PNAS RESEARCH ARTICLE | BIOCHEMISTRY OPEN ACCESS

Human endogenous retrovirus-K (HERV-K) reverse transcriptase (RT) structure and biochemistry reveals remarkable similarities to HIV-1 RT and opportunities for HERV-K–specific inhibition

Eric T. Baldwin[a,1], Matthias Götte[b], Egor P. Tchesnokov[b], Eddy Arnold[c,d], Margit Hagel[a], Charles Nichols[e], Pam Dossang[e], Marieke Lamers[e,f], Paul Wan[e], Stefan Steinbacher[g], and Donna L. Romero[a]

Edited by Stephen Goff, Columbia University Irving Medical Center, New York, NY; received January 6, 2022; accepted April 28, 2022

Human endogenous retroviruses (HERVs) comprise nearly 8% of the human genome and are derived from ancient integrations of retroviruses into the germline. The biology of HERVs is poorly defined, but there is accumulating evidence supporting pathological roles in diverse diseases, such as cancer, autoimmune, and neurodegenerative diseases. Functional proteins are produced by HERV-encoded genes, including reverse transcriptases (RTs), which could be a contributor to the pathology attributed to aberrant HERV-K expression. To facilitate the discovery and development of HERV-K RT potent and selective inhibitors, we expressed active HERV-K RT and determined the crystal structure of a ternary complex of this enzyme with a double-stranded DNA substrate. We demonstrate a range of RT inhibition with antiretroviral nucleotide analogs, while classic nonnucleoside analogs do not inhibit HERV-K RT. Detailed comparisons of HERV-K RT with other known RTs demonstrate similarities to diverse RT families and a striking similarity to the HIV-1 RT asymmetric heterodimer. Our analysis further reveals opportunities for selective HERV-K RT inhibition.

antiretroviral drugs | drug design | mobile elements | repeat biology | repeatome

Significance

A large percentage of the human genome is composed of repetitive elements that are relics of past viral infections. Expression of these human endogenous retroviruses (HERVs) is associated with a variety of diseases, including cancer; however, causality remains to be established. A subset of these HERVs express proteins with reverse transcriptase (RT) activity. This has inspired several clinical studies of antiviral RT inhibitors for indications in which HERV expression is associated with disease. We have determined the X-ray structure of an HERV reverse transcriptase. This structure clarifies the reasons for poor inhibition by 3TC (lamivudine) and lack of inhibition by nonnucleoside inhibitors nevirapine and efavirenz. This structure will enable the design of selective HERV-K RT tools for drug target validation.

Fifty-four percent of the human reference genome is comprised of repeating sequences, the "repeatome." There is evidence that some of these repetitive elements have important functions during development (1) and emerging evidence that such repeats may contribute to disease processes (2, 3). Among these, human endogenous retroviruses (HERVs) are a class of repetitive elements that account for 8% of the genomic DNA with a genetic organization that follows the pattern of long terminal repeat retroviruses of which HIV-1 is a well-known distant family member (4). HERVs are derived from exogenous retroviral infections of germline cells, which are now propagated by normal Mendelian inheritance. Most of the copies of HERV elements in the genome have accumulated inactivating mutations and none are known to propagate replication-competent viruses in humans. However, some of the more recently integrated HERVs contain intact genes that are transcribed and translated into proteins, including reverse transcriptase (RT) (5). Since disease-associated RT activity can arise from diverse sources [exogenous viruses (6), long-interspersed nuclear elements, or a variety of HERVs (7)], pharmacologically specific tools are required to dissect repeat biology and attribute disease causality.

HERVs play a role in early development by rewiring the gene regulatory network of the preimplantation embryo (8). HERV expression appears to be a hallmark of the undifferentiated state, the acquisition of phenotypic plasticity, and stem cell character (9): traits associated with aggressive cancer and poor patient outcomes. HERV expression is tightly controlled in normal adult tissues but is reported to be aberrantly expressed in cancer (10), inflammatory diseases (11), neurological diseases (2), aging (12), and viral disease (13, 14). There are numerous reports of up-regulation of HERV-K (HML-2 [human endogenous MMTV-like] subtype)-derived messenger RNA (mRNA) and protein in a variety of solid and liquid tumor types (3, 5, 15–17). The disease association with endogenous retroviruses and the expression of HERV-encoded proteins during disease states suggests that antiretroviral therapy could be explored in the management of these conditions.

Although direct inhibition of HERVs by HIV-1 RT inhibitors has been reported (18–20), there has been no unambiguously positive clinical trial outcome with these inhibitors in non-HIV indications to date. Several clinical studies (https://clinicaltrials.gov/) are completed or on-going that employ various marketed combination antiviral products, including RT inhibitors, for the treatment of cancer, bone loss, primary biliary cholangitis, Aicardi-Goutières syndrome, psoriasis, multiple sclerosis, Alzheimer's

Author contributions: E.T.B., M.G., E.A., M.H., P.W., and D.L.R. designed research; E.P.T., C.N., P.D., M.L., P.W., and S.S. performed research; E.T.B. analyzed data; and E.T.B. wrote the paper with contributions from all the authors.

Competing interest statement: All the authors are either direct employees of ROME Therapeutics or are consultants or contractors to ROME Therapeutics.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: ebaldwin@rometx.com.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2200260119/-/DCSupplemental.

Published June 30, 2022.

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

disease, and amyotrophic lateral sclerosis (ALS). In some of these studies HERV RTs are the proposed therapeutic target. However, the level of HERV RT target engagement that can be achieved with these HIV drugs is unclear and may be insufficient; thus, there is a need for more potent selective HERV RT inhibitors to test therapeutic hypotheses.

To facilitate the development of HERV-K RT inhibitors, we determined the crystal structure of HERV-K HML-2 RT as a ternary complex with double-stranded DNA (dsDNA) and dNTP at 2.6-Å resolution. The structure reveals that this enzyme has a striking similarity to HIV-1 RT, even including the formation of an analogous asymmetric dimer. Interestingly, HERV-K RT is a homodimer, whereas HIV-1 RT is a heterodimer wherein the ribonuclease H (RNase H) domain of the catalytically inactive subunit (p51) is proteolytically removed. We provide a detailed comparison between HIV-1 and HERV-K RTs at the structural and biochemical levels. The collective data support the design of HERV-K RT-specific molecules for the potential treatment of diseases where the aberrant expression of HERV-K RT has been implicated.

Results

A previously documented active form of full-length HERV-K RT (21) was expressed for crystallization and structure determination. Size-exclusion chromatography (SEC) revealed that HERV-K RT activity was associated with two peaks approximately consistent with the expected retention volumes for monomeric and dimeric species. Unlike HIV-1 RT, neither peak showed evidence of proteolysis consistent with HIV-1 protease (PR)-mediated cleavage of the full-length HIV-1 RT (p66) to remove the RNase H domain and form a p51 truncated subunit. The apparent dimer fraction was advanced into crystallization and a slightly different construct into biochemical evaluation (*Materials and Methods*).

Biochemical Evaluation of Three Classes of HIV-1 RT Inhibitors: Nucleoside RT Inhibitors, Nonnucleoside RT Inhibitors, and Foscarnet. To assess whether the recombinant purified HERV-K RT was capable of reverse transcription, we monitored nucleotide incorporation events during DNA polymerization by HERV-K RT on a DNA/RNA primer/template (Figs. 1 and 2). Polymerase enzymatic activity was assessed using gel-based assays following the pattern of our previous work (22). We used HIV-1 RT DNA polymerization activity as a reference. Both enzymes were capable of extending the 19-nt DNA primer along the RNA template to generate 40-nt full template-length products (*SI Appendix*, Fig. S1*B*, lane 0). In addition, we tested three different classes of known HIV-1 RT inhibitors to characterize the similarities and differences between these RT enzymes.

Three common nucleoside RT inhibitors (NRTIs) were evaluated. Zidovudine-TP (AZT-TP, azidothymidine triphosphate), lamivudine-TP (3TC-TP), and carbovir-TP (CBV-TP) are the active triphosphorylated metabolites of the corresponding NRTIs (Fig. 1) [carbovir is the active metabolite of abacavir (23)]. These NRTIs lack a 3′-hydroxyl group; therefore, once incorporated into the primer during DNA synthesis, they act as obligate chain-terminators for subsequent nucleotide incorporations. All three triphosphate drugs inhibited both HIV-1 and HERV-K RTs, however with differing potencies (Fig. 1*C*). The normalized IC_{50} for 3TC-TP was 40-fold higher against HERV-K as compared to HIV-1 RT, while the corresponding difference in IC_{50} value for CBV-TP and AZT-TP was only sevenfold.

We also compared the patterns of inhibition of RT activity by HIV non-NRTI (NNRTI) nevirapine (NVP) and efavirenz (EFV) (Fig. 2). As expected, addition of increasing concentrations of NVP or EFV to the reactions with HIV-1 RT caused a signal reduction in the bands corresponding to the 40-nt product of the DNA synthesis and accumulation of the signal in the bands corresponding to the unextended 19-nt primer. In contrast, neither of the NNRTIs inhibited the RT activity of HERV-K RT.

Finally, we found that the pyrophosphate mimic foscarnet (phosphonoformic acid, PFA) inhibited DNA synthesis by HIV-1 and HERV-K RTs and induced similar patterns of polymerase pausing (*SI Appendix*, Fig. S1*B*). While the extent of the PFA inhibition was more pronounced with HIV-1 RT, both enzymes were completely inhibited by 100 μM PFA at the level of the full template-length product formation.

