shape-dependent.

Here we show that dsRBD1 preferentially binds to the PBS-segment RNA by recognizing the TLE and PAS stems which are not co-axial stacked in the three-way junction structure. ITC data demonstrate that the dsRBD1: PBS-segment interaction is of higher affinity compared to binding to a straight A-form helical RNA structure (hairpin-control, Figure 6C–E). K5, N6, Y9, K54, K55, K29 and N30 are involved in H-bonds with RNA in both the crystal structure of dsRBD1: (GC)<sub>10</sub> complex and our dsRBD1: PBS-segment model. In addition, K16, T18, Y21, K29, N23 and E41 participate in H-bonds with PBS-segment, which explains the  $K_d$  and enthalpy differences between dsRBD1 binding to a straight helical RNA and PBS-segment. Some of these residues are unique in the dsRBD family (89), and therefore they are likely to contribute to the shape-dependent recognition by dsRBD1.

In summary, we show that the structure of the PBS-segment prior to tRNA annealing folds into a three-way junction structure that is important for viral infectivity. The biophysical studies of the interactions between RHA dsRBD1 and the PBS-segment provides a plausible explanation for the recruitment of a host factor beneficial for the late stage virus replication. Although we cannot exclude the possibility of other yet-to-be discovered molecular interactions involving PBS-segment, our solution structure of the PBS-segment provides the structural basis for further investigation of the biomolecular interactions that occur on the scaffold of the PBS-segment.

#### DATA AVAILABILITY

Atomic coordinates for the PBS-segment RNA structure have been deposited into the RSCB Protein Data Bank under the accession number 7LVA. Chemical shifts have been deposited to the BMRB under accession number 30868. SAXS data for PBS-segment have been deposited to the SASBDB under accession number SASDJU7.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

**FUNDING**

National Institutes of Health [R01AI150460 to X.H., K.B.-L., R01GM109980, R35GM136409 to X.Z., R01GM063732, R01GM117059, R35GM134919 to S.J.C.]. Funding for open access charge: National Institutes of Health Grant [R01AI150460].

*Conflict of interest statement.* None declared.

