Structure of HIV-1 reverse transcriptase/d4TTP complex: Novel DNA cross-linking site and pH-dependent conformational changes

Sergio E. Martinez ^{1,2} Joseph D. Bauman,¹ Kalyan Das ² and Eddy Arnold^{1*}

¹Center for Advanced Biotechnology and Medicine, and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey, 08854

²Rega Institute for Medical Research and Department of Microbiology and Immunology, KU Leuven, Leuven, 3000, Belgium

Received 24 September 2018; Accepted 19 November 2018

DOI: 10.1002/pro.3559

Published online 22 December 2018 proteinscience.org

Abstract: Stavudine (d4T, 2',3'-didehydro-2',3'-dideoxythymidine) was one of the first chain-terminating nucleoside analogs used to treat HIV infection. We present the first structure of the active, triphosphate form of d4T (d4TTP) bound to a catalytic complex of HIV-1 RT/dsDNA template-primer. We also present a new strategy for disulfide (S–S) chemical cross-linking between N⁶ of a modified adenine at the second overhang base to I63C in the fingers subdomain of RT. The cross-link site is upstream of the duplex-binding region of RT, however, the structure is very similar to published RT structures with cross-linking to Q258C in the thumb, which suggests that cross-linking at either site does not appreciably perturb the RT/DNA structures. RT has a catalytic maximum at pH 7.5. We determined the X-ray structures of the I63C-RT/dsDNA/d4TTP cross-linked complexes at pH 7, 7.5, 8, 8.5, 9, and 9.5. We found small (~0.5 Å), pH-dependent motions of the fingers subdomain that folds in to form the dNTP-binding pocket. We propose that the pH-activity profile of RT relates to this motion of the fingers. Due to side effects of neuropathy and lipodystrophy, use of d4T has been stopped in most countries, however, chemical modification of d4T might lead to the development of a new class of nucleoside analogs targeting RNA and DNA polymerases.

Keywords: d4T (stavudine); HIV-1 reverse transcriptase; DNA cross-linking to protein; pH dependence of catalysis

Introduction

Six classes of drugs are in clinical use for the treatment of HIV/AIDS, which act at four distinct points in the life cycle of the virus.¹ Two classes block the HIV particle from entering a host cell, each by

different mechanisms. Fusion inhibitors are peptides that block formation of a six-helix bundle of gp41, a viral membrane protein essential for the fusion of the infecting matured HIV particle membrane with that of the host cell. Entry inhibitors are antagonists that block binding of the HIV gp120 envelope protein to CCR5, a co-receptor for the viral infection. Integrase inhibitors block the integration step of a viral dsDNA (double-stranded DNA) to the host DNA by binding at the active site of HIV integrase. Another enzyme drug target is the HIV protease, which is essential for maturation of HIV particles after budding off from infected cells; the HIV structural proteins and enzymes are synthesized as Gag and Gag-Pol polyproteins, which

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Institute of Allergy and Infectious Diseases AI027690; Grant sponsor: National Institutes of Health R37 AI027690.

*Correspondence to: Eddy Arnold, Department of Chemistry and Chemical Biology, Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ 08854. E-mail: arnold@cabm.rutgers.edu

are processed by the protease into the mature components. This cleavage step is the target of protease inhibitors that occupy the active site.

An important clinical intervention consists of blocking the reverse transcription stage of HIV replication that consists of copying its single-stranded, 10-kb RNA genome into dsDNA. Minus strand synthesis of the HIV genome is initiated with a host cell RNA primer that is packaged into the budding virion. A cellular tRNA^{Lys3} binds to the primer-binding sequence (PBS) near the 5' end of the RNA genome and begins a complex series of steps that end in a dsDNA copy of the RNA genome for integration into a host cell chromosome. The DNA synthesis is carried out by the viral enzyme reverse transcriptase (RT), the target of the remaining two classes of HIV drugs, nucleos(t)ide RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs).² NNRTIs are hydrophobic compounds having a variety of chemical scaffolds, and an NNRTI binds to a transient pocket ~10 Å from the active site. NNRTIs are allosteric inhibitors that act by displacing the 3' end of the primer ~5 Å away from the active site.³

