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

El bieta Nowak[1], **Jennifer T. Miller**[2], **Marion K. Bona**[2], **Justyna Studnicka**[1], **Roman H. Szczepanowski**[3], **Jakub Jurkowski**[1], **Stuart F. J. Le Grice**[2], and **Marcin Nowotny**[1]

[1]Laboratory of Protein Structure, International Institute of Molecular and Cell Biology, Warsaw, Poland [2]Reverse Transcriptase Biochemistry Section, HIV Drug Resistance Program, Frederick National Laboratory, Frederick, MD, USA [3]Biophysics Core Facility, International Institute of Molecular and Cell Biology, Warsaw, Poland

Abstract

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). However, while 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. More strikingly, our structure and biochemical data suggest that the RNase H and DNA polymerase activities are contributed by individual subunits of the homodimer.

INTRODUCTION

Retrotransposons are mobile genetic elements that replicate through an RNA intermediate, and are divided into two groups, depending on the presence of flanking long-terminal repeat (LTR) sequences. Retrotransposons represent one of the most potent forces shaping the architecture of eukaryotic genomes [1]. For example ~40% of the human genome is derived from retroelements with 8% corresponding to the LTR class [2], while in maize, ~75% of the genome is derived from retroelements, mainly of the LTR class [3]. Retroviruses, such as

human immunodeficiency virus (HIV), evolved from LTR elements through acquisition of an envelope gene, allowing egress from infected cells to initiate a subsequent round of infection [4].

Ty3 element of *Saccharomyces cerevisiae* belongs the Gypsy family [5,6] and its RT is perhaps the most extensively characterized LTR-retrotransposon enzyme with respect to its enzymatic activities [7,89], and the architecture of nucleic acid duplexes with which it interacts [10–15]. Although structural motifs mediating substrate recognition and catalysis generally resemble those of vertebrate retroviral RTs, a notable difference between the Ty3 and retroviral enzymes is separation of its DNA polymerase and RNase H active sites by ~13 bp [9], as opposed to 17–18 bp for lentiviral and gammaretroviral enzymes. While the structural basis for such spatial separation is established for HIV-1 RT [16,17], the origin of the shorter distance for Ty3 RT is difficult to rationalize based on the retroviral structures. Ty3 RT lacks the connection, or tether, between its DNA polymerase and RNase H domains. Structural similarity between this subdomain of HIV-1 RT (which lacks the catalytic carboxylates) and its RNase H domain originally suggested the latter arose through domain duplication, while an alternative theory proposes the functional RNase H domain was acquired from a source outside the LTR retrotransposons [18].

Another well characterized LTR element from *Saccharomyces cerevisiae* is Ty1 of the Copia-like group, which is more closely related to retroviruses. However, the polypurine tract (PPT) primers for (+) strand synthesis for both Copia and Gypsy family 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* [19,20] and *in vitro* [9,21], and it has also been proposed that a Ty3 RT-integrase fusion protein participates in reverse transcription *in vivo* [19,22].

Despite extensive biochemical characterization of LTR-retrotransposon RTs, detailed structural information is lacking. Therefore we set out to perform the structural characterization of 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 the spatial separation between the DNA polymerase and RNase H active sites, in addition to phenotypic mixing experiments, suggests 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 Materials and Methods. Selenomethionine-substituted protein was purified by immobilized metal affinity, ion exchange and gel permeation chromatography. Purified enzyme was co-crystallized with a 16 bp RNA-DNA hybrid containing a 2 nt 5′ overhang in the RNA strand, sequence of which corresponded to the Ty3 PPT with the cognate RNase H cleavage site located 12 nt from the 3′-end. 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 [9]. The structure was solved by single wavelength anomalous diffraction method and refined at 3.1 Å resolution (Table 1, Fig 1a) to an R_{free} of 29.6%. Sample experimental electron density maps are shown in Supplementary Figure 1a, b.

Although Ty3 RT was previously reported as monomeric in solution in the absence of nucleic acid [12], 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 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 it is less ordered.

For ease of comparison, we labeled secondary structure elements using the scheme of our previous work on XMRV RT [23] (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 [23,24]. 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, supporting the hypothesis that evolution of retroviral RTs from LTR retrotransposon enzymes involved converting their RNase H domain to a "connector" with loss of catalytic function and recruitment of a new RNase H1 domain [18].

Ty3 RT subunits A and B are identical in sequence and structures of individual subdomains are very similar. Their pairwise superpositions result in low root-mean-square deviations (rmsds) of the positions of pairs of C-α atoms: 0.5 Å for 95 C-α atom pairs of fingers subdomains, 0.9 Å for 116 pairs of palm subdomains, 1.3 Å for 54 pairs of thumb subdomains, and 1.0 Å for 80 pairs of RNase H domains. Both fingers-palm fragments are also structurally similar (rmsd of 2.6 Å over 227 pairs of C-α atoms). However, pronounced differences are apparent in positioning of the RNase H and thumb subdomains (Fig 1b). Positioning of the RNase H domain between the two subunits can be accommodated by a large, ~90° rotation around an axis going roughly through the contact point between the subunit A palm and thumb. Consequently, the subunit B RNase H domain is positioned between its fingers and palm, blocking the DNA polymerase substrate binding cleft, inducing displacement of the thumb subdomain from the palm and its rotation 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 [25].

The subunit interface of the Ty3 RT dimer is quite polar, involving two main contact points. The first is formed by inserting 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; 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 region involves both RNase H domains (Fig 1d): Arg413(A) interacts with the backbone of His68(B), Thr452(A) and His417(A) with Arg441(B), and Asp448(A) with

Ser429(B) and Arg442(B). Arg413 and Arg442 are conserved among other Gypsy retroelements (Supplementary Fig 2), the latter of which may be also essential for interaction with the DNA backbone.