Distance between Polymerase and RNase H Active Sites. The RT-associated RNase H domain cleaves the RNA template of DNA/RNA replication intermediates. Here we employed an RNase H activity assay to compare cleavage patterns of HERV-K RT and HIV-1 RT. The DNA/RNA primer/template facilitates formation of a polymerase-competent complex with specific interactions between the 3′ end of the primer and the polymerase active site (24). The RT enzyme shuttles between pre- and post-translocational states that exist by definition immediately after nucleotide incorporation and before incorporation of the next nucleotide. The partitioning of the RT DNA/RNA complex between the two conformations affects the precise location of the RNase H active site on its RNA template substrate. The pretranslocated conformation (i.e., the N-complex) can be trapped in the presence of foscarnet (24–26). Under these conditions, the RNase H cleavage products are equivalent for HIV-1 RT and HERV-K RT, indicating that the two enzymes possess conformations with the same distance between polymerase and RNase H active sites (*SI Appendix*, Figs. S2–S4).

Determination of the HERV-K RT Structure. The structure of a ternary complex of HERV-K RT with dsDNA and ddTTP was solved by single anomalous dispersion (SAD) from multiple mercury sites (X144) from a Hg-dCTP derivative in space group $P6_1$ (*SI Appendix*, Table S1). This map allowed the positioning of subdomains of 1MU2.PDB (HIV-2 RT has a slightly higher percent identity to HERV-K RT over HIV-1 RT, 28.7% vs. 27.5%) (27) into the electron density map using Coot. After rebuilding the initial model, which consisted of a single ternary complex (protein dimer, dsDNA, and ddTTP), the new model for HERV-K RT was used to solve by molecular replacement a ternary complex (X105) with dsDNA (at higher resolution, 2.6 Å), which was then subjected to refinement and subsequent detailed analysis (*SI Appendix*, Table S1). The ternary complex was prepared for crystallization by combining HERV-K RT with dsDNA and ddTTP. Since the dsDNA primer/template was not terminated with a dideoxynucleotide, we expected that incubation with ddTTP would result in a homogeneously terminated primer strand in a binary complex. Before crystallization setup, this complex was further modified to include mixed dNTPs to create a ternary complex. During structure refinement, it was obvious from the electron density maps that two products were present in these maps. The data were reprocessed in space group $P3_1$ and two different ternary assemblies built into the asymmetric unit. Refinement

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

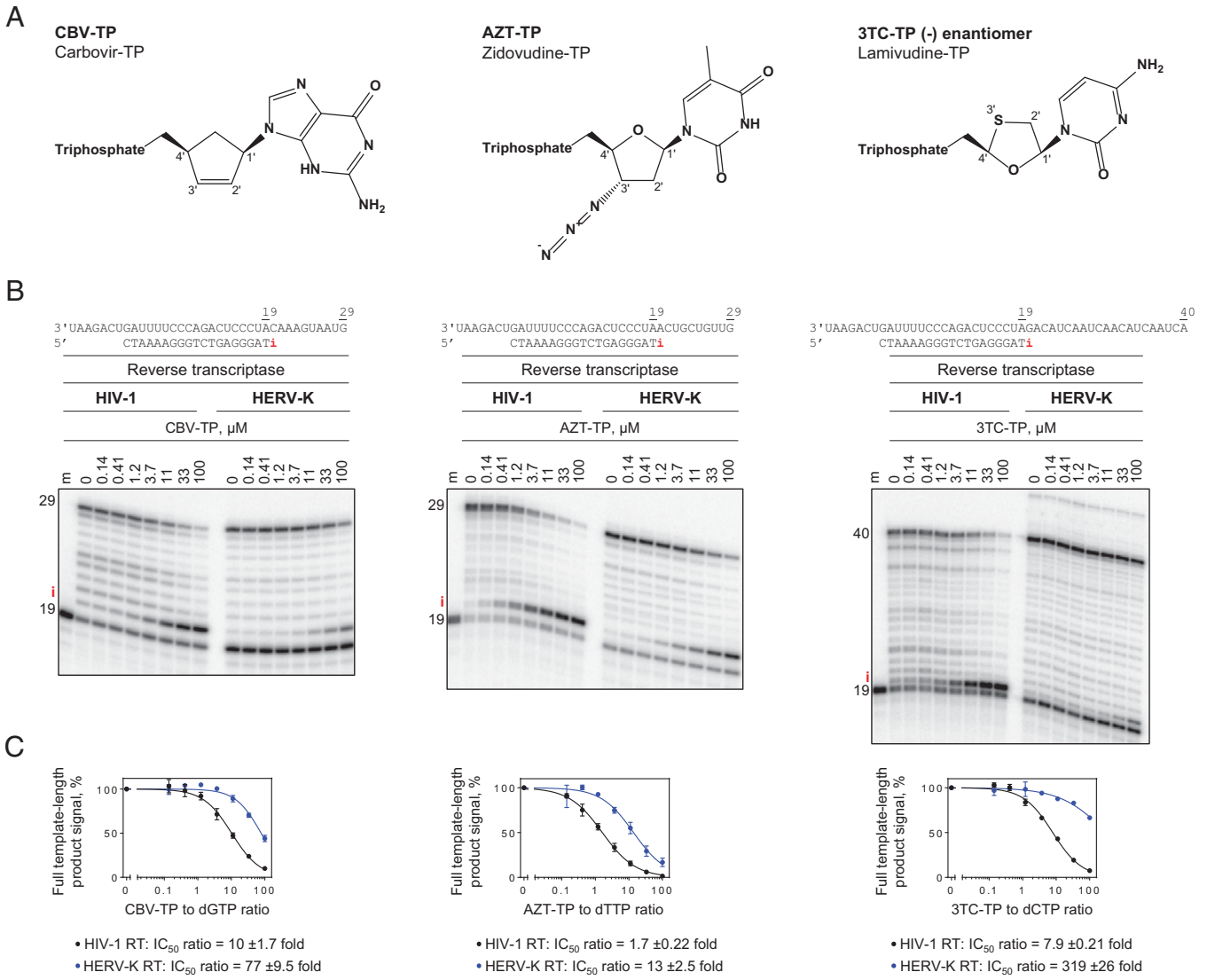

Fig. 1. Nucleotide analog-dependent inhibition of the DNA polymerization activity of HIV-1 and HERV-K RTs. (*A*) Chemical structures of nucleotide analogs CBV-TP, AZT-TP, and 3TC-TP. (*B*) Denaturing PAGE migration patterns of the products of DNA polymerization activity of purified HIV-1 and HERV-K RTs on a DNA/RNA primer/template as shown on top of each subpanel. The numbering on top of the RNA template indicates the length (nucleotide) of the primer (19 nt) and of the full template-length product (29 or 40 nt). The 5′-end of the primer is radiolabeled with ^{32}P to facilitate the detection of products. The red "i" indicates the unique site for the incorporation of the respective nucleotide analog. DNA polymerization activity was monitored at 37 °C for 10 min in the presence of dNTP mixture, $MgCl_2$, and the indicated concentrations of nucleotide analogs. Lane "m" illustrates the migration pattern of the 5′-end radiolabeled 19-nt long primer. (*C*) Graphical representation and data analysis used to determine the IC_{50} values of the nucleotide analogs shown in *B*. The experiments were repeated at least three times to determine the average and SD values.

using the twin law (-h-k l) was successful. The resulting asymmetric unit consists of two ternary complexes, where the first primer is terminated by G in the P-site with dTTP in the N-site and the second primer is terminated by T without 3′-OH in the P-site and dCTP in the N-site (*SI Appendix*, Fig. S9).

Comparison of HERV-K and Other RT Structures. HERV-K RT is an asymmetric homodimer in the ternary complex with subdomain arrangement and subunit interactions that are remarkably similar to HIV-1 RT despite the overall low sequence identity (*SI Appendix*, Fig. S5). Molecule A (molA) of HERV-K RT adopts the active polymerase/RNase H configuration of the p66 subunit of HIV-1, while molecule B (molB) adopts an alternative folding similar to HIV-1 p51. In HERV-K RT, the RNase H domain of molB is disordered and not present in the final refined model. We confirmed that the crystals contained a homodimer with no evidence of p51-like cleavage of the RNase H domain (as observed in HIV-1) (*SI Appendix*, Fig. S5). The subdomain organization of HERV-K RT follows that of HIV-1. The polymerase domain is composed of three subdomains: fingers (residues 20 to 96 and 124 to 166), palm (residues 97 to 123 and 167 to 252), and thumb (residues 253 to 335). The polymerase domain is linked to the RNase H domain by the connection subdomain (residues 336 to 456) and the C-terminal domain is the RNase H domain (residues 457 to 596). The color convention (blue-fingers, red-palm, green-thumb, yellow-connection, and orange-RNase H), as well as the secondary structure nomenclature of Jacobo-Molina et al. (28) (Fig. 3 and *SI Appendix*, Fig. S6), are employed here.