All NRTIs are nucleoside or nucleotide mimics that lack a 3'-hydroxyl, and act as chain terminators of DNA synthesis. They can include other chemical modifications. NRTIs have been a component in nearly all anti-retroviral therapy (ART) treatment regimens to date, and typically two NRTIs are used in the most widely used regimens. d4T (2',3'-didehydro-2',3'-dideoxythymidine) was the fourth drug, and also the fourth NRTI approved for HIV treatment in 1994, following AZT (3'-azido-3'-deoxythymidine) in 1987, ddI (2',3'-dideoxyinosine) in 1991, and ddC (2'-3'-dideoxycytidine) in 1992. NRTIs are phosphorylated in host cells to the triphosphate form, which are substrates for RT. Following a WHO (World Health Organization) guideline change in 2010,⁴ d4T has been phased out primarily due to a high rate of lipodystrophy and peripheral neuropathy, which are side effects from long-term use.^{5,6} Another disadvantage is its low inhibitory activity in some cell types of the central nervous system.⁷ However, one of d4T's advantages is that its incorporation efficiency (k_{pol}/K_d) into the primer strand is comparable to that of the natural dTTP.⁸ Chemical modifications to the d4T backbone might help to design new NRTIs that minimize toxicity while retaining efficacy.

We report here the first structure of a ternary complex of RT with a dsDNA template-primer and d4TTP (2',3'-didehydro-2',3'-dideoxythymidine triphosphate) at the N site (incoming nucleotide-binding site). Homogeneous samples of RT/DNA complexes have been prepared for many structural studies using disulfide cross-linking of a Q258C mutated residue at the base of the thumb of RT to the sixth base of DNA from the primer terminus^{9,10}; the cross-linked base is a thiol-modified G which is located in the minor

groove of dsDNA [Fig. 1(A)]. Here we report an efficient and less expensive cross-linking strategy to form a covalent link between I63C in the fingers sub-domain to a thioethyl-modified N⁶ atom of an adenine in the single-stranded template overhang. This cross-linking site is outside of duplex-binding portion of the DNA-binding cleft, which extends from the polymerase to RNase H active site. Additionally, this chemically modified adenine can be used to cross-link to the major groove in a protein/DNA complex where a nearby cysteine is available.

The cross-linking chemistry used here was inspired by a report from Corn and Berger of the cross-linking of the catalytic domain of the *Escherichia coli* DnaG primase to a single-stranded DNA template.¹¹ The structure of the domain was unknown and was known to bind weakly to ssDNA. The template had been synthesized with an O⁶-phenyl-deoxyinosine (dI) nucleotide. This convertible base was modified by replacement of the O⁶-phenyl group with cystamine (NH₂CH₂CH₂SSCH₂CH₂NH₂), leaving an adenine base with a two-carbon linker attached at N⁶, followed by a disulfide bond that can exchange with a reduced thiol group on the protein [Fig. 1(A)]. Using this chemistry, a 2.35 Å resolution X-ray crystal structure was obtained for DnaG primase bound to a single-stranded DNA template.¹² We adapted this approach to HIV-1 RT to see if it could be an efficient alternative to custom-synthesized DNA oligomers that incorporate a custom-synthesized guanosine phosphoramidite with a cross-linking group at the N² position. The crystals of the new I63C cross-linked complex were found to be robust and withstand a wide pH range. Crystals were grown at pH 9.5 and could be brought down to pH 7 by step soaking without loss of diffraction resolution, to 2.5 Å in the best case presented here. We have used these crystals to determine six structures, with cross-linking to the template second base overhang and bound d4TTP, over the pH range 7.0–9.5. A small motion of the fingers domain in these structures correlates with the pH dependence of polymerase catalytic activity. We have also attempted to ascertain the structural basis for the similar incorporation efficiency of d4TTP and dTTP. In another set of experiments, we have determined one structure by cross-linking I63C to the first-base overhang, which base pairs with the incoming d4TTP and is very similar to the previous six structures.