RNase H domain

RNase H and retroviral connection subdomains adopt the RNase H fold, the most important element of which is the five-stranded, central β-sheet [26]. The first three strands are longer, running antiparallel to each other, while the last two are shorter and parallel to the first. The fold also contains two or three α-helices on one side of the central sheet and a single helix on the other. Comparing cellular RNases H1 and closely related retroviral RNase H domains (collectively referred to as 'cellular RNases H1') with the Ty3 RNase H and retroviral connection subdomains highlights two main differences (Fig 2, Supplementary Figures 3 and 4). Firstly, there is a deletion of ~10 residues between the first two strands of the central β-sheet of the LTR-retrotransposon enzyme, shortening the first strand in its C-terminus (Fig 2a, c). A second difference is arrangement of α-helices between strands 4 and 5. Supplementary Note provides a detailed comparison of substrate-binding residues between cellular and Ty3 RNases H. 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 cannot be identified in Ty3 RNase H, especially those forming the "phosphate-binding pocket" [27,28].

Since neither RNase H active site of the dimer interacts with the RNA, an important mechanistic question is which Ty3 RT subunit contributes RNase H activity and the conformational changes 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 [28] and assumes that the Ty3 RNase H active site interacts with nucleotides –13 or –12, the preferred cleavage sites in 3′-end directed cleavage mode. Bringing the active site of the subunit A or B RNase H domain 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, evidenced by major conformational differences between subunits A and B. Subunit B RNase H domain is located closer to the scissile phosphate, and its movement (likely together with its thumb subdomain) could be accommodated by a ~40 Å translation without invoking severe clashes and 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 its thumb subdomain A, implying that subunit B RNase H domain contributes activity, a postulate 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 [17] although their exact nature is difficult to predict and elucidation of a similar issue for HIV-1 RT has been achieved only recently with crystallography[17].

The requirement for conformational changes conducive to substrate cleavage implies that RNA hydrolysis would be infrequent, agreeing with published experiments examining

RNase H activity concurrent with DNA synthesis [9]. Hydrolysis was rare during DNA synthesis occurring primarily after the enzyme reached 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 RT. For HIV-1 RT, the RNA-DNA substrate must undergo unwinding to allow RNase H cleavage [17]. For monomeric XMRV RT, the RNase H domain is tethered to its connection by a flexible linker, and thus very mobile. Only in the presence of RNA-DNA it becomes transiently organized on the substrate [23]. This feature of RNase H domains possibly regulates 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 minor groove width 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 represent a good model of the PPT structure. Its orientation in the structure would correspond to (–)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 [17,23]. At the resolution of our structure (3.1 Å) subtle changes in nucleic acid conformation may be not be apparent, but we favor the notion that Ty3 PPT recognition reflects dynamic properties of the duplex, possibly lower flexibility, rather than pre-existing 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 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 suggested 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 [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, explaining the extended DNase I footprint and indicating that the subunit B thumb could further stabilize RNA-DNA beyond interactions observed here.

Fig. 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 –5 of the DNA strand (positions –12 to –8 relative to the PPT-U3 junction, as used in ref [11]). The second region comprises interactions between DNA nucleotides –10 to –14 (–4 to +2 relative to the PPT-U3 junction) and residues from both RNase H domains. This bipartite substrate interface supports biochemical data obtained with modified Ty3 PPT substrates containing nucleoside analogs designed to either enhance flexibility or increase rigidity of the hybrid [11]. 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 to 11 nucleotides upstream towards the 5′ end of the RNA strand, corresponding to the portion of the RNA-DNA forming extensive interactions with the subunit A thumb. Supporting our structure, nucleotides between these regions were 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) and characterized previously for monomeric gammaretroviral RTs [23,29] and the heterodimeric HIV-1 enzyme [30,31]. 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 DNA strand 3′-OH is located at the DNA polymerase active site of subunit A, whose configuration resembles that of RTs from retroviruses [16,23] (Supplementary Fig 6), with key carboxylate residues coordinating two divalent metal ions [7]. The fact that Ty3 RT subunit A polymerase domain and active site are superimposable with HIV-1 RT demonstrates that this subunit likely 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 [16,23,32]. 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 was 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 also predicted involvement of thumb residues Tyr298(A) and Gly294(A) with DNA nucleotides –3 and –4, supporting and extending mutagenesis analysis of HIV-1 RT [33–36]. 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 [12], Ty3 RT eluted as monomer in the absence of substrate, with a molecular weight of 53.1 kDa *vs* the expected 54.6 kDa (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 the protein monomer and RNA-DNA hybrid (Fig 4a). The molecular weight of this complex, was 119 kDa *vs* the 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 Ala substitutions in the region involved in dimer formation, namely Arg140 and Arg203 (Fig 1c), and a second with Ala substitutions of Arg441 and Arg442. Arg441(B) and Arg442(B) participate in dimer formation (Fig 1d), while 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 subunit B residues participate in the substrate interface, while in subunit A they are located distal from the RNA-DNA binding cleft. Therefore, our experiments should only assess 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 MW 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 on a template derived from the 5′-terminal region of 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), 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 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 following initial phosphodiester bond formation. 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 recessed 3′ DNA terminus to monitor 3′ – end-directed cleavages when the DNA 3′-OH occupies the polymerase active site. As previously reported [12], we observed cleavage products 13 nt downstream of the DNA 3′-end and less prominent products resulting from an internal cleavage mode ~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 based on 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, while 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 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 important differences between Ty3 and HIV-1 RT. Firstly, for the LTR-retrotransposon enzyme, substrate binding is a pre-requisite to dimerization, while the lentiviral enzyme is a stable dimer in its absence [24,37,38]. Secondly, only the p66 subunit of the HIV-1 RT heterodimer contains a copy of the RNase H domain, thus 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, thumb, together with connection or RNase H domains of these subunits or proteins can be superimposed with an rmsd of 2.1 Å (320 C-α atom pairs) for Ty3 *vs* XMRV and 2.9 Å (237 C-α atom pairs) for Ty3 *vs* HIV-1 (PDB ID: 1RTD [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 due 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 surprisingly similar. Moreover, their arrangement is strikingly analogous (Fig 5c, d). The p51 connection is rotated and placed between its palm and fingers analogous to the Ty3 subunit B RNase H domain. Since 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 β-sheets 8, 9 and 10 at the C-terminus of the palm subdomain are unfolded in HIV-1 while 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 due to the presence of its RNase H domain, as described previously [23]. 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, towards 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 of an asymmetric homodimer induced by substrate binding. A ~13nt separation between the 3'-end of the DNA and the RNase H active site observed for Ty3 RT in biochemical experiments was difficult to rationalize based on retroviral RT structures, because the active site of the RNase H modeled on 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 [39]. Although we observe no major distortion of the RNA-DNA 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 which has been detected in virus-like particles [19], which may facilitate folding of RT.