A detailed analysis of the HERV-K and HIV-1 RT structures is summarized in *SI Appendix* (notes and *SI Appendix*, Figs. S10–S15). Comparison of HERV-K RT to other RT structures reveals several common motifs and a few features that are unique to HERV-K RT. The HERV-K connection subdomain is more similar to the connection subdomain from the monomeric Moloney murine leukemia virus (MoMLV) RT. HERV-K and

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

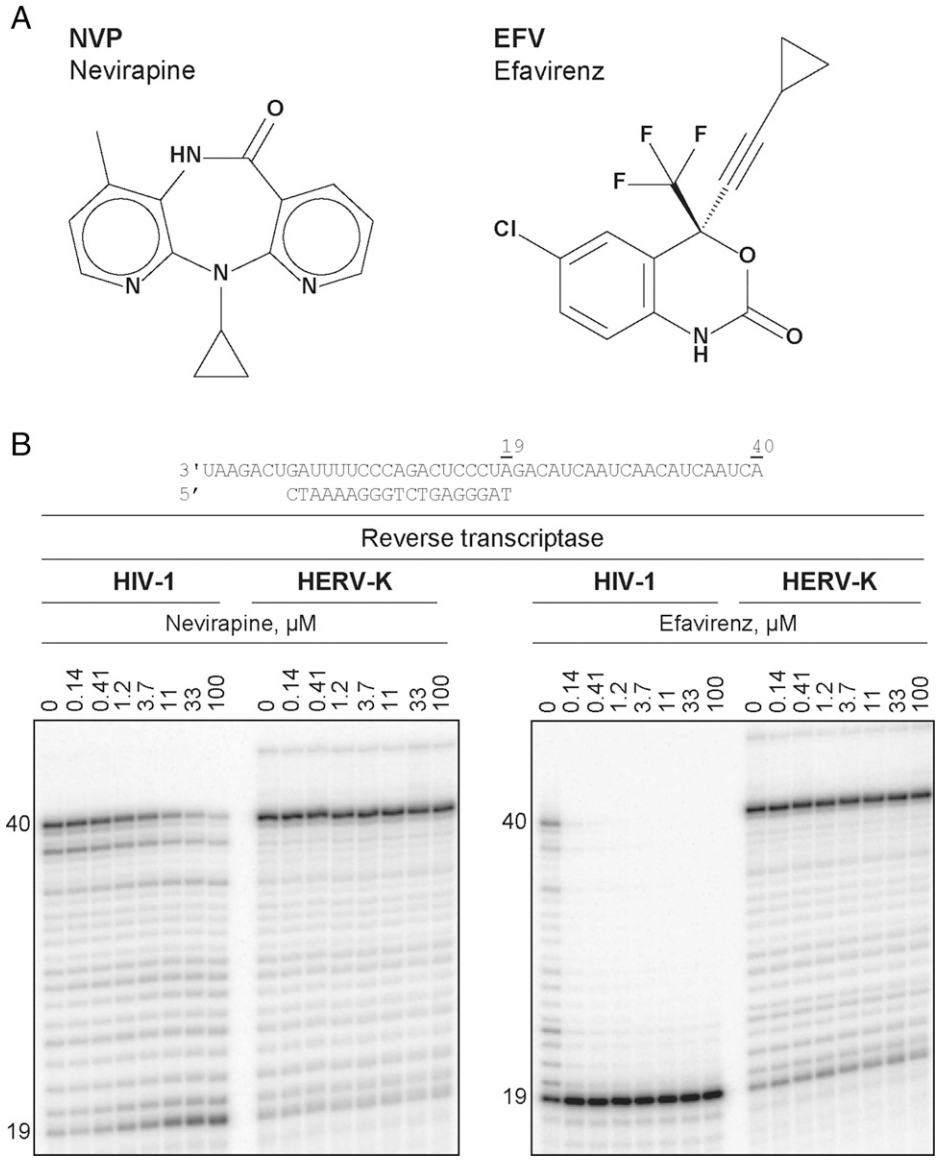

Fig. 2. NVP- and EFV-dependent inhibition of the DNA polymerization activity of HIV-1 and HERV-K RTs. (*A*) Chemical structures of NVP and EFV. (*B*) Denaturing PAGE migration patterns of the products of DNA polymerization activity of purified HIV-1 and HERV-K RTs on a DNA/RNA primer/template as shown on top of the panel. The numbering on top of the RNA template indicates the length (nucleotide) of the primer (19 nt) and of the full template-length product (40 nt). The 5′-end of the primer is radiolabeled with ^{32}P to facilitate the detection of products. DNA polymerization activity was monitored at 37 °C for 10 min in the presence of dNTP mixture, $MgCl_2$, and the indicated concentrations of NVP or EFV. In the HERV-K RT experiments in Fig. 1*C* and in these gels (*B*), a uniform band is present at the top of the gel. We attribute this to a DNA form that did not enter the gel.

MoMLV have the αK helix in common with HIV-1, but also have an additional helix αL1 (Fig. 4*B*: *Left* molA, *Right* molB). The HERV-K molB connection has an additional helix (αK0) that precedes the dominant αK helix of the connection subdomain and is similarly positioned to the equivalent αK0 helix of MoMLV RT (Fig. 4 *B*, *Right*). The HERV-K molB connection subdomain terminates with a unique helix (αL2), which makes hydrophobic interactions with the molB thumb (Fig. 4*C*). This terminal helical interaction with the thumb is very similar to the yeast Ty3 RT, where a helix from the RNase H domain of the Ty3-p51–like subunit contacts the thumb (Fig. 4*D*). In HERV-K there is also a unique short β-sheet interaction between the molA connection subdomain and the RNase H domain (Fig. 4*A*). Finally, the palm subdomain of molB terminates with a prominent and HERV-K unique helix (αF2) that extends toward the thumb of molB (Fig. 3*C*). Thus, while HERV-K RT includes all the subdomain features common to other RT enzymes, it seemingly shares a few motifs with MoMLV and Ty3 RT and displays some unique secondary structural elements.

As might be expected given the catalytic mechanism of RTs, the interactions between the enzyme, primer/template nucleic acid, and dNTP in the ternary complex of HERV-K RT reflect similar if not identical interactions as observed in HIV-1 RT catalytic complexes. The polymerase active site (N-site) is where the incoming dNTP binds base pairs with the template strand cognate base, and a dNMP nucleotide is incorporated into the primer strand. One Mg^{+2} ion is observed in this site coordinated with the triphosphate and with the catalytic aspartic acid residues (D121, D196, and D197). Activation of the primer strand 3′-OH leads to the formation of a covalent bond with the incoming dNTP and loss of pyrophosphate. Polymerase side chains K77, R83, Q162, and K230 contact the phosphates of the dNTP substrate, while R83 also makes an interaction with the face of the nucleotide base (*SI Appendix*, Fig. S7*A*). On the primer side, the interactions with the primer strand are

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

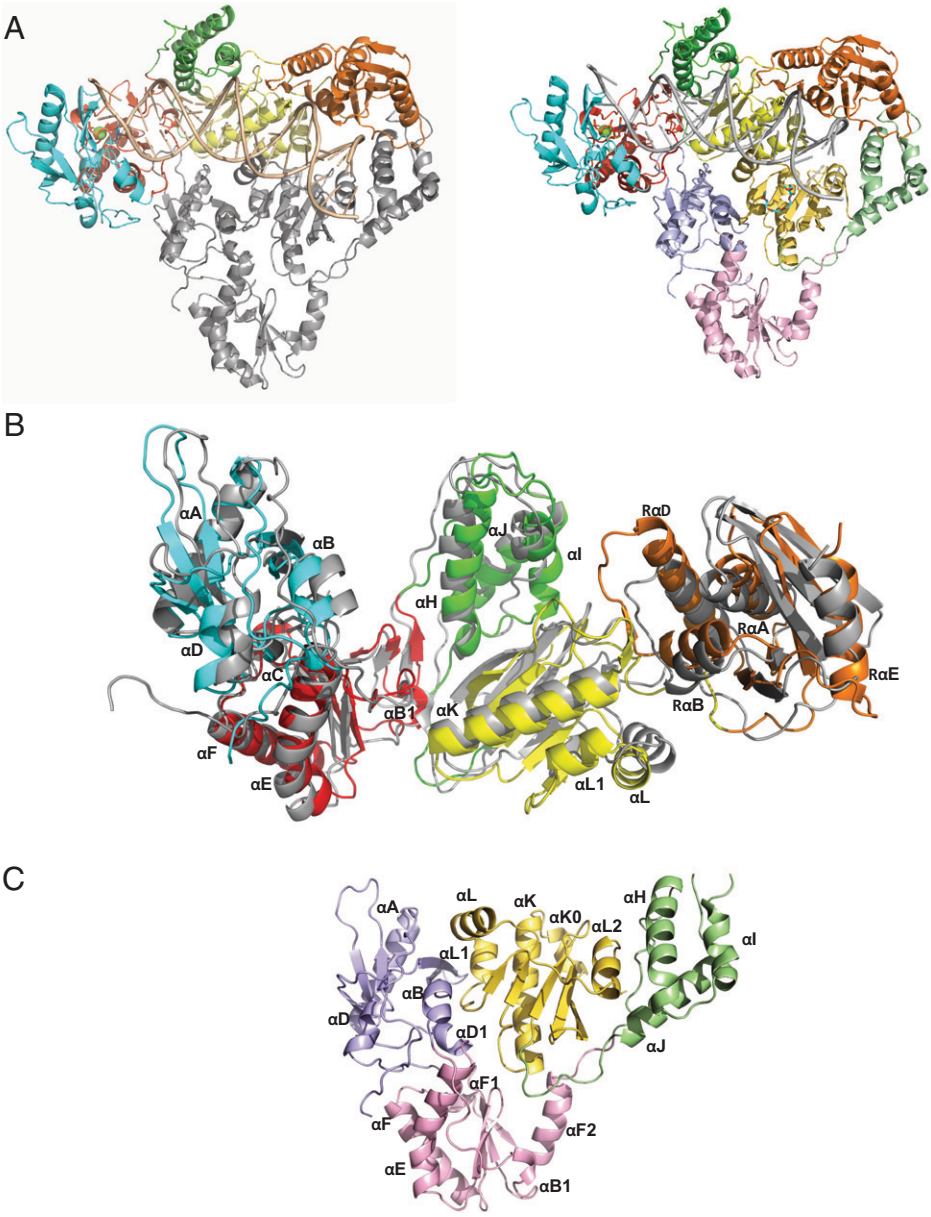

Fig. 3. Overview of HERV-K RT structure. (*A*) HERV-K RT with the molA subdomains colored by convention and the molB subunit colored in gray (*Left*). On the *Right*, molA and molB are colored by subdomain. MolA: fingers (blue), palm (red), thumb (green), connection (yellow), RNase H (orange). MolB: fingers (periwinkle), palm (pink), thumb (light green), connection (mustard). (*B*) Superimposition of HIV-1 (5TXL.PDB) (in gray) upon HERV-K RT molA with an RMSD of 2.67 Å for 381 Cα atoms. HERV-K mol A: fingers (blue), palm (red), thumb (green), connection (yellow), RNase H (orange). The α-helices of HERV-K are indicated and the full secondary structure analysis is found in *SI Appendix*, Fig. S6. (*C*) MolB: fingers (periwinkle), palm (pink), thumb (light green), connection (mustard). The α-helices of HERV-K are indicated and the full secondary structure assignments are found in *SI Appendix*, Fig. S6. αF2 of the molB palm is a unique helix in HERV-K. The equivalent stretch in molA is a coil which is similar to HIV-1 RT.