Results

Preparation of HIV-1 RT cross-linked at I63C to N⁶ dA in template

We inspected RT/DNA structures for potential cysteine mutants that might be within reacting distance of a two-carbon linker to the N⁶ atom of adenine, either in the single-stranded template overhang, or in the

Figure 1. Chemistry of disulfide cross-linking reaction of a protein cysteine residue to modified guanine or adenine in dsDNA, as applied to HIV-1 RT. (A) For the guanine, a three-carbon linker is covalently bound to the N² atom. At the other end of the linker is a disulfide bond, followed by an ethylamine leaving group. For the adenine used in this report, a two-carbon linker is covalently bound to the N⁶ atom. The leaving group is also ethylamine. The thiol group of the cysteine exchanges with the leaving group, leaving a cross-linked complex of protein and DNA. (B) The two cross-linking p66 sites, Q258C and I63C, are marked in the RT/DNA/d4TTP structure at pH 7.0, cross-linked at I63C in the fingers subdomain to N⁶ dA at the template second base overhang, with a two-carbon (two methylene groups) linker. In other cross-linked RT structures reported in the literature, cross-linking is from Q258C in the thumb to N² dG in a base-paired guanine. The distance between the sulfur atoms in Q258C and I63C is ~23 Å. On the template, the first base overhang (base-pairs with the thymine base of d4TTP) and second base overhang are indicated by -1 and -2, respectively. (C) Active site for the post-translocation HIV RT structure at pH 7.0 with cross-linking from I63C to the N⁶ dA in the second template base overhang (PDB 6AMO). The disulfide and two-carbon linker between the sulfur atom of I63C and N² of adenine has no ordered electron density and is indicated by the dashed yellow line. The template strand is green, and the primer strand [terminated with dideoxyguanosine monophosphate, in the P site (primer 3'-terminal nucleotide-binding site)] is magenta. The incoming d4TTP sits at the N site. Active site residues shown include D110, D185, and D186, which coordinate the metal B ion (Mg²⁺). The R72 and K65 side chains form salt bridges with the alpha and gamma phosphates, respectively. The β3–β4 hairpin (Residues 59–76) containing R72 and K65 has poor density, but nearly all of it has some density at lower contour levels. (D) 2Fo–Fc electron density for bound d4TTP in the active site. The map is contoured at 1σ.

major groove. The extended length of the cross-linker from the N⁶ adenine atom to the cysteine S_γ atom is approximately 5.4 Å. The major groove is nearly devoid of any protein contacts, but I63 in the fingers subdomain appeared to be a candidate for cysteine replacement and cross-linking to either the first or second template overhang base [Fig. 1(B) and (C)]. Using PDB 1RTD⁹ as a reference (RT/DNA/dTTP), after modeling I63C and adjusting the χ^1 angle, the S_γ atom is ~5.2 Å from the N⁶ atom of the first-base overhang (adenine, which base-pairs with an incoming dTTP) and ~6 Å from the second-base overhang. A modified adenine with two-carbon linker and the

disulfide bond at N⁶ position at either position looked favorable, however, the second-base overhang has substantial room to swing closer to I63C whereas, the position of the first-base overhang is restricted for base-pairing with an incoming dTTP only.

Oligomers with the convertible nucleotide O⁶-phenyl-dI are commercially available (e.g., Midland Certified Reagent Co., USA). Those are custom-synthesized as a regular DNA oligomer, with the O⁶-phenyl-dI phosphoramidite used at the desired position in the sequence. The cost of oligo synthesis is the same as for the regular N-1 oligo, with an additional charge for the added position with O⁶-phenyl-dI. The

