Architecture of an HIV-1 reverse transcriptase initiation complex

Kevin P. Larsen1,2,4, Yamuna Kalyani Mathiharan1,4, Kalli Kappel1, Aaron T. Coey1,2, Dong-Hua Chen2, Daniel Barrero2, Lauren Madigan2, Joseph D. Puglisi2, Georgios Skiniotis2,3 & Elisabetta Viani Puglisi2

Reverse transcription of the HIV-1 RNA genome into double-stranded DNA is a central step in viral infection and a common target of antiretroviral drugs. The reaction is catalysed by viral reverse transcriptase (RT) that is packaged in an infectious virion with two copies of viral genomic RNA each bound to host lysine 3 transfer RNA (tRNAlys), which acts as a primer for initiation of reverse transcription. Upon viral entry into cells, initiation is slow and non-processive compared to elongation. Despite extensive efforts, the structural basis of RT function during initiation has remained a mystery. Here we use cryo-electron microscopy to determine a three-dimensional structure of an HIV-1 RT initiation complex. In our structure, RT is in an inactive polymerase conformation with open fingers and thumb and with the nucleic acid primer–template complex shifted away from the active site. The primer binding site (PBS) helix forms between tRNAlys and HIV-1 RNA, whereas the viral RNA forms two helical stems positioned above the RT active site, with a linker that connects these helices to the RNase H region of the PBS. Our results illustrate how RNA structure in the initiation complex alters RT conformation to decrease activity, highlighting a potential target for drug action.

During the initiation phase of reverse transcription, RT must bind productively to the viral RNA–tRNAlys complex and then navigate a highly-structured 5′ region of the HIV-1 genome. Critical elements within the viral RNA and host tRNA that are necessary for efficient initiation have been identified. RT pauses at discrete locations, is generally slowed during initiation compared to elongation, and can bind the viral RNA–tRNAlys primer site in different orientations. A rich body of structural data on RT, a heterodimer of p51 and p66 subunits, have shown how its polymerase and RNase H domains interact with DNA–DNA and RNA–RNA duplexes in the absence and presence of antiviral drugs. Lacking, however, are structures that reflect initiation, showing how RT binds to a large bimolecular viral RNA–tRNAlys complex.

We used cryo-electron microscopy (cryo-EM) complemented by biochemical and biophysical experiments to determine the molecular architecture of an HIV-1 reverse transcriptase initiation complex (RTIC). The RTIC was formed using a 101-nucleotide fragment of HIV-1 genomic RNA that encompasses the primer binding site and additional RNA elements required for efficient initiation of reverse transcription. A binary vRNA–tRNA complex was formed with human tRNA that contained a specific cross-linkable site and additional RNA elements required for efficient initiation of reverse transcription. After extending the tRNA primer by one dideoxy-nucleotide to achieve the highest crosslinking efficiency, we generated the cross-linked vRNA–tRNA–RT ternary complex and purified it from free RT and RNA (Fig. 1c, Extended Data Fig. 1a–c). Crosslinking did not affect the global activity of the RTIC. The final cross-linked HIV-1 RTIC had equivalent total activity in incorporation of the next dNTP as an un-crosslinked initiation complex, with rates that are only threefold slower, and is strongly inhibited by nevirapine, a non-nucleoside RT inhibitor (NNRTI) that works through conformational modulation of RT2. Two-dimensional class averages of RTIC clearly showed the RT core as well as protruding RNA densities (Extended Data Fig. 2d). Three-dimensional classification of particle projections revealed substantial conformational variability in the apex of RNA densities (Extended Data Fig. 3a). Owing to this segmental flexibility, we obtained a low-resolution (8.0 Å) reconstruction that best describes the global architecture of RTIC (Fig. 2a, Extended Data Fig. 3b). This EM-density map, encompassing protein and all RNA regions, was of sufficient quality to visualize the tRNA and vRNA, thereby enabling us to position approximately the RNA structures located outside the RT binding cleft. In addition, we obtained a 4.5 Å map by masking out the dynamic peripheral RNA elements and focusing the particle classification and structure refinement on the RT, primer binding site (PBS) helix in the cleft, and additional helical tRNA density (Fig. 3, Extended Data Fig. 3). This higher-resolution map allowed us to describe the conformation of RT and the RNA inside the binding cleft (Fig. 3, Extended Data Fig. 4). An independent 8.2 Å cryo-EM reconstruction of the RTIC was determined in low salt and Mg2+ and revealed a very similar global conformation for the complex (Extended Data Fig. 5), suggesting that the RTIC architecture has limited salt dependence. Models were constructed using the 8.0 Å map to define the global RTIC architecture and the 4.5 Å map to define the structural features of the RTIC core and active site. While the 4.5 Å map provided sufficient resolution to orient the PBS helix of the RTIC, the orientation of the peripheral RNA helical elements of the vRNA and tRNA into the 8.0 Å map was more subjective and relied on iterative Rosetta24 modelling using an accepted secondary structure from past biochemical and biophysical data (see Methods).

The overall RTIC structure shows the RT core with RNA double-helical density within the binding cleft that spans from the active site to the RNase H domain. The helical RNA in the cleft corresponds to the HIV-1 PBS helix formed between nucleotides (nts) 59–76 in the tRNA
Specifically, the D and anticodon stems from nts 10–44 rearrange to form a continuous helical structure, which fits the observed density in the 8.0 Å global map, a long continuous helical RNA density is observed to extend away from the RNase H domain (Fig. 2b). The nucleotide identities of positions 201–203 are highly conserved among recorded HIV-1 sequences, with 70% for 201 and >96% for 202/203, suggesting that this is a common structural feature. In the 8.0 Å global map, a long continuous helical RNA density is observed to extend away from the RNase H domain (Fig. 2b). Accordingly, we propose that the 5′ end of the tRNA (nts 1–54) reverts to form a secondary structure with a contiguous helix (Fig. 2c). Specifically, the D and anticodon stems from nts 10–44 rearrange to form a continuous helical structure, which fits the observed density far better than the three-way junction observed in the free initiation complex (Extended Data Fig. 6).

The helical refolded tRNA domain is connected by a single-stranded connection loop to a 7-bp helix (H1) involving the 5′ (nts 125–131) and 3′ (nts 217–223) termini of the viral RNA construct (Fig. 2b, c) and containing the conserved primer activation signal (PAS) sequence. H1 and the connection loop form a bridge between the RNA located in the RNase H domain and that located near the active site of RT. A three-way RNA junction is formed by the PBS, H1 and a second helical stem loop (H2) comprising nts 134–178 of HIV-1 viral RNA. Density consistent with single-stranded RNA connections H2 to the PBS in the active site. The relative strength (indicative of stability) of the EM density for H1, the connection loop, and the apical regions of H2 differs among several of our low-resolution classes, as do their orientations with respect to the base of H2 and the PBS (Extended Data Fig. 7a, b). For classes that contain strong density of these RNA features, similar models fit these maps by treating the helical RNA elements as rigid units around flexible junction regions (Extended Data Fig. 7c).