via the same motifs but often involving alternative residues in HERV-K RT when compared to HIV-1 RT. The exception is W276/W266 from the base of the thumb subdomain, which contacts the deoxyribose ring of the P-2 site. Of note, the YMDD sequence in the palm of HIV-1 RT contains catalytic residues D184 and D185, and is equivalent to YIDD in HERV-K RT. The YIDD motif (I195) is positioned beneath the deoxyribose ring of the P-site and the primer grip is positioned near the P-1 site (Fig. 5*A* and *SI Appendix*, Fig. S7*B*). In HIV-1 RT, methionine to isoleucine or valine mutations are associated with drug resistance to 3TC and other nucleotide analogs (29–31). On the template side, G163/G152 is positioned beneath the P-site template deoxyribose ring, while P168/P157 is positioned underneath the P-1 template deoxyribose ring. Several conserved hydrophobic amino acids (W38, F73, I75, and L85) are found to interact with the single-stranded template overhang. Interestingly, these residues from the finger's subdomain are completely conserved between HIV and HERV-K RT (*SI Appendix*, Fig. S7*C*). Previous work in HIV-1 shows that at least one of these (F61) is involved in polymerase translocation. In HIV-1 RT F61A (HERV-K RT F73) causes a strong bias to the posttranslocational complex and is resistant to PFA (32). Finally, the RNase H active sites of HIV-1 and HERV-K RT contain several identical residues and are arranged very similarly to the respective polymerase active sites (*SI Appendix*, Fig. S7*D*).

NRTI Activity and Structural Differences between HIV-1 and HERV-K RT. The active site residues of HIV-1 RT and HERV-K RT are very similar, but important differences impact NRTI

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

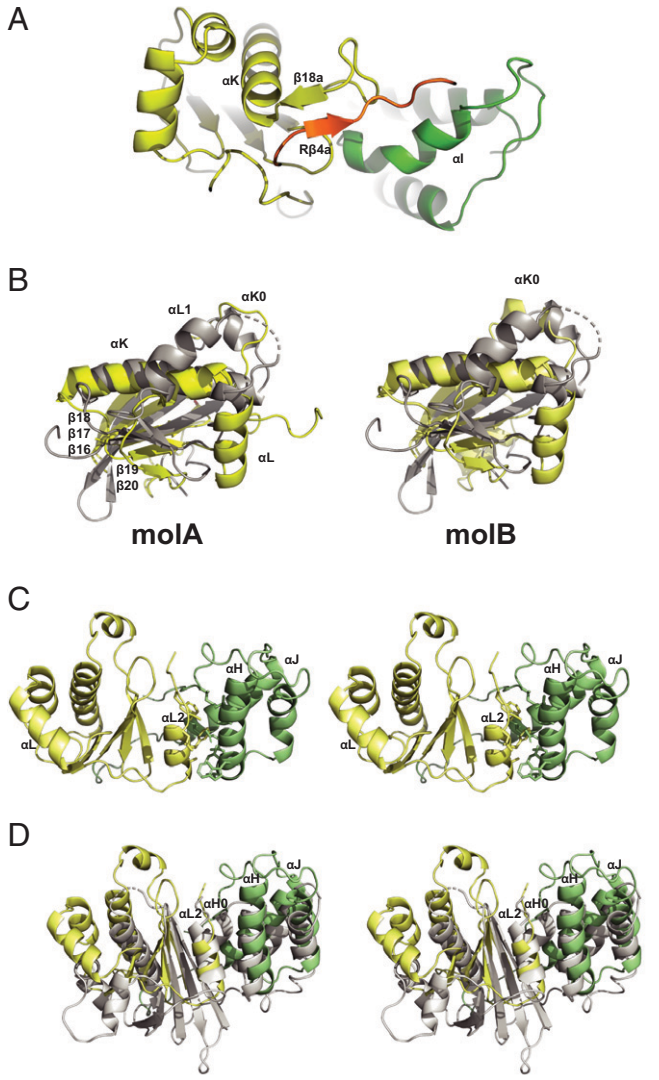

Fig. 4. Secondary structural elements from HERV-K RT that are notable. (*A*) A unique short stretch of β-sheet composed of a strand from molA connection subdomain (β18a, see *SI Appendix*, Fig. S6) and a strand from molA RNase H domain (Rβ4a). Since the RNase H domain of molB is not experimentally observed, this interaction is not present in molB. (*B*) The connection subdomain of molA and molB compared to the connection subdomain (gray) from the monomeric enzyme MoMLV RT (4MH8.PDB). Yellow ribbons model on (*Left*) is the connection domain from HERV-K RT molA. MoMLV has a similar αL1 helix but displays a short αL helix. MoMLV also has a short helix that precedes αK (αK0), which is not present in molA of HERV-K. However, HERV-K does have the equivalent αK0 helix in molB (*Right*). These connection features of MoMLV are also true of the related structure of XMRV RT (4HKQ.PDB). (*C*) HERV-K RT molB connection subdomain includes a C-terminal helix. This helix, αL2, packs against the molB thumb subdomain especially αH and αJ. A stereoview shows the hydrophobic residues contributing to the interface. (*D*) The αL2 helix of HERV-K RT demonstrates a similar interaction to the molB thumb as is observed in Ty3 RT (4OL8.PDB). A stereoview of the HERV-K RT (mustard and light green) with the Ty3 RT in gray illustrates the similarity. The Ty3 RNase H domain interaction with the thumb is equivalent to the connection subdomain/thumb interaction of HERV-K molB.

susceptibility. The residues in HIV-1 that surround the azido group of AZT (33) (3KLG.PDB) are identical between HIV-1 and HERV-K RT, with the exception of the gatekeeper (see below). AZT maintains respectable potency against HERV-K RT (13 μM) and Y115F is not a common driver of AZT resistance in HIV-1 (https://hivdb.stanford.edu/). The gatekeeper [selects against NTPs with 2′-OH (34, 35)] is F126 in HERV-K and Y115 in HIV-1 RT. Furthermore, the central hydrophobic residue of the HIV-1 YMDD loop is isoleucine 195 in HERV-K (YIDD). It is known from HIV-1 RT drug-resistance studies that these two HERV-K equivalent residues (F126 and I195) can modulate potency of NRTIs. We can therefore surmise that these two differences play a role in the differential sensitivity of HERV-K RT to nucleoside drugs used to treat HIV-1 infection. In particular, YIDD is a known resistance mutation in HIV-1 RT for 3TC (36) and results in a greater than 100-fold loss in potency (30), which is completely consistent with our biochemical data in HERV-K RT. This loss in potency is believed to result from the steric clash between the oxathiolane ring of 3TC and the bulky β-branched amino acid, I184 (HIV-1 RT) (36). Furthermore, the lower potency of CBV-TP in HERV-K could also be explained by the corresponding HIV-1 drug-resistance data. Abacavir resistance of about eightfold arises from the combination of Y115F and M184V/I mutations (37). One could speculate that the oxygen to carbon substitution in the carbocycle results in steric clash with HIV-1 V/I184 (as in 3TC). It also has been observed that tyrosine is more tolerant of unfavorable aromatic stacking interactions (e.g., with the abacavir cyclopentene ring) than phenylalanine (38). In summary, the NRTI potency differences between HIV-1 and HERV-K can be rationalized from the HIV-1 drug-resistance mutational data.

Previous kinetic studies with HIV-1 RT in combination with site-specific footprinting have shown that PFA inhibits DNA synthesis with a bias toward pretranslocation (25). PFA inhibition patterns obtained with HIV-1 RT and HERV-K RT show similar “hot spots” for inhibition, which suggests a similar mechanism of action. However, inhibition of HERV-K RT requires higher concentrations of PFA. Structures of HIV-1 RT with DNA and PFA in the pretranslocated complex show that K65 and R72 interact with the inhibitor (39). The structurally equivalent residues in HERV-K RT are K77 and R83.

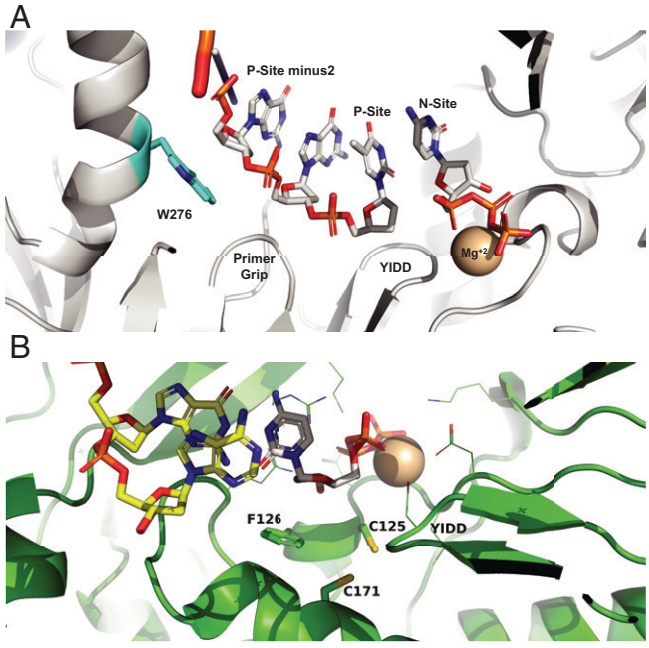

Fig. 5. Selected HERV-K surface sites that represent potential drug discovery opportunities. (*A*) The W276 (blue carbons) site is positioned at the base of the thumb subdomain and supports the ribose ring of the P-minus-2 nucleotide. The region directly below the tryptophan side chain and stretching into the plane of this view is quite large and has several hydrophobic residues. Motion of the thumb could be restricted by an allosteric inhibitor binding in this location. (*B*) Located directly beneath the N-site ribose ring are two Cys residues (molA C125 and C171). The sulfur atoms of these side chains are nicely positioned for covalent modification by various chemical warheads.

