

# Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry

Elżbieta Nowak<sup>1</sup>, Jennifer T Miller<sup>2</sup>, Marion K Bona<sup>2</sup>, Justyna Studnicka<sup>1</sup>, Roman H Szczepanowski<sup>3</sup>, Jakub Jurkowski<sup>1</sup>, Stuart F J Le Grice<sup>2</sup> & Marcin Nowotny<sup>1</sup>

Retrotransposons are a class of mobile genetic elements that replicate by converting their single-stranded RNA intermediate to double-stranded DNA through the combined DNA polymerase and ribonuclease H (RNase H) activities of the element-encoded reverse transcriptase (RT). Although a wealth of structural information is available for lentiviral and gammaretroviral RTs, equivalent studies on counterpart enzymes of long terminal repeat (LTR)-containing retrotransposons, from which they are evolutionarily derived, is lacking. In this study, we report the first crystal structure of a complex of RT from the *Saccharomyces cerevisiae* LTR retrotransposon Ty3 in the presence of its polypurine tract-containing RNA-DNA hybrid. In contrast to its retroviral counterparts, Ty3 RT adopts an asymmetric homodimeric architecture whose assembly is substrate dependent. Moreover, our structure and biochemical data suggest that the RNase H and DNA polymerase activities are contributed by individual subunits of the homodimer.

Retrotransposons are mobile genetic elements that replicate through an RNA intermediate. They are divided into two groups, depending on whether flanking LTR sequences are present. Retrotransposons represent one of the main forces shaping the architecture of eukaryotic genomes<sup>1</sup>. For example, ~40% of the human genome is derived from retroelements, with 8% corresponding to the LTR class<sup>2</sup>, while in maize ~75% of the genome is derived from retroelements, mainly of the LTR class<sup>3</sup>. Retroviruses, such as human immunodeficiency virus (HIV), evolved from LTR elements through acquisition of an envelope gene, which allowed their egress from infected cells to initiate a subsequent round of infection<sup>4</sup>.

The Ty3 element of *S. cerevisiae* belongs to the Gypsy family<sup>5,6</sup>, and its RT is perhaps the most extensively characterized LTR retrotransposon enzyme with respect to its enzymatic activities<sup>7–9</sup> and the architecture of nucleic acid duplexes with which it interacts<sup>10–15</sup>. Although the structural motifs mediating substrate recognition and catalysis generally resemble those of vertebrate retroviral RTs, a notable difference between the Ty3 and retroviral enzymes is the separation of Ty3's DNA polymerase and RNase H active sites by ~13 bp (ref. 9), as opposed to 17 or 18 bp for lentiviral and gammaretroviral enzymes. Although the structural basis for such spatial separation is established for HIV-1 RT<sup>16,17</sup>, the origin of the shorter distance for Ty3 RT is difficult to explain from the retroviral structures. Ty3 RT lacks the connection, or tether, between its DNA polymerase and RNase H domains. Structural similarity between the RNase H subdomain of HIV-1 RT and its connection subdomain (which lacks the catalytic carboxylates found in the RNase H domain) originally suggested the former arose through domain duplication, while an alternative theory proposes that the

functional RNase H domain was acquired from a source outside the LTR retrotransposons<sup>18</sup>.

Another well-characterized LTR element from *S. cerevisiae* is Ty1 of the Copia-like group, which is more distantly related to retroviruses than Ty3. The polypurine tract (PPT) primers for plus-strand synthesis for both the Copia and Gypsy families differ in length and composition from retroviral PPTs. LTR retroelement PPTs generally contain shorter, less homogeneous tracts of purines, implying differences in PPT recognition. LTR retrotransposon PPTs are accurately processed by their cognate RT *in vivo*<sup>19,20</sup> and *in vitro*<sup>9,21</sup>, and it has also been proposed that a Ty3 RT-integrase fusion protein participates in reverse transcription *in vivo*<sup>19,22</sup>.

Despite extensive biochemical characterization of LTR retrotransposon RTs, detailed structural information is lacking. We therefore set out to structurally characterize Ty3 RT, and we report here the first structure of a retrotransposon RT in complex with its cognate PPT RNA-DNA hybrid at 3.1-Å resolution. The active enzyme is an asymmetric homodimer of 55-kDa subunits that associate in the presence of the nucleic acid substrate. Modeling of the spatial separation between the DNA polymerase and RNase H active sites, in addition to phenotypic mixing experiments, suggest that DNA polymerase and RNase H catalytic activities reside in separate subunits.

## RESULTS

### Overall structure

Details of RT purification, crystallization and structure solution can be found in the Online Methods. Selenomethionine-substituted protein was purified by immobilized metal-affinity, ion-exchange and

<sup>1</sup>Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, Warsaw, Poland. <sup>2</sup>Reverse Transcriptase Biochemistry Section, HIV Drug Resistance Program, Frederick National Laboratory, Frederick, Maryland, USA. <sup>3</sup>Biophysics Core Facility, International Institute of Molecular and Cell Biology, Warsaw, Poland. Correspondence should be addressed to M.N. (mnowotny@iimcb.gov.pl) or S.F.J.L.G. (legrices@mail.nih.gov).

Received 4 December 2013; accepted 11 February 2014; published online 9 March 2014; doi:10.1038/nsmb.2785

**Table 1** Data collection and refinement statistics for Ty3 RT–RNA–DNA complex crystals

	Ty3 RT SeMet
<b>Data collection</b>	
Space group	$P2_1 2_1 2$
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	320.7, 75.1, 108.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Resolution (Å)	50.0–3.1 (3.29–3.1) <sup>a</sup>
$R_{\text{merge}}$	0.10 (0.85)
$I / \sigma I$	11.9 (2.1)
$CC_{1/2}$ <sup>b</sup>	99.8 (72.9)
Completeness (%)	99.6 (98.4)
Redundancy	5.1 (5.1)
<b>Refinement</b>	
Resolution (Å)	3.1
No. reflections	90,832
$R_{\text{work}} / R_{\text{free}}$	22.7 / 29.6
No. atoms	14,105
Protein	12,734
Ligand / ion	1360 / 10
Water	1
<i>B</i> factors	131.7
Protein	129.3
Ligand / ion	153.5 / 158.0
Water	79.7
r.m.s. deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.069

The data collection statistics are based on a single crystal.

<sup>a</sup>Values in parentheses are for highest-resolution shell. <sup>b</sup> $CC_{1/2}$ , correlation coefficient between the average intensities in two parts of the unmerged data, each with a random half of the measurements of each unique reflection<sup>41</sup>.

gel-permeation chromatography. Purified enzyme was cocrystallized with a 16-bp RNA–DNA hybrid containing a 2-nt 5′ overhang in the RNA strand, the sequence of which corresponded to the Ty3 PPT with the cognate RNase H cleavage site located 12 nt from the 3′ end of the DNA. In such a substrate, positioning the 3′ end of the DNA strand in the polymerase catalytic center locates the biologically relevant PPT–U3 junction within the RNase H active site<sup>9</sup>. The structure was solved by the single-wavelength anomalous diffraction method and refined at 3.1 Å resolution (Table 1 and Fig. 1a) to an  $R_{\text{free}}$  of 29.6%. Sample experimental electron density maps are shown in Supplementary Figure 1a,b.

Although Ty3 RT has previously been described as monomeric in solution in the absence of nucleic acid<sup>12</sup>, early construction of the atomic model suggested that the biological unit in our crystals was an asymmetric homodimer in complex with an RNA–DNA hybrid (Fig. 1a). We hereafter designate the dimer subunits A and B. Two essentially identical copies of dimer–substrate complex are present in the asymmetric unit (I and II; Supplementary Fig. 1c). Complex II (chains E–H) has higher *B*-factors and less well-defined electron densities, indicating that it is less ordered.