**REFERENCE**

- Coffin, J.M., Hughes, S.H. and Varmus, H. (1997) In: *Retroviruses*. Cold Spring Harbor Laboratory Press, NY.
- De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M.C., Roques, B. and Darlix, J.L. (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6472–6476.
- Hargittai, M.R., Gorelick, R.J., Rouzina, I. and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site. *J. Mol. Biol.*, **337**, 951–968.
- Cen, S., Khorchid, A., Gabor, J., Rong, L., Wainberg, M.A. and Kleiman, L. (2000) Roles of Pr55(gag) and NCp7 in tRNA(3)(Lys) genomic placement and the initiation step of reverse transcription in human immunodeficiency virus type 1. *J. Virol.*, **74**, 10796–10800.
- Seif, E., Niu, M. and Kleiman, L. (2015) In vitro SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology*, **12**, 40.
- Beerens, N., Groot, F. and Berkhout, B. (2001) Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. *J. Biol. Chem.*, **276**, 31247–31256.
- Beerens, N. and Berkhout, B. (2002) The tRNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and processive elongation of reverse transcription. *J. Virol.*, **76**, 2329–2339.
- Beerens, N. and Berkhout, B. (2002) Switching the in vitro tRNA usage of HIV-1 by simultaneous adaptation of the PBS and PAS. *RNA*, **8**, 357–369.
- Beerens, N., Jepsen, M.D., Nechyporuk-Zloy, V., Kruger, A.C., Darlix, J.L., Kjems, J. and Birkefeld, V. (2013) Role of the primer activation signal in tRNA annealing onto the HIV-1 genome studied by single-molecule FRET microscopy. *RNA*, **19**, 517–526.
- Jones, C.P., Saadatmand, J., Kleiman, L. and Musier-Forsyth, K. (2013) Molecular mimicry of human tRNA<sup>Lys</sup> anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. *RNA*, **19**, 219–229.
- Cen, S., Khorchid, A., Javanbakht, H., Gabor, J., Stello, T., Shiba, K., Musier-Forsyth, K. and Kleiman, L. (2001) Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. *J. Virol.*, **75**, 5043–5048.
- Javanbakht, H., Cen, S., Musier-Forsyth, K. and Kleiman, L. (2002) Correlation between tRNA<sup>Lys</sup> aminoacylation and its incorporation into HIV-1. *J. Biol. Chem.*, **277**, 17389–17396.
- Guo, F., Cen, S., Niu, M., Javanbakht, H. and Kleiman, L. (2003) Specific inhibition of the synthesis of human lysyl-tRNA synthetase results in decreases in tRNA(Lys) incorporation, tRNA(3)(Lys) annealing to viral RNA, and viral infectivity in human immunodeficiency virus type 1. *J. Virol.*, **77**, 9817–9822.
- Cen, S., Javanbakht, H., Niu, M. and Kleiman, L. (2004) Ability of wild-type and mutant lysyl-tRNA synthetase to facilitate tRNA(Lys) incorporation into human immunodeficiency virus type 1. *J. Virol.*, **78**, 1595–1601.
- Goldschmidt, V., Ehresmann, C., Ehresmann, B. and Marquet, R. (2003) Does the HIV-1 primer activation signal interact with tRNA<sup>3</sup>(Lys) during the initiation of reverse transcription? *Nucleic Acids Res.*, **31**, 850–859.
- Roy, B.B., Hu, J., Guo, X., Russell, R.S., Guo, F., Kleiman, L. and Liang, C. (2006) Association of RNA helicase A with human immunodeficiency virus type 1 particles. *J. Biol. Chem.*, **281**, 12625–12635.
- Bolinger, C., Sharma, A., Singh, D., Yu, L. and Boris-Lawrie, K. (2010) RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res.*, **38**, 1686–1696.
- Xing, L., Niu, M. and Kleiman, L. (2012) In vitro and In vivo analysis of the interaction between RNA helicase A and HIV-1 RNA. *J. Virol.*, **86**, 13272–13280.
- Boeras, I., Song, Z., Moran, A., Franklin, J., Brown, W.C., Johnson, M., Boris-Lawrie, K. and Heng, X. (2016) DHX9/RHA binding to the PBS-segment of the genomic RNA during HIV-1 assembly Bolsters virion infectivity. *J. Mol. Biol.*, **428**, 2418–2429.
- Brady, S., Singh, G., Bolinger, C., Song, Z., Boeras, I., Weng, K., Trent, B., Brown, W.C., Singh, K., Boris-Lawrie, K. et al. (2019) Virion-associated, host-derived DHX9/RNA helicase A enhances the processivity of HIV-1 reverse transcriptase on genomic RNA. *J. Biol. Chem.*, **294**, 11473–11485.
- Damgaard, C.K., Andersen, E.S., Knudsen, B., Gorodkin, J. and Kjems, J. (2004) RNA interactions in the 5' region of the HIV-1 genome. *J. Mol. Biol.*, **336**, 369–379.
- Abbink, T.E. and Berkhout, B. (2003) A novel long distance base-pairing interaction in human immunodeficiency virus type 1 RNA occludes the Gag start codon. *J. Biol. Chem.*, **278**, 11601–11611.
- Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.*, **6**, e96.
- Jones, C.P., Cantara, W.A., Olson, E.D. and Musier-Forsyth, K. (2014) Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3395–3400.
- Kenyon, J.C., Prestwood, L.J., Le Grice, S.F. and Lever, A.M. (2013) In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.*, **41**, e174.
- Lange, M.J. and Burke, D.H. (2014) Screening inhibitory potential of anti-HIV RT RNA aptamers. *Methods Mol. Biol.*, **1103**, 11–29.
- Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.
- Lee, B.M., De Guzman, R.N., Turner, B.G., Tjandra, N. and Summers, M.F. (1998) Dynamical behavior of the HIV-1 nucleocapsid protein. *J. Mol. Biol.*, **279**, 633–649.
- Alam, K.K., Chang, J.L., Lange, M.J., Nguyen, P.D.M., Sawyer, A.W. and Burke, D.H. (2018) Poly-target selection identifies broad-spectrum RNA aptamers. *Mol. Ther. Nucleic Acids*, **13**, 605–619.
- Gremminger, T., Song, Z., Ji, J., Foster, A., Weng, K. and Heng, X. (2020) Extended interactions between HIV-1 viral RNA and tRNA(Lys3) are important to maintain viral RNA Integrity. *Int. J. Mol. Sci.*, **22**, 58.
- Mbisa, J.L., Delviks-Frankenberry, K.A., Thomas, J.A., Gorelick, R.J. and Pathak, V.K. (2009) Real-time PCR analysis of HIV-1 replication post-entry events. *Methods Mol. Biol.*, **485**, 55–72.
- Rinaldi, A.J., Suddala, K.C. and Walter, N.G. (2015) Native purification and labeling of RNA for single molecule fluorescence studies. *Methods Mol. Biol.*, **1240**, 63–95.
- 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.
- Johnson, B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.*, **278**, 313–352.
- Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. et al. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
- Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.*, **25**, 495–503.

37. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.*, **42**, 342–346.
38. Guntert, P. and Buchner, L. (2015) Combined automated NOE assignment and structure calculation with CYANA. *J. Biomol. NMR*, **62**, 453–471.
39. Tolbert, B.S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D.A. and Summers, M.F. (2010) Major groove width variations in RNA structures determined by NMR and impact of <sup>13</sup>C residual chemical shift anisotropy and <sup>1</sup>H-<sup>13</sup>C residual dipolar coupling on refinement. *J. Biomol. NMR*, **47**, 205–219.
40. Svergun, D., Barberato, C. and Koch, M.H.J. (1995) CRYSOLO - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.*, **28**, 768–773.
41. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
42. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Marius Clore, G. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
43. Xu, X., Zhao, P. and Chen, S.J. (2014) Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One*, **9**, e107504.
44. Cao, S., Giedroc, D.P. and Chen, S.J. (2010) Predicting loop-helix tertiary structural contacts in RNA pseudoknots. *RNA*, **16**, 538–552.
45. Cao, S. and Chen, S.J. (2011) Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. *RNA*, **17**, 2130–2143.
46. Cao, S. and Chen, S.J. (2012) Predicting kissing interactions in microRNA-target complex and assessment of microRNA activity. *Nucleic Acids Res.*, **40**, 4681–4690.
47. Xu, X., Qiu, L., Yan, C., Ma, Z., Grinter, S.Z. and Zou, X. (2017) Performance of MDockPP in CAPRI rounds 28–29 and 31–35 including the prediction of water-mediated interactions. *Proteins*, **85**, 424–434.
48. Huang, S.Y. and Zou, X. (2014) A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.*, **42**, e55.
49. 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.
50. Webb, B. and Sali, A. (2016) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci*, **86**, 2.9.1–2.9.37.
51. Prabu, J.R., Muller, M., Thomae, A.W., Schussler, S., Bonneau, F., Becker, P.B. and Conti, E. (2015) Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA-ATP coupling. *Mol. Cell*, **60**, 487–499.
52. Hubbard, S. and Thornton, J. (1993). In: *Department of Biochemistry Molecular Biology*. University College London.
53. Keane, S.C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hoscic, A., Florwick, A., Santos, J. et al. (2015) RNA structure. Structure of the HIV-1 RNA packaging signal. *Science*, **348**, 917–921.
54. Barton, S., Heng, X., Johnson, B.A. and Summers, M.F. (2013) Database proton NMR chemical shifts for RNA signal assignment and validation. *J. Biomol. NMR*, **55**, 33–46.
55. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
56. Gao, F., Robertson, D.L., Carruthers, C.D., Li, Y., Bailes, E., Kostrikis, L.G., Salminen, M.O., Bibollet-Ruche, F., Peeters, M., Ho, D.D. et al. (1998) An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). *J. Virol.*, **72**, 10234–10241.
57. Goldschmidt, V., Paillart, J.C., Rigour, M., Ehresmann, B., Aubertin, A.M., Ehresmann, C. and Marquet, R. (2004) Structural variability of the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.*, **279**, 35923–35931.
58. Isel, C., Marquet, R., Keith, G., Ehresmann, C. and Ehresmann, B. (1993) Modified nucleotides of tRNA(3Lys) modulate primer/template loop-loop interaction in the initiation complex of HIV-1 reverse transcription. *J. Biol. Chem.*, **268**, 25269–25272.
59. Coey, A., Larsen, K., Puglisi, J.D. and Viani Puglisi, E. (2016) Heterogeneous structures formed by conserved RNA sequences within the HIV reverse transcription initiation site. *RNA*, **22**, 1689–1698.
