Supporting Information

Structure of HIV-1 reverse transcriptase cleaving RNA in an RNA/DNA hybrid

Lan Tian[a,1], Min-Sung Kim[a,b,1], Hongzhi Li[c], Jimin Wang[d] and Wei Yang[a,2]

[a.] Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

[b.] Integrative Bioscience and Biotechnology, Pohang University of Science and Technology, Pohang, Gyeongbuk 37673, Republic of Korea.

[c.] City of Hope, 1500 East Duarte Road, Duarte, CA 91010

[d.] Richards Center for Structural Biology, Yale University, 266 Whitney Ave, New Haven, CT 06520

[1.] These authors contribute equally.

[2.] Correspondence and requests for materials should be addressed to W.Y. (wei.yang@nih.gov).

Table of Contents

1. General methods

Expression and purification of recombinant HIV-1 RT and cellular RNases H

Coding sequences of the p66 and p51 subunits of HIV-1 RT were PCR amplified using the pRT plasmid kindly provided by Drs. Jenny Miller and Stuart Le Grice (NCI, NIH) (1) and cloned into pRSFDuet-1 for co-expression. Each subunit has an N-terminal His_6-tag followed by the PreScission protease cleavage sequence. The presence of His_6-tags increases the p51 and p66 expression level in BLR(DE3) *E. coli* cells (Novagen) after IPTG induction. Protein was purified as described previously (2), and the excess single subunit was removed by MonoS chromatography. The catalytic cores of *B. halodurans* (Bh) and human (Hs) RNase H1 were purified as described (3, 4).

Preparation of RNA/DNA hybrids for co-crystallization with RT

RNA oligonucleotides were purchased from Dharmacon, and those containing an abasic analog were HPLC purified. DNA oligonucleotides (trityl-on) were purchased from the Facility for Biotechnology Resources at FDA and purified by reverse-phase chromatography with an R3 column before detritylation. By design, each RNA/DNA hybrid (Fig. 1A) has a nick in the RNA strand and is thus composed of two RNA oligos (upstream and downstream, near the polymerase and RNase H active site, respectively) and one DNA. The three components were mixed at equal molar ratio and annealed in a thermocycler by heating at 75°C for 10 min and slow cooling to 4°C at a rate of 1°C/min. Co-crystallization was carried out as described previously (2).

Structure determination and refinement

Diffraction data were collected at the SER-CAT ID-22 beamline at the Advance Photon Source (APS) in Argonne National Laboratory and processed using HKL2000 (5) and XDS (6). Structures were determined using Molecular Replacement in CCP4 (7), refined using PHENIX (8), and modeled in COOT (9). All structural figures were generated using PyMol (www.pymol.com).

RNA cleavage assay

For both RNase H and polymerase activity assays, the abasic site at the templating position (R1) (Fig. 1A) was replaced with a uracil (R1U). The downstream RNA oligos were ^{32}P-

labeled at the 5´ end and annealed with R1U (upstream RNA) to the compliment DNA at a 1 : 1.1 : 1.1 molar ratio, so unlabeled strands were 10% more than the labeled strand. For the single turnover analysis of RNA hydrolysis, 40 nM RNA/DNA hybrid was incubated with 60 nM RT in a mixing buffer (0.1 mg/ml BSA, 50 mM Tris·HCl (pH 8.0), 50 mM KCl, 1 mM DTT, 0.5 mM EDTA) at 16°C for 10 minutes in the presence or absence of 100 μM EFV or 500 μM dAMPNPP. Reactions were initiated by addition of 5 mM $MgCl_2$ and terminated at the indicated time points by mixing with 2X volume of stop solution (90% v/v formamide, 0.025% SDS, and 50 mM EDTA). After heat denaturation, reaction products were resolved on 25% polyacrylamide TBE-Urea gels, visualized on a Typhoon PhosphorImager, and quantified using ImageQuant software. Time course data were fitted to a one-phase association equation with non-linear regression using Prism 7.

DNA synthesis assay

DNA oligos were 5´ 6-FAM labeled and annealed with upstream and downstream RNA oligos at a 1:1.2:1.2 molar ratio. For the single turnover DNA synthesis assay, 300 nM RNA/DNA hybrids were incubated with 450 nM D498N mutant RT (inactive RNase H) (2) in the mixing buffer at 16°C for 10 minutes. The reaction was initiated by addition of 5 mM $MgCl_2$ and 100 μM dATP and terminated at indicated time points by mixing with the stop solution. Reaction products were analyzed as described above.