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

However, the A114S mutation in HIV-1 RT is associated with resistance to PFA and the equivalent residue in HERV-K RT is C125 (40). The similarity of C125 to the HIV-1 resistant A114S could help to explain the reduced inhibitory effect of PFA in HERV-K RT.

Analysis of Druggable Sites. We analyzed the HERV-K RT structure for potential druggable sites using a combination of visual inspection, knowledge of heavy metal-binding locations, SiteMap [small molecule binding hot spot identification tool (41)], analysis of the static structure, normal modes, molecular dynamics with SiteMap scoring, and mixed solvent molecular dynamics (41). We also examined the analogous allosteric inhibitor binding locations from HIV-1 RT (42–48). The NNRTI-binding site of HIV-1 RT is located between the YMDD loop and the primer grip. This location in HERV-K RT is well-packed and does not appear to be favorable for compound binding (*Discussion*). We also examined a site beneath the YIDD loop that is analogous to the NNRTI-adjacent site (43). Interestingly, this location does seem less well-packed and retains many of the identical residues of the equivalent HIV-1 RT site. Normal modes and molecular dynamics indicate structural breathing in this location and the appearance of solvent pathways to the buried location. This site is also interesting because it is connected to a constellation of cysteine residues at the base of the YIDD loop (molA C200, C211, and C189) that are partially solvent-exposed and C200 was one of the binding sites of our covalent mercury derivative. Another site revealed by SiteMap surrounds the W276 residue at the base of the thumb (Fig. 5*A*). This is an equivalent residue to W266 of HIV-1 RT where the inhibitor DAVP-1 was shown to bind (47). In HERV-K this location is more expansive with a large hydrophobic solvent-exposed surface below the tryptophan. Virtual screening of these locations could be an attractive source of lead matter.

Finally, the N-site itself in HERV-K affords a novel opportunity for covalent drug discovery. The floor of the N-site is uniquely occupied by two cysteine residues (molA C125 and C171). These residues are positioned such that covalent warheads from nucleoside or nonnucleoside inhibitors could form covalent interactions and freeze HERV-K RT in an irreversible complex with primer/template (Fig. 5*B*). We have modeled several such compounds and are pursuing this opportunity for important tool compounds to better understand HERV-K biology. Interestingly, the molB C171 location is also of interest. In HERV-K, this is a strong SiteMap hot spot with considerable volume. The molB C171 location was a mercury derivative binding site and is at the base of a deep surface pocket (*SI Appendix*, Fig. S8). We have tested this mercury compound in our enzyme assay and found that it is an inhibitor, validating the notion of cysteine-directed inhibition of HERV-K RT.

Discussion

Over the past few decades, a number of RT structures have been solved from class I [gammaretrovirus (49)], class II [lentivirus (28, 50)], and class III [spumaretrovirus (51, 52)] retroviruses. The structure of the yeast retrotransposon Ty3-gypsy RT has also been determined (53). These families of RTs employ a variety of reported assemblages to accomplish the retrotranscription task. We have extended these observations to the endogenous betaretrovirus family member HERV-K, which clusters with class II retroviruses. The first structure of an endogenous retroviral RT revealed an unexpected asymmetric homodimer quaternary structure.

Mouse mammary tumor virus (MMTV) is the prototypical betaretrovirus (54) with 45% sequence identity to HERV-K (HML-2) RT, and MMTV RT has been expressed, purified, and shown to be monomeric (55). The RT sequences of HERV-K and HIV-1 are only 27.5% identical and HERV-K lacks the tryptophan motif of HIV-1, including W401, which is important for assembly of HIV-1 dimers (56, 57), leading to a prediction that HERV-K RT would be monomeric. However, during purification of HERV-K RT, we observed the formation of dimers and we began to explore the possibility that HERV-K RT might form a dimeric structure similar to that found in lentiviruses, like HIV-1 RT (p66/p51). We used foscarnet to trap a defined ternary complex to measure the distance from the HERV-K polymerase active site to the RNase H active site (*SI Appendix*, Fig. S4) and the results were consistent with an HIV-like arrangement for HERV-K RT. Therefore, we concluded that HERV-K RT was likely a stable homodimer.

Retroviruses produce polycistronic mRNAs that are translated into polyprotein precursors, which are eventually processed by viral protease to form the mature proteins and enzymes. In HIV-1 and HERV-K, the RT enzyme is expressed as part of the Gag-Pol polyprotein, which encodes structural proteins (Gag) and enzymes (Pol: RT, protease, and integrase). Since active HIV-1 protease is a symmetric homodimer, various proposals have circulated regarding the nature of the dimeric forms of Gag-Pol and, specifically, whether the RT dimer was symmetric or asymmetric in advance of protease maturation. The prevailing view is that HIV-1 Gag-Pol homodimer is an asymmetric dimer (58–60), in which the RT dimer resembles the mature HIV-1 RT p66/p51 heterodimer. We hypothesized that the HERV-K RT homodimer may be similar to the dimeric forms of other RTs; however, the symmetry remained an open question (61, 62). Our structure of HERV-K RT is completely consistent with an asymmetric homodimer that closely resembles the mature asymmetric heterodimer of HIV-1 RT. While the p51-like subunit (molB) of the HERV-K RT ternary complex is not proteolytically processed (*SI Appendix*, Fig. S5), the RNase H domain of that chain is disordered in our structure. Potentially the second RNase H domain remains compact and flexibly tethered to the dimer in much the same way as the disordered RNase H domain of Xenotropic murine leukemia virus-related virus (XMRV) and MoMLV RTs (49, 63). Our structure is also consistent with the asymmetric dimer cryoelectron microscopy structure of HIV-1 Pol polyprotein that has recently been determined (60), in which the RT region is arranged very similarly to the asymmetric mature RT p66/p51 heterdimer.

HERV-K RT is not yet a validated drug target with strong causal connection to disease. Since there are many copies of HERV-K in the genome, target validation based on genetic manipulation is impractical. Therefore, the discovery of selective small-molecule HERV-K RT inhibitors that could be used in pharmacological target validation studies is critical. This was a primary motivation for our structure determination of HERV-K RT, enzymatic evaluation of known RT inhibitors, and analysis of potentially druggable sites. Our data suggest that structure-based discovery of HERV-K–specific molecules is feasible. In particular, we note the region around W276 (47) and the floor of the N-site, which includes two cysteine residues (C125 and C171). These are positioned to be reactive (Fig. 5*B*) and provide opportunities for development of covalent inhibitors (64–67).

Studies of NRTIs—such as AZT, 3TC, or abacavir—in cell-based assays are complicated by the requirement for biotransformation to the triphosphate and by having a range of known cellular targets. Thus, cell-based experiments should be evaluated carefully before attributing a direct effect to a specific target. In the present

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

work we have shown that HERV-K RT is directly inhibited by abacavir (CBV-TP is the active metabolite) and AZT-TP. Studies exist where effects of abacavir or AZT in cells or animals could be attributed to HERV-K RT. For example, the current clinical trial in ALS using Triumeq (abacavir, 3TC, and dolutegravir) (68) could plausibly provide HERV-K RT-mediated effects via inhibition by abacavir, but not 3TC. We find that 3TC is a poor inhibitor of HERV-K RT, in contrast to its potent HIV-1 RT inhibitory activity.

In contrast to previous reports (18–20), our data clearly show that HERV-K (HML-2) RT enzyme is not inhibited by NVP or EFV (Fig. 1). We examined the structures of HIV-1, HIV-2, and HERV-K RT to determine if our biochemical observations were consistent with the structural requirements for NVP or EFV inhibition in HIV-1 RT. The NNRTI-binding site in HIV-1 RT is located between the YMDD loop and the primer grip, and is usually only experimentally observed to be in an open state in the presence of a bound NNRTI compound. Computational methods support the transient opening of this site to accommodate NNRTI ligands (69). The reordering of hydrophobic side chains results in pocket formation, trapping of the compound, displacement of the primer grip, and a 5.5-Å shift of the 3′ end of the primer away from the polymerase active site (70). The HIV-1 RT complex with NVP is unlikely to accommodate ordered binding of dNTP at the N-site (71). The corresponding site between YIDD and the primer grip in HERV-K RT is nicely packed and our normal modes and molecular dynamics calculations do not reveal a pocket that could be a hot spot for compound binding. Resistance patterns for NVP and EFV are well known from studies of HIV-1 and HIV-2. In HIV-2, Y181I and Y188L display equivalent locations in the two structures (27); however, the loss of these aromatic side chains in HIV-2 (and mutants of HIV-1) eliminate ring-stacking interactions that make major contributions to the binding energy of NVP to HIV-1 RT (72). We note that HERV-K RT, like HIV-2 RT, lacks these aromatic residues that are critical for NNRTI binding. Hence, it is unlikely that these NNRTIs would be potent against HERV-K RT. Additionally, L100I, L234I, and K103N mutations in HIV-1 RT are associated with resistance to NNRTIs. These known resistance-mediating side chains are present in HERV-K RT, which also helps to explain the lack of HERV-K RT inhibition by NVP and EFV.