For ease of comparison, we labeled secondary structure elements using the scheme from our previous work on XMRV RT<sup>23</sup> (Fig. 1a). Numbers were added to letter designations for additional helices of the Ty3 structure. Subunit A shares the overall architecture of retroviral RTs whose structures have been determined<sup>23,24</sup>. The DNA polymerase domain has the topology of a right hand, with the ‘palm’ subdomain housing the active site, the ‘fingers’ stabilizing the RNA template strand and the ‘thumb’ interacting mainly with the

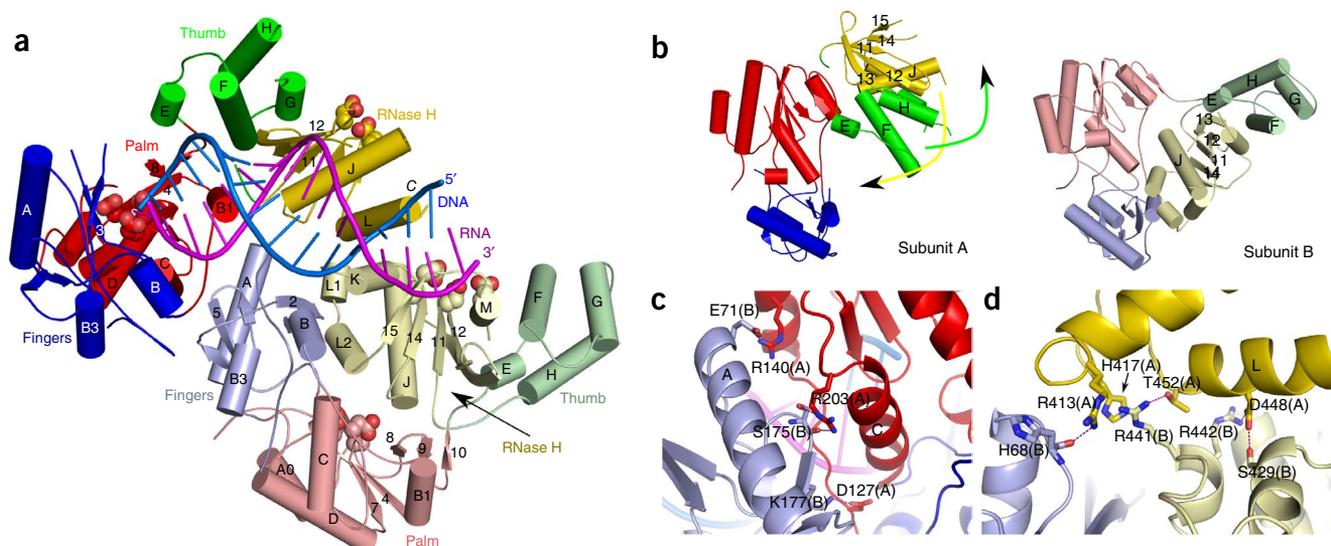
DNA strand. In contrast, the position of the Ty3 RNase H domain corresponds with that of the retroviral connection subdomain. This supports the hypothesis that evolution of retroviral RTs from LTR retrotransposon enzymes involved conversion of their RNase H domain to a ‘connector’ with loss of catalytic function and recruitment of a new RNase H1 domain<sup>18</sup>.

Ty3 RT subunits A and B are identical in sequence, and the structures of individual subdomains are very similar. Their pairwise superpositions result in low r.m.s. deviations for the positions of pairs of C $\alpha$  atoms: 0.5 Å for 95 C $\alpha$  atom pairs in the fingers subdomains, 0.9 Å for 116 pairs in the palm subdomains, 1.3 Å for 54 pairs in the thumb subdomains and 1.0 Å for 80 pairs in the RNase H domains. Both of the fingers–palm fragments are also structurally similar (r.m.s. deviation of 2.6 Å over 227 pairs of C $\alpha$  atoms). However, pronounced differences are apparent in the positioning of the RNase H and thumb subdomains (Fig. 1b). The difference in the position of the RNase H domain in the two subunits corresponds to a large, ~90° rotation around an axis running roughly through the contact point between the palm and thumb of subunit A. Consequently, the subunit B RNase H domain is positioned between its fingers and palm, blocking the DNA polymerase substrate binding cleft and causing the thumb subdomain to be displaced from the palm and rotated relative to the RNase H domain. Surprisingly, the conformation of subunit B resembles that of p51 HIV-1 RT, which lacks an RNase H domain<sup>25</sup>.

The subunit interface of the Ty3 RT dimer is highly polar, involving two main contact points. The first is formed by insertion of the subunit B fingers between the palm and RNase H domains of subunit A (Fig. 1a). Prominent interactions in this area involve (i) Arg203 (subunit A) and Ser175 (subunit B; hereafter, letters in parentheses represent dimer subunits), (ii) Asp127(A) and Lys177(B), and (iii) a salt bridge between Arg140(A) and Glu71(B) (Fig. 1c). The other contact region involves both RNase H domains (Fig. 1d): Arg413(A) interacts with the backbone of His68(B), Thr452(A) and His417(A) interact with Arg441(B), and Asp448(A) interacts with Ser429(B) and Arg442(B). Arg413 and Arg442 are conserved among other Gypsy retroelements (Supplementary Fig. 2), and Arg442 may be also essential for interaction with the DNA backbone.

### RNase H domain

The RNase H domain and the retroviral connection subdomains adopt the RNase H fold, the most important element of which is the five-stranded, central  $\beta$ -sheet<sup>26</sup>. The first three strands are longer and run antiparallel to one another, whereas the last two are shorter and parallel to the first. The fold also contains two or three  $\alpha$ -helices on one side of the central sheet and a single helix on the other. Cellular RNase H1 enzymes and closely related retroviral RNase H domains (collectively referred to as ‘cellular RNase H1 enzymes’) have two main differences from the Ty3 RNase H and retroviral connection subdomains (Fig. 2 and Supplementary Figs. 3 and 4). First, there is a deletion of about ten residues between the first two strands of the central  $\beta$ -sheet of the LTR retrotransposon enzyme, shortening the first strand in its C terminus (Fig. 2a,c). A second difference is the arrangement of  $\alpha$ -helices between strands 4 and 5. The Supplementary Note provides a detailed comparison of substrate-binding residues between cellular and Ty3 RNase H enzymes. The Ty3 RNase H active site resembles cellular enzymes, likely functioning through the same mechanism (Supplementary Fig. 5a). However, equivalents of many residues mediating substrate binding in bacterial, human and HIV-1 RNase H1 could not be identified in Ty3 RNase H, especially those forming the phosphate-binding pocket<sup>27,28</sup>.