RNA/DNA hybrid-binding assays

RNA/DNA hybrid (10 nM) used for DNA synthesis assay were mixed with RT in the mixing buffer plus 1 mM $CaCl_2$, and incubated at 16°C for 10 minutes. Fluorescent polarization was recorded at 488 nm absorption and 515 nm emission using CLARIOStar (BMG Labtech). Data was fitted to a one-site specific binding equation with non-linear regression and basal values using Prism 7.

Screening for HIV-1 RT inhibitors

Virtual screening of the NCI DTP library (https://dtp.cancer.gov) targeting the interfacial cavity between p51 and p66 identified 96 compounds. Of these, 46 were available and obtained from NCI free of charge. Each compound was dissolved in DMSO as 20 mM stock solution and

stored at -20°C. These compounds were used without further purification or structural characterization.

Initial screening

Initial screening for HIV-1 RT inhibitors was carried out using fluorescence-based binding assays. The RNA is a continuous strand consisting of both R1U and R54 sequences plus four extra U's at the 3´ end (5´ AUGAUGGCCACAAUAACUAGUGGCAUAUUUU-6-FAM), and 3´-labeled with 6-FAM (Dharmacon). The DNA oligo (5´ TATGCCACTAGTTATTGTGGCC) is complementary to the RNA and forms a 22 bp RNA/DNA hybrid. RT (80 nM) was first incubated with 0.5 mM of each compound in 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.1 mg/ml BSA and 1 mM DTT at 25°C for 20 minutes. An equal volume of 40 nM RNA/DNA hybrid was added and incubated at 16°C for another 10 minutes. Addition of 1 mM $CaCl_2$ (for binding assay) or $MgCl_2$ (for cleavage assay) to the mixture allowed the binding and binding and cleavage to complete. Reactions were stopped by adding 4 mM EDTA. Fluorescent polarization was recorded as described above. Compounds inhibiting either binding or cleavage were selected for further analysis.

Characterization by RT binding, Polymerase and RNase H Activity assays

HIV-1 RT (200 nM) was incubated with each compound (showing inhibitory effects in the initial screening) at various concentrations from 5 μM to 0.5 mM at 25°C for 20 minutes. The protein-compound mixture was divided for incubation with an equal volume of either 100 nM RNA/DNA hybrid (3´-FAM-labeled RNA) for binding and RNase H activity assays, or with 100 nM RNA/DNA hybrid with 5´-FAM-labeled DNA for the polymerase activity assay. Products for RNA hydrolysis and DNA synthesis were resolved by TBE-Urea gel following by fluorescent scan using Typhoon.