Although the RTIC is active in the addition of the next dNTP (Extended Data Fig. 1d, e, h), the complex adopts an inactive conformation in which the position of the tRNA primer terminus within the palm subdomain is shifted approximately 13 Å away from the active site of RT, reminiscent of nucleic acid–RT complexes bound with an NNRTI (Fig. 4a, Extended Data Fig. 9b). As observed in RT structures bound with NNRTI, the primer grip (B12–B13–B14 sheet) is displaced at the 3′ end (Extended Data Fig. 8a). In this experiment, observation of a high FRET state would indicate H1 formation. In the buffer conditions used for cryo-EM imaging, we find that more than 95% of RTIC molecules are in a stable, high-FRET state, indicating that H1 forms for a surface-immobilized RTIC at room temperature (Extended Data Fig. 8b, c).

The electron micrograph density accounts for the majority of vRNA and tRNA structure. Density corresponding to the upper HIV helix 2 stem loop is missing, suggesting that it is partially disordered. c, Proposed secondary structure of the vRNA–tRNA bound within the RTIC. The majority of vRNA helices are well accounted for in the density, with the exception of the apical portions of helix 2 (faded). Additional base pairs (boxed) between the vRNA and the tRNA, which extend the PBS helix, are consistent with the continuous helix that spans the RT binding cleft. The tRNA has refolded and adopted an extended helical conformation.
translocated, with the dC77–G181 pair in the nucleotide acceptor site (N-site) (Fig. 4). The PBS helix is also lifted about 6.0 Å away from the palm and connection domains. The path of the viral RNA template and the base of H2 come into close contact with the residues in the fingers domain. The fingers domain of RT adopts a semi-open conformation similar to that of RT structures bound to nucleic acids that lack an incoming nucleotide20,22 (Fig. 4a, Extended Data Fig. 9). On the basis of previous mechanistic studies of RT enzymology8,29, we conclude that the RTIC here is blocked in a pre-translocation conformation for the primer–template complex (Fig. 4c). Unlike NNRTI-bound complexes, the RTIC is functional and can incorporate the next dNTP, suggesting that there is conformational plasticity within the RT active site. Although RT contacts the RNA substrate using similar domains as in previously determined RT–nucleic acid complexes11–18, the extent of these interactions appears different. The thumb and RNase H domains make the vast majority of observed RNA contact in the RTIC, with a substantial loss of potential interactions in the palm and connection subdomains, consistent with decreased RT–RNA affinity in the initiation complex. The loss of RT–RNA contacts in the palm subdomain20 arises from displacement of the tRNA primer terminus away from the active site (Fig. 4a). Although the RTIC structure is not at sufficient resolution to identify specific protein–RNA contacts, there appear to be additional RT–RNA interactions involving the fingers domain with the vRNA template-strand and H2. The sterically bulky vRNA helices immediately adjacent to the fingers region form a wedge that hinders proper accommodation of the PBS into the cleft and leads to loss of RT–RNA contacts in the cleft and displacement of tRNA 3′ end (Fig. 4b); this is likely to inhibit translocation of the PBS helix to enable efficient and rapid incorporation of the next dNTP.

The architecture of the vRNA–tRNA complex in the initiation complex explains previous experimental results on the role of RNA in initiation8–10. The observed RNA conformation is consistent with chemical probing and enzymatic mapping on similar binary vRNA–tRNA and ternary complexes, which were previously interpreted in terms of tRNA–viral RNA pairings13,14,25. No additional interactions between the vRNA and tRNA occur beyond the extended PBS helix at the +1 stage of initiation (Fig. 2b), consistent with biochemical results on similar HIV-1 subtype-B sequences13,14. Notably absent is any PAS–anti-PAS interaction between HIV-1 nts 123–130 and tRNA nts 48–55 (Fig. 1a), which has been implicated in RT initiation and shown to form dynamically in the absence of the RTIC11,12,27,30. The formation and positions of vRNA H1 and H2 are consistent with their proposed function as barriers during initiation13,14,15,19. The conserved connection loop, bridging RNA within the RNase H domain back to H1, may help to position the vRNA helices in the proper orientation for binding of RT to the tRNA 3′ terminus. The HIV-1 genomic RNA from the MAL isolate, commonly used in past initiation studies, maintains many of these sequence elements with an added 23-nt insertion in the connection loop that may engage in additional interactions. Both H1 and H2 are required for efficient initiation of reverse transcription, and their displacement and unfolding are required for reverse transcription to proceed29,11,13,18; melting of H2 during initiation occurs after addition of the sixth nucleotide19.

The vRNA 3′ end in the RTIC forms an elongated helical structure compatible with an alternative predicted fold21 that involves an extended PBS helix. The vRNA helix and three-way junction of H1, H2 and PBS (Extended Data Fig. 7). Such plasticity is likely to be essential for the RTIC to proceed to elongation.

Our results suggest a model of RT initiation in which RNA structure regulates RT activity. The vRNA and tRNA form a dynamic RNA complex, in which the tRNA refolds to form a metastable conformation of its 5′ region. The ability to refold in this way could explain the use of vRNA in HIV-1 initiation. Although RT contacts the PBS in the cleft, the disrupted palm subdomain contacts between the PBS and RT explain the poor affinity of RT for the vRNA–tRNA complex. Within the framework of the standard dNTP incorporation mechanism29, RT in this +1 initiation complex adopts a pre-translocation conformation with an open active site and improper positioning of nucleic acid for catalysis (Fig. 4c). The vRNA helices, whose orientation hinders productive binding to RT, must be displaced and/or unfolded for the tRNA primer terminus to reposition within the active site such that the RT fingers can clamp down on an incoming nucleotide (Fig. 4b). The dissociation of RT during initiation is rapid, and competes with forward polymerization reactions8,9. RT may dissociate and rebind to the vRNA–tRNA to reposition the primer terminus into the active site. In this pathway, RT rebinding could facilitate melting of downstream RNA structures that hinder translocation. The necessity for these rearrangements during early stages of initiation is likely to explain the low processivity of initiation and the observed pauses that control the start of HIV-1 replication8,9. The single-stranded, A-rich connection loop bridging the 3′ end of the vRNA PBS to H2 may position the vRNA helical in the global map (Fig. 2).
helices properly and allow conformational communication with the RT RNase H domain and refolded tRNA. As reverse transcription proceeds, structural rearrangements in vRNA and tRNA must occur to favour the transition to productive elongation. Thus, the initiation complex is likely to change progressively as initiation proceeds, and may be specifically vulnerable to inhibition by drugs. Higher-resolution structural views of these different states, and dynamics to link them together, will be needed to elucidate further the steps of initiation and underlying RNA conformations that regulate early steps in HIV-1 infection.

Online content
Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0055-9.

Received: 18 September 2017; Accepted: 19 March 2018; Published online 25 April 2018.

1. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. A detailed model of reverse transcription and tests of crucial aspects. Cell 18, 93–100 (1979).
2. Sarafianos, S. G. et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J. Mol. Biol. 385, 693–713 (2009).
3. Baltimore, D. RNA-dependent DNA polymerase in viroms of RNA tumour viruses. Nature 226, 1209–1211 (1970).
4. Temin, H. M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226, 1211–1213 (1970).
5. Paillart, J. C., Shehu-Xhilaga, M., Marquet, R. & Mak, J. Dimerezation of retroviral RNA genomes: an inseparable pair. Nat. Rev. Microbiol. 2, 461–472 (2004).
6. Huang, Y. et al. Incorporation of excess wild-type and mutant tRNA(3Lys) into human immunodeficiency virus type 1. J. Virol. 68, 7676–7683 (1994).
7. Ratner, L. et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313, 277–284 (1985).
8. Lanchy, J. M. et al. Contacts between reverse transcriptase and the primer and template govern the transition from initiation to elongation of HIV-1 reverse transcription. J. Virol. 76, 273–283 (2002).
9. Goldschmidt, V. et al. Structural variability of the initiation complex of HIV-1 reverse transcription. J. Biol. Chem. 276, 31247–31256 (2001).
10. Isel, C., Ehresmann, C. & Marquet, R. Initiation of HIV reverse transcription. Virus Res. 2, 213–243 (2010).
11. Beerens, N. & Berkhourt, B. The RNA primer activation signal in the human immunodeficiency virus type 1 genome is important for initiation and productive elongation of reverse transcription. J. Virol. 76, 2329–2339 (2002).
12. Beerens, N., Groot, F. & Berkhourt, B. Initiation of HIV-1 reverse transcription is regulated by a primer activation signal. J. Biol. Chem. 276, 10386–10394 (2001).
13. Goldschmidt, V., Ehresmann, C., Ehresmann, B. & Marquet, R. Does the HIV-1 primer activation signal interact with tRNA? J. Virol. 76, 2329–2339 (2002).
14. Goldschmidt, V. et al. Structural variability of the initiation complex of HIV-1 reverse transcription. J. Biol. Chem. 279, 35923–35931 (2004).
15. Goldschmidt, V. et al. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. J. Biol. Chem. 277, 42333–42342 (2002).
16. Isel, C. et al. Structural basis for the specificity of the initiation of HIV-1 reverse transcription. EMBO J. 18, 1038–1048 (1999).
17. Iwai, Y., Rosen, A. E., Guo, J., Musier-Forsyth, K. & Levin, J. G. Efficient initiation of HIV-1 reverse transcription in vitro. Requirement for RNA sequences downstream of the primer binding site abrogated by nucleocapsid protein-dependent primer-template interactions. J. Biol. Chem. 278, 14185–14195 (2003).
18. Liang, C. et al. The importance of the A-rich loop in human immunodeficiency virus type 1 reverse transcription and infectivity. J. Virol. 71, 5750–5757 (1997).
19. Liu, S., Harada, B. T., Miller, J. T., Le Grice, S. F. & Zhuang, X. Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. Nat. Struct. Mol. Biol. 17, 1452–1460 (2010).
20. Huang, H., Chopra, R., Verdone, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282, 1669–1675 (1998).
21. Jacobo-Molina, A. et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc. Natl Acad. Sci. USA 90, 6320–6324 (1993).
22. Li, A., Gong, S. & Johnson, K. A. Rate-limiting pyrophosphate release by HIV reverse transcriptase improves fidelity. J. Biol. Chem. 281, 26554–26565 (2006).
23. Puglisi, E. V. & Puglisi, J. D. Secondary structure of the HIV reverse transcription initiation complex by NMR. J. Mol. Biol. 410, 863–874 (2011).
24. Puglisi, E. V. & Puglisi, J. D. Secondary structure of the HIV reverse transcription initiation complex formed by conserved RNA sequences within the HIV reverse transcription initiation site. RNA 22, 1689–1698 (2016).
25. Watts, J. M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711–716 (2009).
26. Foley, B. et al. HIV Sequence Compendium 2013 (Los Alamos National Laboratory, Los Alamos, 2013).
27. Peisley, A. & Skiniotis, G. 2D projection analysis of GPCR complexes by negative stain electron microscopy. Methods Mol. Biol. 1335, 29–38 (2015).
28. Das, K., Martinez, S. E., Bauman, J. D. & Arnold, E. HIV-1 reverse transcriptase improves fidelity. J. Mol. Biol. 410, 711–716 (2011).
29. Peisley, A. & Skiniotis, G. 2D projection analysis of GPCR complexes by negative stain electron microscopy. Methods Mol. Biol. 1335, 29–38 (2015).
30. Li, A., Gong, S. & Johnson, K. A. Rate-limiting pyrophosphate release by HIV reverse transcriptase improves fidelity. J. Biol. Chem. 281, 26554–26565 (2006).
31. Puglisi, E. V. & Puglisi, J. D. Secondary structure of the HIV reverse transcription initiation complex formed by conserved RNA sequences within the HIV reverse transcription initiation site. RNA 22, 1689–1698 (2016).
32. Puglisi, E. V. & Puglisi, J. D. Secondary structure of the HIV reverse transcription initiation complex formed by conserved RNA sequences within the HIV reverse transcription initiation site. RNA 22, 1689–1698 (2016).
33. Watts, J. M. et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature 460, 711–716 (2009).
34. Foley, B. et al. HIV Sequence Compendium 2013 (Los Alamos National Laboratory, Los Alamos, 2013).
35. Yeaman, S. L., Johnson, K. A., Vlahov, D., Raffa, R. & Allander, V. W. First-line antiretroviral therapy and HIV-1 reverse transcriptase: implications for drug resistance. Science 282, 1669–1675 (1998).
36. Latorraca for discussions and assistance with the Sherlock cluster. Supported by National Institutes of Health grant GM082545 to E.V.P., T32-GM008294 (Molecular Biophysics Training Program) to K.P.L., A.T.C. and K.K., National Science Foundation Graduate Research Fellowship Program (DGE-114747) to A.T.C and K.K., and Gabilian Stanford Graduate Fellowship to K.K. We thank Stanford University and the Stanford Research Computing Center for providing the Sherlock cluster resources. Additional calculations were performed on the Stanford BioX3 cluster, supported by NIH Shared Instrumentation Grant 1S10RR02664701.

Reviewer information Nature thanks N. Sluis-Cremer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.P.L., Y.K.M. and D.-H.C. acquired preliminary cryo-EM data and performed initial cryo-EM map calculations. Y.K.M. acquired cryo-EM data and obtained the 3D reconstructions shown in the main manuscript. K.P.L. acquired Mg\textsuperscript{2+} cryo-EM data and performed corresponding cryo-EM map calculations. A.T.C. purified the vRNA used for single-molecule experimentation and performed the single-molecule experiments. K.P.L., D.B. and L.M. performed all vRNA and RT sample preparations. K.P.L. performed all \textsuperscript{32P}-dUTP incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation. K.K. performed the vRNA–tRNA model building with input from K.P.L. and E.V.P. incorporation assays. D.B. performed the RT activity assays. K.P.L. designed the purification scheme and purified the RTIC used in all experimentation.
**METHODS**

**Sample preparation.** HIV-1 vRNA constructs were prepared by in vitro transcription with T7 RNA polymerase as previously described\(^2,7,28\). Transcripts were denatured in 8 M urea and purified on a sequencing PAGE gel. Gel extraction was performed using 0.3 M ammonium acetate. Following ethanol precipitation, the RNA was dissolved in 10 mM Bis-Tris propane, pH 7.0, 10 mM NaCl and stored at \(-20^\circ\text{C}\). The crosslinkable RNA\(^3\) was chemically synthesized, 5′-biocytin and washed with 300 mM NaCl to remove the free vRNA–tRNA complexes. The mixture was heated to 90°C and slow cooled to room temperature. The vRNA–tRNA complex was purified from higher order and unannealed monomer species using a Superdex 200 (26/600) gel filtration column with 10 mM Bis-Tris propane, pH 7.0, 100 mM NaCl. The presence of a single species was confirmed with native PAGE and samples were concentrated on a Vivaspin 20 10,000 MWCO concentrator. Samples were stored at \(-20^\circ\text{C}\) and exhibited minimal aggregation over time.