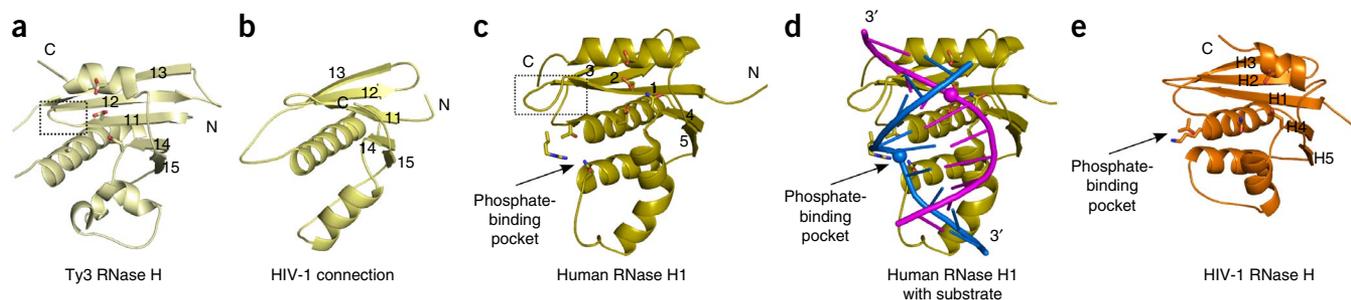


**Figure 1** Overall structure of Ty3 RT and the dimer interface. **(a)** Cartoon representation of Ty3 RT in complex with an RNA-DNA hybrid substrate. Protein subdomains are colored blue for fingers, red for palm and green for thumb, and the RNase H domain is in yellow. Lighter shades of the same colors are used for subunit B. Secondary structure elements are labeled (numbers for strands and letters for helices) using the same scheme as in our previous work on XMRV RT<sup>23</sup>. Residues forming the DNA polymerase and RNase H active sites are shown as spheres. **(b)** Comparison of the structures of subunits A and B, colored as in **a**. Arrows indicate the movements of the thumb and RNase H domains that transform their conformation in subunit A to that of subunit B. **(c,d)** Residues involved in dimer formation. Protein structure is colored as in **a**.

As neither of the dimer's RNase H active sites interacts with the RNA, an important mechanistic question is which Ty3 RT subunit contributes RNase H activity and what conformational changes are necessary to support this. **Supplementary Figure 5b** depicts a catalytic interaction of the RNase H domain with the substrate. This was prepared using the human RNase H1 complex structure<sup>28</sup> and assumes that the Ty3 RNase H active site interacts with nucleotide -13 or -12, the preferred cleavage sites in 3' end-directed cleavage mode. Bringing the active site of the RNase H domain from subunit A or B into the proximity of the RNA backbone would necessitate a substantial conformational change. Such large changes of the palm-fingers arrangement relative to the thumb-RNase H fragment are possible, as evidenced by major conformational differences between subunits A and B. The subunit B RNase H domain is located closer to the scissile phosphate, and its movement (probably together with its thumb subdomain) could be accommodated by a translation of ~40 Å without invoking severe clashes, while preserving dimerization contacts of the palm and fingers subdomains. A corresponding rearrangement of subunit A RNase H would disrupt the dimer structure and eliminate

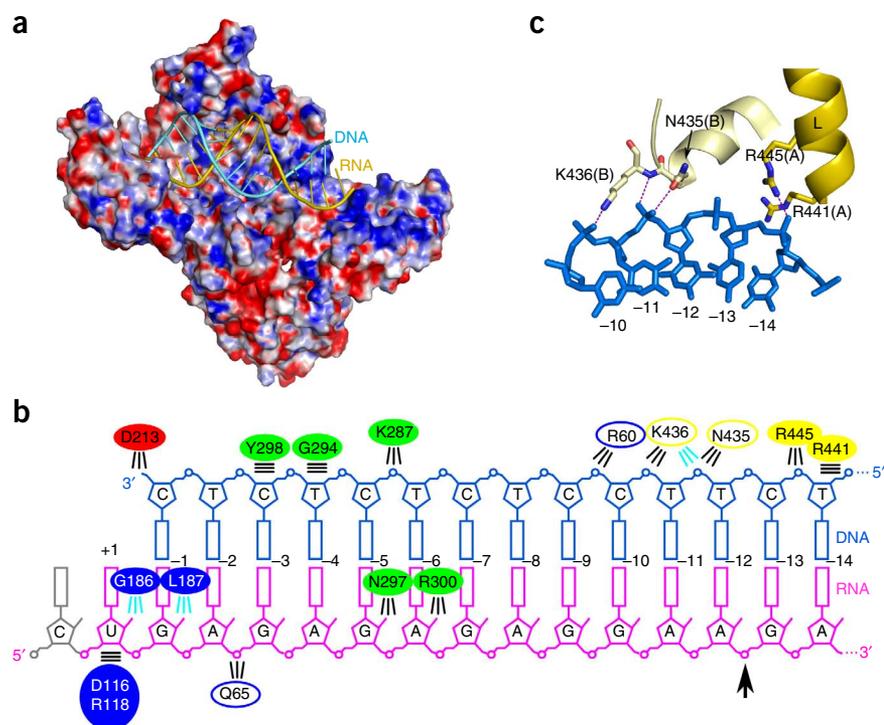
critical contacts between the substrate and thumb subdomain of subunit A. This implies that subunit B's RNase H domain contributes activity, an idea that is supported by biochemical data presented below. Conformational changes of the protein could induce substrate deformation similar to that observed with HIV-1 RT<sup>17</sup>, although the exact nature of these changes is difficult to predict and a similar issue for HIV-1 RT has been elucidated only recently with crystallography<sup>17</sup>.

The requirement for conformational changes conducive to substrate cleavage implies that RNA hydrolysis would be infrequent, in agreement with published experiments examining RNase H activity concurrent with DNA synthesis<sup>9</sup>. Hydrolysis is rare during DNA synthesis, occurring primarily after the enzyme reaches the end of the substrate, possibly providing sufficient time for rearrangement into an RNase H-competent mode. Infrequent and transient interactions of the RNase H domain with the substrate emerge as a common element of the mechanism of RTs. For HIV-1 RT, the RNA-DNA substrate must be unwound to allow RNase H cleavage<sup>17</sup>. In monomeric XMRV RT, the RNase H domain is tethered to its connection by a



**Figure 2** Comparison of RNase H and connection subdomains. **(a–e)** Cartoon representations of Ty3 RNase H domain **(a)**, HIV-1 connection subdomain **(b)**, human RNase H1 protein **(c)**; protein chain from PDB 2QK9 (ref. 28)), human RNase H1 with bound RNA-DNA substrate **(d)**; PDB 2QK9 (ref. 28)) and HIV-1 RNase H domain **(e)**; PDB 1RTD (ref. 16)). Strands of the central  $\beta$ -sheet are labeled, and residues forming the active site and phosphate-binding pocket are shown as sticks. The C-terminal region of the first  $\beta$ -strand, which differs in length between Ty3 RNase H and cellular enzymes, is indicated with a dotted box.

**Figure 3** Substrate binding. (a) Ty3 RT structure colored according to surface potential (red, negative; blue, positive;  $\pm 15$  kTe). Nucleic acid is shown in cyan (DNA) and yellow (RNA). (b) Diagram of protein–nucleic acid interactions. Arrow indicates the PPT-U3 junction (the preferred site of RNase H cleavage). The 5' RNA nucleotide not observed in the structure is shown in gray. Ovals are colored by protein domain as in **Figure 1**; solid ovals denote subunit A and empty ovals subunit B. Parallel horizontal lines indicate van der Waals interactions. Diagonal and vertical lines indicate interactions mediated by the protein's backbone (cyan) or side chains (black). (c) Interactions between RNase H domains (yellow) and the DNA strand (blue). Interacting protein residues are shown as sticks.



flexible linker and thus is very mobile. Only in the presence of RNA–DNA does it become transiently organized on the substrate<sup>23</sup>. This feature of RNase H domains may regulate their function in specialized cleavage events during primer generation and removal.