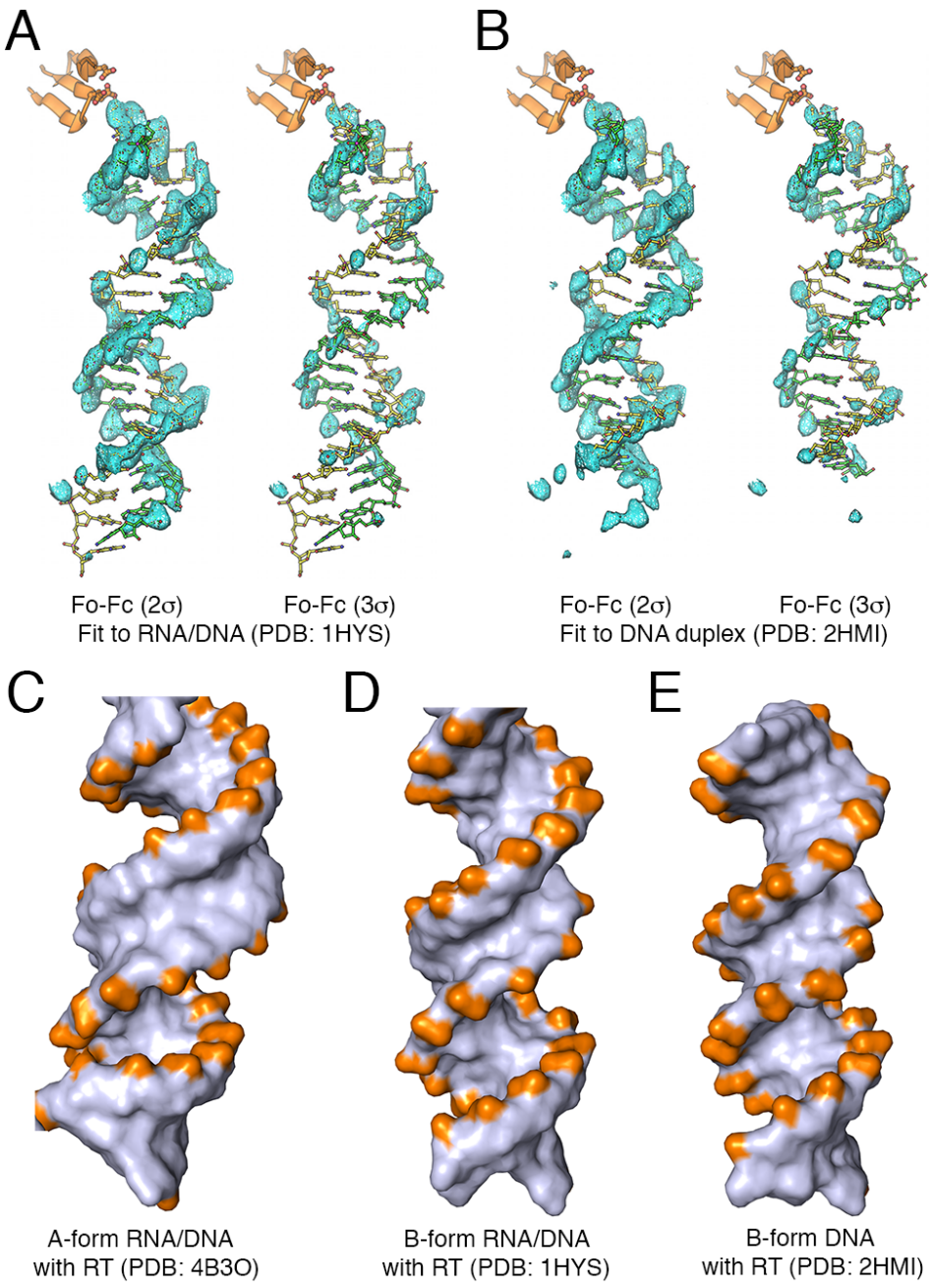


Fig. S1. The B-form RNA/DNA hybrid in 1HYS (PDB accession code). (*A, B*) Unbiased Fo-Fc maps calculated with all nucleic acids omitted from 1HYS and contoured at 2σ (left) and 3σ (right) are superimposed with the RNA/DNA hybrid of 1HYS (10) (A), or DNA duplex of 2HMI (11) (B). There is little density for the bases, but the backbone densities fit the B-form DNA duplex well. (*C-E*) Side-by-side comparison of the RNA/DNA hybrid and DNA duplex complexed with HIV-1 RT. The A-form RNA/DNA hybrid reported by Lapkouski et al. (PDB: 4B3O) (2) (*C*), the B-form RNA/DNA hybrid reported by Sarafianos et al. (PDB: 1HYS) (*D*) (10), and the DNA duplex complexed with HIV-1 RT (*E*) by Ding et al. (PDB: 2HMI) (11) are shown as space-filing model and the phosphate groups are highlighted in orange.