While 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 [40]. 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 more open subunit A-like conformation, allowing dimerization. Another unique and, to our knowledge, unprecedented feature of Ty3 RT structure is the fact that its asymmetry allows for the separation of enzymatic activities between subunits and bringing the RNase H domain into position to mediate hydrolysis. Whether this applies to related LTR-retrotransposon enzymes remains to be determined.

ONLINE METHODS

Protein purification

Ty3 RT was cloned into an expression vector with a 3C protease cleavage site between His-tag and the protein. His$_6$-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 rpm 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 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, 10% glycerol and applied to a Superdex 200 size exclusion column (GE Healthcare). Peak fractions were concentrated using 10 kDa cut-off centricon (Milipore) 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′-CAGGTGTCGCACTCTG-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 16 bp duplex portion and 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 step-wise addition of 50% glycerol to the crystallization drop to a final concentration of 25% and flash frozen in liquid N_2.

X-ray diffraction data for the selenomethionine crystal (at Se peak wavelength 0.979Å) was collected at beamline 14.1 at BESSY on MAR 225CCD detector at 100 K. Data were processed and scaled by XDS [42]. Data collection statistics are given in Table 1. The structure was solved by single anomalous diffraction (SAD) method using AutoSol module of Phenix [43]. Iterative building with COOT [44] was performed, and refinement was performed in Phenix with TLS (Translation-Libration-Screw). R-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 27 bp double-stranded region (RNA: 5′-AACAGAGUGCGACACCUGAUUCCAUGACU and DNA: 5′-AGTCATGGAATCAGGTGTCGCACTCTG). 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 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 minutes 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 $MgCl_2$, 5 mM DTT and 10% glycerol. DNA synthesis was initiated by adding enzyme to a final concentration of 400 nM and allowed to proceed at 30°C for the indicated times. Aliquots were quenched with an equal volume of 7 M urea and 1X TBE, heated to 95°C for 2 minutes, and polymerization products fractionated by denaturing polyacrylamide gel electrophoresis. 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-UCAUGCCCUGCUAGCUACUCGAUAUGGCAAUAAGACUCCA) was hybridized to DNA (28-mer: 5′- TGGAGTCTTATTGCCATATCGAGTAGCT) in 10 mM Tris (pH 7.8), 25 mM KCl and annealed by heating to 85°C for 3 minutes 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 $MgCl_2$, 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 rpm 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 [45] and continuous sedimentation coefficient distribution c(s) model, covering range of 0.1 – 10 S.

Biophysical parameters of the buffer: density ρ = 1.00639 g/cm^3 (4°C), viscosity η= 0.01567 poise (4°C) and proteins: partial specific volume V-bar = 0.7418 cm^3/g (20°C), and V-bar =0.7352 cm^3/g (4°C), were calculated using SEDNTERP software (version 1.09, http://www.jphilo.mailway.com/download.htm).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

1. Cordaux R & Batzer MA The impact of retrotransposons on human genome evolution. Nat Rev Genet 10, 691–703 (2009). [PubMed: 19763152]

2. Lander ES et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001). [PubMed: 11237011]

3. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y & Bennetzen JL The paleontology of intergene retrotransposons of maize. Nat Genet 20, 43–5 (1998). [PubMed: 9731528]

4. Malik HS, Henikoff S & Eickbush TH Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res 10, 1307–18 (2000). [PubMed: 10984449]

5. Llorens C et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 39, D70–4 (2011). [PubMed: 21036865]

6. Sandmeyer SB, Aye M & Menees TM Ty3: a position-specific, gypsylike element in Saccharomyces cerevisiae. (eds. Craig NL, Craigie R, Gellert M & Lambowitz AM) 663–682 (ASM Press, Washington DC, 2002).

7. Bibillo A, Lener D, Klarmann GJ & Le Grice SF Functional roles of carboxylate residues comprising the DNA polymerase active site triad of Ty3 reverse transcriptase. Nucleic Acids Res 33, 171–81 (2005). [PubMed: 15647500]

8. Bibillo A, Lener D, Tewari A & Le Grice SF Interaction of the Ty3 reverse transcriptase thumb subdomain with template-primer. J Biol Chem 280, 30282–90 (2005). [PubMed: 15944162]

9. Lener D, Budihas SR & Le Grice SF Mutating conserved residues in the ribonuclease H domain of Ty3 reverse transcriptase affects specialized cleavage events. J Biol Chem 277, 26486–95 (2002). [PubMed: 11994277]

10. Brinson RG et al. Probing anomalous structural features in polypurine tract-containing RNA-DNA hybrids with neomycin B. Biochemistry 48, 6988–97 (2009). [PubMed: 19449839]

11. Dash C, Marino JP & Le Grice SF Examining Ty3 polypurine tract structure and function by nucleoside analog interference. J Biol Chem 281, 2773–83 (2006). [PubMed: 16306041]