DNA oligomer can be ordered on a small synthesis scale, such as 0.2 μmole , that makes it feasible for preliminary experiments. Before use, the linker must be attached by reacting a free base amine with the O⁶. A two-carbon linker was chosen in part because it can be derived from cystamine, an inexpensive and commercially available compound. It was also the linker used in the primase study.¹¹ A three-carbon analog of cystamine could be used, but this compound (3,3'-dithiobispropylamine; homocystamine) is not commercially available. It was custom synthesized at one point in the synthesis of the N² dG phosphoramidite.¹³ Linkers of any number of carbons could, in principle, be generated if the H₂N(CH₂)_nSS(CH₂)_nNH₂ amine compound used as the source is available, where $n = 2$ for the cystamine used in this study. However, as described in the Materials and Methods section, the two-carbon linker successfully cross-links to yield the desired RT/DNA complex.

Structure of wild-type RT/DNA/d4TTP complex

The cross-linked RT139A (I63C) cocrystallized with bound d4TTP in the P2₁ space group and with cell dimensions similar to those of previously reported Q258C RT/nucleic acid cross-linked complexes.¹⁴ While Q258C RT/DNA crystals (binary or ternary complex) grow at approximately pH 6.8–7.6, I63C RT/DNA/d4TTP crystals only grow over the pH range 9.5–10.5; attempts to grow at a lower pH generated only spherules. Apart from different cross-linking positions, the Q258C RT (RT127A) has the D498N mutation. With Q258C RT, cross-linking is to the sixth base, starting at the 3' terminus of the extended primer. With I63C RT in this study, the cysteine has been cross-linked to first or second base of the template overhang. The quality of the 2.5 Å electron density is shown for the active site [Fig. 1(D)], and template-primer at pH 7.0, with cross-linking at the template second base overhang [Fig. 2(A)]. The fingers subdomain is flexible, and therefore, the electron density for the disulfide bond and cross-linker from I63C to the adenine is not discernible, indicative of disorder [Fig. 1(C)]. The I63C RT/DNA complex crystallized with the same space group symmetry and cell parameters as Q258C RT/DNA complex. A superposition was done in Chimera¹⁵ ("match" command) between C α atoms of p66 and p51 (Chains A and B, respectively) of the pH 7.0 I63C RT/DNA/d4TTP, and the same chains of Q258C RT/DNA/AZTTP (PDB 3V4I) [Fig. 2(B)]; for all matching 961 C α atom pairs in p66 and p51, the RMSD is only 0.67 Å. At the fingers subdomain, there is a small motion, but the fingers subdomain is the most flexible region that has the weakest density in RT in this crystal form. However, high flexibility of the fingers is essential for dNTP binding and catalysis in the crystal.³ The RMSD of all 41 matching pairs of phosphorus

Figure 2. Cross-linking at Q258C or I63C gives identical structures. (A) Electron density of pH 7.0 RT/DNA/d4TTP structure (PDB 6AMO, I63C cross-linking to second base template overhang). 2.5 Å 2Fo-Fc density map contoured at 1 σ for the template-primer accompanying a ribbon diagram for p66 (Chain A). The incoming d4TTP at the N site and its bound Mg²⁺ are shown at the bottom right. The side chains of sites for cross-linking cysteine mutations Q258C (thumb) and I63C (fingers) are indicated in cyan. Density for the template-primer continues from there to beyond the RNase H domain. Fingers subdomain, blue; palm, red; thumb, green; connection, yellow; RNase H domain, orange; p51, gray. (B) Superposition of the RT heterodimer (Chains A and B, p66 and p51) from the N⁶ dA cross-linked (second template overhang base) pH 7.0 structure with bound d4TTP (PDB 6AMO), and the same chains from the N² dG cross-linked structure from PDB 3V4I with bound AZTTP. The superposition was based on the C α atoms of p66 and p51. The N⁶ dA structure ribbons are indicated in blue (p66) and magenta (p51). The N² dG structure ribbons are indicated in red (p66) and green (p51). (C) Close up view of the thumb subdomain in the superposition, with the same color scheme for protein. Helices H, I, and J that make up the thumb superimpose very well. Only the template-primer for PDB 3V4I is shown, along with the cross-link between Q258C in helix H, and the modified guanine (Residue 817 in Chain P [primer]).

