Structure of HIV-1 RT/dsRNA initiation complex prior to nucleotide incorporation

Kalyan Das[a,b,c,d,1,2], Sergio E. Martinez[a,b,c,d,1], Jeffrey J. DeStefano[e,f], and Eddy Arnold[a,b,2]

[a]Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ 08854; [b]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854; [c]Rega Institute, The Katholieke Universiteit Leuven, 3000 Leuven, Belgium; [d]Department of Microbiology, Immunology and Transplantation, The Katholieke Universiteit Leuven, 3000 Leuven, Belgium; [e]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742; and [f]Maryland Pathogen Research Institute, University of Maryland, College Park, MD 20742

Edited by John M. Coffin, Tufts University, Boston, MA, and approved February 19, 2019 (received for review August 16, 2018)

The initiation phase of HIV reverse transcription has features that are distinct from its elongation phase. The first structure of a reverse transcription initiation complex (RTIC) that trapped the complex after incorporation of one ddCMP nucleotide was published recently [Larsen KP, et al. (2018) *Nature* 557:118–122]. Here we report a crystal structure of a catalytically active HIV-1 RT/dsRNA complex that mimics the state of the RTIC before the first nucleotide incorporation. The structure reveals that the dsRNA-bound conformation of RT is closer to that of RT bound to a nonnucleoside RT inhibitor (NNRTI) and dsDNA; a hyperextended thumb conformation helps to accommodate the relatively wide dsRNA duplex. The RNA primer 3′ end is positioned 5 Å away from the polymerase site; however, unlike in an NNRTI-bound state in which structural elements of RT restrict the movement of the primer, the primer terminus of dsRNA is not blocked from reaching the active site of RT. The observed structural changes and energetic cost of bringing the primer 3′ end to the priming site are hypothesized to explain the slower nucleotide incorporation rate of the RTIC. An unusual crystal lattice interaction of dsRNA with its symmetry mate is reminiscent of the RNA architecture within the extended vRNA–tRNA^{Lys3} in the RTIC. This RT/dsRNA complex captures the key structural characteristics and components of the RTIC, including the RT conformational changes and interactions with the dsRNA primer-binding site region, and these features have implications for better understanding of RT initiation.

reverse transcription initiation | RTIC | double-stranded RNA structure | DNA polymerase | HIV reverse transcriptase

HIV-1 reverse transcriptase (RT) is responsible for copying the retroviral single-stranded RNA (ssRNA) genome into viral double-stranded DNA (dsDNA). While copying the ∼10-kb HIV-1 ssRNA genome to make the viral dsDNA, RT uses RNA- and DNA-dependent DNA polymerization to incorporate ∼20,000 nucleotides. During most of the dsDNA synthesis, RT binds a dsDNA or an RNA–DNA template–primer (1, 2). The mechanism of reverse transcription initiation, however, is different from that of the elongation phase because initiation requires an external primer. Like other retroviruses, HIV-1 recruits a specific host tRNA that anneals to the primer-binding site (PBS) of the viral genome (3). The PBS sequence is different for different members of the retrovirus family, and thereby, a specific retrovirus recruits a specific cognate tRNA. HIV recruits tRNA^{Lys3} to initiate DNA polymerization, and tRNA^{Lys3} is packaged into newly assembled viruses (4, 5).

The 5′ untranslated region (UTR) of the HIV genome contains the PBS sequence, which is 18 nucleotides long. The PBS sequence is complementary to the 3′ end of tRNA^{Lys3}, and both anneal to form an 18-mer dsRNA. The annealed 18-mer dsRNA should bind the nucleic acid-binding cleft of RT to start minus-strand DNA synthesis from the 3′ end of the tRNA^{Lys3}. Apart from 18-mer dsRNA, additional structural elements of tRNA and vRNA interact to form a compact complex (6, 7). The overall structure of the vRNA–tRNA^{Lys3} complex and its interactions with RT beyond the PBS–tRNA duplex appear to be critical for stabilizing the RT initiation complex (RTIC) (8) and for reverse transcription initiation. Other tRNAs, when mutated to have sequence complementarity with PBS, are not efficient for HIV-1 reverse transcription initiation (9, 10). Posttranscriptional modifications of tRNA^{Lys3} have been reported to affect the formation of transcription initiation-competent vRNA–tRNA^{Lys3} complexes (11, 12). Nucleotide incorporation for initiation is almost 50-fold slower than that for reverse transcription elongation, and the processivity of RT during elongation is four orders of magnitude higher in comparison with initiation (13). The published kinetic studies indicate that slower nucleotide incorporation by the RTIC could be due to a different structural state of the initiation complex compared with elongation complexes and the lower binding affinity of RT to dsRNA. The structural state might be responsible for a lower rate of catalytic incorporation, and the lower binding affinity of dsRNA to RT would reduce the stability of the RTIC compared with RT elongation complexes. However, the stability and catalytic efficiency of an RTIC in vivo might be improved by the involvement of HIV nucleocapsid (NC) protein (14, 15), and other viral and cellular components.

Structures of HIV-1 RT have been studied extensively for the past 25 y since the first RT structures were reported (16, 17). Structures of RT have revealed various functional and conformational states that have helped in (*i*) explaining multiple unique functionalities of this enzyme, (*ii*) understanding the molecular mechanisms of drug resistance, and (*iii*) assisting drug design (18).

Significance

RT is a key enzyme in the life cycle of HIV, and is targeted by multiple antiviral drugs. Although for most of its function RT binds a dsDNA or RNA–DNA template–primer substrate, initiation of reverse transcription involves binding of dsRNA. The current study presents a structure of an RT/dsRNA complex that has the basic components of a reverse transcription initiation complex (RTIC). The unique structural features help understand the significantly slower rate of nucleotide incorporation by an RTIC compared with a catalytically efficient reverse transcription elongation complex. This complex may help in designing new experiments for understanding the intricate and slow process of reverse transcription initiation.