12. Rausch JW et al. Interaction of p55 reverse transcriptase from the Saccharomyces cerevisiae retrotransposon Ty3 with conformationally distinct nucleic acid duplexes. J Biol Chem 275, 13879–87 (2000). [PubMed: 10788512]

13. Turner KB et al. SHAMS: combining chemical modification of RNA with mass spectrometry to examine polypurine tract-containing RNA/DNA hybrids. RNA 15, 1605–13 (2009). [PubMed: 19535461]

14. Yi-Brunozzi HY, Brabazon DM, Lener D, Le Grice SF & Marino JP A ribose sugar conformational switch in the LTR-retrotransposon Ty3 polypurine tract-containing RNA/DNA hybrid. Journal of the American Chemical Society 127, 16344–5 (2005). [PubMed: 16305191]

15. Yi-Brunozzi HY et al. High-resolution NMR analysis of the conformations of native and base analog substituted retroviral and LTR-retrotransposon PPT primers. Chemistry & biology 15, 254–62 (2008). [PubMed: 18355725]

16. Huang H, Chopra R, Verdine GL & Harrison SC Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282, 1669–75 (1998). [PubMed: 9831551]

17. Lapkouski M, Tian L, Miller JT, Le Grice SF & Yang W Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. Nat Struct Mol Biol 20, 230–6 (2013). [PubMed: 23314251]

18. Malik HS & Eickbush TH Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11, 1187–97 (2001). [PubMed: 11435400]

19. Kirchner J & Sandmeyer SB Ty3 integrase mutants defective in reverse transcription or 3'-end processing of extrachromosomal Ty3 DNA. J Virol 70, 4737–47 (1996). [PubMed: 8676501]

20. Wilhelm M, Heyman T, Friant S & Wilhelm FX Heterogeneous terminal structure of Ty1 and Ty3 reverse transcripts. Nucleic Acids Res 25, 2161–6 (1997). [PubMed: 9153316]

21. Wilhelm M, Uzun O, Mules EH, Gabriel A & Wilhelm FX Polypurine tract formation by Ty1 RNase H. The Journal of biological chemistry 276, 47695–701 (2001). [PubMed: 11595735]

22. Nymark-McMahon MH, Beliakova-Bethell NS, Darlix JL, Le Grice SF & Sandmeyer SB Ty3 integrase is required for initiation of reverse transcription. J Virol 76, 2804–16 (2002). [PubMed: 11861848]

23. Nowak E et al. Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. Nucleic Acids Res (2013).

24. Kohlstaedt LA, Wang J, Friedman JM, Rice PA & Steitz TA Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256, 1783–90 (1992). [PubMed: 1377403]

25. Mous J, Heimer EP & Le Grice SF Processing protease and reverse transcriptase from human immunodeficiency virus type I polyprotein in Escherichia coli. J Virol 62, 1433–6 (1988). [PubMed: 2450209]

26. Nowotny M Retroviral integrase superfamily: the structural perspective. EMBO Rep 10, 144–51 (2009). [PubMed: 19165139]

27. Nowotny M, Gaidamakov SA, Crouch RJ & Yang W Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell 121, 1005–16 (2005). [PubMed: 15989951]

28. Nowotny M et al. Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. Mol Cell 28, 264–76 (2007). [PubMed: 17964265]

29. 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, 613–32 (2000). [PubMed: 10669612]

30. Boyer PL, Ferris AL & Hughes SH Cassette mutagenesis of the reverse transcriptase of human immunodeficiency virus type 1. J Virol 66, 1031–9 (1992). [PubMed: 1370546]

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, 27666–73 (1999). [PubMed: 10488107]

32. Sarafianos SG et al. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J 20, 1449–61 (2001). [PubMed: 11250910]

33. Beard WA 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, 28091–7 (1994). [PubMed: 7525566]

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, 19516–23 (1995). [PubMed: 7543900]

35. Bebenek K et al. A minor groove binding track in reverse transcriptase. Nat Struct Biol 4, 194–7 (1997). [PubMed: 9164459]

36. Powell MD et al. Residues in the alphaH and alphaI 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, 19885–93 (1999). [PubMed: 10391934]

37. North TW, Cronn RC, Remington KM, Tandberg RT & Judd RC Characterization of reverse transcriptase from feline immunodeficiency virus. J Biol Chem 265, 5121–8 (1990). [PubMed: 1690735]

38. Thomas DA & Furman PA Purification and kinetic characterization of equine infectious anemia virus reverse transcriptase. Biochem Biophys Res Commun 180, 1365–71 (1991). [PubMed: 1719980]

39. Rausch JW & Le Grice SF 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. Int J Biochem Cell Biol 36, 1752–66 (2004). [PubMed: 15183342]

40. Swapna LS, Srikeerthana K & Srinivasan N Extent of structural asymmetry in homodimeric proteins: prevalence and relevance. PloS one 7, e36688 (2012). [PubMed: 22629324]

41. Karplus PA & Diederichs K Linking crystallographic model and data quality. Science 336, 1030–3 (2012). [PubMed: 22628654]

REFERENCES FOR ONLINE METHODS

42. Kabsch W Xds. Acta Crystallogr D Biol Crystallogr 66, 125–32 (2010). [PubMed: 20124692]

43. Afonine PV, Grosse-Kunstleve RW & Adams PD 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, 2126–32 (2004). [PubMed: 15572765]