### Substrate binding

The PPT RNA–DNA hybrid in our structure adopts a conformation intermediate between A- and B-form duplex, and the width of the minor groove is between 9 and 10.4 Å. The substrate comprises the entire PPT sequence along with four residues from the U3 region and should therefore provide an accurate model of the PPT structure. Its orientation in the structure would correspond to minus-strand DNA extension with possible simultaneous generation of the 3' end of the PPT primer. We detected no major structural deformations of the hybrid, which superimposes well with the random RNA–DNA hybrids in structures recently reported for HIV-1 and XMRV RT<sup>17,23</sup>. At the resolution of our structure (3.1 Å), subtle changes in nucleic acid conformation may not be apparent, but we favor the notion that Ty3 PPT recognition reflects dynamic properties of the duplex, possibly lower flexibility, rather than preexisting deformations. Such dynamic properties may mediate conformational changes required for RNase H cleavage.

In the Ty3 RT complex structure, the hybrid is accommodated in a mostly positively charged cleft of the dimer. Its lower portion is defined by both of the fingers subdomains and the subunit B RNase H domain, while the top comprises the subunit A palm, thumb and RNase H domains (**Fig. 3a**). Footprinting studies have suggested that Ty3 RT protects template nucleotides –1 to –24 (numbering relative to the polymerase active site is used throughout, unless specified otherwise) and primer nucleotides –1 to –25 (ref. 12). Although the hybrid in our structure is shorter than this footprint (crystallization trials with longer hybrids were unsuccessful), when a longer duplex is modeled the extended region passes very close to, and could interact with, the positively charged region of the subunit B thumb. This explains the extended DNase I footprint and indicates that the thumb of subunit B could further stabilize the RNA–DNA substrate beyond interactions observed here. **Figure 3b** provides details of the protein–substrate interactions, identifying two main regions. The first involves contacts between the DNA polymerase domain of subunit A and nucleotides +1 to –6 (positions –13 to –8 relative to the PPT–U3 junction, as in ref. 11). The second region comprises interactions between DNA nucleotides –9 to –14 (–4 to +2 relative

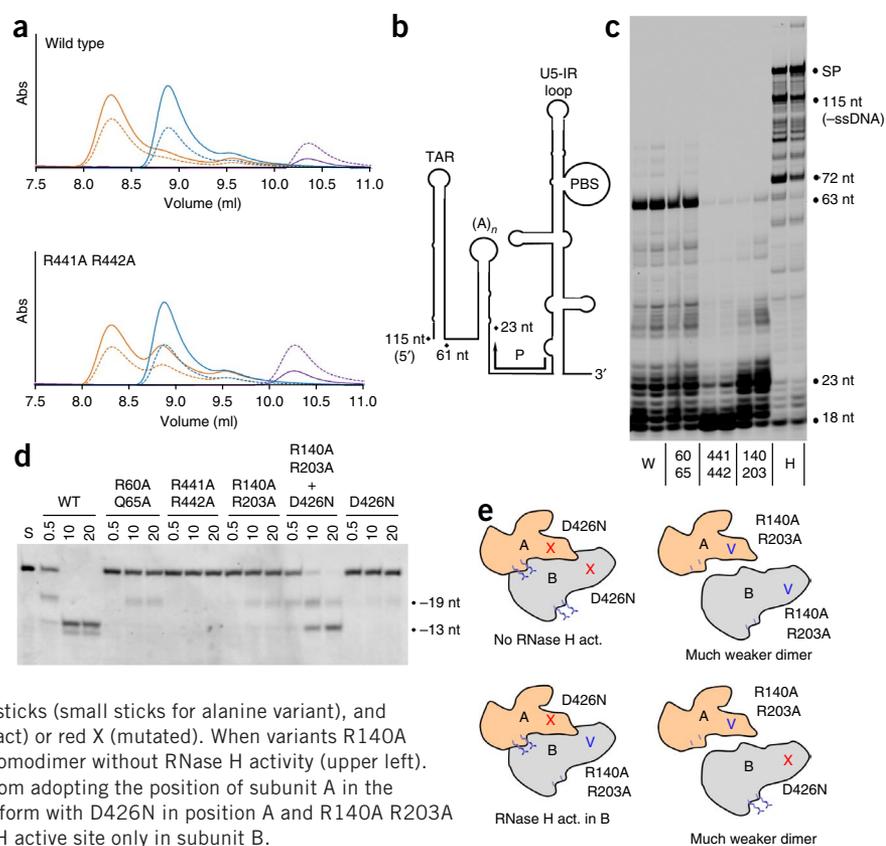
to the PPT–U3 junction) and residues from both RNase H domains. This bipartite substrate interface is consistent with biochemical data obtained with modified Ty3 PPT substrates containing nucleoside analogs designed to either enhance flexibility or increase rigidity of the hybrid<sup>11</sup>. These experiments identified two regions important for precise RNase H–mediated cleavage: namely, around the scissile bond defining the PPT–U3 junction, which would form interactions with both RNase H domains, and 8–11 nucleotides upstream toward the 5' end of the RNA strand, corresponding to the portion of the RNA–DNA hybrid that interacts extensively with the subunit A thumb. Supporting our structure is the fact that nucleotides between these regions are more tolerant to modification, showing they do not form contacts with the protein.

Template nucleotide +1, which would base-pair with the incoming dNTP, is stabilized by interactions of its 2'-OH group with the backbone of Gly186(A) (**Fig. 3b**). Nucleotide +1 is also stabilized by a 'pin' structure comprising the side chains of Arg118(A) and Asp116(A), which has been characterized previously for monomeric gammaretroviral RTs<sup>23,29</sup> and the heterodimeric HIV-1 enzyme<sup>30,31</sup>. 2'-OH groups of the RNA also form interactions with thumb residues Asn297(A) and Arg300(A) and the backbone of fingers residue Leu187(A).

The 3'-OH on the DNA strand is located at the DNA polymerase active site of subunit A, whose configuration resembles that of RTs from retroviruses<sup>16,23</sup> (**Supplementary Fig. 6**), with key carboxylate residues coordinating two divalent metal ions<sup>7</sup>. The fact that the polymerase domain and active site of Ty3 RT subunit A are superimposable with HIV-1 RT demonstrates that this subunit probably contributes polymerase activity. However, one difference in the Ty3 polymerase active site is the residue stabilizing the base of the incoming dNTP (which is absent in our structure). This is well conserved among retroviral RTs (Gln151 in HIV-1), but is replaced by Phe185(A) in Ty3 RT (**Supplementary Fig. 3**).

Upstream of the active site, the DNA strand forms extensive interactions with helix F of the subunit A thumb (**Fig. 1a**), which for retroviral RTs is inserted into the minor groove of the hybrid<sup>16,23,32</sup>.

**Figure 4** Biochemical experiments. (a) GF experiments with hybrid 3 and wild-type Ty3 RT (top) or R441A R442A variant (bottom). Traces are shown in purple for protein, blue for RNA-DNA hybrid and orange for the mixture. Dotted lines represent  $E_{280}$  and solid line  $E_{260}$ . (b) Schematic representation of the HIV-1 genome region used to examine RNA-dependent DNA polymerase activity. The position of the DNA primer (P) is indicated, together with major pause sites. PBS, primer-binding site; TAR, transactivation response element; (A)<sub>n</sub>, poly(A) hairpin; U5-IR, unique 5' inverted repeat. (c) DNA polymerization assays. Products of 10- and 20-min reactions are shown. The major polymerase-stalling products are marked at right. SP, self-priming product; W, wild-type Ty3 RT; H, HIV-1 RT. Numbers below lanes denote the positions of alanine substitutions in Ty3 RT. (d) RNase H activity assays. Lane S contained uncleaved substrate with fluorescently end-labeled RNA. Hydrolysis was examined at 0.5, 10 and 20 min (the lanes are labeled accordingly). The cleavage sites relative to the 3' end of the DNA are indicated. Uncropped images from c,d can be found in **Supplementary Figure 8**. (e) Cartoon of the phenotypic mixing experiment. Arg140 and Arg203 are shown as blue sticks (small sticks for alanine variant), and the RNase H active site is marked with a blue V (intact) or red X (mutated). When variants R140A R203A and D426N are mixed, D426N can form a homodimer without RNase H activity (upper left). R140A R203A substitutions preclude this variant from adopting the position of subunit A in the dimer (right diagrams); however, a mixed dimer can form with D426N in position A and R140A R203A in position B (lower left) and with the intact RNase H active site only in subunit B.