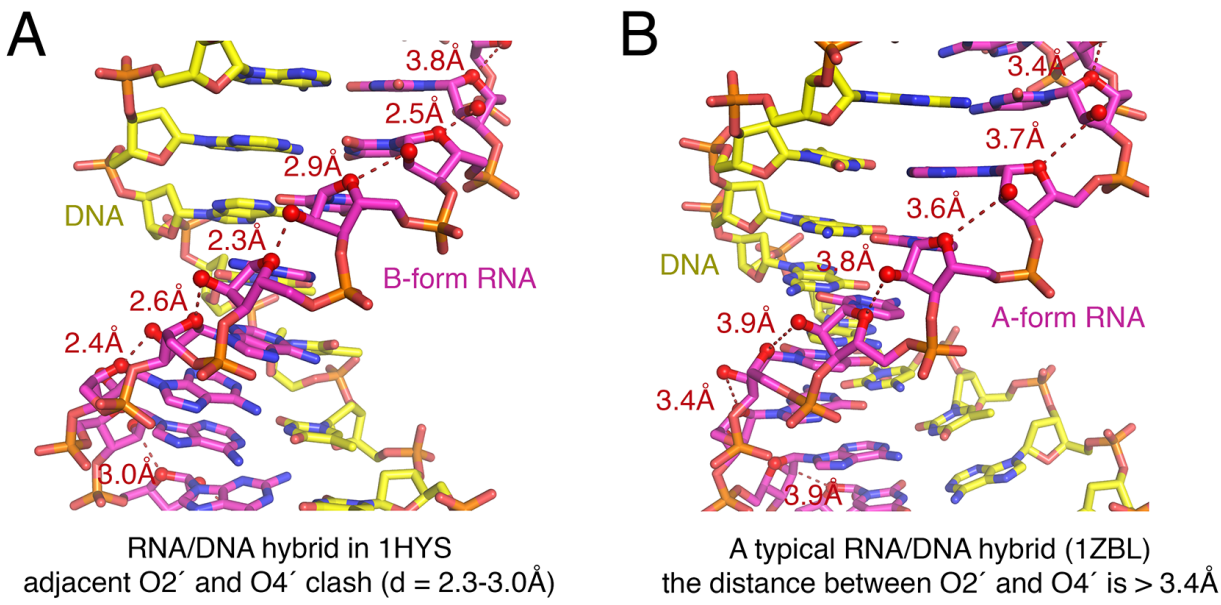


Fig. S2. Severe clashes in the B-form RNA/DNA hybrid of 1HYS. (*A*) The hybrid in 1HYS is shown as pink (RNA) and yellow (DNA) sticks. The oxygen and nitrogen atoms are colored red and blue. The close contacts between O2´ and O4´ of the adjacent nucleotides are marked. (*B*) The distances between O2´ and O4´ in a typical A-form RNA/DNA hybrid (PDB: 1ZBL) are shown.

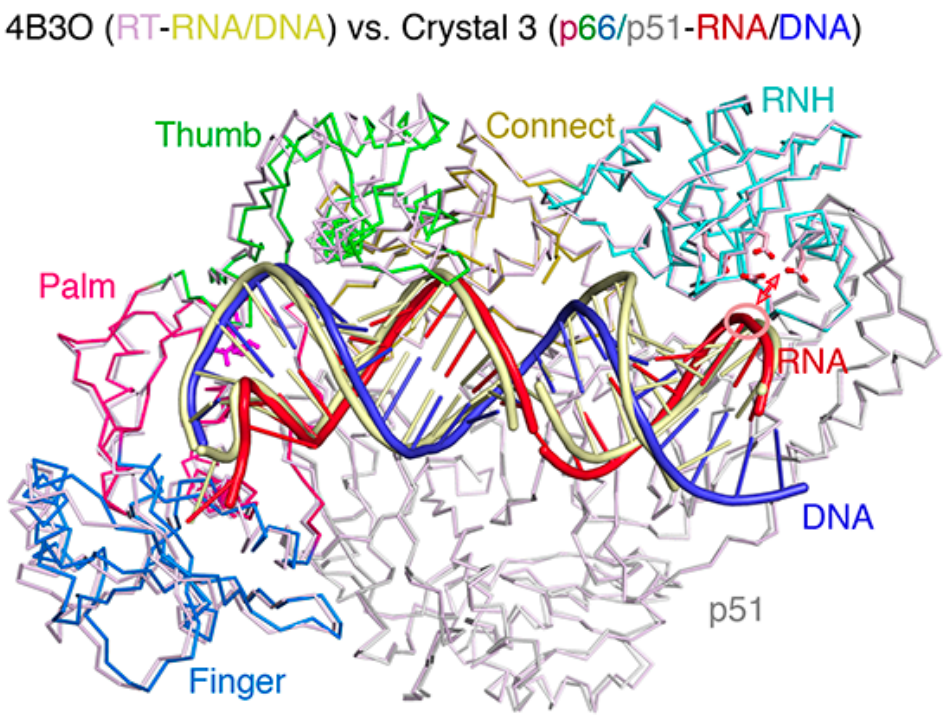


Fig. S3. Comparison of two inactive RT-RNA/DNA hybrid complex structures. The previously published structure (PDB: 4B3O, light pink RT and pale yellow RNA/DNA) is superimposed with Crystal 3 reported here (the p66 subunit is shown in multiple colors and RNA/DNA in red and blue). Despite discernible changes in the thumb and finger domains, the RNA strand remains outside of the RNase H active site in both structures because of disfavored substrate sequence.