45. Schuck P Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78, 1606–19 (2000). [PubMed: 10692345]
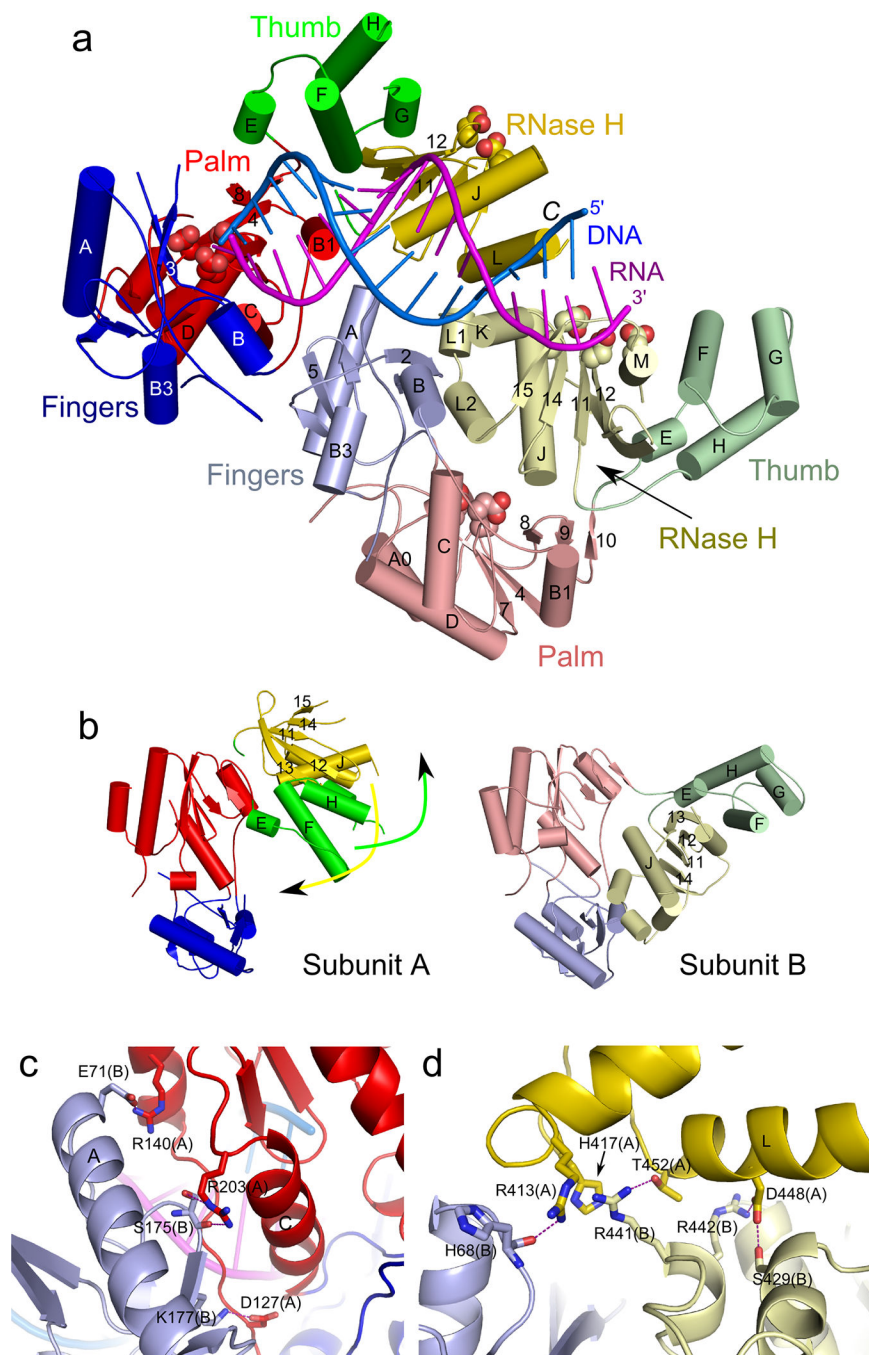
Figure 1. Overall structure of Ty3 RT and the dimer interface.
(a) Cartoon representation of Ty3 RT in complex with an RNA-DNA hybrid substrate. Protein subdomains are colored blue for fingers, red for palm, green for thumb, and the RNase H domain is in yellow. Lighter shades of the same colors are used 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 [23]. Residues forming the DNA polymerase and RNase H active sites are shown as spheres. (b) Comparison of the structures of subunits A and B. Arrows indicate the movements of thumb and RNase H domains that