Author contributions: K.D. and E.A. designed research; K.D., S.E.M., J.J.D., and E.A. performed research; S.E.M. and J.J.D. contributed new reagents/analytic tools; K.D. and E.A. analyzed data; and K.D. and E.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.wwpdb.org (PDB ID code 6HAK).

[1]K.D. and S.E.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: kalyan.das@kuleuven.be or arnold@cabm.rutgers.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1814170116/-/DCSupplemental.

Published online March 22, 2019.

Structural study of the RTIC has been challenging, and the recent outstanding single-particle cryo-EM structure by Puglisi and coworkers (19) provided the first glimpse of RTIC architecture. A modified $tRNA^{Lys3}$ containing a cross-linkable N2-thioalkyl guanine (G) base at the sixth position from the 3′ end (20) was annealed to the PBS region of a 101-nucleotide long 5′ UTR section of vRNA, and the thioalkylguanine of the annealed tRNA–vRNA formed a disulfide cross-link with Q258C mutant RT, yielding a homogenous and stable RTIC complex for structural study. The cryo-EM study generated an experimental density map [Electron Microscopy Data Bank (EMDB)-7032] for the RT/vRNA–$tRNA^{Lys3}$ complex at 8-Å resolution and a 4.5-Å resolution density map (EMDB-7031) for the core RT/dsRNA (PBS region); a final model containing a dsRNA and an RT poly-Ala backbone was generated (PDB ID 6B19) by fitting the structural elements of RT to the 4.5-Å map. The structures revealed the basic architecture and several important features of the RTIC; however, structural and functional studies of different states of the RTIC are required for gaining broader insights into the mechanism of reverse transcription initiation.

The cross-linking of dsRNA to RT was performed with primer extension (19), i.e., the designed RTIC is catalytically active. Given that the protein–RNA cross-link should not permit translocation, the incorporated 2′,3′-ddCMP was expected to occupy the position that corresponds to the nucleotide-binding site (N site) (Fig. 1*A*). The published RTIC structure revealed that the ddCMP was incorporated at the primer 3′ end; unexpectedly, however, the primer 3′ end is positioned away from both the priming (P) and N sites. One important question is: How can the RTIC incorporate the first nucleotide if the primer end is not positioned for catalysis? Here we report a crystal structure of an HIV-1 RT/dsRNA catalytic complex at 3.95-Å resolution in which the 3′ end of the primer RNA strand is expected to be at the P site (Fig. 1*B*), a state competent for incorporation of the first nucleotide.

Results and Discussion

Structural Features of the RT/dsRNA Complex. We designed a 23-mer template/17-mer primer dsRNA to mimic the PBS region of the duplex in the vRNA–$tRNA^{Lys3}$ complex. The dsRNA has a cross-linkable thioalkyl-modified G base on the template opposite to the sixth base from the primer 3′ end (Fig. 1*B*). Compared with the PBS–$tRNA^{Lys3}$ duplex sequence, the C and G nucleotides for the sixth base pair were switched between the template and primer for template cross-linking. The dsRNA was cross-linked to RT through a disulfide bond between the modified G and the mutated thumb residue C258 in the Q258C mutant RT (20, 21). The cross-linked complex was shown to be catalytically competent for polymerization by incorporating α-^{32}P dCTP (*SI Appendix*, Fig. S1). The RT/dsRNA complex crystallized in a new crystal form with $P4_32_12$ space group symmetry, in contrast to the usual $P2_1$ crystal form that we routinely obtain for the same Q258C mutant RT cross-linked to a dsDNA or an RNA–DNA substrate (22, 23). The RT/dsRNA complex was crystallized with