Tyr298(A) and Gly294(A) form van der Waals interactions with the sugar-phosphate backbone of DNA nucleotides -3 and -4, respectively, while Lys287(A) interacts with the phosphate group of DNA nucleotide -5, and Asn297(A) forms an additional hydrogen bond with the 2'-OH of RNA nucleotide -5. An important structural residue is Phe292(A), located on the side opposite the substrate interface and stabilizing helix F. Interactions mediated by the thumb subdomain support previous biochemical studies showing the importance of Phe292, Gly294 and Tyr298 (ref 8). Among several substitutions, G294A RT is the most affected in the absence of a heparin trap, indicating its critical contribution to this component of the interface. Experiments with LNA-substituted nucleic acids have also predicted interactions of thumb residues Tyr298(A) and Gly294(A) with DNA nucleotides -3 and -4, supporting and extending findings from mutagenesis analysis of HIV-1 RT<sup>33-36</sup>. The next substrate region interacting with protein involves DNA nucleotides -10 to -13, which contact both RNase H domains (Fig. 3b,c). Arg441(A), Arg445(A), Asn435(B) and Lys436(B) mediate these interactions with the DNA backbone (Fig. 3c).

### Biochemical characterization

To confirm substrate-induced dimerization, we coupled high-resolution gel filtration (GF) with multi-angle light scattering (MALS) to determine the molecular weight of nucleoprotein complexes (Fig. 4a). As shown previously<sup>12</sup>, Ty3 RT eluted as monomer in the absence of substrate, with a molecular weight of 53.1 kDa, compared with the expected 54.6 kDa for a monomer (Fig. 4a). When mixed with a 27-bp RNA-DNA hybrid containing a 2-nt RNA 5' overhang (hybrid 3), the nucleoprotein complex eluted much earlier than either the protein monomer or the RNA-DNA hybrid alone (Fig. 4a). The molecular weight of this complex was 119 kDa, compared with a calculated

value of 126 kDa for a dimer interacting with hybrid 3. Finally, analytical ultracentrifugation (AUC) sedimentation velocity experiments with the Ty3 RT-hybrid 3 complex also indicated formation of a 2:1 protein-nucleic acid complex (Supplementary Fig. 7).

We next prepared three Ty3 RT mutants. The first contained dual alanine substitutions in the region involved in dimer formation, at Arg140 and Arg203 (Fig. 1c), and the second had alanine substitutions of Arg441 and Arg442. Arg441(B) and Arg442(B) participate in dimer formation (Fig. 1d), whereas Arg441(A) and Arg442(A) are located close to the DNA backbone and Arg441(A) participates in substrate binding (Fig. 3c). The third variant contained dual substitutions in novel substrate contacts mediated by Arg60 and Gln65. These residues in subunit B participate in the substrate interface, whereas in subunit A they are located distal from the RNA-DNA binding cleft. Therefore, our experiments should assess only their role in subunit B.

We first examined the oligomeric state of the Ty3 RT variants with substitutions in the dimer interface. Although mutant R140A R203A was unstable in GF experiments, AUC indicated it failed to form dimers in the presence of hybrid 3 (Supplementary Fig. 7). When R441A R442A RT was mixed with hybrid 3, GF-MALS (Fig. 4a; measured molecular weight of 105 kDa) and AUC (Supplementary Fig. 7) indicated a mild defect in dimer formation. Enzymatic activity of Ty3 RT variants was next examined. RNA-dependent DNA polymerase activity was evaluated with a template derived from the 5'-terminal region of the HIV genome that forms extensive secondary structures (Fig. 4b). Wild-type Ty3 RT (lane W) and mutant R60A Q65A were less processive than HIV-1 RT (lane H), as evidenced by transient pausing at the base of the poly(A) hairpin and an inability to resolve the TAR hairpin (Fig. 4c). Despite this, both enzymes displayed similar activity, indicating that substrate contacts mediated by Arg60(B) and Gln65(B) are not essential for processivity and strand-displacement

**Figure 5** Structural comparison of Ty3 and HIV-1 RTs. (a,b) Comparison of Ty3 and HIV-1 RTs. The HIV-1 RNase H domain is shown in orange. (c,d) Comparison of subunit B of Ty3 RT with p51 of HIV-1 RT. The region comprising short  $\beta$ -strands that are unfolded in HIV-1 RT is indicated with a dotted box.

activity. In contrast, mutant R140A R203A showed a strong processivity defect, with polymerization products accumulating at the base of the poly(A) hairpin. Lastly, for R441A R442A Ty3 RT the major product was a single-nucleotide extension of the primer, possibly indicating an inability to release pyrophosphate after formation of the initial phosphodiester bond. In conclusion, Ty3 RT dimerization and substrate contacts identified in our crystal structure are required for efficient polymerization.

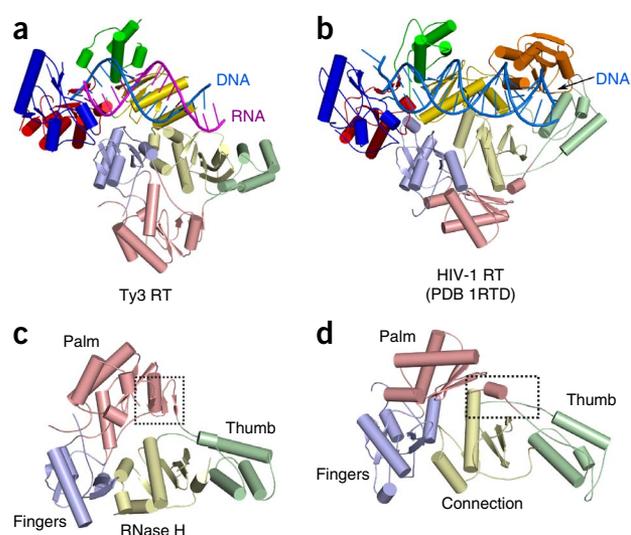
RNase H activity was evaluated on a hybrid with a recessed 3' DNA terminus to monitor 3' end-directed cleavage when the DNA 3'-OH occupies the polymerase active site. As previously reported<sup>12</sup>, we observed products of cleavage 13 nt downstream of the DNA 3' end and less prominent products resulting from an internal cleavage mode  $\sim$ 19 nt from the DNA 3' end (Fig. 4d). R60A Q65A RT showed reduced RNase H activity (Fig. 4d), supporting our notion that subunit B residues contacting the substrate are important for RNase H activity. This can be explained with the assumption that subunit B RNase H undergoes a conformational change to allow substrate cleavage. Arg60(B) and Gln65(B) would not change position and would contribute to substrate stabilization during and after the conformational change. RNase H activity of R140A R203A and R441A R442A RTs was also severely affected. Therefore, residues identified as participating in dimer formation and substrate binding are important for RNase H activity.