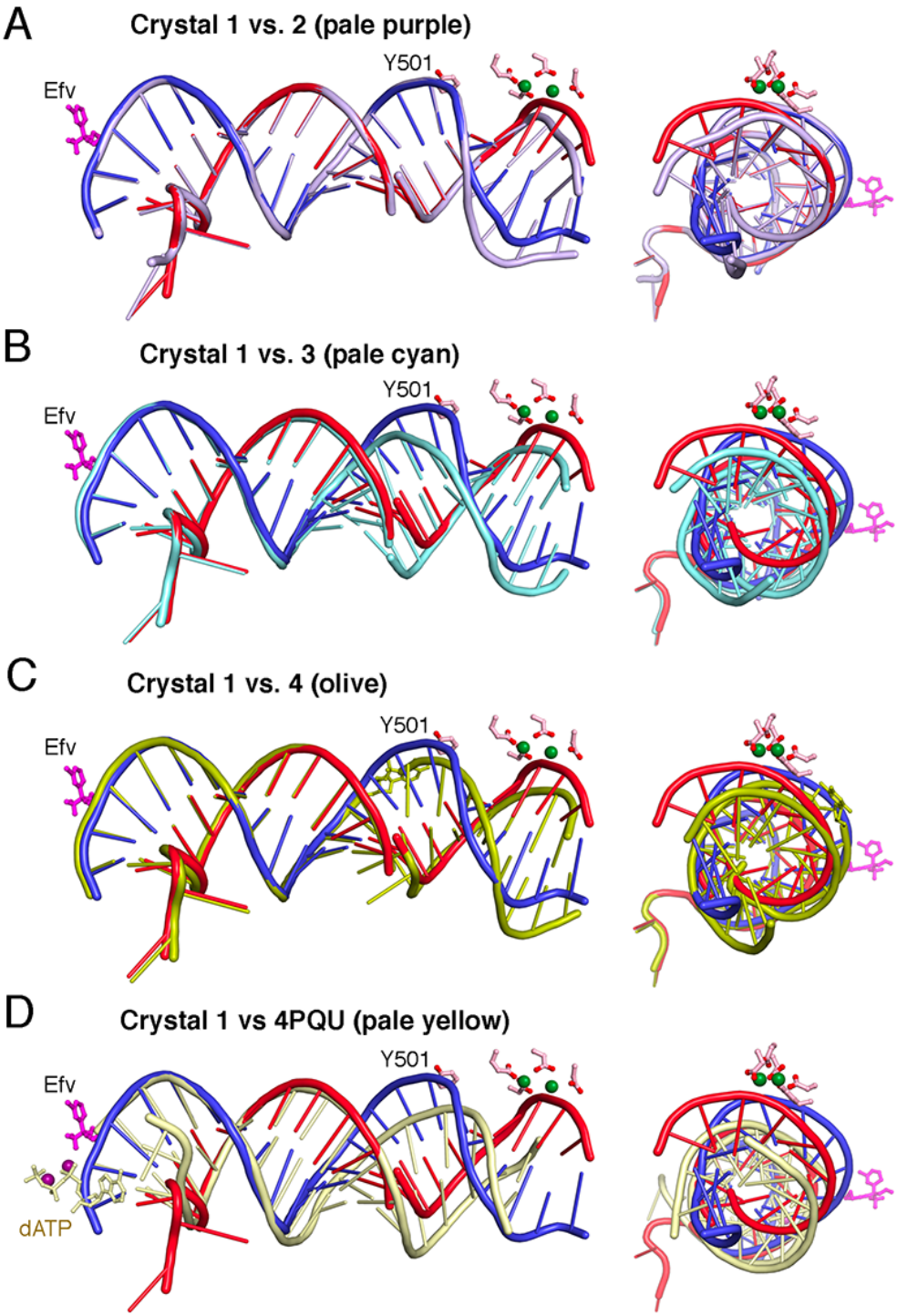
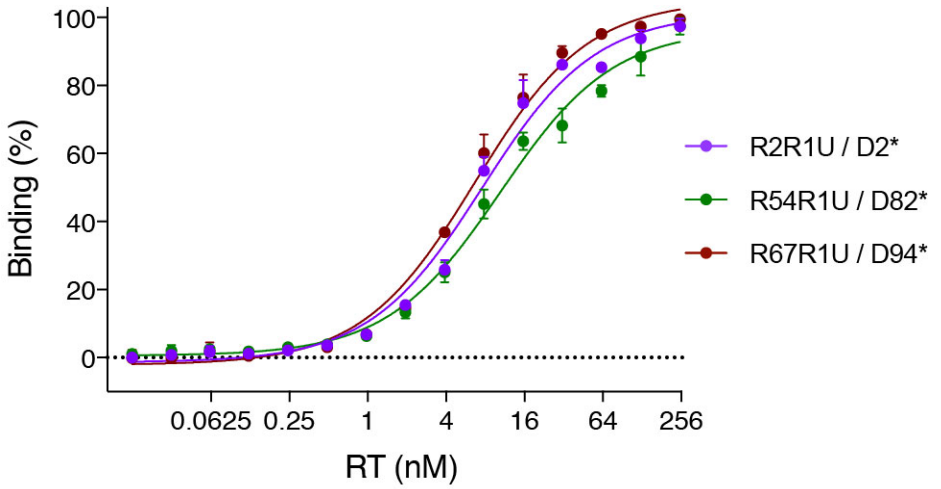