60. Brown, J.D., Kharytonchyk, S., Chaudry, I., Iyer, A.S., Carter, H., Becker, G., Desai, Y., Glang, L., Choi, S.H., Singh, K. et al. (2020) Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science*, **368**, 413–417.
61. Huthoff, H., Bugala, K., Barciszewski, J. and Berkhout, B. (2003) On the importance of the primer activation signal for initiation of tRNA(lys3)-primed reverse transcription of the HIV-1 RNA genome. *Nucleic Acids Res.*, **31**, 5186–5194.
62. Ooms, M., Cupac, D., Abbink, T.E., Huthoff, H. and Berkhout, B. (2007) The availability of the primer activation signal (PAS) affects the efficiency of HIV-1 reverse transcription initiation. *Nucleic Acids Res.*, **35**, 1649–1659.
63. Heng, X., Kharytonchyk, S., Garcia, E.L., Lu, K., Divakaruni, S.S., LaCotti, C., Edme, K., Telesnitsky, A. and Summers, M.F. (2012) Identification of a minimal region of the HIV-1 5'-leader required for RNA dimerization, NC binding, and packaging. *J. Mol. Biol.*, **417**, 224–239.
64. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. and Marquet, R. (1995) Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.*, **247**, 236–250.
65. Brule, F., Marquet, R., Rong, L., Wainberg, M.A., Roques, B.P., Le Grice, S.F., Ehresmann, B. and Ehresmann, C. (2002) Structural and functional properties of the HIV-1 RNA-tRNA(Lys)3 primer complex annealed by the nucleocapsid protein: comparison with the heat-annealed complex. *RNA*, **8**, 8–15.
66. Sleiman, D., Barraud, P., Brachet, F. and Tisne, C. (2013) The Interaction between tRNA(Lys) 3 and the primer activation signal deciphered by NMR spectroscopy. *PLoS One*, **8**, e64700.
67. Huang, Y., Wang, J., Shalom, A., Li, Z., Khorchid, A., Wainberg, M.A. and Kleiman, L. (1997) Primer tRNA<sub>3</sub>Lys on the viral genome exists in unextended and two-base extended forms within mature human immunodeficiency virus type 1. *J. Virol.*, **71**, 726–728.
68. Lu, K., Heng, X. and Summers, M.F. (2011) Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.*, **410**, 609–633.
69. Fu, Q. and Yuan, Y.A. (2013) Structural insights into RISC assembly facilitated by dsRNA-binding domains of human RNA helicase A (DHX9). *Nucleic Acids Res.*, **41**, 3457–3470.
70. Turnbull, W.B. and Daranas, A.H. (2003) On the value of  $\epsilon$ : can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.*, **125**, 14859–14866.
71. Ranji, A., Shkriabai, N., Kvaratskhelia, M., Musier-Forsyth, K. and Boris-Lawrie, K. (2011) Features of double-stranded RNA-binding domains of RNA helicase A are necessary for selective recognition and translation of complex mRNAs. *J. Biol. Chem.*, **286**, 5328–5337.
72. Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A. et al. (2011) NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science*, **334**, 242–245.
73. 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.
74. Roldan, A., Warren, O.U., Russell, R.S., Liang, C. and Wainberg, M.A. (2005) A HIV-1 minimal gag protein is superior to nucleocapsid at in vitro annealing and exhibits multimerization-induced inhibition of reverse transcription. *J. Biol. Chem.*, **280**, 17488–17496.
75. Guo, F., Saadatmand, J., Niu, M. and Kleiman, L. (2009) Roles of Gag and NCp7 in facilitating tRNA(Lys)(3) annealing to viral RNA in human immunodeficiency virus type 1. *J. Virol.*, **83**, 8099–8107.
76. Jones, C.P., Datta, S.A., Rein, A., Rouzina, I. and Musier-Forsyth, K. (2011) Matrix domain modulates HIV-1 Gag's nucleic acid chaperone activity via inositol phosphate binding. *J. Virol.*, **85**, 1594–1603.