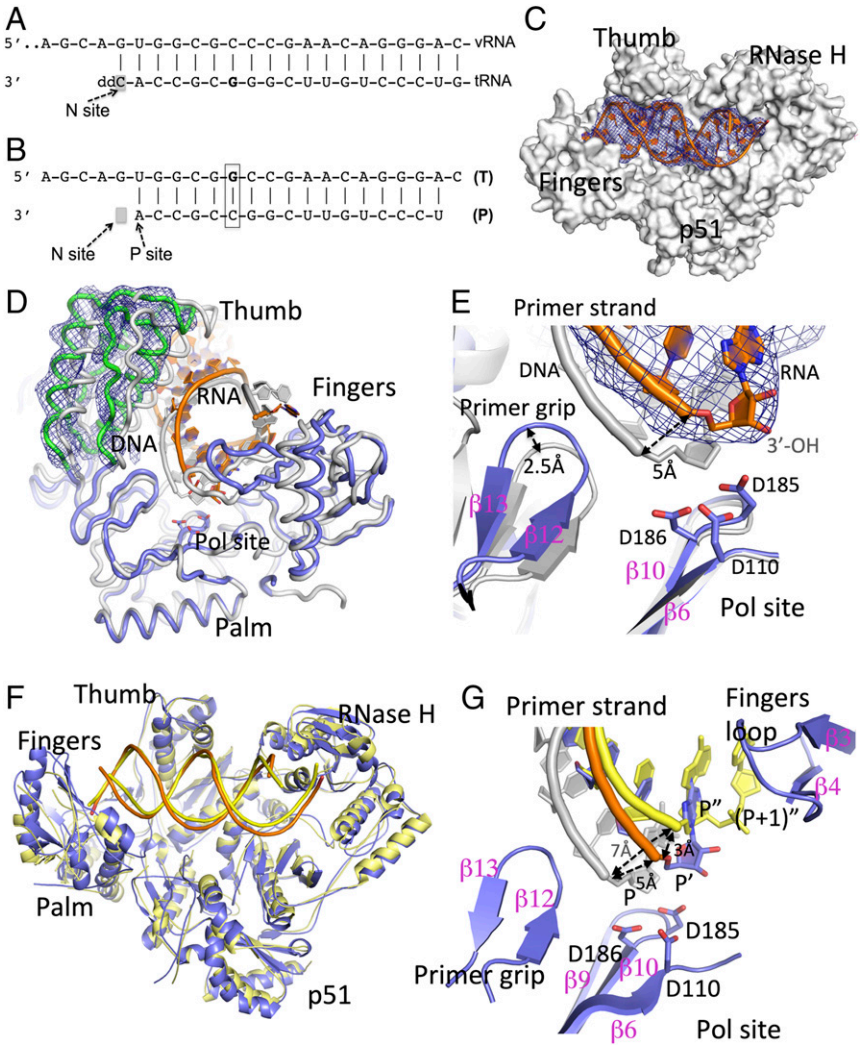

Fig. 1. Structural features of the RT/dsRNA complex. (*A*) Nucleotide sequence of the annealed PBS region of vRNA–$tRNA^{Lys3}$ in the RTIC for which the cryo-EM structure was reported (19). The vRNA–$tRNA^{Lys3}$ was cross-linked to RT at "**G**" and primer extended with catalytic incorporation of a ddCMP. (*B*) Nucleotide sequence of dsRNA used in the current study; P, primer strand and T, template. The template **G** was cross-linked to the Q258C position of RT without primer extension; the RNA primer 3′ end is expected to be at the priming (P) site. For the template cross-linking, the G and C positions of the sixth base pair were switched compared with the naturally annealed PBS–$tRNA^{Lys3}$ sequence. (*C*) The 3.95-Å resolution (2|Fo|−|Fc|) electron density map contoured at 1.4σ defines the track of dsRNA in the binding cleft of RT. (*D*) Superposition of the RT/dsRNA structure (blue protein, green thumb, and orange RNA) with the RT/dsDNA reverse transcription elongation complex (PDB ID 5TXL; gray protein and DNA). The thumb of the RT/dsRNA structure, as defined by the electron density, is in a hyperextended position. (*E*) A close-up view of the polymerase active site shows that the 3′ end of the RNA primer is located about 5 Å away from the DNA 3′ terminus, which is at the P site; the primer grip is also repositioned. (*F*) Superposition of the RT/dsRNA structure (blue protein, orange RNA) on the structure of the RTIC core (PDB ID 6B19; yellow protein and RNA) shows high structural resemblance. (*G*) Relative locations of the primer 3′ ends: P′ of RT/dsRNA (blue/orange), (P + 1)″ of cryo-EM RTIC N complex (yellow), and P of RT/dsDNA (gray) structures. The incorporated ddCMP is at (P + 1)″ and the nucleotide corresponding to P′ of RT/dsRNA is located at P″ in the RTIC structure. The primer 3′ end has to reach the P site for nucleotide incorporation by RT.

two copies of the complex in the asymmetric unit; however, copy 1 (with chain IDs A, B, T, and P for p66, p51, template, and primer, respectively) has better electron density for the dsRNA (Fig. 1*C* and *SI Appendix*, Fig. S2), and all of the analyses are based on copy 1 unless otherwise stated. The X-ray data and refinement statistics are listed in *SI Appendix*, Table S1.

The dsRNA has significant conformational differences with dsDNA or RNA–DNA in their RT-bound states. RT conformationally adapts to accommodate different types of nucleic acid duplexes: dsRNA, dsDNA, or RNA–DNA. In the RT/dsRNA structure, the p66 thumb subdomain is in a hyperextended state (Fig. 1*D*) compared with RT elongation complexes (represented by the structure of RT/dsDNA/dATP complex; PDB ID 5TXL). Unexpectedly, the 3′ end of the RNA primer is located at a position P′, which is about 5 Å away from the P site of an RT elongation complex (Fig. 1*E*). The two independent copies of RT/dsRNA in the asymmetric unit are highly superimposable, and also superimpose well on the 4.5-Å resolution cryo-EM structure of RT/dsRNA (N complex); the superimposed thumb subdomains are at a hyperextended position, and the dsRNAs follow a very similar track (Fig. 1*F*). However, the nucleotide corresponding to the RNA primer 3′ end is at the P″ position in the cryo-EM RT/dsRNA (N complex); the P″ position is about 3 Å away from the P′ position and about 7 Å away from the P site (Fig. 1*G*). Neither the primer-terminal nucleotide at the P″ nor at the P′ position would permit nucleotide incorporation, and the primer 3′ end has to move to the P site. This mispositioning of the primer terminus away from the P site may correlate with the significantly slower rate of incorporation of a nucleotide by the initiation complex vs. an elongation complex of RT, as discussed later.

Interactions Between dsRNA and RT. The RNA duplex has adopted largely A-form geometry and has less conformational flexibility than dsDNA and RNA–DNA duplexes. An overlay of dsRNA, dsDNA, and RNA–DNA duplexes based on the alignment of RNase H domains shows two distinct conformations and paths (Fig. 2*A*; stereoview in *SI Appendix*, Fig. S3*A*): (*i*) dsRNAs in the RT/dsRNA and RTIC core structures represent the reverse transcription initiation state, and (*ii*) the RNA–DNA and dsDNA-bound structures including the nevirapine-inhibited RT/dsDNA/nevirapine complex represent the elongation state. Based on the RNase H superposition, it is clear that RT undergoes a significant structural rearrangement to accommodate the dsRNA compared with an elongation substrate; the primer grips are about 8 Å apart and the polymerase active sites are about 7 Å apart when dsRNA-bound RT and dsDNA-bound RT are aligned based on the RNase H superposition.