HIV-1 RT was expressed in Escherichia coli strain BL21(DE3). Two expression vectors, one containing p66 and ampicillin resistance and the other containing p51 and kanamycin resistance, were constructed. The C terminus of p66 contains an unstructured linker and a six-histidine tag. A cysteine mutation for crosslinking was introduced into helix H of p66 (Q528C)\(^7,28\). The protein used in this study also has the C202S mutation, introduced in the small RT domain and the FATQ motif, introduced to eliminate RNase H activity as RT has been shown to cleave dsRNA when stalled for long periods\(^20,32\). Cell pellets were lysed through sonication and the enzyme was purified by gravity Ni-nitritriaticic acid (Ni-NTA) affinity chromatography, followed by size-exclusion chromatography using a Superdex 200 (26/600). The His\(_{6}\) tag was cleaved by thrombin digestion overnight. The cleaved protein was re-applied to a Ni-NTA column to remove the uncleaved His\(_{6}\) tag. This was followed by an additional final size-exclusion chromatography step. The protein was stored at 4°C in 300 mM NaCl, 50 mM Tris, pH 8.0, 5 mM 3-mer.

The RTIC was prepared by mixing RT and vRNA–tRNA complex at 2 μM each, respectively, in a buffer containing 25 mM NaCl, 25 mM KCl, 5 mM MgCl\(_{2}\), 50 mM Tris, pH 7.5, 100 μM dTCP (or dCTP if used for 3′ incorporation assays). The mixture was allowed to crosslink overnight at room temperature. The complex was purified by anion-exchange chromatography with a linear gradient. This was followed by a size-exclusion chromatography step to remove any higher-molecular-weight aggregates. The purity and homogeneity of the final complex were assessed by SDS–PAGE (under non-reducing conditions) and size-exclusion chromatography (on a Superose 6 10/300 GL column).

Amino-GMP-labelled viral RNA for single-molecule experimentation was transcribed as previously described but with nucleotide concentrations of 1 mM ATP, CTP and UTP and 0.5 mM GMP. The vRNA sequence is identical to that used in the cryo-EM experiments, but contains an additional unstructured sequence on the 3′ end for immobilization and oligonucleotide hybridization purposes and an additional GGU on the 5′ end for labelling purposes. 5′-Amino-G-monophosphate (GMP), purchased from TriLink Biotechnologies, was added to the reaction at a concentration of 1 mM ATP, CTP and UTP. The vRNA was dissolved in 10 mM Bis-Tris propane (pH 7.0), 75 mM NaCl. The RNA was then separated from template DNA and free NTPs by size-exclusion chromatography (ENRIC SEC 650 10 × 300) in 100 mM sodium phosphate buffer (pH 8.2), 75 mM NaCl. Purified amino-G-MP-labelled RNA was concentrated to 1 μM and labelled using NHS chemistry with 1,000-fold excess cyanine dye (Lumiprobe). Excess dye was removed by passage over a 10DG desalting column in 10 mM Bis-Tris propane (pH 7.0), 75 mM NaCl. The RNA was then annealed to template DNA and free NTPs by size-exclusion chromatography (ENRIC SEC 650 10 × 300) purification to buffer exchange the labelled vRNA. Labelling efficiency was calculated by measuring the absorbance values of the labelled species at both 260 nm (RNA absorbance) and 550 nm (Cy3 absorbance). These absorbance values were used to calculate the concentrations of the RNA and the Cy3 dye. Using the ratio between these two values, we estimate that our 5′ labelling efficiency is approximately 70%.

Dye-labelled vRNA–tRNA complexes were heat-annealed and purified as previously described. The single-molecule RTIC complex was prepared as stated above, but with a several modifications. To simplify the purification, the his-tag was kept on the p66 subunit of RT. The RTIC was then applied to a Ni-NTA column and washed with 300 mM NaCl to remove the free vRNA–tRNA complexes. The RTIC was eluted from the column. Synthetic dye labelled RNA (5′-biotinial-5′-AUCUCUCCCU-AUCCCdC-3′ (TriLink)) were annealed to the complex at 37°C for 5 min in tenfold molar excess. Excess oligonucleotides and free RT were rinsed away during TIRF slide preparation. The above protocol was also performed for a dye-labelled vRNA–tRNA–only control, but skipped the RTIC complex formation, purification, and free RT rinse.

**Single-molecule FRET experiments.** Single-molecule FRET experiments were performed using a prism-based total internal reflection instrument with a diode-pumped solid-state 532-nm laser as previously described\(^2,7,33–38\). This includes the use of an oxygen scavenging system (protocatechuic 3,4-dioxygenase (PCD) and 3-carboxylic acid (PCA)) and a triplet state quencher (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Troxol)) to reduce aberrant dye behaviours. The laser power measured 50 mW at the prism. The fluorescence signal was recorded with an exposure of 100 ms per frame for 5 min at room temperature. FRET traces were manually analysed using home-written scripts in MATLAB (MathWorks)\(^2\). This analysis began by using a colocalization script to select only spots that exhibited both Cy3 donor and Cy5 acceptor fluorescence (under donor–only excitation conditions). Such colocalization allowed us to eliminate partially labelled molecules. Next, FRET traces were manually inspected to eliminate cases with multiple single-dye photobleaching events (multiple molecules) or traces that exhibited poor dye photophysics (Extended Data Fig. 8d). After this manual inspection, the final dataset used for analysis included only traces in which both dyes exhibited clear single photobleaching events to ensure reliable data (Extended Data Fig. 8c). For our single-molecule experiment, 708 traces were selected through colocalization. After manual inspection and elimination of poor traces, 480 traces were used for the final analysis. In addition to the RTIC experiment, a control experiment using a dye-labelled vRNA–tRNA was performed to assess the FRET population without RT. We found that in the absence of RT, a lower population of low-FRET state molecules with FRET efficiency 0.3 existed, but this state was not observed upon binding and cross-linking of RT (data not shown).

**Negative-stain EM.** We applied 3.5 μL of 0.1 μM RTIC sample onto glow-discharged carbon-coated grids, and blotted and stained them with 1% uranyl formate according to standard protocols\(^2\). Negative–stained grids were imaged on an FEI Morgani at 100 kV.