Our structure, moreover, implies that enzymatic activities of Ty3 RT reside in different subunits of the dimer. Although the homodimeric nature of Ty3 RT complex makes it challenging to verify this notion biochemically, we exploited the dimerization defect of mutant R140A R203A. Both Arg140 and Arg203 are critical to the dimer interface of subunit A, whereas in B they are distal from the dimer or substrate interface. Therefore, when R140A R203A is mixed with RNase H-deficient protein (D426N, which we used for crystallization), only two out of four possible dimer combinations should form, namely a D426N homodimer, lacking RNase H activity, and a mixed dimer with subunit B contributing R140A R203A (Fig. 4e). If RNase H activity derives from subunit B, such a mixed dimer should be active. When R140A R203A and D426N variants were mixed at an equimolar ratio, RNase H activity was rescued (Fig. 4d), confirming that the DNA polymerase and RNase H activities of Ty3 RT reside in different subunits.

### Comparison of retroviral and LTR retrotransposon RTs

We document here key differences between Ty3 and HIV-1 RT. First, for the LTR retrotransposon enzyme, substrate binding is a prerequisite to dimerization, whereas the lentiviral enzyme is a stable dimer in the absence of substrate<sup>24,37,38</sup>. Second, only the p66 subunit of the HIV-1 RT heterodimer contains a copy of the RNase H domain, so both enzyme activities reside in one subunit.

Topologically, however, the two enzymes are surprisingly similar. **Supplementary Figure 3** aligns Ty3 RT subunit A and HIV-1 RT p66 (fingers-palm-thumb-connection). Their structures, as well as that of monomeric XMRV RT, are quite similar (Fig. 5a,b). The fingers, palm and thumb, together with the connection or RNase H domains of these subunits or proteins can be superimposed with an r.m.s. deviation of 2.1 Å (320 C $\alpha$  atom pairs) for Ty3 compared with XMRV, and



2.9 Å (237 C $\alpha$  atom pairs) for Ty3 compared with HIV-1 (PDB 1RTD (ref. 16)). Differences are relatively minor, including (i) an N-terminal extension in Ty3 RT, (ii) altered trajectory of the protein backbone between Thr201(A) and Arg206(A) of the Ty3 RT palm owing to deletion between helix C and strand 6, and (iii) the absence of thumb helix E in the HIV-1.

When Ty3 RT subunit B is compared with the HIV-1 p51 subunit, structures of individual subdomains are also similar. Moreover, their arrangement is strikingly analogous (Fig. 5c,d). p51's connection is rotated and placed between its palm and fingers, analogous to the Ty3 subunit B RNase H domain. As the dimeric organization of HIV-1 RT is well-documented, this further supports the notion that our structure represents the physiological architecture of Ty3 RT and that DNA polymerase activity is the property of subunit A. There are, however, several notable differences between HIV-1 RT p51 and Ty3 RT subunit B. The p51 palm subdomain has a different position, and the palm-fingers module cannot be superimposed between the two subunits of the dimer as well as for Ty3 RT. Moreover, the p51 thumb is further from its palm in order to accommodate the larger p66 subunit and, in particular, its RNase H domain. This larger separation of palm and thumb requires that short  $\beta$ -sheets 8, 9 and 10 at the C terminus of the palm subdomain are unfolded in HIV-1, whereas their structure is maintained in Ty3 RT.

When Ty3 RT subunit A, HIV-1 p66 and XMRV RT are superimposed, the trajectories of the nucleic acid substrates are very similar for XMRV and Ty3 enzymes. The substrate of HIV-1 RT passes further away from the connection subdomain owing to the presence of its RNase H domain, as described previously<sup>23</sup>. Overall, substrate interactions around the DNA polymerase active site are conserved between the three enzymes, and equivalent residues can be identified in each protein. However, toward the RNase H domains the substrate interfaces involve different sets of residues.

### DISCUSSION

We present the first crystal structure of a retrotransposon RT, revealing an unanticipated architecture featuring an asymmetric homodimer induced by substrate binding. A separation of  $\sim$ 13 nt between the 3' end of the DNA and the RNase H active site, which was observed for Ty3 RT in biochemical experiments, was difficult to rationalize using retroviral RT structures because the active site of the RNase H modeled on the HIV-1 p66 or XMRV connection subdomains would be facing away from the substrate. Dimerization thus offers an

elegant explanation for the shorter distance between the polymerase and RNase H active sites.

Previous studies have proposed that structural deformations protect the PPT from RNase H cleavage<sup>39</sup>. Although we observed no major distortion of the RNA-DNA substrate in our structure, subtle alterations may be responsible for its special features. Further structures of Ty3 RT with random-sequence RNA-DNA and PPT bound at different registers should shed light on this issue. Another interesting question is the role of the described Ty3 RT-integrase fusion that has been detected in virus-like particles<sup>19</sup>, which may facilitate folding of RT.

Although the overall conformations of Ty3 and HIV-1 RT are comparable, a critical difference is the homodimeric nature of the former. Homodimers with the high degree of asymmetry observed here are rare<sup>40</sup>. Important questions to be addressed are the conformation of substrate-free monomeric Ty3 RT and the mechanism of substrate-induced dimerization. It is likely that the more compact, subunit B-like conformation is preferred in the absence of the nucleic acid. Substrate binding could then stabilize a more open, subunit A-like conformation, allowing dimerization. Another distinctive and, to our knowledge, unprecedented feature of the Ty3 RT structure is that its asymmetry allows for the separation of enzymatic activities between subunits and brings the RNase H domain into position to mediate hydrolysis. Whether this applies to related LTR retrotransposon enzymes remains to be determined.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Protein Data Bank: Coordinates have been deposited with accession code [4OL8](#).

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

## ACKNOWLEDGMENTS

We thank W. Yang for critical reading of the manuscript, I. Ptasiwicz for excellent technical assistance, the staff of beamline 14-1 at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung for assistance with data collection, and J.M. Bujnicki for help in the preparation of the sequence alignments. This work was supported by grants from the Polish National Science Center (contract no. N N301 439738 to M.N.) and the FP7 HEALTHPROT project (contract no. 229676 to M.N.). S.F.J.L.G. is supported by the Intramural Research Program of the National Cancer Institute, US National Institutes of Health and by federal funds from National Institutes of Health contract no. HHSN261200800001E (M.K.B.). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government. M.N. is a recipient of the Foundation for Polish Science 'Ideas for Poland' award. The research of M.N. was supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute. The research was performed using Centre for Preclinical Research and Technology infrastructure (European Union project POIG.02.02.00-14-024/08-00).