Fig. S4. Structural variations of RNA/DNA hybrids complexed with HIV-1 RT. (*A*) The RNA/DNA hybrids in Crystal 1 (red and blue) and Crystal 2 (light purple) after superposition of the RT structure. The active site residues and Y501 of RT are also shown. (*B-D*) Comparison of the RNA/DNA hybrids in Crystal 1 (red and blue) with Crystal 3 (light blue) (B), or Crystal 4 (olive) (C), and with RNA/DNA hybrid engaging in DNA synthesis (PDB: 4PQU in pale yellow) (D).



RNA/DNA	R2R1U/D2*	R54R1U/D82*	R67R1U/D94*
Kd(nM)	7.6±0.6	10.4±0.8	6.7±0.4

Fig. S5. Fluorescence anisotropy-based binding assays of HIV-1 RT to three RNA/DNA hybrids. The RNA strand is composed of two portions, and R1U contains an uracil instead of an abasic analog as in R1. DNAs were 6-FAM labeled at the 5´-end as indicated by the asterisk.

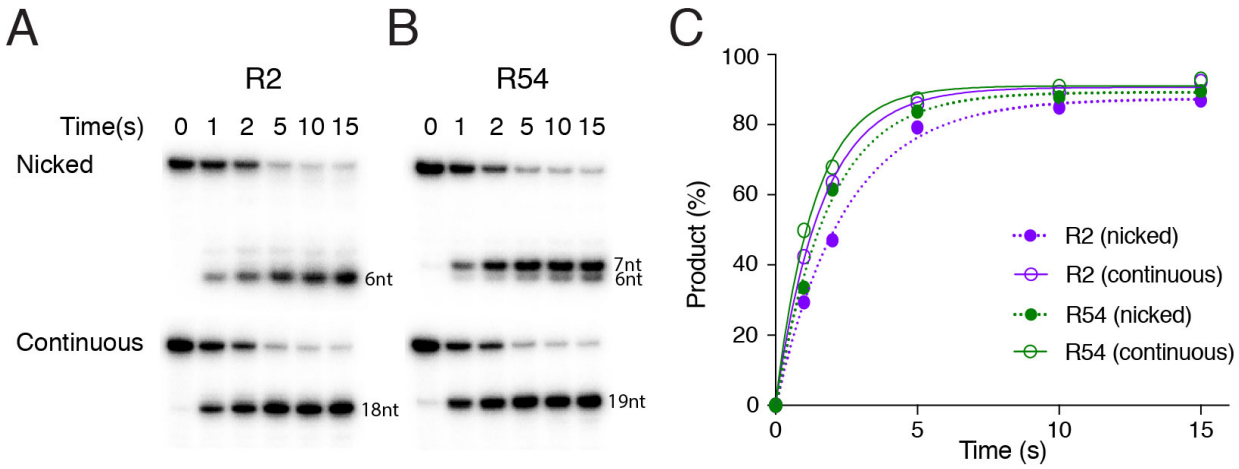


Fig. S6. Comparison of RNA hydrolysis by HIV-1 RT with or without a nick in RNA. (A-B) R2 and R54 sequences represent the non-preferred and preferred substrate sequences, respectively. The RNA/DNA hybrids are the same as described in Fig. 1 except for an additional 7 bps extending beyond the RNase H domain. The RNA strands for cleavage were 5´-^{32}P labeled. The assays were performed in solution as described in Fig. 1D. (C) The relative RNase H activities on four different hybrids are plotted. The preferred cleavage sites are the same with or without a nick on RNA, and the RNase H activity is slightly reduced in the presence of a nick.