**Cryo-EM data acquisition.** RTIC complex in high monovalent salt buffer (300 mM NaCl, 10 mM Tris-HCl pH 8.0) containing 0.2-0.25% (w/v) beta-ocyt glycol (β-G) was applied to glow-discharged lacey carbon grids (Quantifoil R2/2, 200 mesh) and subsequently vitrified using a FEI Vitrobot. Frozen hydrated samples were imaged on an FEI Titan Krios at 300 kV with a Gatan K2 Summit direct detection camera in counting mode with 200 ms exposure per frame. Forty frames per micrograph were collected at a magnification of 29,000 ×, corresponding to 1 Å per pixel at the specimen level. In total, 4,209 micrographs were collected at defocus values ranging from −1.3 to −2.5 μm. The movie frames were motion–corrected and dose-weighted by MotionCor2\(^2\) and CTF parameters were estimated by CTFFIND4\(^2\). RTIC complex in low monovalent salt buffer and Mg\(^2+\) (75 mM NaCl, 2 mM MgCl\(_{2}\), 10 mM Tris-HCl pH 8.0) containing 0.2% (w/v) β-G was applied to glow-discharged lacey carbon grids (Ems, 200 mesh, Copper) and subsequently vitrified using a Leica EM GP. Frozen hydrated samples were imaged on a Tecnai F20 at 200 kV with a Gatan K2 Summit direct detection camera in counting mode with 200 ms exposure per frame. Sixty frames per micrograph were collected at a magnification of 29,000 ×, which corresponds to 1.28 Å per pixel at the specimen level. In total, 898 micrographs were collected at defocus values ranging from −2.0 to −3.0 μm and a dose rate of 8.0 electrons per pixel per second. The micrograph movies were motion–corrected and dose-weighted as above, and CTF parameters were estimated by GCTF\(^4\). **Cryo-EM data processing.** Cryo-EM data for the 8.0 and 4.5 Å maps were processed using Relion\(^2,42–44\). 675,688 particle projections were semi-automatically picked from the motion–corrected micrographs, and sorted through subsequent rounds of reference-free 2D classification. 444,374 particle projections belonging to classes with well-defined RT and RNA features were selected for further processing (Extended Data Fig. 3a). An initial 3D model was obtained using VIPER\(^45\) based on the selected 2D classes, and used for 3D classification in Relion\(^2,42–44\) (Extended Data Fig. 3a). Because particle alignment was affected by the flexible protruding RNA, we used a mask and focused the alignment on RT and PBS alone. 167,906 particle projections sorted to 3D classes displaying all features of RT and PBS were selected for subsequent 3D classifications. To further improve the quality of RT/PBS core, one more round of 3D classification with finer angular sampling was executed; particles from two classes with well-defined secondary structure densities were combined and the 3D structure was refined to a resolution of 4.5 Å. For global RTIC maps (including the flexible protruding RNA) a 3D classification without mask using the 167,906 particle projections subset was performed; eight classes were considered, where two classes were dominating most of the RNA protrusions was refined to a resolution of 8.0 Å. The resolution reported is according to the 0.143 gold standard" Fourier shell correlation (FSC) criterion (Extended Data Fig. 3b). The 4.5 and 8.0 Å maps were corrected for the
modulation transfer function (MTF) of K2 direct detection camera at 300 kV and then sharpened using B factors of −250 and −200 Å 2, respectively, during the post-processing step (Extended Data Table 1). Local resolution was estimated using Relion (Extended Data Fig. 3c).

Cryo-EM data for the 8.2 Å Mg 2+ map were processed using Relion. 148,523 particle projections were semi-automatically picked from motion-corrected micrographs and sorted through subsequent rounds of reference-free 2D classification. 125,615 particle projections belonging to classes with well-defined RT and RNA features were selected for further processing (Extended Data Fig. 3e). An initial 3D model was obtained using EMAN2 based on selected 2D classes 46 and used for 3D classification. The resolution reported is according to the 0.143 ‘gold standard’ FSC criterion (Extended Data Fig. 3f). The maps were corrected for the modulation transfer function (MTF) of K2 direct detection camera at 200 kV and then sharpened using a B factor of −200 during the post-processing step (Extended Data Table 1).

Model building and refinement. The crystal structure of RT bound to a DNA–RNA duplex, with nucleic acid substrate removed, was used as a starting model for RT 22. After manually fitting the main-chain backbone of RT into its distinct density, four regions of EM density corresponding to RNA were apparent. The most notable region of density is the well-formed RNA helix in the cleft of RT, which corresponds to the PBS helix of the vRNA–tRNA complex. The complete model of the vRNA–tRNA complex was built piecewise, and iteratively, using the Rosetta FARFAR method 24. Initial, models of the first seven base pairs of the PBS helix (vRNA residues 181–187 and tRNA residues 71–77 with the extended dC, originally modelled as RNA for simplicity and later edited to a dD in Coot 47) were built with FARFAR, then clustered. The centres of the ten most populated clusters were fit into the density using the colores tool in the Situs package 48. The resulting models were manually inspected and selected on the basis of fit to the density and proximity of G71 to C258 on RT (base and residue involved in crosslinking). This helix was extended out to nineteen base pairs (to include vRNA residues 187–199 and tRNA residues 59–70), models were again clustered, and the cluster centres were fit into the density. The model that best fit the 4.5 Å cryo-EM density was selected. This RT–PBS model, called the RTIC core, was then refined using Phenix real-space-refinement 49 with secondary structure restraints in place for the RNA and protein. To further restrain the model during refinement, the N2-cysteamine–deoxyguanosine was inserted into the tRNA sequence and a loose disulfide bond constraint with C258 was used during refinement (this was later reverted to a dG as there was no density for the linker atoms). The model was visually inspected and manually adjusted in Coot 47. Protein residues lacking EM density, the vast majority of which were located in the fingers and palm subdomains, were removed after comparison to prior models of RT (Extended Data Fig. 4). Owing to insufficient resolution, large regions of RT did not exhibit reasonable density for sidechains. Therefore the RT model was truncated to a main-chain backbone before final inspection and the description of the PDB region. The RTIC core structure was validated using Molprobity 50. This refined RTIC core model (Fig. 3) served as the anchor point for orienting the vRNA and tRNA portions of the global RTIC model.

Two additional regions of RNA density were located near the fingers subdomain of RT (Fig. 2a). As the crosslinking method used to form the complex harnesses RT polymerase activity, these regions of density correspond to the template vRNA helices. We traced the vRNA template strand out of the active site, allowing us to confidently orient the base of vRNA H1. Models of vRNA H2 (residues 134–178) were built with FARFAR 24 based on the consensus secondary structure from past biochemical and biophysical data 28,51. These models were clustered and fit into the density.