transform its conformation to that of subunit B. (c, d) Residues involved in dimer formation. Protein structure colored as in (a).
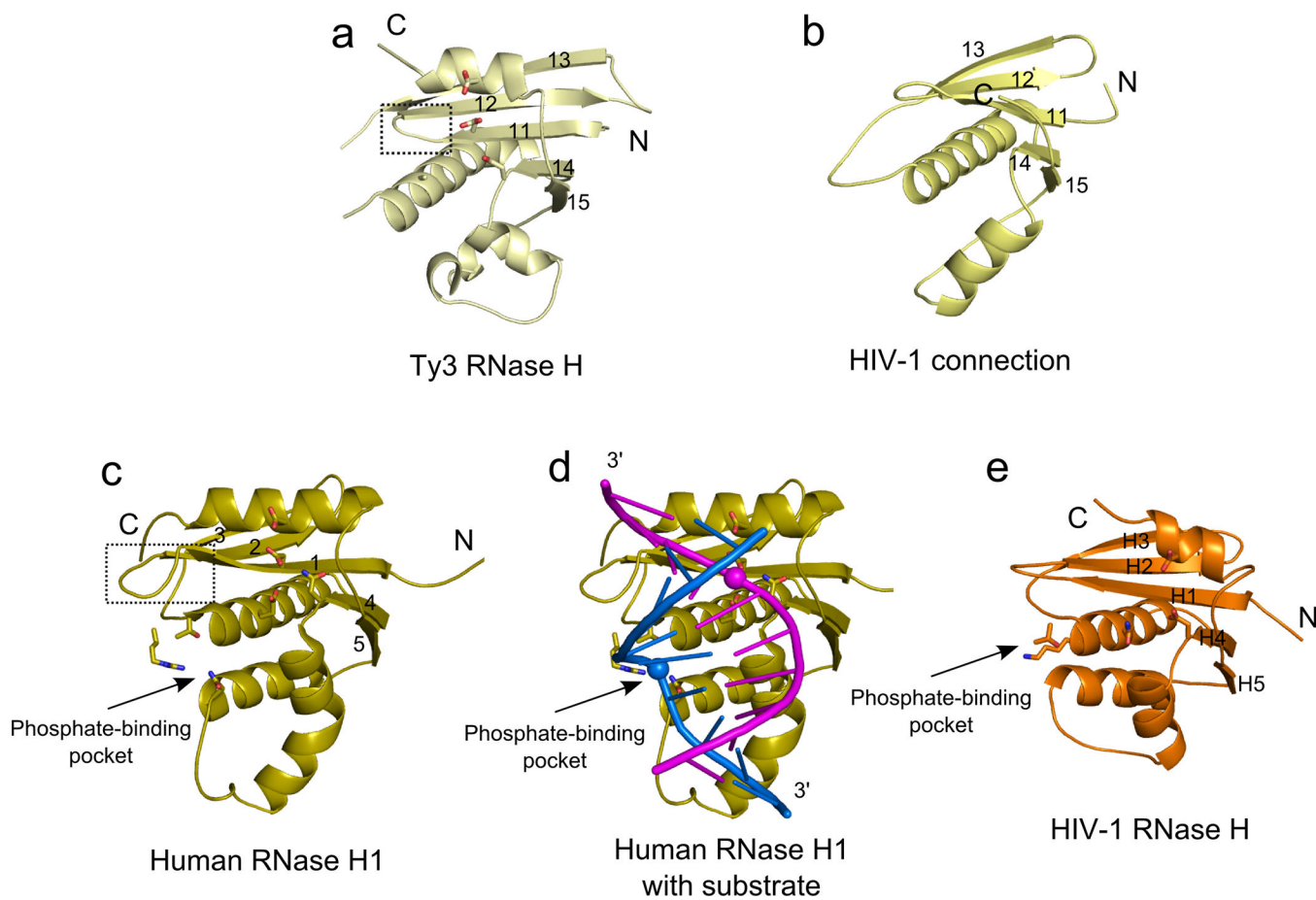
Figure 2. Comparison of RNase H and connection subdomains.
(a) Cartoon representation of Ty3 RNase H domain, (b) HIV-1 connection subdomain, (c) human RNase H1 (protein alone) (PDB ID: 2QK9 (ref 28)), (d) human RNase H1 with bound RNA-DNA substrate (PDB ID: 2QK9 (ref 28)), and (e) HIV-1 RNase H domain (from PDB ID: 1RTD [16]). Strands of the central β-sheet are labeled and residues forming the active site and phosphate-binding pocket are shown as sticks. The C-terminal region of the first β-strand, which differs in length between Ty3 RNase H and cellular enzymes, is indicated with a dashed box.
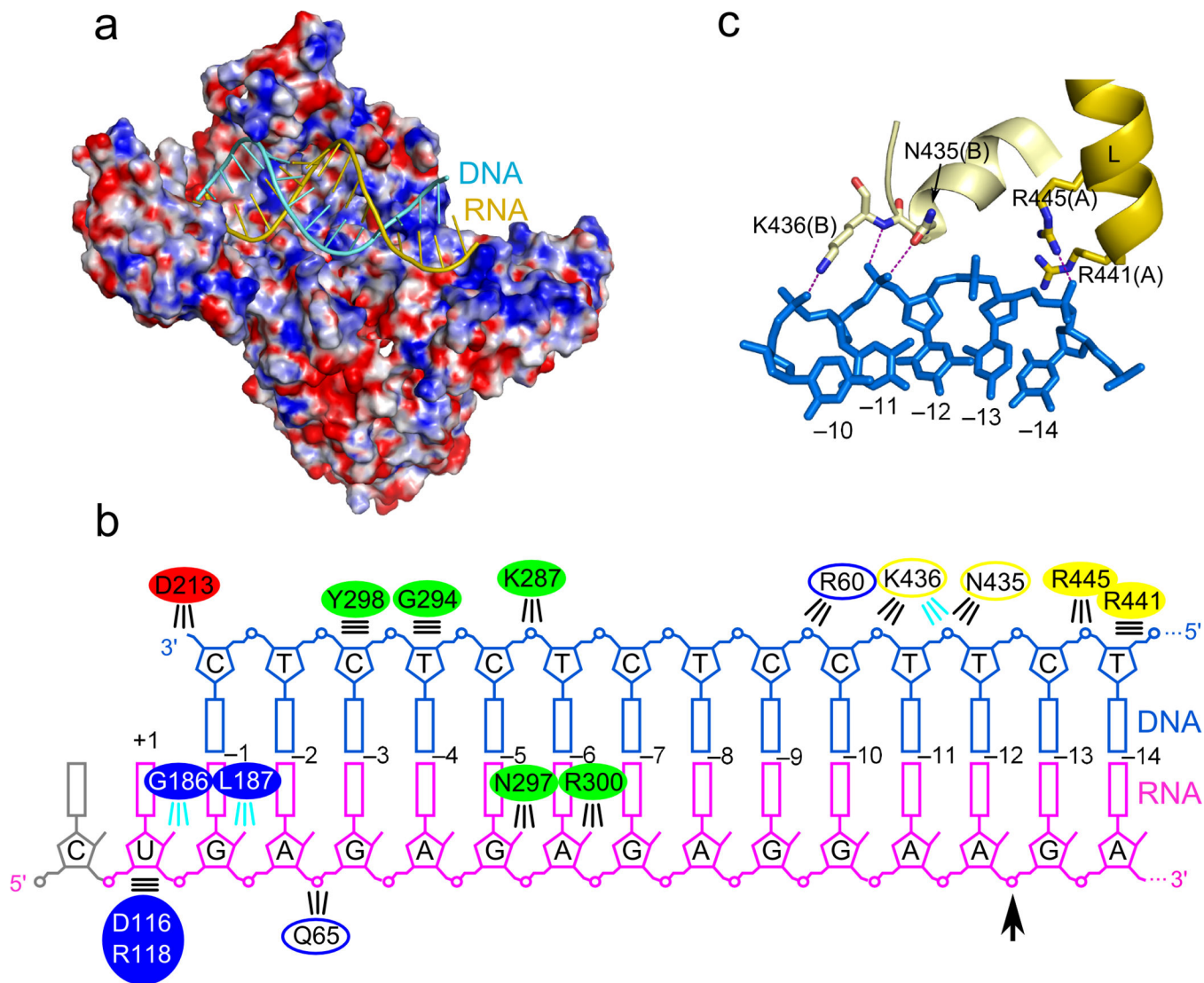