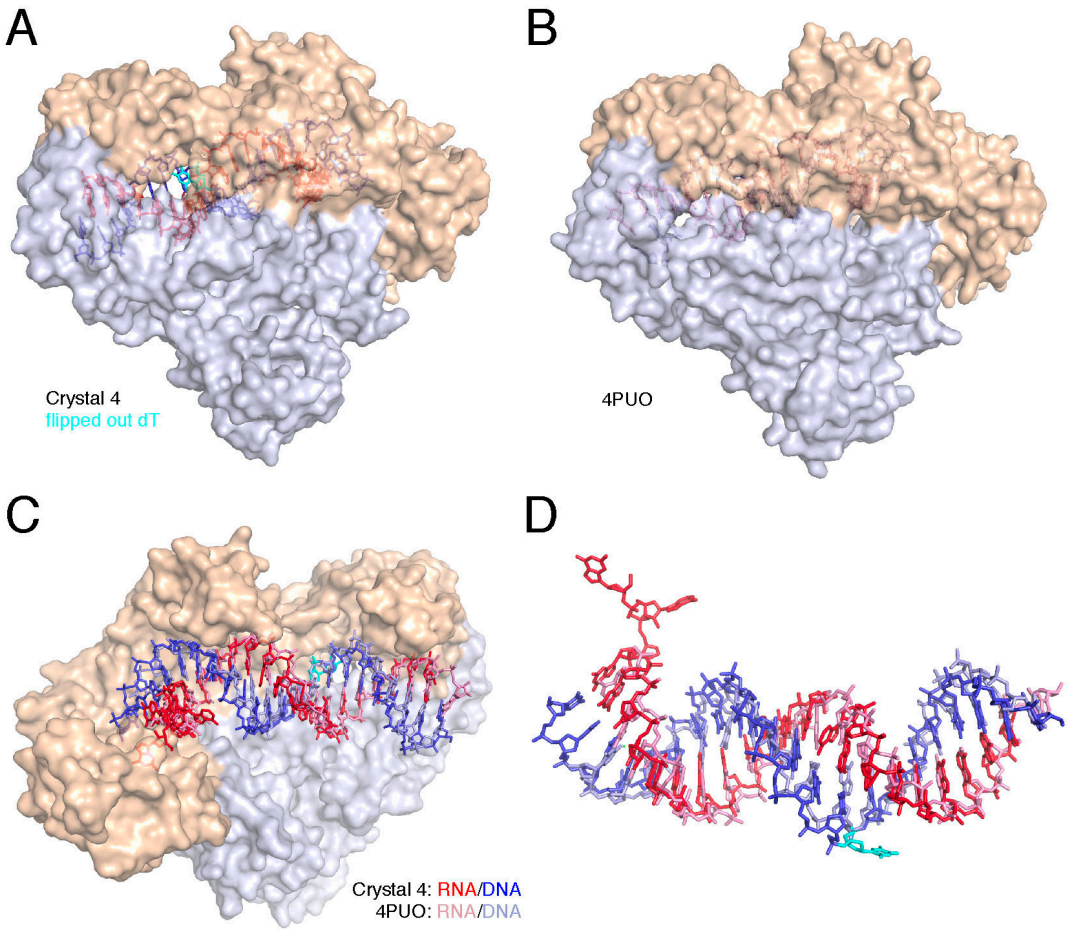


Fig. S7. Comparison of RNase H-inactive HIV-1 RT structures. (*A*) Crystal 4 contains a polypurine-tract sequence and thus the RNase H active site is not engaged. A DNA base is flipped out to the p66-p51 interfacial cavity. (*B*) Soaking of crystals of HIV-1 RT-RNA/DNA hybrid in Nevirapine (12) doesn't induce RNA degradation mode (PDB: 4PUO). The interfacial cavity is tiny. (*C, D*) Superposition of p51 subunits of Crystal 4 and 4PUO reveals the similarity of the similar structures of RNA/DNA hybrids, both of which are outside of RNase H active site.