After positioning vRNA H2, only one region of RNA density near the fingers remained unaccounted. This density, which was continuous with the H2 density, corresponds to vRNA H1 (residues 125–131 and 217–223). This density also connects to RNA located in the cleft of RT near the RNase H domain. This suggests that the connection loop may contact H1 and contribute to the density observed in this region. To confirm the presence of H1 in the RTIC, single-molecule experiments were performed in which a FRET pair was placed on the 5’ and 3’ ends of the helix. We find that in our imaging conditions, 85–95% of molecules exist in a stable conformation, high FRET, consistent with H1 formation. vRNA helix 1 was modelled as an ideal RT (2 μM) were also preincubated for 20 min under the same conditions, but with dCTP in order to fully incorporate the first nucleotide before dTTP incorporation. Incorporation reactions were started by adding a mixture of α,β-ATP–dTTP (50 μM), and dTTP (50 μM). Reactions were quenched at a range of times from 1 s to 4 h with the addition of EDTA and SDS loading buffer. The reactions were run on an 4–20% SDS–PAGE gel, dried, and exposed for 18 h on a phosphoimager screen (Molecular Dynamics) and imaged with a Storm 860 (Molecular Dynamics). Bands were quantified using ImageQuant. Intensity was normalized to the highest band intensity for the individual time course assays after background subtraction (set to 1). All time course assays were reliably reproduced and the slow reactions required no special equipment 4. Plotting and curve fitting was done using IgorPro. For NNRTI experiments, 1 μM nevirapine was added to the pre-reaction incubation mixture of the RTIC.

Relative total incorporation assay. Reactions were performed as described above. RTIC reaction mixtures were quenched at 1 h and the free RT and vRNA–tRNA were quenched at 30 min. Samples were quantified as described above. Incorporation was normalized to the average free RT + vRNA–tRNA band intensity (set to 100%). Relative total incorporation assays were done in triplicate.

Reverse transcriptase assay. vRNA–tRNA complexes were purified as described above. We used a 4:1 mixture of the 5’ end-labeled reactions were pre-incubated at 37 °C for 5 min in 50 mM Tris–HCl (pH 8.0), 50 mM KCl, 6 mM MgCl 2, and 5 mM MnCl 2 at a vRNA–tRNA concentration of 200 nM and RT concentration of 3 μM. Reactions were initiated by the addition of a dNTP mixture.
that brought the final individual dNTP concentrations to 100 μM. Reactions were performed in triplicate and quenched at 30 min with EDTA (50 mM). Samples were denatured in a formamide loading buffer, heated for 5 min at 95 °C, and loaded on an 8.5% polyacrylamide gel that was pre-run for 2 h. Samples were run for 3 h at 120 W before imaging with a Typhoon Trio (Amersham Biosciences).

Fully extended and unextended primer bands were quantified using ImageQuant. Percent primer extension was calculated and normalized to wild-type RT.

Data availability. All other data are available from the corresponding author upon reasonable request.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Cryo-EM maps of the global RTIC, the core of the RTIC, and the global RTIC with MgCl2 have been deposited in the Electron Microscopy Data Bank under accession codes EMDB-7032, EMDB-7031 and EMDB-7540. The coordinates of the RTIC core model have been deposited in the Protein Data Bank under accession code 6B19. The global RTIC model is available as Supplementary Data. All other data are available from the corresponding author upon reasonable request.

32. Götte, M. et al. HIV-1 reverse transcriptase-associated RNase H cleaves RNA/RNA in arrested complexes: implications for the mechanism by which RNase H discriminates between RNA/RNA and RNA/DNA. EMBO J. 14, 833–841 (1995).

33. Marshall, R. A., Dorywalska, M. & Puglisi, J. D. Irreversible chemical steps control intersubunit dynamics during translation. Proc. Natl Acad. Sci. USA 105, 15364–15369 (2008).

34. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. Biophys. J. 94, 1826–1835 (2008).

35. Johansson, M., Chen, J., Tsai, A., Kornberg, G. & Puglisi, J. D. Sequence-dependent elongation dynamics on macrolide-bound ribosomes. Cell Rep. 7, 1534–1546 (2014).

36. O’Leary, S. E., Petrov, A., Chen, J. & Puglisi, J. D. Dynamic recognition of the mRNA cap by Saccharomyces cerevisiae eIF4E. Structure 21, 2197–2207 (2013).

37. Aitken, C. E. & Puglisi, J. D. Following the intersubunit conformation of the ribosome during translation in real time. Nat. Struct. Mol. Biol. 17, 793–800 (2010).

38. Chen, J., Tsai, A., Petrov, A. & Puglisi, J. D. Nonfluorescent quenchers to correlate single-molecule conformational and compositional dynamics. J. Am. Chem. Soc. 134, 5734–5737 (2012).

39. Zheng, S. Q. et al. MotioCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017).

40. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221 (2015).

41. Zhang, K. Grif: Real-time CTF determination and correction. J. Struct. Biol. 193, 1–12 (2016).

42. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530 (2012).

43. Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. J. Struct. Biol. 189, 114–122 (2015).

44. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. Methods Enzymol. 579, 125–157 (2016).

45. Penczek, P. A., Grassucci, R. A. & Frank, J. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53, 251–270 (1994).

46. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. J. Struct. Biol. 157, 38–46 (2007).

47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).

48. Wrapp, T., et al.cryo-microscopy of SARS-CoV-2 at 2.4 Å resolution. Nature 579, 71–76 (2020).

49. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).

50. Lecuit, M. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, W375–W383 (2007).

51. Lavender, C. A., Gorelick, R. J. & Weeks, K. M. Structure-based alignment and consensus secondary structures for three HIV-related RNA genomes. PLoS Comput. Biol. 11, e1004230 (2015).

52. Petses, S. V. et al. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).

53. Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. & Marquet, R. Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/ tRNA(3Lys) (template/primer). J. Mol. Biol. 247, 236–250 (1995).