To assess the extent of RT interactions with the different nucleic acid duplexes along the binding cleft, we plotted the number of interatomic distances (≤4.5 Å) between each nucleotide and RT in the RT/dsRNA complex (Fig. 2*B*) and compared them with those in the RT/dsDNA/nevirapine, RT/dsDNA, and RT/RNA–DNA complex structures (Fig. 2 *C*–*E*). This analysis reveals that the 5′ end of the RNA template has the most interactions with the p66 fingers subdomain, which is consistent with the well-ordered electron density for this region (*SI Appendix*, Fig. S3*B*). The remaining portions of the dsRNA interact less extensively with RT compared with the interactions of nucleic acid duplexes in RT elongation complexes including the nevirapine-bound one. Some of the key differences in the RT/dsRNA complex structure are (*i*) no close interactions of dsRNA with the primer grip, (*ii*) relatively weak interaction of the template with the thumb-connection region (*SI Appendix*, Fig. S3*C*), and (*iii*) poor interaction of the RNA primer with the RNase H primer grip (*SI Appendix*, Fig. S3*D*). The local alignments in *SI Appendix*, Fig. S3 *C* and *D* show that the RT interacts less favorably with dsRNA compared with elongation substrates. A comparison of the relative B factors of RT-bound dsRNA vs. dsDNA also correlates with the extent of their interactions with RT by showing higher

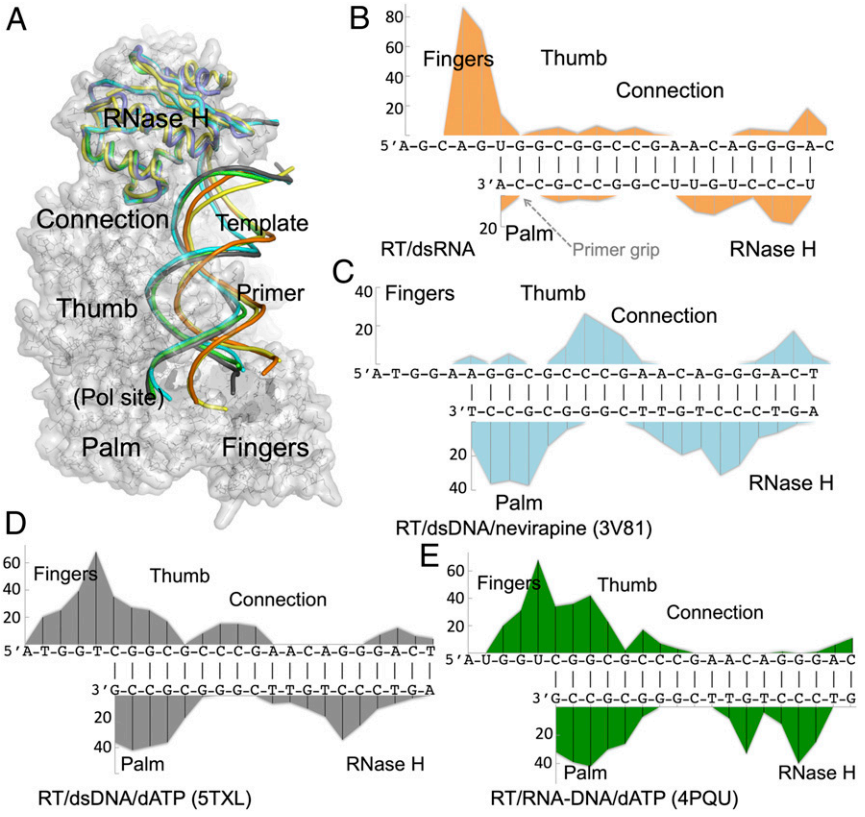

Fig. 2. Interactions of dsRNA with HIV-1 RT. (*A*) Superposition of the structures of the RT/dsRNA complex (blue RNase H, orange dsRNA), RT/dsRNA in the RTIC (PDB ID 6B19; yellow), RT/RNA–DNA/dATP (PDB ID 4PQU; green), RT/dsDNA/dATP (PBD ID 5TXL; gray), and RT/dsDNA/NNRTI (PDB ID 3V81; cyan) based on aligning the RNase H domains. The superposition shows two distinct nucleic acid tracks: (*i*) one for the RT/dsRNA and RTIC initiation complex structures, and (*ii*) the other for the remaining three structures that represent RNA- and DNA-dependent elongation complexes and NNRTI-inhibited elongation complex. Plot of number of interactions (nonhydrogen atom distances ≤4.5 Å) of individual nucleotides with RT in the RT/dsRNA (*B*), RT/dsDNA/nevirapine (*C*), RT/dsDNA/dATP (*D*), and RT/RNA–DNA/dATP (*E*) complexes. Sequence and number of distances for the template strand nucleotides are at the *Top* and for the primer strand nucleotides are at the *Bottom* in each plot; the *y* axes represent the number of interactions (distances ≤4.5 Å) of individual nucleotides with RT. Some of the important structural differences are shown in *SI Appendix*, Fig. S3.

B factors for most of dsRNA and lower B factors for the parts of dsDNA that interact with RT (*SI Appendix*, Fig. S4).

The Hyperextended Thumb of RT Is a Requirement for Initiation of Reverse Transcription. Pairwise comparisons of structures of RT complexes show that the RT conformation in the current RT/dsRNA complex closely resembles that of RT in RT/dsDNA/nonnucleoside RT inhibitor (NNRTI) complex (PDB ID 3V81; Fig. 3*A* and *SI Appendix*, Fig. S5); the p66 palm, thumb, and connection subdomains and most of p51 superimpose well, whereas, the fingers and RNase H deviate the most when the two structures are aligned. The RTIC cryo-EM publication (19) also reported the resemblance with the 3V81 structure, and consequently, the authors used the 3V81 structure as a template for fitting the RT model to the density (EMBD-7031).