Figure 3. Substrate binding.
(a) Ty3 RT structure colored according to surface potential (red negative, blue positive – +/–
15 kTe). Nucleic acid is shown in cyan (DNA) and yellow (RNA). (b) Diagram of protein –
nucleic acid interactions. Arrow indicates the PPT-U3 junction (the preferred site of RNase
H cleavage). The 5′ RNA nucleotide not observed in the structure is shown in gray. The
ovals are colored according to protein domains using the color scheme of Figure 1. Solid
ovals denote subunit A and empty ovals subunit B. Parallel lines indicate van der Waals
interactions. Interactions mediated by the backbone of the protein are shown in cyan and
side chains in black. (c) Interactions between RNase H domains and the DNA strand.
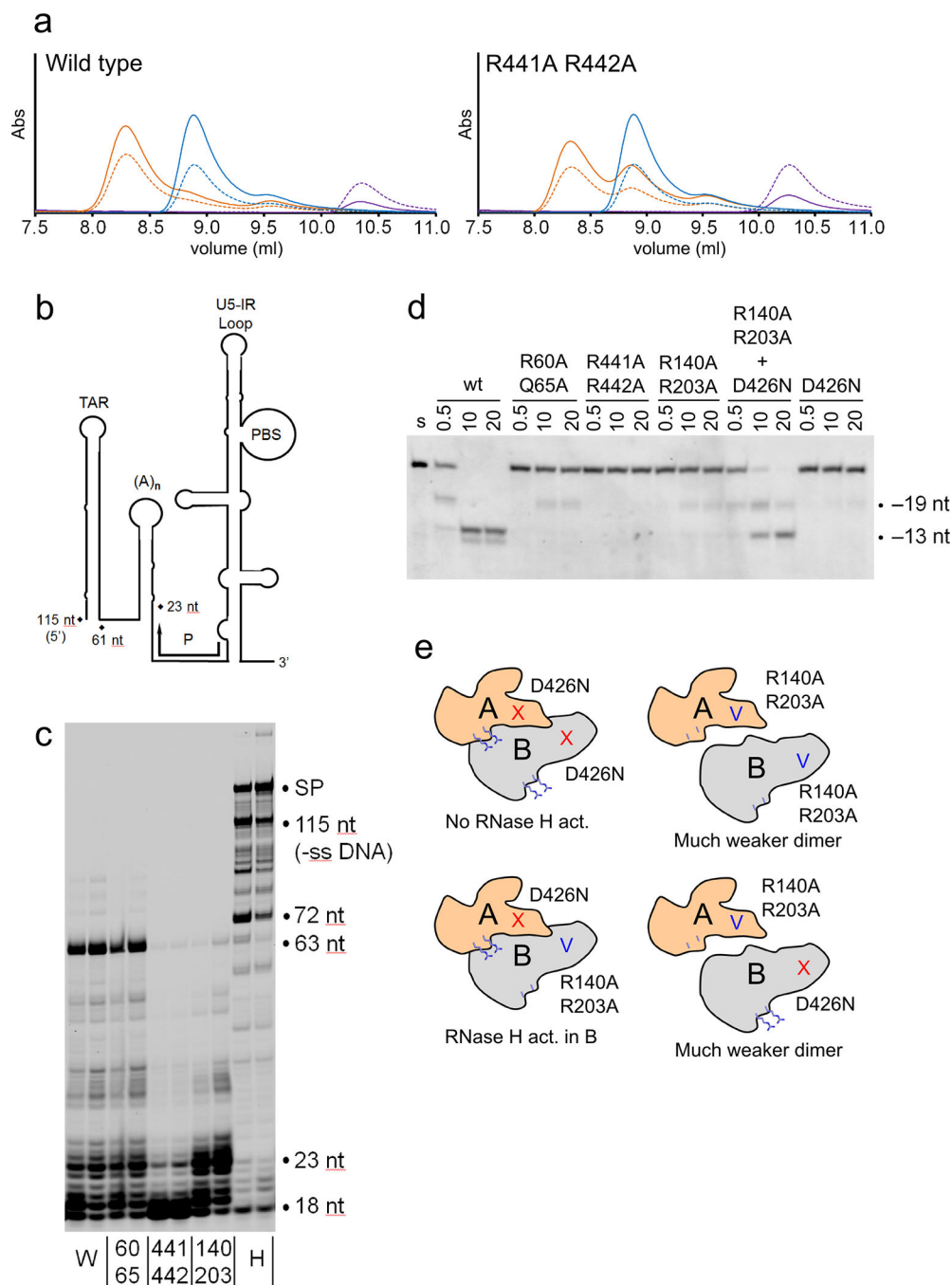
a