## AUTHOR CONTRIBUTIONS

E.N. obtained crystals of Ty3 RT-substrate complex; E.N. and M.N. solved and analyzed the structure; J.T.M., M.K.B. and J.S. performed biochemical experiments; E.N. and R.H.S. performed biophysical protein characterization; J.J. prepared the expression construct and conducted initial crystallization experiments; S.F.J.L.G. and M.N. designed the research and wrote the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Cordaux, R. & Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **10**, 691–703 (2009).
- Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y. & Bennetzen, J.L. The paleontology of intergenic retrotransposons of maize. *Nat. Genet.* **20**, 43–45 (1998).
- Malik, H.S., Henikoff, S. & Eickbush, T.H. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **10**, 1307–1318 (2000).
- Llorens, C. *et al.* The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* **39**, D70–D74 (2011).
- Sandmeyer, S.B., Aye, M. & Menees, T.M. Ty3: a position-specific, gypsylike element in *Saccharomyces cerevisiae*. in *Mobile DNA II* (eds. Craig, N.L., Craigie, R., Gellert, M. & Lambowitz, A.M.) 663–682 (ASM Press, Washington, DC, 2002).
- Bibillo, A., Lener, D., Klarmann, G.J. & Le Grice, S.F. Functional roles of carboxylate residues comprising the DNA polymerase active site triad of Ty3 reverse transcriptase. *Nucleic Acids Res.* **33**, 171–181 (2005).
- Bibillo, A., Lener, D., Tewari, A. & Le Grice, S.F. Interaction of the Ty3 reverse transcriptase thumb subdomain with template-primer. *J. Biol. Chem.* **280**, 30282–30290 (2005).
- Lener, D., Budihas, S.R. & Le Grice, S.F. Mutating conserved residues in the ribonuclease H domain of Ty3 reverse transcriptase affects specialized cleavage events. *J. Biol. Chem.* **277**, 26486–26495 (2002).
- Brinson, R.G. *et al.* Probing anomalous structural features in polypurine tract-containing RNA-DNA hybrids with neomycin B. *Biochemistry* **48**, 6988–6997 (2009).
- Dash, C., Marino, J.P. & Le Grice, S.F. Examining Ty3 polypurine tract structure and function by nucleoside analog interference. *J. Biol. Chem.* **281**, 2773–2783 (2006).
- Rausch, J.W. *et al.* Interaction of p55 reverse transcriptase from the *Saccharomyces cerevisiae* retrotransposon Ty3 with conformationally distinct nucleic acid duplexes. *J. Biol. Chem.* **275**, 13879–13887 (2000).
- Turner, K.B. *et al.* SHAMS: combining chemical modification of RNA with mass spectrometry to examine polypurine tract-containing RNA/DNA hybrids. *RNA* **15**, 1605–1613 (2009).
- Yi-Brunozzi, H.Y., Brabazon, D.M., Lener, D., Le Grice, S.F. & Marino, J.P. A ribose sugar conformational switch in the LTR-retrotransposon Ty3 polypurine tract-containing RNA/DNA hybrid. *J. Am. Chem. Soc.* **127**, 16344–16345 (2005).
- Yi-Brunozzi, H.Y. *et al.* High-resolution NMR analysis of the conformations of native and base analog substituted retroviral and LTR-retrotransposon PPT primers. *Chem. Biol.* **15**, 254–262 (2008).
- Huang, H., Chopra, R., Verdine, 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).
- Lapkouski, M., Tian, L., Miller, J.T., Le Grice, S.F. & Yang, W. Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. *Nat. Struct. Mol. Biol.* **20**, 230–236 (2013).
- Malik, H.S. & Eickbush, T.H. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res.* **11**, 1187–1197 (2001).
- Kirchner, J. & Sandmeyer, S.B. Ty3 integrase mutants defective in reverse transcription or 3'-end processing of extrachromosomal Ty3 DNA. *J. Virol.* **70**, 4737–4747 (1996).
- Wilhelm, M., Heyman, T., Friant, S. & Wilhelm, F.X. Heterogeneous terminal structure of Ty1 and Ty3 reverse transcripts. *Nucleic Acids Res.* **25**, 2161–2166 (1997).
- Wilhelm, M., Uzun, O., Mules, E.H., Gabriel, A. & Wilhelm, F.X. Polypurine tract formation by Ty1 RNase H. *J. Biol. Chem.* **276**, 47695–47701 (2001).
- Nymark-McMahon, M.H., Beliakova-Bethell, N.S., Darlix, J.L., Le Grice, S.F. & Sandmeyer, S.B. Ty3 integrase is required for initiation of reverse transcription. *J. Virol.* **76**, 2804–2816 (2002).
- Nowak, E. *et al.* Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res.* **41**, 3874–3887 (2013).
- Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A. & Steitz, T.A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
- Mous, J., Heimer, E.P. & Le Grice, S.F. Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in *Escherichia coli*. *J. Virol.* **62**, 1433–1436 (1988).
- Nowotny, M. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10**, 144–151 (2009).
- Nowotny, M., Gaidamakov, S.A., Crouch, R.J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016 (2005).
- Nowotny, M. *et al.* Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell* **28**, 264–276 (2007).
- Najmudin, S. *et al.* Crystal structures of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase complexed with nucleic acid: functional implications for template-primer binding to the fingers domain. *J. Mol. Biol.* **296**, 613–632 (2000).

30. Boyer, P.L., Ferris, A.L. & Hughes, S.H. Cassette mutagenesis of the reverse transcriptase of human immunodeficiency virus type 1. *J. Virol.* **66**, 1031–1039 (1992).
31. Kim, B. *et al.* New human immunodeficiency virus, type 1 reverse transcriptase (HIV-1 RT) mutants with increased fidelity of DNA synthesis. Accuracy, template binding, and processivity. *J. Biol. Chem.* **274**, 27666–27673 (1999).
32. Sarafianos, S.G. *et al.* Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449–1461 (2001).
33. Beard, W.A. *et al.* Structure/function studies of human immunodeficiency virus type 1 reverse transcriptase. Alanine scanning mutagenesis of an alpha-helix in the thumb subdomain. *J. Biol. Chem.* **269**, 28091–28097 (1994).
34. Bebenek, K. *et al.* Reduced frameshift fidelity and processivity of HIV-1 reverse transcriptase mutants containing alanine substitutions in helix H of the thumb subdomain. *J. Biol. Chem.* **270**, 19516–19523 (1995).
35. Bebenek, K. *et al.* A minor groove binding track in reverse transcriptase. *Nat. Struct. Biol.* **4**, 194–197 (1997).
36. Powell, M.D. *et al.* Residues in the  $\alpha$ H and  $\alpha$ I helices of the HIV-1 reverse transcriptase thumb subdomain required for the specificity of RNase H-catalyzed removal of the polypurine tract primer. *J. Biol. Chem.* **274**, 19885–19893 (1999).
37. North, T.W., Cronn, R.C., Remington, K.M., Tandberg, R.T. & Judd, R.C. Characterization of reverse transcriptase from feline immunodeficiency virus. *J. Biol. Chem.* **265**, 5121–5128 (1990).
38. Thomas, D.A. & Furman, P.A. Purification and kinetic characterization of equine infectious anemia virus reverse transcriptase. *Biochem. Biophys. Res. Commun.* **180**, 1365–1371 (1991).
39. Rausch, J.W. & Le Grice, S.F. 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. *Int. J. Biochem. Cell Biol.* **36**, 1752–1766 (2004).
40. Swapna, L.S., Srikeerthana, K. & Srinivasan, N. Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. *PLoS ONE* **7**, e36688 (2012).
41. Karplus, P.A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).

## ONLINE METHODS

**Protein purification.** Ty3 RT was cloned into an expression vector with a 3C protease cleavage site between the His-tag and the protein. His<sub>6</sub>-tagged RT with the RNase H-inactivating substitution D426N was expressed in *E. coli* and purified by immobilized metal affinity, ion exchange, and gel exclusion chromatography. The His-tag was removed by overnight incubation with 3C protease. Purified protein eluted from the gel filtration column at a volume expected for the monomeric form. Protein for SAD phasing was expressed in selenomethionine-containing media in *E. coli* BL21(DE3) Magic cells. Cells were induced with 0.4 M IPTG, grown overnight at 18 °C, harvested by centrifugation and lysed by sonication in buffer containing 20 mM HEPES (pH 7.0), 250 mM NaCl, 20 mM imidazole, 10% glycerol and 5 mM β-mercaptoethanol (buffer A). The lysate was clarified by centrifugation at 40000 r.p.m. and loaded onto a 5 ml Ni-NTA (HiTrap, GE Healthcare) column equilibrated in buffer A. After washing with buffer A containing 40 mM imidazole, protein was eluted with buffer A containing 300 mM imidazole. Following ammonium sulfate precipitation, protein was dissolved in 20 mM HEPES (pH 7.0), 50 mM NaCl, 10% glycerol and 1 mM DTT (buffer B), applied to a Mono S column (GE Healthcare) and eluted with a linear gradient of NaCl from 0.1 to 0.5 M. RT-containing fractions were precipitated with ammonium sulfate, dissolved in buffer containing 20 mM HEPES (pH 7.0), 100 mM KCl, 1 mM DTT and 10% glycerol, and applied to a Superdex 200 size exclusion column (GE Healthcare). Peak fractions were concentrated using a 10 kDa cut-off centricon (Millipore) to 15 mg/ml.