54. Isel, C. et al. Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA3Lys. EMBO J. 15, 917–924 (1996).
Extended Data Fig. 1 | Purification and activity of RTIC. a, Initial anion-exchange purification of the RTIC away from free RT and vRNA–tRNA. This purification was repeated for each sample (>10) used in the manuscript, with only slight variations in the chromatogram. b, Polishing step using size-exclusion chromatography purification of the RTIC after anion exchange. This purification was repeated for each sample used in the manuscript (>10), with only slight variations in the chromatogram. c, A final 10% native TBE gel on the purified components. RT barely enters the gel under these running conditions. The RTIC runs as a single band, but trace amounts of free vRNA and/or vRNA–tRNA complex are sometimes present. This native gel is a representative result that was repeated independently for all purified RTIC samples used in the paper (>10). d, Autoradiograph image illustrating that the RTIC is capable of incorporating an incoming $\alpha$-32P-dTTP nucleotide when extended and purified using dCTP instead of ddCTP. This gel is a representative result that was repeated independently for crosslinked and uncrosslinked samples (>6 independently prepared samples) used in dTTP incorporation assays. e, The RTIC incorporates $\alpha$-32P-dTTP at roughly 89% efficiency compared to the free components after reaching a plateau. Values are mean ± s.d. ($n = 3$ independent experiments) with normalization to total incorporation of free RT + vRNA–tRNA reactions. f, Autoradiograph image showing that the incorporation of dTTP is inhibited in the presence of nevirapine (NNRTI). Images have been adjusted to allow identification of the NNRTI-inhibited band. This gel is a representative result that was repeated independently for crosslinked and uncrosslinked samples (3 samples each). g, Relative activities, judged by primer usage, of wild-type, Q258C, and Q258C/E478Q reverse transcriptase mutants used in this study. Values are mean ± s.d. ($n = 3$ independent experiments) with normalization to the primer usage of wild-type RT. h, RTIC (triangles), RTIC with NNRTI (circles) or vRNA–tRNA + excess RT (squares) reactions were initiated by addition of $\alpha$-32P-dTTP and quenched at different time points. Data were fit using the relationship for the free vRNA–tRNA + RT reaction: Intensity = $A(1-e^{-kt_{pol}}) + B(1-e^{-kt_{slow}})$. Data were fit using the relationship for the RTIC (with or without NNRTI) reaction: Intensity = $B(1-e^{-kt_{slow}})$ where $A$ and $B$ represent the amplitude of the fast and slow processes, respectively, $k_{pol}$ is the apparent extension rate constant, and $k_{slow}$ is the rate of the slow process. The second relationship was used for the RTIC data, as the slow process appears to dominate incorporation when the vRNA–tRNA substrate is crosslinked to RT. The best fits were obtained with: $A = 0.7166$ AU, $k_{pol} = 0.1078$ s$^{-1}$, $B = 0.2754$, $k_{slow} = 0.01002$ s$^{-1}$ for the vRNA–tRNA + excess RT; $B = 0.9808$, $k_{slow} = 0.003140$ s$^{-1}$ for the RTIC; and $B = 1.095$, $k_{slow} = 0.0001714$ s$^{-1}$ for the RTIC with NNRTI. $k_{slow}$ is about 3.19 times slower for crosslinked RTIC than for un-crosslinked components. Assays were independently repeated three times to ensure reproducibility.
Extended Data Fig. 2 | Representative negative-stain EM images, cryo-EM images, and 2D averages of the RTIC. 
a. Representative negative-stain EM image of HIV RTIC reveals a mono-disperse sample that is free of aggregates. Approximately a dozen images were taken of each sample before cryo-EM grid preparation to ensure sample quality.
b. Cryo-EM image of RTIC without β-OG. The long chains correspond to RNA from the complex with very few particles resembling the protein.
c. Cryo-EM image of RTIC with β-OG. Single particles corresponding to the complex appear similar to the negative-stain visualization. All 5,107 images used in both cryo-EM datasets have a similar appearance with slight differences in particle density.
d. Representative 2D averages of RTIC complex from the cryo-EM data collected with β-OG. Both datasets exhibit very similar 2D classes.

Results are reproducible in the absence of β-OG (>10 samples tested).
Extended Data Fig. 3 | Data processing workflow for RTIC complex.

**a.** Data processing workflow for the 8.0 Å global and 4.5 Å core maps.

**b.** Gold standard FSC curve of RTIC core and global maps.

**c.** The final 4.5 Å map is coloured according to local resolution estimated by Relion.

**d.** Angular distribution of particle projections. The length of each projection direction is proportional to the number of assigned particles.

**e.** Data processing workflow for the 8.2 Å global Mg²⁺ map.

**f.** Gold standard FSC curve of RTIC Mg²⁺ global map.
Extended Data Fig. 4 | Quality of the cryo-EM density for the core RTIC map. a, View of HIV-1 RT from the front. The subdomains of RT are coloured. Underneath the main RTIC view, each subdomain of RT, plus the p51 subunit, is shown fit into the 4.5 Å map. b, View of HIV-1 RT from the polymerase active site side. The subdomains of RT are coloured. Underneath the main RTIC view, each subdomain of RT, plus the p51 subunit, is shown fit into the 4.5 Å map. In a, b, regions of protein, namely loops and linkers, that lacked sufficient density were removed after comparison with previously published structures of RT. These regions are indicated by dotted lines and are most commonly found in the finger and palm subdomains. c, Representative regions of 4.5 Å map fitted with protein secondary structure that display densities for side chains. A view of the PBS helix fit into the 4.5 Å map is also shown; phosphates of the RNA backbone are partially resolved. Regions are coloured with respect to the main text models.
Extended Data Fig. 5 | Mg$^{2+}$ global map views and structure comparison. 

a. Side and top views of the 8.2 Å global map at different density thresholds. The orientation of the peripheral vRNA and tRNA elements is within the variability seen among the different RTIC conformers. 

b. A model of the RTIC built into the Mg$^{2+}$ density using the main text global RTIC model. vRNA and tRNA helices were treated as rigid bodies derived from main text model (see Extended Data Fig. 6 and Methods). 

c. Comparison of the global RTIC model RNA (grey) with the Mg$^{2+}$ model RNA (coloured). All three regions of RNA structure (H1, H2, and tRNA) differ in the Mg$^{2+}$ model, but are adequately described by rigid body movements of the RNA helical elements taken from the global RTIC model. Both H1 and H2 represent a substantial structural barrier to initiation. 

d. Partial accommodation of H1 into high monovalent salt classes 3, 4 and 7.
Extended Data Fig. 6 | Low-resolution tRNA density and fold comparison. a, Top and side views of the elongated helical tRNA density observed in the low-resolution global map of the RTIC. b, Top and side views of the vRNA–tRNA model generated using the hypothesized elongated tRNA helical fold. The tRNA model fits the long helical density well. Corresponding secondary structure is in d. c, Top and side views of the vRNA–tRNA model generated using previously hypothesized tRNA secondary structures that have the anticodon and D-stem loops independently folded. Corresponding secondary structure is in e. d, Secondary structure depiction of the new vRNA–tRNA and canonical clover-leaf fold of the tRNA. The different domains are coloured and correspond with the models in panels b and c. e, Secondary-structure depiction of the old vRNA–tRNA fold with independent anticodon and D-stem loops. The domains are coloured and correspond with the model in c and clover-leaf fold of the tRNA in d.
Extended Data Fig. 7 | Peripheral RNA heterogeneity of the RTIC conformers. a, Tiled views of eight conformations emerging from 3D classification of RTIC. Each class is numbered and class 7 was used for the global RTIC reconstruction. b, Superposition of the eight classes from a. The main areas of RNA heterogeneity are focused on the orientations of vRNA H2, H1 and the connection loop, and the tRNA. With no stabilizing protein contacts, vRNA H2, H1, and the tRNA sample a wide range of conformations, limiting the resolution of the global map. c, Additional RTIC models built into classes 3 (tan) and 4 (blue). The models for the tRNA, vRNA H1, and vRNA H2 were all derived from the global RTIC model and treated as rigid bodies for model building. The connecting loop was not built in these models as the density for this region was not clear in these maps, though there is reasonable density to model a loop near H1. Junctions between the helices serve as hinges that allow movement of the independent domains. The main text global RTIC model (grey) is included as a comparison. d, The vRNA and tRNA helices treated as rigid bodies for modelling are shown in bold. Hinge points for each helix are highlighted with grey circles and serve as points of flexibility for the RTIC.
Extended Data Fig. 8 | Single-molecule experimentation and analysis.