Wild type

R441A R442A

b

d

c

e

Figure 4. Biochemical experiments.

(a) GF experiments with hybrid 3 and wild-type Ty3 RT (left panel) or R441A R442A variant (right panel). Traces are shown in purple for protein, blue for hybrids and orange for the mixture. Dashed line represents E_{280} and solid line E_{260}. (b) Schematic representation of HIV-1 genome 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 - trans-activation response element. (c) DNA polymerization assays. Products of 10 and 20 minute reactions are shown. The major polymerase stalling products are marked on

the right of each panel. SP – self-priming product. (d) RNase H activity assays. Lane 's' contained uncleaved substrate with fluorescently end-labeled RNA. Hydrolysis was examined at 0.5, 10, 20 min (the lanes are labeled accordingly). The cleavage sites relative to the 3′ end of the DNA are indicated. (e) Cartoon of the phenotypic mixing experiment. Arg140 and Arg230 are schematically shown with blue sticks (small sticks for Ala variant) and the position of RNase H active site with blue 'V' (intact) or red 'X' (mutated). When variants R140A R203A and D426N are mixed, D426N can form a homodimer without the 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 diagram) and with the intact RNase H active site only in subunit B. Uncropped images can be found in Supplementary Figure 8.
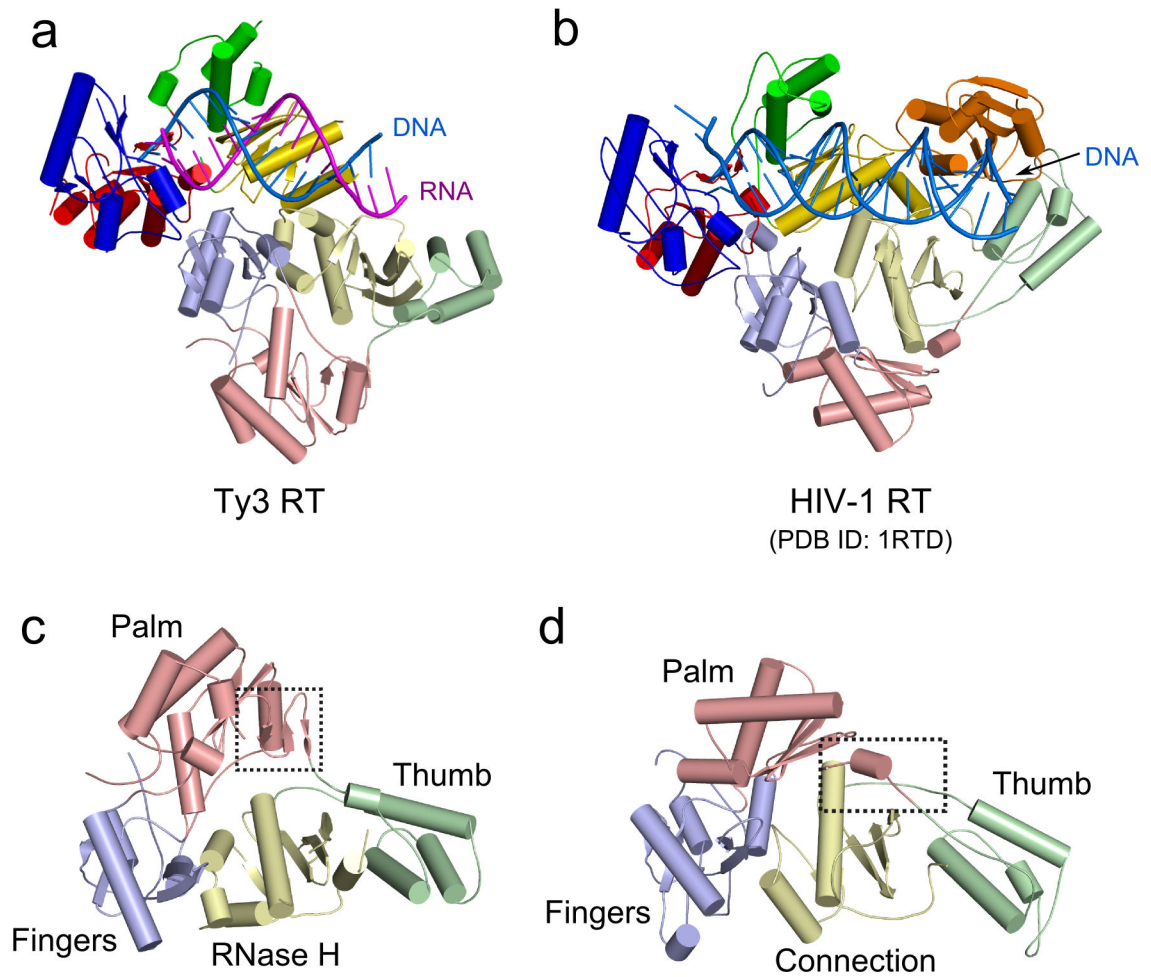
Figure 5. Comparison of Ty3 and HIV-1 RT (PDB ID: 1RTD [16]).
Comparison of the structures of Ty3 (a) and HIV-1 (PDB ID: 1RTD [16]) (b) RTs. The HIV-1
RNase H domain is shown in orange. (c, d) Comparison of subunit B of Ty3 RT with p51 of
HIV-1 RT. The region comprising short β-strands that are unfolded in HIV-1 RT is indicated
with a dashed box.

Table 1.

Data collection and refinement statistics of Ty3 RT - RNA-DNA complex crystals

	Ty3 RT SeMet
Data collection	
Space group	$P2_1\,2_1\,2$
Cell dimensions	
a, b, c (Å)	320.7, 75.1, 108.3
α, β, γ (°)	90, 90, 90
Resolution (Å)	5.0–3.1 (3.29–3.1) [*]
R_{merge}	0.10 (0.85)
$I/\sigma I$	11.9 (2.1)
$CC_{1/2}$ [**]	99.8 (72.9)
Completeness (%)	99.6 (98.4)
Redundancy	5.1 (5.1)
Refinement	
Resolution (Å)	3.1
No. reflections	90,832
R_{work} / R_{free}	22.7/29.6
No. atoms	14,105
Protein	12,734
Ligand/ion	1360/10
Water	1
B 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 is based on a single crystal

[*] Values in parentheses are for highest-resolution shell.

[**] $CC_{1/2}$ - 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 [41]