Materials and Methods

Expression, Purification, Crystallization, and Structure Determination of HERV-K RT. An N-terminal tobacco etch virus (TEV) cleavable 6-His tag construct of HERV-K RT (1–596) MAHHHHHHDY DIPTTENLYF QGK (1) thru A(596) AAC63291 (21) was codon-optimized, synthesized (Genscript), and cloned into pFastBac-1 for expression in baculovirus-infected Sf9 cells. A single glycine residue is expected as a cloning artifact in the final product. Recombinant protein expression was conducted on a 6-L scale following standard protocols for 72 h. Cells were harvested, lysed by sonication, and clarified by centrifugation. An initial immobilized metal-affinity chromatography (IMAC) batch capture and elution was followed by the removal of the His_6 tag on the amino terminus with TEV protease (at 1:10 [wt/wt] for 18 h at 4 °C). Following cleavage, the His-tag and uncleaved HERV-K RT were removed by negative IMAC. Cleaved HERV-K RT was further purified using a Superdex 26/60 200-pg (prep grade) gel-filtration column run at 2 mL/min (50 mM Tris pH 8.0, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM TCEP [Tris(2-carboxyethyl)phosphine] and 5% glycerol). The SEC column elution revealed two peaks that likely represented monomer and dimeric forms of HERV-K RT. The dimer peak was collected for crystallization and was buffer exchanged into higher salt buffer (50 mM Tris pH 8, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM TCEP and 5% [vol/vol] glycerol), concentrated to 6.4 mg/mL, flash frozen, and stored at −80 °C.

Template: 3′-dCdAdAdAdGdGdGdAdAdAdGdCdGdAdAdAdGdTdCdCdAdGdGdG

Primer: 5′-dGdTdTdTdCdCdCdTdTdTdCdGdCdTdTdTdCdAdGdG

These oligonucleotides for crystallization were purchased from Integrated DNA Technologies (IDT). The sequence of the template strand was based upon the HIV-1 tRNA primer-binding site (U5-PBS from the HIV-1 reference genome, NC_001802.1, nucleotides 195 to 218).

Frozen aliquots of HERV-K RT were thawed on ice, diluted to 4 mg/mL with higher salt buffer, mixed with 150 μM ddTTP and a 1.2:1 molar ratio of dsDNA (1.2 duplex:1 protein dimer), and incubated for 2 h. The complex was then diluted 1:1 with low-salt (50 mM NaCl) buffer to drop the overall NaCl concentration to 175 mM. The ternary complex was then reconcentrated to 7.5 mg/mL and used to set up crystallizations. The initial hit for HERV-K RT was obtained from an LMB screen (Molecular Dimensions) condition #E6 (10% PEG400, 50 mM MES pH 6.0, 2 mM $MgCl_2$ and 100 mM KCl), yielding a shower of small crystals. These crystals were soft, difficult to handle and generally diffracted to between 3- and 3.5-Å resolution.

Optimization screening was conducted using both grid screens varying pH and (PEG400), and a wide range of additive screens blending other components with the original condition. A reduction in nucleation and significant improvement in crystal size and visual quality was observed with the addition of several organic solvents, most notably 1,4-dioxane. All subsequent target crystallizations were set up as a single row of drops with well solution comprising LMB condition E6 plus variable amounts of dioxane.

Cryoprotection and crystal hardening regimes were also extensively screened, and it was found that a 5-min soak with well solution supplemented with mixtures of DMSO and 1,4-dioxane was both an effective cryoprotectant and also hardened the crystals, making them much easier to manipulate.

The X105 crystallization was set up using a three-drop Swissci crystallization plate and drops were dispensed by a Mosquito robot with 180 nL protein/dsDNA/ddTTP (prepared as above), then 20 nL 10 mM dNTPs (mixed dATP, dTTP, dGTP, and dCTP; Promega), where the latter was expected to occupy the N-site after translocation of the ddTTP terminated primer, then 200-nL well solution, and finally 20 nL of freshly prepared seedstock. X105 grew in LMB condition E6 +20% 1,4-dioxane, and the resulting crystal was ∼120 μm in diameter. Soaking solution was prepared as follows: 4 μL 1,4-dioxane, 2 μL dNTPs (10 mM each, 0.5-mM final concentration), 8 μL 4-bromo-pyrazole dissolved in DMSO (10 mM, 2 mM final concentration), 26-μL LMB screen condition #E6. Two microliters of this solution were layered onto the crystal drop and the plate was resealed for 5 min. Crystal X105 was then mounted using a 150-μM round litholoop (MiTeGen), immersed in liquid nitrogen, and sent to Diamond for data collection.

Crystal X144 was set up using three-drop Swissci crystallization plates and drops were dispensed by a Mosquito robot with 180-nL protein/dsDNA/ddTTP, then 20 nL 10 mM 5-mercuryacetate-2′-deoxycytidine-5′-triphosphate (Jena Bioscience), then 200-nL well solution, and finally 20 nL of freshly prepared seedstock. X144 grew in the LMB condition E6 +10% 1,4-dioxane condition and the resulting crystal was ∼80 μm in diameter. Soaking solution was prepared as follows: 4 μL 1,4 dioxane, 2 μL 10 mM 5-mercuryacetate-2′-deoxycytidine-5′-triphosphate; 8 μL DMSO, 26 μL LMB screen condition #E6. Two microliters of this solution were layered onto the crystal drop and the plate was resealed for 5 min. Crystal X144 was then mounted using a 100-μM diameter litholoop, immersed in liquid nitrogen, and sent to Diamond for data collection.

Data for both crystals were collected at Diamond beamline IO3 at the wavelength indicated in (*SI Appendix*, Table S1) using a Dectris EIGER2 XE 16M detector. Data for X144 was processed with DIALS (73) and data for X105 data were processed using AUTOPROC with STARANISO anisotropy correction (74). All datasets were auto-indexed and processed as $P6_1$ at Diamond, and X105 data were subsequently reprocessed in space group $P3_1$ since the electron density map was consistent with a twin (see below). The Matthews coefficient was most consistent with two molecules of HERV-K in the asymmetric unit ($P6_1$). The self-rotation function was flat in $P3_1$ or $P6_1$ with no evidence of twofold noncrystallographic symmetry that might be consistent with a symmetric homodimer.

Diffraction data for SAD phasing were processed in space group $P6_1$. Data between 64 and 3.0 Å were included. The positions of 12-Hg atoms were identified with SHELXD (75) and refined using PHASER (76). The figure of merit (FOM) for 41,840 reflections was 0.158. Density modification was performed with PARROT (77), which increased the FOM to 0.408. SHELXD, PHASER and PARROT were all implemented in the CCP4 suite (77). The phase quality decreased

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

significantly at around 4.5-Å resolution. The resulting electron density showed only a few clearly identifiable helical secondary structure elements but allowed domain-wise positioning of a starting model (1MU2.PDB) (27) using Coot (78). The structural model was rebuilt with Coot and refined with REFMAC5 (78).

The X105 data in $P6_1$ was refined with CCP4 suite (77) initially with the asymmetric dimer of HERV-K RT, dsDNA, the P-site occupied by dideoxy terminated thymidine and the N-site occupied by dCTP and a Mg^{2+} ion. Difference density on bases and the presence of a strong positive peak on the 3′-OH of the P-site nucleotide suggested that the primer ends with guanine in the P-site and ddTTP in the N-site. This alternative was built and refined but did not result in a clean difference map. The data were reprocessed in $P3_1$ and rerefined with REFMAC5 as a twin (-h,-k,l) with two ternary complexes in the asymmetric unit (HERV-K RT assembly1–unreacted dsDNA with dTTP in the N-site and HERV-K RT assembly2 –the primer terminated with dideoxythymidine and the N-site occupied with dCTP) (*SI Appendix*, Fig. S9). The coordinates have been deposited in the protein data bank as 7SR6.PDB.

HERV-K RT crystals were collected, dried and dissolved in SDS/PAGE loading buffer and the resulting protein was visualized (*SI Appendix*, Fig. S5). There is no evidence for proteolytic processing of the RNase H domain, as occurs with HIV-1 RT to produce the p51 subunit. Molecular weight markers were Invitrogen SeeBlue Plus 2 prestained protein standard markers from Thermo Fisher Scientific.

During initial purification of HERV-K RT protein for crystallization studies, various experimental attempts to facilitate the formation of an HIV-1 RT-like heterodimer were carried out without apparent proteolytic breakdown and release of the RNase H. HERV-K RT crude lysate from BV was incubated for a week at 4 °C and the His-captured RT was resolved on SDS/PAGE. Similarly purified HERV-K RT was incubated with both *Escherichia coli* and HEK293 crude lysates for up to 1 wk. Conversion of HERV-K RT to a heterodimer was not observed.

All figures were prepared with PyMol (79, 80) unless otherwise noted, and images exported with the highest-quality settings, ray trace with transparent background, and wall-eyed stereo where appropriate. *SI Appendix*, Fig. S14 *B* and *C* was created with Maestro (41). Sequence comparisons were carried out with Jalview (81) and Cobalt (82). Residue numbers refer to the HERV-K RT sequence unless otherwise noted.

HERV-K RT Activity Assays and Measurement of the Polymerase to RNase H Active Site Distance.

Protein expression and purification. The pFastBac-1 (Invitrogen) plasmid with the codon-optimized synthetic DNA sequences (GenScript) coding for HERV-K (AAC63291) and HIV-1 (NC_001802.1) RTs were used as a starting material for protein expression in insect cells (Sf9, Invitrogen). RTs were expressed as a part of a polyprotein containing TEV, HERV-K (or HIV-1) protease at its N terminus (TEV_protease-HERV-K_protease-HERV-K_RT, or TEV_protease-HIV-1_protease-HIV-1_RT). Each component of the polyprotein was separated by the TEV cleavage sites (ENLYFQG) to facilitate the posttranslational processing, which results in a single glycine residue at the N terminus of the HERV-K RT as a cloning artifact. We employed the MultiBac (Geneva Biotech) system for protein expression in insect cells (Sf9, Invitrogen) according to previously published protocols (83, 84). HERV-K and HIV-1 RT were purified using Ni-NTA affinity chromatography of the eight-histidine tag located at the C terminus of the RTs and according to the manufacturer's specifications (Thermo Scientific).