RT has been known to require large domain movements for interconverting between different functional states (18). It has been shown in the past that (*i*) the thumb subdomain is folded into the nucleic acid-binding cleft in apo-RT structures, (*ii*) the thumb opens up to bind a dsDNA or an RNA–DNA duplex in an RT elongation complex, and (*iii*) binding of an NNRTI causes hyperextension of the thumb. A highly superimposable palm–thumb region of RT/dsRNA and RT/dsDNA/nevirapine complexes suggests that RT may need to attain a state with the hyperextended thumb as a requirement for the formation of the RTIC. The NNRTI-binding pocket region is open the most in the RT/dsRNA complex compared with the pocket region in any other RT structure without a bound NNRTI. In general, displacement of the primer grip and flipping of the Y181 and Y188 side chains are associated with formation of the binding pocket for a typical NNRTI. In the RT/dsRNA structure, the primer grip has an open conformation, and the side chains of Y181 and Y188 are disordered (*SI Appendix*, Fig. S6); most aromatic residues (F, Y, and W) in the structure have associated density. A Cα superposition of the palm subdomains of RT/dsRNA and 3V81 suggests that repositioning of the W229 side chain is the major adjustment required to accommodate the NNRTI nevirapine in the pocket of the RT/dsRNA complex (Fig. 3*B*). Apart from dsRNA-bound RT structures, no functional complex of RT has a hyperextended thumb conformation. Despite high resemblance of the RT conformations between RT/dsRNA and RT/dsDNA/nevirapine complexes, the nucleic acid tracks and interactions (Fig. 2 and *SI Appendix*, Fig. S5) are significantly different in the two structures. The primer 3′ ends in both structures do not reach the polymerase active site and are also about 3.5 Å apart from each other. The paths of the nucleic acids leading from the superimposed thumbs to their respective primer 3′ ends are significantly different in the two structures (Fig. 3 and *SI Appendix*, Fig. S3*C*) and, as discussed below, this difference may help explain why the RTIC is capable of incorporating nucleotides, whereas an RT/dsDNA/NNRTI complex is not.

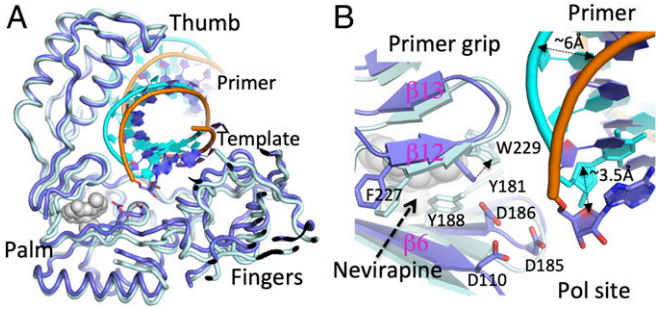

Fig. 3. dsRNA-bound RT closely resembles RT in the RT/dsDNA/nevirapine complex. (*A*) The polymerase domains of the RT/dsRNA and RT/dsDNA/nevirapine structures as aligned in *SI Appendix*, Fig. S5 shows high structural resemblance of palm and thumb subdomains, and an almost open NNRTI pocket in the RT/dsRNA structure despite no inhibitor being present. However, the tracks of the nucleic acids leading from thumb to primer 3′ ends are significantly different. (*B*) A close-up view of the polymerase site and NNRTI pocket shows that the primer 3′ ends in both structures are displaced away from the polymerase site, and also about 3.5 Å apart. Only minimal structural arrangements of the NNRTI pocket region of RT/dsRNA structure are needed to accommodate nevirapine; the major adjustment required is the repositioning of W229 as indicated by an arrow. The side chains of Y181 and Y188 have disordered electron density (*SI Appendix*, Fig. S6), and therefore are not included in the RT/dsRNA structure.

Primer Grip Interaction and Positioning of the RNA Primer 3′ End for Nucleotide Incorporation. A common structural state represented by the RTIC core (19) and two independent copies of RT/dsRNA complex in the crystallographic asymmetric unit, having significantly different crystal contacts, may be attributed to a functional state. However, the mispositioning of the primer 3′ end with respect to the catalytic triad (D110, D185, and D186) in these structures apparently is due to the large conformational difference of a dsRNA compared with a typical dsDNA or RNA–DNA template–primer elongation substrate of RT. The conserved "primer grip" plays a critical role by interacting with the −2 nucleotide (analogous to the nucleotide position 75 of $tRNA^{Lys3}$ in the RTIC) of the primer adjacent to the polymerase site, and the interaction is essential for catalytic incorporation of a nucleotide and for translocation of nucleic acid that permits incorporation of the following nucleotide. The primer grip maintains its contacts with the primer in all RT elongation complex structures, including the nonproductive nevirapine-bound RT/dsDNA structure (Fig. 2). In the RT/dsDNA/nevirapine complex, the primer terminus is shifted away from the catalytic triad by repositioning of the primer grip (22). The −2 primer nucleotide backbone and the conserved residues M230 and G231 of the primer grip are less than 4 Å apart in RNA–DNA and dsDNA-bound RT elongation complexes (Fig. 4*A*) and in the RT/dsDNA/nevirapine complex (Fig. 4*B*). Unexpectedly, this distance is about 7 Å in the RT/dsRNA complex structure (Fig. 4*C*), suggesting that the dsRNA-bound conformation of RT is not optimally configured to position the RNA primer 3′ end at the P site for nucleotide incorporation.

Due to a lack of stabilizing interactions between the RNA primer and the primer grip, the RNA primer 3′ end is likely to have higher mobility. Also, the primer grip is known to be highly dynamic in terms of its position relative to the polymerase site (24), and it is conceivable that the primer grip can occupy multiple locations in the RTIC. In configurations where the RNA primer and primer grip develop favorable interactions analogous to those observed in RT elongation complexes, the primer grip could position the RNA primer terminus in a catalytically competent fashion. Incorporation of ddCMP into the RTIC (19), apparently would have involved such a structural arrangement.