(P) atoms in the template-primer is 0.94 Å. In addition, an examination of the thumb in this superimposed pair shows very close alignment of the three helices making up the thumb [Fig. 2(C)]. Therefore, cross-linking at either thumb or fingers has a

negligible to no effect on the structure of RT in complex with a template-primer.

Binding of d4TTP to wild-type RT/DNA complex

Despite the early, widespread use of d4T in HIV treatment, only two entries in the PDB contain d4T in any form, but none in complex with RT. In PDB 1Z4Q, human mitochondrial deoxyribonucleotidase, catalytically inactivated with the D41N mutation, is in complex with d4T monophosphate.¹⁶ The second structure 1F3F is of nucleoside diphosphate kinase, bound to d4TTP as a phosphate donor.¹⁷ Both enzymes have functions, folds, and catalytic reactions different from those of RT, where the β and γ phosphates are released as a pyrophosphate molecule in the catalytic reaction of DNA polymerization.

In RT/DNA/d4TTP structure, the inhibitor binds at the N site in a manner seen in other structures of RT ternary complexes, e.g. the dTTP bound in PDB 1RTD [Fig. 1(C)]. The thymine base of d4TTP forms a Watson–Crick base pair with the adenine first base overhang on the template strand and base-stacks with the guanine of the ddG at the primer terminus. All three phosphate groups contribute to the octahedral coordination of the metal B ion (Mg^{2+}), the remaining chelation coming from the carboxylate groups of D110 and D185, and the carbonyl oxygen of V111. Other contacts to the triphosphate group include the side chains of K65 and R72. The d4TTP stereochemistry was derived from the small molecule structure of d4T crystallized with molecules of the solvent 1-methyl-2-pyrrolidone (CSD KONNOQ, CCDC 145541);¹⁸ the 2',3'-didehydroribose ring is planar because of the double bond between C2' and C3'.

A pH-dependent conformational change of the fingers subdomain may regulate catalysis

We attempted to ascertain if there was any conformational change in the RT/DNA/d4TTP heterodimer over the pH range from 7.0 to 9.5. For the five pH steps 7.5, 8, 8.5, 9, and 9.5, C α atoms from the palm subdomain (Residues 85–119, 151–243) of p66 subunit (Chain A or Chain C) were superimposed on the corresponding pH 7.0 chain, using the least squares fit “match” command in Chimera.¹⁵ Plots were made of the distance between the two C α atoms in each matching pair of residues (Figs. S1 and S2). Above pH 7.0, this superposition reveals a clear rigid body tilting of the fingers away from the palm and away from the bound d4TTP that sits at the active site in the palm (Figs. 3, S1, and S2). The fingers subdomain is a distinct folding domain which comprises two segments of residues, 1–84 and 120–150, and alternating segments from the palm subdomain, 85–119 and 151–243. There are two copies of the complex in the asymmetric unit, and for both p66 chains (A and C), at all higher pHs, the palm superimposes on pH 7.0

Figure 3. Superposition of pH 7.0 RT/DNA/d4TTP (PDB 6AMO, magenta) and pH 9.5 (PDB 6AVM, green) structures on the palm subdomains shows a small pH-dependent tilt of the fingers subdomain away from the dNTP-binding site at pH 9.5. For a sense of scale, the distance between C α carbons is 3.8 Å.

with differences averaging ~ 0.1 Å. The deviation of the fingers at higher pHs ranges up to approximately eight times this for the first fingers segment, and 12 times for the second segment. For clarity, plots that zoom in on the region only up to 10 residues past the fingers subdomain are shown (Figs. S3 and S4). In Chain A (Fig. S3), a spike covers the short $\beta 3$ – $\beta 4$ hairpin (also known as fingers loop; Residues 59–76). The $\beta 3$ – $\beta 4$ hairpin closes down to form the dNTP-binding pocket at the N-site. Because of its structural flexibility, the electron density for the fingers loop is weak. The backbone conformation is therefore less reliable, accounting for this spike. Nevertheless, the rigid body movement of the fingers subdomain as a whole is clear. In the first fingers segment, pH 9.5 and 9.0 have the largest deviations away from the palm in both Chains A and C. In the second segment, pH 8.5 and 9.0 have the largest deviation in Chain A, and 9.0 and 9.5 in Chain C. In both segments, a more complex pattern relates the lower pHs with peak height.