This method of HERV-K RT preparation afforded another opportunity to observe an HERV-K protease-mediated cleavage of HERV-K RT to an HIV-1 RT-like heterodimer. Following our established pattern for HIV-1 RT production with the HIV-1 protease, active protease is produced and cleaves HIV-1 RT to produce the expected heterodimer. The HERV-K RT produced in a similar manner results in a uniformly homogenous full-length enzyme, suggesting that a HERV-K protease-sensitive site may not exist in HERV-K RT between the connection subdomain and the RNase H domain.

Chemicals. AZT-TP and 3TC-TP were purchased from TriLink. CBV-TP was purchased from NuBlocks. PFA was purchased from Sigma-Aldrich. EFV and NVP were purchased from MedChemExpress.

DNA primer/RNA template duplexes. The sequence of the template strand was based upon the HIV-1 tRNA primer-binding site (U5-PBS from the HIV-1 reference genome, NC_001802.1, nucleotides 140 to 162).

All nucleic acid sequences were purchased from Dharmacon. For RT assays, the 5′-end of the DNA primer was labeled with ^{32}P using [γ-^{32}P]-ATP (PerkinElmer) and T4 polynucleotide kinase (Thermo Fisher Scientific) according to the manufacturer's protocol. For RNase H assays, the 3′-end of the RNA template was labeled with fluorescein (Dharmacon). DNA/RNA primer/templates were heat-annealed in 25 mM Tris·HCl (pH 8) and 50 mM NaCl by heating the samples at 95 °C for 10 min followed by a slow cooling down to the ambient temperature over ∼2 h.

RT activity assays. RT activity assays consisted of preincubating ∼60 nM HERV-K RT (or ∼20 nM HIV-1 RT) with 30 nM of annealed DNA/RNA 5′-end-radiolabeled-primer/template hybrid and the indicated concentrations of a given compound (PFA, NVP, EFV, CBV-TP, AZT-TP, or 3TC-TP) in the presence of 1 μM dNTP mixture, 0.5 mM EDTA, and 25 mM Tris·HCl buffer (pH 8) for 10 min at 37 °C. Nucleotide incorporation (DNA synthesis) was initiated by the addition of 2.5 mM $MgCl_2$. The reactions (15 μL) were incubated for 10 min at 37 °C and then stopped by the addition of 15 μL of formamide/EDTA (25 mM) mixture and incubated at 95 °C for 10 min. Three-microliter reaction samples were subjected to denaturing 8 M urea 20% PAGE to resolve products of the DNA synthesis followed by signal quantification (ImageQuant 5.2, GE Healthcare Bio-Sciences) through phosphorimaging (Amersham Typhoon 5, Cytivia). The signal corresponding to the full template-length product of DNA synthesis in the presence of a compound was normalized (%) to the signal obtained in the absence of the compound. The normalized values of DNA synthesis were plotted versus compound concentrations divided by the concentration of the competing natural dNTP and fitted to a log(inhibitor) versus normalized-response (variable slope) equation using GraphPad Prism 7.0 (GraphPad Software) to determine the IC_{50} values for the inhibition of DNA synthesis by a given compound. These IC50 values are independent of the concentration of the competing natural dNTP and, therefore, reported as fold ratios.

RNase H activity assays. RNase H activity assays consisted of preincubating ∼1 μM HERV-K RT (using the calculated extinction coefficient of 78430 M-1 cm-1 for a single molecule of RT; or ∼0.21 μM HIV-1 RT) with 30 nM of annealed DNA/RNA primer/3′-end-fluorescein-labeled-template in the presence of 1 μM dNTP mixture, 0.5 mM EDTA, 25 mM Tris·HCl buffer (pH 8), and the indicated concentrations of foscarnet for 10 min at 37 °C. RNA cleavage was initiated by the addition of a mixture containing 5 mM $MgCl_2$ and 2 g/L heparin. The reactions (15 μL) were incubated for 1 min at 37 °C and then stopped by the addition of 15 μL of formamide/EDTA (25 mM) mixture and incubated at 95 °C for 10 min. Three-microliter reaction samples were subjected to denaturing 8 M urea 20% polyacrylamide gel electrophoresis to resolve the products of the RNase H activity.

Data Availability. The atomic coordinates have been deposited in the Protein Data Bank (PDB) (PDB ID code 7SR6) (85).

ACKNOWLEDGMENTS. We thank Dennis Zaller and Menachem Fromer (ROME Therapeutics) for editing of the initial manuscript; William Sinko for running various simulations (ROME Therapeutics); Anja Jestel for coordination of research activities at Proteros; Shipra Bijpuria (Syngene) for final refinement of the structure; Stefano Pernigo and Philip Leonard for structure refinement contributions; Rowan Halls-Kass for protein resupplies; and Thomas Walpole (Charles River Laboratory) for manuscript editing.

Author affiliations: [a]ROME Therapeutics, Boston, MA 02215; [b]Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, AB T6G 2E1, Canada; [c]Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ, 08854; [d]Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ, 08854; [e]Charles River Laboratory, Chesterford Research Park, Saffron Walden CB10 1XL, United Kingdom; [f]DomainEx, Chesterford Research Park, Saffron Walden CB10 1XL United Kingdom; and [g]Proteros Biostructures GmbH, 82152 Planegg-Martinsried, Germany

1. C. Chang, P. T. Chen, G. D. Chang, C. J. Huang, H. Chen, Functional characterization of the placental fusogenic membrane protein syncytin. *Biol. Reprod.* **71**, 1956–1962 (2004).
2. P. Küry *et al.*, Human endogenous retroviruses in neurological diseases. *Trends Mol. Med.* **24**, 379–394 (2018).
3. E. Dervan, D. D. Bhattacharyya, J. D. McAuliffe, F. H. Khan, S. A. Glynn, Ancient adversary–HERV-K (HML-2) in cancer. *Front. Oncol.* **11**, 658489 (2021).
4. P. Medstrand, J. Blomberg, Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: Differential transcription in normal human tissues. *J. Virol.* **67**, 6778–6787 (1993).

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.