The factors such as a faster rate of dissociation of RT from vRNA–$tRNA^{Lys3}$ as suggested by single-molecule studies (25) appear to contribute to the slower rate of nucleotide incorporation by the RTIC. It is likely that the structural state (Fig. 1*F*) observed for RT/dsRNA and RTIC leads either to a catalytic complex in the presence of dCTP and Mg^{2+} ions or to dissociation, considering that an RTIC dissociates ∼200 times faster than an RT elongation complex (13). The positioning of the RNA primer 3′ end away from the polymerase active site as observed in RT/dsRNA structures may also account for the slower rate of nucleotide incorporation by the RTIC. Following the incorporation of the first six nucleotides, the nucleic acid stretch between the thumb and the polymerase site (*SI Appendix*, Fig. S3*C*) becomes an RNA–DNA duplex with a conformation similar to that in an RNA-dependent DNA polymerization elongation complex; the initiation complex switches to an elongation complex with ∼3,400-fold increase in nucleotide incorporation efficiency

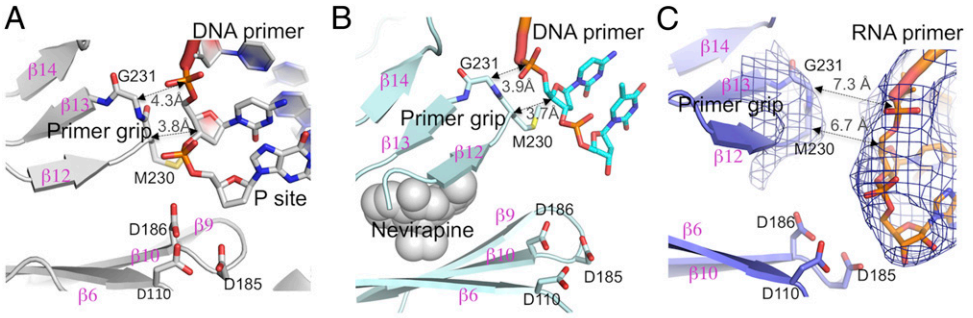

Fig. 4. Positioning of the primer 3′ end and the primer grip. (*A*) The primer grip in an RT/dsDNA elongation complex structure interacts with the DNA primer strand and helps position its 3′ end at the P site. (*B*) The interaction between the primer grip and DNA primer remains intact in the RT/dsDNA/nevirapine structure; however, the bound nevirapine molecule restricts the primer grip to a hyperextended position, and consequently the primer 3′ end is locked in a nonproductive position for polymerization. (*C*) The 3′ end of the RNA primer is located away from the P site, the primer grip has no noticeable interaction with the RNA primer, and the primer grip is at a pocket-open position, however not locked. The (2|Fo|−|Fc|) electron density map defines the relative positions of the primer grip and primer strand in the RT/dsRNA complex. Unlike in the RT/dsDNA/nevirapine structure, the structural flexibility of the primer grip and primer strand would permit the RNA 3′ end to reach the P site for nucleotide incorporation in an RTIC complex.

(26). A systematic study of the structures following the successive incorporation of each of the first six DNA nucleotides will elucidate the conformational switching from the initiation to elongation state of RT.

In contrast to the primer terminus repositioning in an RTIC as discussed, the position of the primer terminus in the RT/dsDNA/nevirapine complex is constrained by the combination of (*i*) restricted primer grip position by a bound NNRTI and (*ii*) sustained interaction between primer and primer grip (Fig. 4). As a consequence of bound NNRTI, the locked primer grip acts as a structural barrier that prevents the primer 3′ end from reaching the P site for nucleotide incorporation (22).

Low Binding Affinity of dsRNA with HIV-1 RT. Experimentally, formation of a stable RTIC or RT/dsRNA complex for structural study has only been successful by cross-linking the RNA to the RT p66 thumb at the middle of the nucleic acid-binding cleft (*SI Appendix*, Fig. S7*A*). In fact, in another attempt to obtain the structure of RT/dsRNA, we cross-linked the template RNA overhang with the amino acid residue C63 on the fingers subdomain of I63C mutant RT. The structure of a catalytic elongation complex of I63C mutant RT cross-linked to the template overhang of a dsDNA was determined successfully at 2.5-Å resolution (27). An analogous experiment with a dsRNA led successfully to the formation of an RT/dsRNA cross-linked product (*SI Appendix*, Fig. S7*B*), and the complex crystallized in a new P1 crystal form that diffracted to 4-Å resolution. However, the structure did not show binding of the dsRNA in the nucleic acid-binding cleft (*SI Appendix*, Fig. S7*C*); the electron density clearly showed the thumb position closed down and occupying the cleft. Based on the analysis of RT–nucleic acid interactions in different RT structures (Fig. 2), it appears that the fingers cross-linking with the template overhang would not enhance the relatively weak interactions of dsRNA with RT in the cleft. In contrast, the thumb cross-link places the dsRNA within the cleft and the weak RT–dsRNA interactions define the track of the dsRNA in the cleft. Significant differences in the binding of thumb cross-linked dsRNA vs. thumb cross-linked dsDNA or RNA–DNA imply that the minor groove cross-linking to the sixth base with a flexible linker has no or minimal impact on the nucleic acid interactions with RT that predominantly involve the sugar-phosphate backbone, and that the cross-linked complexes are functionally relevant.

Importance of tRNA–vRNA Interactions Beyond the PBS Duplex in the RTIC. In the current structure of RT/dsRNA cross-linked to Q258C RT, the new crystal form exhibits noteworthy crystal lattice interactions. Between the two copies of the RT/dsRNA complex in the asymmetric unit of the crystal, the dsRNA of copy 1 has better density and lower B factors compared with weaker density and higher B factors for copy 2 dsRNA (*SI Appendix*, Fig. S2). Analysis of the crystal symmetry contacts revealed an unusual weak dsRNA:dsRNA interaction involving sugar-phosphate backbone atoms only (*SI Appendix*, Fig. S8*A*), where the minimum distance between any two atoms from the symmetry-related dsRNAs is >5 Å. This type of interaction has not been observed in crystal packing of any RT complex. We placed the copy 1 RT/dsRNA and the interacting symmetry-related dsRNA into the published 8-Å cryo-EM density of the RTIC (EMDB-7032) by fitting the RT/dsRNA complex to the density map. Interestingly, the dsRNA (symmetry) occupied the density that is located just above the RT/dsRNA part in EMBD-7032 (*SI Appendix*, Fig. S8*B*); this piece of density has been assigned to helix 1 (H1) and the connecting loop of vRNA (19). The observed RNA:RNA interaction in the current structure is a surprise mimic of the interaction between two structural segments within the vRNA–$tRNA^{Lys3}$ complex. This distance is also analogous to that observed in ternary folds of tRNA structures, in which diffuse Mg^{2+} ions appear to stabilize the negatively charged backbones in close proximity (28). Apparently, diffuse Mg^{2+} ion-mediated RNA:RNA backbone interaction facilitates the crystal packing in the RT/dsRNA structure. This type of RNA:RNA interaction may contribute to form a compact vRNA–$tRNA^{Lys3}$ structure and improve the binding of the PBS stretch of dsRNA duplex to RT in the RTIC (29, 30). Posttranscriptional modifications influence folding and structures of tRNAs (31), and various posttranscriptional modifications of $tRNA^{Lys3}$ and extended template–primer interactions are important for efficient initiation and the transition from initiation to elongation of HIV-1 reverse transcription (32). Future studies may reveal how the interplay among the structural elements helps in stabilizing an intact structure for vRNA–$tRNA^{Lys3}$ to form the RTIC.