Discussion

Template overhang dA crosslink with I63C RT

A high similarity of RT/DNA/dNTP structures cross-linked to Q258C or I63C shows that the latter is a viable cross-linking strategy. At a considerably lower cost, it may make feasible other applications where crystals of this cross-linked complex could be produced and used in large numbers, such as fragment screening and crystal engineering. Although the dA N⁶ itself remains constant, the rest of the sequence of the cross-linking oligomer can be varied with much greater ease than for a fixed N² dG cross-linking oligomer at the sixth position. In the former case, the

sequence is specified when ordering from the manufacturer, and addition of the C2 linker is relatively simple. In the latter case, the N² dG phosphoramidite has to be synthesized through a non-standard procedure, and the subsequent oligomer itself also must be synthesized through a non-standard procedure. The cross-linking reaction conditions are very similar between the two. Cross-linking, and its optimization by varying reaction parameters, can be easily monitored in the same way by monitoring the p66 band shift on SDS-PAGE.

Interactions of d4TTP

Despite having a flat ribose ring of d4TTP, which is the key conformational difference between d4TTP and dTTP, a comparison of K_i (d4TTP) and K_m (dTTP) shows that d4TTP is as efficient a substrate as dTTP.¹⁹ Both are also incorporated into template-primer (RNA/DNA or DNA/DNA) by RT at nearly the same catalytic efficiency (k_{pol}/K_d)⁸. For the RNA template, the catalytic efficiency is 1.8 versus 2.1 $\mu\text{M}^{-1} \text{s}^{-1}$ for d4TTP and dTTP, respectively. d4TTP has a 4.3-fold lower k_{pol} (14 vs. 60 s^{-1}), but the K_d reflects 3.6-fold higher affinity (78 vs. 28 μM). The slower k_{pol} may be due to the C2'–C3' double bond, preventing the five-membered ring from assuming the 3'-endo conformation that is found in both bound dNTP and the 3'-primer end, i.e. in a nucleotide before and after incorporation. In RT-bound dNTPs, the 3'-endo conformation is locked through a hydrogen bond between the 3' hydroxyl oxygen and the main-chain amide nitrogen of Y115. Our recent structure of dATP and ddATP ternary complexes with wild-type and Q151M RTs showed the ring in 3'-endo conformation for dATP, but 3'-exo for ddATP, which lacks the 3' hydroxyl.²⁰ The k_{pol} is slower for ddNTP compared to dNTPs, possibly because of this difference in ribose ring conformations. The flat conformation of the sugar ring of d4TTP differs from that of ddATP and dTTP, in which C2'–C3' is a single bond. The higher affinity K_d for d4TTP, despite the absence of the 3' hydroxyl, may be due to a favorable stacking of the planar 5-member ring with the aromatic side chain of Y115 (Fig. 4). The two rings are almost stacked with an approximate tilt of 15° between them. The C2' atom in dTTP (PDB 1RTD) appears to protrude down toward that side of the phenol ring of Y115, tilting it by 18–22° in χ_2 compared to d4T (PDB 6AMO). This moves Cε2 further from C1'. Given that the resolution of PDB 1RTD is 3.2 Å versus 2.5 for 6AMO, we also checked the 2.5 Å structure of RT/DNA/dATP (PDB 5TXL). A similar displacement of 14–27° was found.