5. O. Hohn, K. Hanke, N. Bannert, HERV-K(HML-2), the best preserved family of HERVs: Endogenization, expression, and implications in health and disease. *Front. Oncol.* **3**, 246 (2013).
6. M. V. Viola, M. Frazier, L. White, J. Brody, S. Spiegelman, RNA-instructed DNA polymerase activity in a cytoplasmic particulate fraction in brains from Guamanian patients. *J. Exp. Med.* **142**, 483–494 (1975).
7. E. Oricchio *et al.*, Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene* **26**, 4226–4233 (2007).
8. B. Fu, H. Ma, D. Liu, Functions and regulation of endogenous retrovirus elements during zygotic genome activation: Implications for improving somatic cell nuclear transfer efficiency. *Biomolecules* **11**, 829 (2021).
9. E. Balestrieri *et al.*, Human endogenous retrovirus K in the crosstalk between cancer cells microenvironment and plasticity: A new perspective for combination therapy. *Front. Microbiol.* **9**, 1448 (2018).
10. R. F. Downey *et al.*, Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? *Int. J. Cancer* **137**, 1249–1257 (2015).
11. M. Greenig, HERVs, immunity, and autoimmunity: Understanding the connection. *PeerJ* **7**, e6711 (2019).
12. V. Gorbunova *et al.*, The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43–53 (2021).
13. C. Römer, Viruses and endogenous retroviruses as roots for neuroinflammation and neurodegenerative diseases. *Front. Neurosci.* **15**, 648629 (2021).
14. E. Balestrieri *et al.*, Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *EBioMedicine* **66**, 103341 (2021).
15. Y. Gao, X.-F. Yu, T. Chen, Human endogenous retroviruses in cancer: Expression, regulation and function. *Oncol. Lett.* **21**, 121 (2021).
16. M. Garcia-Montojo, T. Doucet-O'Hare, L. Henderson, A. Nath, Human endogenous retrovirus-K (HML-2): A comprehensive review. *Crit. Rev. Microbiol.* **44**, 715–738 (2018).
17. N. Bannert, H. Hofmann, A. Block, O. Hohn, HERVs new role in cancer: From accused perpetrators to cheerful protectors. *Front. Microbiol.* **9**, 178 (2018).
18. R. Tyagi, W. Li, D. Parades, M. A. Bianchet, A. Nath, Inhibition of human endogenous retrovirus-K by antiretroviral drugs. *Retrovirology* **14**, 21 (2017).
19. R. Contreras-Galindo, D. Dube, K. Fujinaga, M. H. Kaplan, D. M. Markovitz, Susceptibility of human endogenous retrovirus type K to reverse transcriptase inhibitors. *J. Virol.* **91**, e01309–e01317 (2017).
20. I. Sciamanna *et al.*, Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. *Oncogene* **24**, 3923–3931 (2005).
21. B. Berkhout, M. Jebbink, J. Zsíros, Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus. *J. Virol.* **73**, 2365–2375 (1999).
22. E. P. Tchesnokov, P. Raeisimakiani, M. Ngure, D. Marchant, M. Götte, Recombinant RNA-dependent RNA polymerase complex of Ebola virus. *Sci. Rep.* **8**, 3970 (2018).
23. J. M. Barbarino, D. L. Kroetz, R. B. Altman, T. E. Klein, PharmGKB summary: Abacavir pathway. *Pharmacogenet. Genomics* **24**, 276–282 (2014).
24. B. Marchand, M. Götte, Site-specific footprinting reveals differences in the translocation status of HIV-1 reverse transcriptase. Implications for polymerase translocation and drug resistance. *J. Biol. Chem.* **278**, 35362–35372 (2003).
25. B. Marchand, E. P. Tchesnokov, M. Götte, The pyrophosphate analogue foscarnet traps the pre-translocational state of HIV-1 reverse transcriptase in a Brownian ratchet model of polymerase translocation. *J. Biol. Chem.* **282**, 3337–3346 (2007).
26. G. L. Beilhartz *et al.*, HIV-1 reverse transcriptase can simultaneously engage its DNA/RNA substrate at both DNA polymerase and RNase H active sites: Implications for RNase H inhibition. *J. Mol. Biol.* **388**, 462–474 (2009).
27. J. Ren *et al.*, Structure of HIV-2 reverse transcriptase at 2.35-A resolution and the mechanism of resistance to non-nucleoside inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14410–14415 (2002).
28. A. Jacobo-Molina *et al.*, Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6320–6324 (1993).
29. R. F. Schinazi *et al.*, Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. *Antimicrob. Agents Chemother.* **37**, 875–881 (1993).
30. C. A. Boucher *et al.*, High-level resistance to (-) enantiomeric 2′-deoxy-3′-thiacytidine in vitro is due to one amino acid substitution in the catalytic site of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob. Agents Chemother.* **37**, 2231–2234 (1993).
31. M. A. Wainberg *et al.*, Clinical correlates and molecular basis of HIV drug resistance. *J. Acquir. Immune Defic. Syndr. (1988)* **6** (suppl 1), S36–S46 (1993).
32. B. Scarth, S. McCormick, M. Götte, Effects of mutations F61A and A62V in the fingers subdomain of HIV-1 reverse transcriptase on the translocational equilibrium. *J. Mol. Biol.* **405**, 349–360 (2011).
33. X. Tu *et al.*, Structural basis of HIV-1 resistance to AZT by excision. *Nat. Struct. Mol. Biol.* **17**, 1202–1209 (2010).
34. P. L. Boyer, S. H. Hughes, Effects of amino acid substitutions at position 115 on the fidelity of human immunodeficiency virus type 1 reverse transcriptase. *J. Virol.* **74**, 6494–6500 (2000).
35. G. Gao, M. Orlova, M. M. Georgiadis, W. A. Hendrickson, S. P. Goff, Conferring RNA polymerase activity to a DNA polymerase: A single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 407–411 (1997).
36. S. G. Sarafianos *et al.*, Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10027–10032 (1999).
37. P. R. Harrigan *et al.*, Resistance profile of the human immunodeficiency virus type 1 reverse transcriptase inhibitor abacavir (1592U89) after monotherapy and combination therapy. CNA2001 investigative group. *J. Infect. Dis.* **181**, 912–920 (2000).
38. R. Anjana *et al.*, Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: A study based on orientation and distance. *Bioinformation* **8**, 1220–1224 (2012).
39. K. Das *et al.*, Conformational states of HIV-1 reverse transcriptase for nucleotide incorporation vs pyrophosphorolysis-binding of foscarnet. *ACS Chem. Biol.* **11**, 2158–2164 (2016).
40. D. Arion, N. Sluis-Cremer, M. A. Parniak, Mechanism by which phosphonoformic acid resistance mutations restore 3′-azido-3′-deoxythymidine (AZT) sensitivity to AZT-resistant HIV-1 reverse transcriptase. *J. Biol. Chem.* **275**, 9251–9255 (2000).
41. T. A. Halgren, Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
42. K.-F. Chan *et al.*, An alternative HIV-1 non-nucleoside reverse transcriptase inhibition mechanism: Targeting the p51 subunit. *Molecules* **25**, 5902 (2020).
43. J. D. Bauman *et al.*, Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. *J. Med. Chem.* **56**, 2738–2746 (2013).
44. J. D. Bauman, J. J. E. K. Harrison, E. Arnold, Rapid experimental SAD phasing and hot-spot identification with halogenated fragments. *IUCrJ* **3**, 51–60 (2016).
45. F. Xavier Ruiz, E. Arnold, Evolving understanding of HIV-1 reverse transcriptase structure, function, inhibition, and resistance. *Curr. Opin. Struct. Biol.* **61**, 113–123 (2020).
46. F. X. Ruiz, A. Hoang, K. Das, E. Arnold, Structural basis of HIV-1 inhibition by nucleotide-competing reverse transcriptase inhibitor INDOPY-1. *J. Med. Chem.* **62**, 9996–10002 (2019).
47. L. Bellucci, L. Angeli, A. Tafi, M. Radi, M. Botta, Unconventional plasticity of HIV-1 reverse transcriptase: How inhibitors could open a connection "gate" between allosteric and catalytic sites. *J. Chem. Inf. Model.* **53**, 3117–3122 (2013).
48. S. Freisz *et al.*, Crystal structure of HIV-1 reverse transcriptase bound to a non-nucleoside inhibitor with a novel mechanism of action. *Angew. Chem. Int. Ed. Engl.* **49**, 1805–1808 (2010).
49. E. Nowak *et al.*, Structural analysis of monomeric retroviral reverse transcriptase in complex with an RNA/DNA hybrid. *Nucleic Acids Res.* **41**, 3874–3887 (2013).
50. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
51. J. J. E. K. Harrison *et al.*, Crystal structure of a retroviral polyprotein: Prototype foamy virus protease-reverse transcriptase (PR-RT). *Viruses* **13**, 1495 (2021).
52. M. Nowacka *et al.*, Structures of substrate complexes of foamy viral protease-reverse transcriptase. *J. Virol.* **95**, e0084821 (2021).
53. E. Nowak *et al.*, Ty3 reverse transcriptase complexed with an RNA-DNA hybrid shows structural and functional asymmetry. *Nat. Struct. Mol. Biol.* **21**, 389–396 (2014).
54. A. Herschhorn, A. Hizi, Retroviral reverse transcriptases. *Cell. Mol. Life Sci.* **67**, 2717–2747 (2010).
55. R. Taube, S. Loya, O. Avidan, M. Perach, A. Hizi, Reverse transcriptase of mouse mammary tumour virus: Expression in bacteria, purification and biochemical characterization. *Biochem. J.* **332**, 807–808 (1998).
56. G. Tachedjian *et al.*, Role of residues in the tryptophan repeat motif for HIV-1 reverse transcriptase dimerization. *J. Mol. Biol.* **326**, 381–396 (2003).
57. G. Tachedjian, H. E. Aronson, S. P. Goff, Analysis of mutations and suppressors affecting interactions between the subunits of the HIV type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6334–6339 (2000).
58. T. Schmidt, C. D. Schwieters, G. M. Clore, Spatial domain organization in the HIV-1 reverse transcriptase p66 homodimer precursor probed by double electron-electron resonance EPR. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 17809–17816 (2019).
59. R. E. London, HIV-1 reverse transcriptase: A metamorphic protein with three stable states. *Structure* **27**, 420–426 (2019).
60. J. J. E. K. Harrison *et al.*, Cryo-EM structure of the pol polyprotein provides insights into HIV maturation. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.10.03.462959. Accessed 1 May 2022.
61. N. G. Sharaf *et al.*, The p66 immature precursor of HIV-1 reverse transcriptase. *Proteins* **82**, 2343–2352 (2014).
62. R. L. Slack *et al.*, Conformational changes in HIV-1 reverse transcriptase that facilitate its maturation. *Structure* **27**, 1581–1593.3 (2019).
63. D. Das, M. M. Georgiadis, The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure* **12**, 819–829 (2004).
64. B. Halford, Covalent drugs go from fringe. *Chem. Eng. News* **98**, 28–33 (2020).
65. M. R. Arkin *et al.*, Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1603–1608 (2003).
66. J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
67. F. Sutanto, M. Konstantinidou, A. Dömling, Covalent inhibitors: A rational approach to drug discovery. *RSC Med Chem* **11**, 876–884 (2020).
68. M. Garcia-Montojo *et al.*, Inhibition of HERV-K (HML-2) in amyotrophic lateral sclerosis patients on antiretroviral therapy. *J. Neurol. Sci.* **423**, 117358 (2021).
69. L. S. Dodda, J. Tirado-Rives, W. L. Jorgensen, Unbinding dynamics of non-nucleoside inhibitors from HIV-1 reverse transcriptase. *J. Phys. Chem. B* **123**, 1741–1748 (2019).
70. R. Esnouf *et al.*, Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat. Struct. Biol.* **2**, 303–308 (1995).
71. K. Das, S. E. Martinez, J. D. Bauman, E. Arnold, HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat. Struct. Mol. Biol.* **19**, 253–259 (2012).
72. C. K. Shih *et al.*, Chimeric human immunodeficiency virus type 1/type 2 reverse transcriptases display reversed sensitivity to nonnucleoside analog inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9878–9882 (1991).
73. M. T. B. Clabbers, T. Gruene, J. M. Parkhurst, J. P. Abrahams, D. G. Waterman, Electron diffraction data processing with DIALS. *Acta Crystallogr. D Struct. Biol.* **74**, 506–518 (2018).
74. C. Vonrhein *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
75. G. M. Sheldrick, A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
76. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
77. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
78. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
79. Schrödinger, The PyMOL molecular graphics system. Version 1.3.r1 (Schrodinger Inc., 2010).
80. W. L. DeLano, *PyMOL Reference Guide* (DeLano Scientific, San Carlos, CA, 2004). pymol.sourceforge.net/.
81. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2–A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
82. J. S. Papadopoulos, R. Agarwala, COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007).
83. I. Berger, D. J. Fitzgerald, T. J. Richmond, Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004).
84. C. Bieniossek, T. J. Richmond, I. Berger, MultiBac: Multigene baculovirus-based eukaryotic protein complex production. *Curr. Protoc. Protein Sci.* **Chapter 5**, Unit 5.20 (2008).
85. E. T. Baldwin, C. Nichols, Human Endogenous Retrovirus (HERV-K) reverse transcriptase ternary complex with dsDNA template Primer and dNTP. Protein Data Bank. http://www.rcsb.org/pdb/explore/explore.do?structureId=7SR6. Deposited 8 November 2021.

Downloaded from https://www.pnas.org by UNIVERSITEIT VAN AMSTERDAM UVA on April 15, 2025 from IP address 145.18.166.139.