Methods

RT–dsRNA Cross-Linking and Purification. The RT construct RT127A containing Q258C and D498N in the p66 subunit and C280S in both p66 and p51 subunits was used for cross-linking with a dsRNA. This RT construct has been used for structural studies of RT elongation complexes with dsDNA and RNA–DNA and also for the RT/dsDNA/nevirapine complex. RT127A was expressed and purified as previously reported (22). Briefly, the RT was expressed in BL21-CodonPlus-RIL cells, induced with 1 mM isopropyl β-D-1-thiogalactopyranoside at an A_{600} of 0.9, followed by expression at 37 °C for 3 h. The cells were sonicated at a power output of ∼45 W with a Misonix 3000 sonicator. The samples were purified using a nickel-nitrilotriacetic acid (Ni-NTA) column according to the manufacturer's recommendations (Qiagen). The final purification step was carried out using a Mono Q column, and the purified RT samples were buffer exchanged into 10 mM Tris pH 8.0, 75 mM NaCl.

The dsRNA (Fig. 1*B*) has the sequence of the vRNA (PBS)–tRNA duplex that is required to bind RT as a part of an RTIC. A 23-mer RNA template (5′-AGCAGUGGCGGCCGAACAGGGAC) bearing a cross-linkable thioalkyl tether

(on G) was custom synthesized by Roger Jones (Rutgers University). A complementary 17-mer RNA primer was purchased from Integrated DNA Technologies. The template and primer strands were annealed to form a dsRNA. A 50-μL volume containing 28.2 nmols of 23-mer RNA template and 31.0 nmols (10% molar excess) of 17-mer RNA primer in 50 mM NaCl and 50 mM Hepes pH 7.0 was put in a liter of water at 95 °C. This was allowed to cool, covered with an ice bucket for 4.5 h until the temperature reached 29 °C. The 23:17-mer dsRNA was cross-linked to RT127A at the mutated Q258C site of p66 (*SI Appendix*, Fig. S7*A*). The following cross-linking reaction was set up in a volume of 1,000 μL: 28.2 nmols of annealed RNA template–primer, 50 mM Hepes pH 7.0, 86 mM NaCl, 0.2 mM $MgCl_2$, 0.1 mM EDTA, 1 mM β-mercaptoethanol, 0.2 mM ddCTP, and 8.3 mg (71 nmols; 2.5-fold molar excess to template–primer) of RT127A; cross-linking of elongation complexes are typically done in the presence of 5 mM $MgCl_2$ and no EDTA. The RT127A was first combined with EDTA, then the other components were added, and the template–primer added last. The mixture was put in a 37 °C water bath for 6 h, and then at 4 °C. Unlike for the formation of previous elongation complexes, no primer extension was observed in the crystal structure of the RT/dsRNA complex due to a considerably low concentration of Mg^{2+} ions and the presence of EDTA. The cross-linked RT/dsRNA complex was purified using Ni-NTA and heparin columns in tandem. An Amicon Ultra-4 Ultracel unit (30-kDa cutoff) was used to exchange the buffer into 75 mM NaCl, 10 mM Tris·HCl, pH 8.0. The protein was concentrated to 10 mg/mL. The final yield of material ready for crystallization was 69 μg. The N-terminal His tag on p51 was not cleaved to avoid further losses of the small amount of complex. An attempt was made to generate an analogous cross-linked complex with a construct identical to RT127A, but with I63C (fingers) (27) instead of Q258C (thumb). The 23-mer template was the same except (*i*) the fourth template base from the 5′ end (adenine) contained a two-carbon cross-linker at N^6, and (*ii*) the G and C at the sixth position from the primer 3′ end are as in the natural PBS sequence. The cross-linking was successful (*SI Appendix*, Fig. S7*B*), the thumb was down and dsRNA appeared to be outside of the nucleic acid-binding cleft and disordered in the crystal structure (*SI Appendix*, Fig. S7*C*).

Crystallization and Data Collection. Sitting drops of 0.5 μL of protein plus 0.4 μL of well solution were set up at 4 °C on Combiclover Jr. plates from Emerald BioStructures. The well solution contained 10–12% PEG 8000 (wt/vol), 50 mM Bis-Tris propane pH 6.8–7.2, 100 mM $(NH_4)_2SO_4$, 5% sucrose (wt/vol), 5% glycerol (vol/vol), and 20 mM $MgCl_2$. The crystals were cryoprotected by dipping in crystallization solution with glycerol concentration elevated to 10%, 15%, and then 20%. An X-ray diffraction dataset was collected from one crystal at the F1 beamline, Cornell High Energy Synchrotron Source (CHESS). The data were processed and scaled using Mosflm and Aimless, respectively, as implemented in CCP4 (33). RT/dsRNA in the current study crystallized in a lattice with $P4_32_12$ space group symmetry compared with a $P2_1$ crystal system for all RT127A elongation complexes, and a P1 crystal form of I63C RT/dsRNA complex (*SI Appendix*, Fig. S7*C*). Structure determination revealed that a dsRNA:dsRNA crystal lattice interaction appears to be responsible for this change in space group. The structure was solved by molecular replacement using the RT model from structure 5XTL as the starting model for the program Phaser as implemented in the Phenix suite of programs (34). The positions and orientations of the protein subdomains/substructures were optimized by rigid body refinement. Cycles of model building and refinement using Coot (35) and Phenix, respectively, developed clear density for dsRNA in copy 1 of the two copies of the complexes in the asymmetric unit. Inclusion of the dsRNA (copy 1) and further refinement and rebuilding improved the difference density for the second dsRNA (copy 2). The structure figures were generated using PyMOL. The structure factors and coordinates are deposited in PDB with accession code 6HAK, and the data processing and refinement statistics are in *SI Appendix*, Table S1.