**Crystallography.** Crystallization trials were prepared for protein alone, as well as in the presence of RNA-DNA hybrids ranging from 14 to 26 bp. HPLC-purified oligonucleotides were purchased from Metabion International AG. Before crystallization, protein was mixed with RNA-DNA hybrid in a 1:1.2 molar ratio and a final protein concentration of 7 mg/ml. Hybrids were produced by annealing either an RNA oligonucleotide, 5'-AACAGAGUGCGACACCUG-3', with a DNA oligonucleotide, 5'-CAGGTGTGCGCACTCTG-3' (hybrid 1), or an RNA oligonucleotide, 5'-CUGAGAGAGAGGAAGAUG-3', with a DNA oligonucleotide, 5'-CATCTTCCTCTCTCTC-3' (hybrid 2). Hybrid 2 corresponds to the Ty3 PPT sequence with the PPT-U3 junction located 12 nt from the 3' end of the DNA strand and is efficiently and specifically cleaved at PPT-U3 by Ty3 RT (not shown). No crystals were obtained with substrates corresponding to hybrid 2, in which the PPT-U3 junction was located 13 nt from the 3' end of the DNA strand.

The first crystals were obtained in the presence of hybrid 1, with a 16 bp duplex portion and a 2 nt 5' RNA overhang in 1.7 M sodium citrate, and diffracted X-rays to only 7 Å resolution. Substituting the random RNA-DNA sequence with the Ty3 PPT sequence (hybrid 2) yielded better quality crystals. The best crystals were obtained by the hanging drop vapor diffusion method and the optimal crystallization condition contained 0.1 M Tris (pH 8.5), 0.2 M ammonium sulfate and 17% PEG 3350. Before data collection, crystals were cryoprotected by stepwise addition of 50% glycerol to the crystallization drop to a final concentration of 25% and flash frozen in liquid N<sub>2</sub>.

X-ray diffraction data for the selenomethionine crystal (at Se peak wavelength 0.979 Å) were collected at beamline 14.1 at BESSY on a MAR 225 CCD detector at 100 K. Data were processed and scaled by XDS<sup>42</sup>. Data collection statistics are given in **Table 1**. The structure was solved by the single anomalous diffraction (SAD) method using the AutoSol module of Phenix<sup>43</sup>. Iterative building with COOT<sup>44</sup> was performed, and refinement was performed in Phenix with TLS (Translation-Libration-Screw).  $R_{\text{free}}$  was calculated with 5% of unique reflections. In the final model, 99.1% of the residues are within the allowed regions of the Ramachandran plot. Structural analyses, including superpositions and structural figures, were prepared in PyMol (<http://www.pymol.org/>). Coordinates of the structure have been deposited in the Protein Data Bank under the accession code 4OL8.

**Substrate binding assays.** For substrate-binding assays we used hybrid 3, with a 27 bp double-stranded region (RNA: 5'-AACAGAGUGCGACACCUGAUUCC

AUGACU and DNA: 5'-AGTCATGGAATCAGGTGTGCGCACTCTG). Ty3 RT was mixed with RNA-DNA hybrids at 2:1 molar ratio and applied to a Superdex 200 column (GE Healthcare) equilibrated in 20 mM HEPES (pH 7.0), 150 mM KCl, 5% glycerol and 1 mM DTT. The eluted species were monitored by  $E_{260}$  and  $E_{280}$ . Molecular weight of those species was determined using the multi-angle light scattering method on Optilab T-rEX and Dawn Heleos II (Wyatt Technology Corporation, USA).

**RNA-dependent DNA polymerase assays.** HIV-1 RNA template, prepared by *in vitro* transcription, was purified by denaturing polyacrylamide electrophoresis, followed by electroelution and precipitation. Purified RNA was mixed with an equimolar amount of a 5' Cy5 labeled DNA oligonucleotide complementary to nt 98–113 of the HIV-1 genome (5'-Cy5-CAGACGGGCACACACTAC; IDT, Coralville, IA) in 10 mM Tris, pH 7.8, 25 mM KCl and annealed by heating to 95 °C for 2 min followed by slow cooling to 4 °C. The polymerization reaction contained 200 nM template-primer, 200 μM dNTPs, 10 mM Tris (pH 7.8), 130 mM NaCl, 9 mM MgCl<sub>2</sub>, 5 mM DTT and 10% glycerol. DNA synthesis was initiated by adding enzyme to a final concentration of 400 nM and was allowed to proceed at 30 °C for the indicated times. Aliquots were quenched with an equal volume of 7 M urea and 1× TBE, heated to 95 °C for 2 min, and polymerization products were fractionated by denaturing PAGE. The gel was imaged on a Typhoon Trio + Imaging system with Image Quant Total Lab software (GE Healthcare, Piscataway, NJ).

**RNase H assays.** RNA and DNA oligonucleotides were purchased from IDT. RNA (40-mer: 5'-Cy5- UCAUGCCCGUCUAGCUACUCGAUAUGGCAAUAAGAC UCCA) was hybridized to DNA (28-mer: 5'- TGGAGTCTTATTGCCATATCG AGTAGCT) in 10 mM Tris (pH 7.8), 25 mM KCl and annealed by heating to 85 °C for 3 min followed by cooling to 4 °C at 0.2 °C per second. The reactions contained 750 nM RNA-DNA, 10 mM Tris (pH 7.8), 150 mM NaCl, 9 mM MgCl<sub>2</sub>, 5 mM DTT and 10% glycerol. Hydrolysis was initiated by adding enzyme to a final concentration of 675 nM (or for the mixture of two variants, 337.5 nM each of R140A R203A and D426N) and proceeded at 30 °C for the indicated times. Samples were processed and visualized as described above. Original images of gels used in this study can be found in **Supplementary Figure 8**.

**Sedimentation velocity.** Sedimentation velocity experiments were performed in a Beckman-Coulter ProteomeLab XL-I analytical ultracentrifuge, equipped with AN-50Ti rotor (8-holes) and 12 mm path length, double-sector charcoal-Epon cells, loaded with 400 μL of samples and 410 μL of buffer (20 mM HEPES-KOH, pH 7.0, 100 mM KCl and 0.5 mM EDTA). WT protein and variants R140A R203A or R441A R442A were mixed with RNA-DNA hybrid at 1:2.5 molar ratio. The experiments were carried out at 4 °C and 48,000 r.p.m. using continuous scan mode and radial spacing of 0.003 cm. Scans were collected in 6 min intervals at 260 nm. The fitting of absorbance versus cell radius data was performed using SEDFIT software, version 14.3e<sup>45</sup> and continuous sedimentation coefficient distribution  $c(s)$  model, covering range of 0.1–10 S.

Biophysical parameters of the buffer: density  $\rho = 1.00639$  g/cm<sup>3</sup> (4 °C), viscosity  $\eta = 0.01567$  poise (4 °C), and proteins: partial specific volume  $V\text{-bar} = 0.7418$  cm<sup>3</sup>/g (20 °C), and  $V\text{-bar} = 0.7352$  cm<sup>3</sup>/g (4 °C), were calculated using SEDNTERP software (version 1.09, <http://www.jphilo.mailway.com/download.htm>).

42. Kabsch, W. Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
43. Afonine, P.V., Grosse-Kunstleve, R.W. & Adams, P.D. The Phenix refinement framework. *CCP4 Newsl.* **42**, contribution 8 (2005).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
45. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).