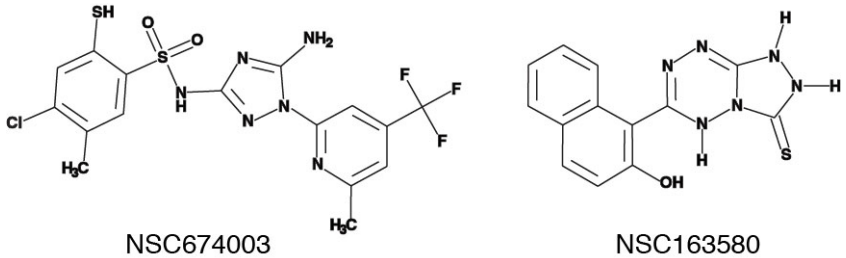


Fig. S8. Diagrams of the two compounds with low inhibitive effects on HIV-1 RT.

Table S1. Crystallographic data collection and refinement statistics

	Crystal 1 (PDB: 6BSH)	Crystal 2 (PDB: 6BSG)	Crystal 3 (PDB: 6BSJ)	Crystal 4 (PDB: 6BSI)
Data collection				
RNA/DNA	R60R54/D91	R60R54/D82	R60R2/D2	R60R67/D94
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Cell dimensions				
a, b, c (Å)	161.8, 161.8, 128.8	162.04, 162.0, 129.2	163.2, 163.2, 129.4	163.9, 163.9, 129.6
Resolution (Å) *	37.2 - 2.65 (2.70 - 2.65)	40.51 - 2.44 (2.53 - 2.44)	41.2 - 2.89 (2.97 - 2.89)	47.3 - 3.25 (3.31 - 3.25)
No. of reflections *	56790 (5625)	72938 (7221)	44876 (4422)	31873 (3135)
Multiplicity *	6.2 (6.0)	7.4 (6.4)	7.4 (7.4)	9.4 (9.3)
Completeness (%)*	99.91 (99.75)	100.00 (100.00)	99.96 (99.98)	99.7 (99.6)
Mean I / σI *	17.94 (2.04)	13.49 (2.14)	12.50 (1.96)	15.22 (1.96)
R_{sym} (%) *	7.6(81.3)	8.1 (81.4)	9.3 (110.0)	8.8(132.7)
Refinement				
No. of reflections	56770	72910	44853	31854
R_{work}/R_{free} (%)	18.92 / 22.73	19.13 / 22.38	20.06/ 25.09	19.28 / 24.08
No. of atoms				
Protein / nucleic acid	7722 / 989	7743 / 989	7650 / 989	7534 / 992
EFV / Ca^{2+}	21 / 2	21/2	21/1	21/2
B-factors				
Protein / nucleic acid	74.7 / 74.6	75.2 / 79.5	92.9 / 104.9	117 / 126.7
EFV / Ca^{2+}	57.5 / 66.8	61.3 / 97.5	76.8 / 120.2	90.3 / 137
R.m.s. deviations				
Bond distances (Å)	0.009	0.007	0.010	0.010
Bond angles (°)	1.049	1.071	1.052	1.165
Ramanchandran plot (%)				
Favored regions	97.0	97.4	95.4	94.7
Allowed regions	3.0	2.6	4.6	5.3
Disallowed regions	0	0	0	0

*Data for the highest resolution shell is shown in parentheses.

References

1. Le Grice SF, Cameron CE, & Benkovic SJ (1995) Purification and characterization of human immunodeficiency virus type 1 reverse transcriptase. *Methods Enzymol* 262:130-144.
2. Lapkouski M, Tian L, Miller JT, Le Grice SF, & Yang W (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat Struct Mol Biol* 20(2):230-236.
3. Nowotny M, Gaidamakov SA, Crouch RJ, & Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121(7):1005-1016.
4. Nowotny M, *et al*. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol Cell* 28(2):264-276.
5. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 276:307-326.
6. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):133-144.
7. Winn MD, *et al*. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-242.
8. Adams PD, *et al*. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
9. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.
10. Sarafianos SG, *et al*. (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J* 20(6):1449-1461.
11. Ding J, *et al*. (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 A resolution. *J Mol Biol* 284(4):1095-1111.
12. Das K, Martinez SE, Bandwar RP, & Arnold E (2014) Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. *Nucleic Acids Res* 42(12):8125-8137.