77. Fujita, H., Ohshima, T., Oishi, T., Aratani, S., Fujii, R., Fukamizu, A. and Nakajima, T. (2005) Relevance of nuclear localization and functions of RNA helicase A. *Int. J. Mol. Med.*, **15**, 555–560.
78. Lv, M., Yao, Y., Li, F., Xu, L., Yang, L., Gong, Q., Xu, Y.Z., Shi, Y., Fan, Y.J. and Tang, Y. (2019) Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in *Drosophila*. *Nucleic Acids Res.*, **47**, 3142–3157.
79. Ankush Jagtap, P.K., Muller, M., Masiewicz, P., von Bulow, S., Hollmann, N.M., Chen, P.C., Simon, B., Thomae, A.W., Becker, P.B. and Hennig, J. (2019) Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1,2 of *Drosophila* helicase Maleless. *Nucleic Acids Res.*, **47**, 4319–4333.
80. Masliah, G., Barraud, P. and Allain, F.H. (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.: CMLS*, **70**, 1875–1895.
81. Jayachandran, U., Grey, H. and Cook, A.G. (2016) Nuclear factor 90 uses an ADAR2-like binding mode to recognize specific bases in dsRNA. *Nucleic Acids Res.*, **44**, 1924–1936.
82. Masliah, G., Maris, C., Konig, S.L., Yulikov, M., Aeschmann, F., Malinowska, A.L., Mabile, J., Weiler, J., Holla, A., Hunziker, J. *et al.* (2018) Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains. *EMBO J.*, **37**, e97089.
83. Partin, A.C., Zhang, K., Jeong, B.C., Herrell, E., Li, S., Chiu, W. and Nam, Y. (2020) Cryo-EM structures of human Drosha and DGCR8 in complex with primary microRNA. *Mol. Cell*, **78**, 411–422.
84. Lazzaretti, D., Bandholz-Cajamarca, L., Emmerich, C., Schaaf, K., Basquin, C., Irion, U. and Bono, F. (2018) The crystal structure of Staufen1 in complex with a physiological RNA sheds light on substrate selectivity. *Life Sci. Alliance*, **1**, e201800187.
85. Bou-Nader, C., Barraud, P., Pecqueur, L., Perez, J., Velours, C., Shepard, W., Fontecave, M., Tisne, C. and Hamdane, D. (2019) Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Res.*, **47**, 3117–3126.
86. Wu, H., Henras, A., Chanfreau, G. and Feigon, J. (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8307–8312.
87. Steff, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B. *et al.* (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, **143**, 225–237.
88. Bottaro, S. and Lindorff-Larsen, K. (2017) Mapping the universe of RNA tetraloop folds. *Biophys. J.*, **113**, 257–267.
89. Fierro-Monti, I. and Mathews, M.B. (2000) Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.*, **25**, 241–246.
90. Davis, I.W., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.*, **32**, W615–W619.
91. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.