a, Secondary structure depiction of the vRNA–tRNA construct used for single-molecule experiments. The labelling scheme is shown, with the Cy3 dye located on the 5′ end of the vRNA helix 1 and Cy5 dye located on an oligonucleotide positioned near the 5′ end of helix 1. The vRNA–tRNA complex was crosslinked to RT for the experiments. b, Ninety-five per cent of the RTIC complexes are in the high FRET, helix 1 formation state (480 traces analysed, see Methods). c, Example trace of the ones used for final FRET analysis. The high FRET state of the RTIC complex, which is attributed to helix 1 formation. Photobleaching events for both Cy5 and Cy3 are indicated. d, Examples of traces removed from final FRET analysis. Traces exhibit the presence of multiple molecules (multiple single-dye photobleaching events) or poor dye behaviour (blinking and quenching).
Extended Data Fig. 9 | Comparison with NNRTI bound and active RT–nucleic acid complexes in the cryo-EM map. All alignments between structures and the RTIC were done using the p51 subunit. a, Comparison of an active conformation RT–nucleic acid structure (pink, 1RTD) with the RTIC core (RT, purple; tRNA primer, red; vRNA template, yellow). The EM map overlay shows the poor fit of the 1RTD model in the fingers, thumb, and primer grip of RT. Deviations of the nucleic acid primer and template of 1RTD away from the RTIC density are also apparent. b, Comparison of an NNRTI-bound RT–nucleic acid structure (dark grey, 3V81) with the RTIC core. The EM map overlay shows the closer fit of the fingers and primer grip regions of RT in the 3V81 model. The thumb region also overlays well, but with slight deviations. Most noticeably, the nucleic acid primer/template in the 3V81 model deviates, although not as dramatically as in 1RTD, from the RTIC core EM density.
Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics

|                        | RTIC Core (EMDB-7031) | RTIC Global (EMDB-7032) | RTIC Global w/MgCl₂ (EMDB-7540) |
|------------------------|------------------------|--------------------------|----------------------------------|
|                        | (PDB-6B19)             | (Supplementary Data file for model) |                                  |
| **Data collection and processing** |                        |                          |                                  |
| Magnification (calibrated) | 50,000                 | 50,000                   | 38,880                           |
| Voltage (kV)            | 300                    | 300                      | 200                              |
| Electron exposure (e⁻/Å²) | 60 and 85              | 60 and 85                | 75                               |
| Defocus range (µm)      | -1.3 to -2.5           | -1.3 to -2.5             | -2.0 to -3.0                     |
| Pixel size (Å)          | 1.0                    | 1.0                      | 1.286                            |
| Symmetry imposed        | C1                     | C1                       | C1                               |
| Initial particle images (no.) | 765,688                | 765,688                  | 148,523                          |
| Final particle images (no.) | 128,153                | 21,520                   | 67,346                           |
| Map resolution (Å)      | 4.5                    | 8.0                      | 8.2                              |
| FSC threshold           | 0.143                  | 0.143                    | 0.143                            |
| **Refinement**          |                        |                          |                                  |
| Initial model used (PDB code) | 3V81                   | 3V81                     | 3V81                             |
| Map sharpening B factor (Å²) | -250                   | -200                     | -200                             |
| Model composition       |                        |                          |                                  |
| Non-hydrogen atoms      | 5,299                  | 8,545                    | 8,545                            |
| Protein residues        | 909                    | 962                      | 962                              |
| RNA nucleotides         | 38                     | 178                      | 178                              |
| R.m.s. deviations       |                        |                          |                                  |
| Bond lengths (Å)        | 0.023                  | 0.003                    | 0.003                            |
| Bond angles (°)         | 1.603                  | 0.83                     | 0.83                             |
| Validation              |                        |                          |                                  |
| MolProbity score        | 2.41                   | 1.92                     | 1.92                             |
| Clashscore              | 15.00                  | 14.25                    | 14.25                            |
| Poor rotamers (%)       | N/A                    | N/A                      | N/A                              |
| Ramachandran plot       |                        |                          |                                  |
| Favored (%)             | 80.46                  | 96.15                    | 96.15                            |
| Allowed (%)             | 19.42                  | 3.85                     | 3.85                             |
| Disallowed (%)          | 0.12                   | 0.00                     | 0.00                             |

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a**
- **Confirmed**

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
- Clearly defined error bars
  - *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

| Data collection       | SerialEM, Leginon, Gatan DigitalMicrograph |
|-----------------------|--------------------------------------------|
| Data analysis         | Relion, CTF2KINDA, GCTF, MotionCor2, VIPER, EMAN2, Phenix, Coot, Rosetta, UCSF Chimera, Igor Pro, Prism 7, Situs, Matlab. |

For manuscripts utilizing custom algorithms or software that are not central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. Git-ub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Cryo-EM maps of the global RTIC, the core of the RTIC, and the global RTIC w/MgCl2 have been deposited in the Electron Microscopy Data Bank under accession codes EMDB-7032, EMDB-7031, and EMDB-7540. The coordinates of the RTIC core model have been deposited in the Protein Data Bank under accession code 6B19. The global RTIC model is available as Supplementary Data. All other data are available from the corresponding author upon reasonable request.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Final sample size for main text cryo-EM dataset was determined when further addition of data did not improve the quality of final maps. Final sample size for the MgCl₂ cryo-EM dataset was determined when the resulting map provided sufficient resolution for comparisons with the main text cryo-EM maps. Single-molecule data was collected as 6000 images (3000 in each fluorescent channel). After analysis of this dataset it was decided that further data collection was not necessary.

Data exclusions

Particles for cryo-EM data processing were excluded in accordance with common practices used during 2D and 3D classification. For single-molecule FRET analysis, colocalization scripts were employed to select only traces that exhibited both donor and acceptor dye fluorescence under donor-only excitation conditions. This colocalization eliminated singly labeled molecules. Additional manual analysis of this dataset was done to remove traces that exhibited multiple single-dye photobleaching events (multiple molecules) and/or poor dye photophysics. See Methods for more details.

Replication

All attempts at replication were successful.

Randomization

Randomization was not relevant to our study and would hamper analysis if employed.

Blinding

Blinding was not applicable to this study. Investigator involvement in data collection and analysis was required for proper cryo-EM and single-molecule data collection and processing.

Materials & experimental systems

Policy information about availability of materials

n/a Involved in the study

- Unique materials
- Antibodies
- Eukaryotic cell lines
- Research animals
- Human research participants

We require information from authors about some experimental systems used in the life sciences. Here, indicate whether each material or system listed is relevant to your study.

If you are not sure if a material or system applies to your research, read the appropriate section below before selecting a response.

Unique materials

Obtaining unique materials

All unique materials are readily available from standard commercial sources (N2-cystamine-dG: TriLink BioTechnologies).

Method-specific reporting

n/a Involved in the study

- ChIP-seq
- Flow cytometry
- Magnetic resonance imaging

We also require information from authors about specific methods used in the life sciences. Here, indicate whether each method listed is relevant to your study.

If you are not sure if a method applies to your research, read the appropriate section below before selecting a response.