ACKNOWLEDGMENTS. We thank Roger Jones (Rutgers University) for the cross-linkable RNA; Steven Tuske, Stefan Sarafianos, Anthony Hoang, and Dmitry Lyumkis for helpful discussions; and the CHESS F1 beamline facility for data collection. This work was supported by NIH MERIT Award R37 AI027690 and the HIVE Center Grant U54 GM103368 (both to E.A.); NIH Grant GM116645 (to J.J.D.); and KU Leuven start-up grant (to K.D.).

1. Katz RA, Skalka AM (1994) The retroviral enzymes. *Annu Rev Biochem* 63:133–173.
2. Hu WS, Hughes SH (2012) HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2:a006882.
3. Mak J, Kleiman L (1997) Primer tRNAs for reverse transcription. *J Virol* 71:8087–8095.
4. Gabor J, Cen S, Javanbakht H, Niu M, Kleiman L (2002) Effect of altering the tRNA(Lys) (3) concentration in human immunodeficiency virus type 1 upon its annealing to viral RNA, GagPol incorporation, and viral infectivity. *J Virol* 76:9096–9102.
5. Kleiman L, Jones CP, Musier-Forsyth K (2010) Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett* 584:359–365.
6. Isel C, et al. (1999) Structural basis for the specificity of the initiation of HIV-1 reverse transcription. *EMBO J* 18:1038–1048.
7. Bilbille Y, et al. (2009) The structure of the human tRNALys3 anticodon bound to the HIV genome is stabilized by modified nucleosides and adjacent mismatch base pairs. *Nucleic Acids Res* 37:3342–3353.
8. Coey AT, et al. (2018) Dynamic interplay of RNA and protein in the human immunodeficiency virus-1 reverse transcription initiation complex. *J Mol Biol* 430: 5137–5150.
9. Das AT, Klaver B, Berkhout B (1995) Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *J Virol* 69:3090–3097.
10. Dupuy LC, Kelly NJ, Elgavish TE, Harvey SC, Morrow CD (2003) Probing the importance of tRNA anticodon: Human immunodeficiency virus type 1 (HIV-1) RNA genome complementarity with an HIV-1 that selects tRNA(Glu) for replication. *J Virol* 77: 8756–8764.
11. Isel C, Marquet R, Keith G, Ehresmann C, 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.
12. Bénas P, et al. (2000) The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. *RNA* 6:1347–1355.
13. Lanchy JM, Ehresmann C, Le Grice SF, Ehresmann B, Marquet R (1996) Binding and kinetic properties of HIV-1 reverse transcriptase markedly differ during initiation and elongation of reverse transcription. *EMBO J* 15:7178–7187.
14. Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K (2010) Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* 7:754–774.
15. Seif E, Niu M, Kleiman L (2015) In virio SHAPE analysis of tRNA(Lys3) annealing to HIV-1 genomic RNA in wild type and protease-deficient virus. *Retrovirology* 12:40.
16. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA (1992) Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256: 1783–1790.
17. Jacobo-Molina A, et al. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci USA* 90:6320–6324.
18. Das K, Arnold E (2013) HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. *Curr Opin Virol* 3:111–118.
19. Larsen KP, et al. (2018) Architecture of an HIV-1 reverse transcriptase initiation complex. *Nature* 557:118–122.
20. Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 282:1669–1675.
21. Sarafianos SG, et al. (2003) Trapping HIV-1 reverse transcriptase before and after translocation on DNA. *J Biol Chem* 278:16280–16288.
22. Das K, Martinez SE, Bauman JD, Arnold E (2012) HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* 19:253–259.
23. 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: 8125–8137.
24. Das K, et al. (2011) Crystal structure of tert-butyldimethylsilyl-spiroaminooxathioledioxide-thymine (TSAO-T) in complex with HIV-1 reverse transcriptase (RT) redefines the elastic limits of the non-nucleoside inhibitor-binding pocket. *J Med Chem* 54:2727–2737.
25. Liu S, Harada BT, Miller JT, Le Grice SF, Zhuang X (2010) Initiation complex dynamics direct the transitions between distinct phases of early HIV reverse transcription. *Nat Struct Mol Biol* 17:1453–1460.
26. Lanchy JM, et al. (1998) Contacts between reverse transcriptase and the primer strand govern the transition from initiation to elongation of HIV-1 reverse transcription. *J Biol Chem* 273:24425–24432.
27. Martinez SE, Bauman JD, Das K, Arnold E (2019) Structure of HIV-1 reverse transcriptase/d4TTP complex: Novel DNA cross-linking site and pH-dependent conformational changes. *Protein Sci* 28:587–597.
28. Draper DE (2004) A guide to ions and RNA structure. *RNA* 10:335–343.
29. Beerens N, Groot F, Berkhout B (2000) Stabilization of the U5-leader stem in the HIV-1 RNA genome affects initiation and elongation of reverse transcription. *Nucleic Acids Res* 28:4130–4137.
30. Westerhout EM, Berkhout B (2007) A systematic analysis of the effect of target RNA structure on RNA interference. *Nucleic Acids Res* 35:4322–4330.
31. Helm M (2006) Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res* 34:721–733.
32. Isel C, et al. (1996) 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.
33. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67:235–242.
34. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
35. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.