Implications for d4T resistance

Like for AZT, thymidine analog mutations (TAMs) primarily emerge in patients treated with d4T. TAMs primarily assist binding of an ATP to RT as an excision substrate to unblock the primer terminus.^{21,22}

Figure 4. To compare the environment of the ribose rings in d4TTP (green) and dTTP (magenta), Chains A from pH 7.0 RT/DNA/d4TTP (PDB 6AMO) and PDB 1RTD were aligned based on the palm subdomain (85–119, 151–242) using the “match” command in chimera. The 3'-endo conformation of dTTP in PDB 1RTD causes C2' to push down toward the phenol ring of Y115 and tilt χ_2 by 18–22°. This results in a smaller hydrophobic contact area elsewhere on the ring.

Structurally, incorporated d4T-MP (or AZT-MP) has to occupy the N site to facilitate excision.^{23,24} The position and interaction of the 3'-azido group appear to favor positioning the AZT-MP at the N-site after incorporation, and therefore AZT is more favored to be excised compared to other NRTIs such as 3TC. Apparently, the ring stacking of d4TMP favors its positioning at the N site. However, in contrast, abacavir, which has a planar pentose ring, is less affected by the TAM mutations. Apparently, a combination of the modifications of the sugar ring for thymine analogs is more susceptible to excision. The Q151M complex and 69-insertion multidrug resistance mutations are also resistant to d4T.

pH dependency of RT/DNA/d4TTP structure

The pH-dependent motion of the fingers subdomain suggests a simple explanation for the optimal pH of 7.5 for catalysis. At pH values above 7.5, closing of the β3–β4 ribbon with R72 over the dNTP might be reduced by tilting of the fingers subdomain away from the bound dNTP substrate. Additional contributors could be a change in protonation of the Mg^{2+} chelating groups, D110, D185, and D186, or of the triphosphate group. Given that the pK_a of the aspartate carboxylate group is ~3.9, it seems likely that all three aspartates would be mostly deprotonated at the pH ≥ 7 used in this study. This potential weakening of R72 and K65 contacts to the alpha and gamma phosphates, across the transition from pH 8.0 to 9.0, is consistent with the pH dependence of HIV reverse transcriptase catalytic activity that drops to 60% and 26%, respectively, from its maximum at pH 7.5.²⁵ Normalized to this pH, the % catalytic activity is ~60,

100, 60, 30, and 26% over pH 7, 7.5, 8, 8.5, and 9, respectively. The reduction of the catalytic activity at higher pH could in part relate to the shifting of the fingers (and with it, the R72 and K65 side chains contacting the α and γ phosphates) away from the active site seen in our pH series of structures, from pH 8.0 to 9.0, and a reduction of the stacking of the R72 guanidinium with the base of the dNTP at the N site. A complex of RT/DNA aptamer with foscarnet (PFA), a pyrophosphate analog, shows that R72 interacts directly with PFA, as it does with both the alpha and beta phosphate groups of dNTPs.²⁶ Other evidence suggests the rate-limiting step in dNTP incorporation is in fact the closing of the β 3– β 4 hairpin over the bound dNTP. The dNTP incorporation rate (k_{pol}) can be separated into a kinetic step consisting of the closing down of the β 3– β 4 hairpin (with K65 and R72) over the bound dNTP, and the subsequent nucleophilic attack resulting in phosphodiester bond formation. Evidence in favor of the conformational change of the β 3– β 4 hairpin as the rate-limiting step comes from studying the incorporation of dNTPs that substitute a sulfur atom for a non-bridging oxygen in the alpha phosphate (dNTP- α -S). The ratio in rate of incorporation is close to 1,^{27–29} but a study of the rate of phosphoester hydrolysis in solution shows a decrease in the rate of 30–100-fold with the sulfur atom substitution.³⁰ Thus, this suggests the rate-limiting step is not the chemical step, but the conformational change of the fingers. Therefore, this interaction is likely to assist with pyrophosphate release. An increasing distance and weakening of this interaction at higher pH could slow down the release of pyrophosphate, lowering the incorporation rate (k_{pol}), and the catalytic activity. Another factor may potentially be a lower affinity of the N site for dNTP with the β 3– β 4 hairpin closing less tightly over it.

Among the strongest evidence that the pH-dependent motion of the fingers subdomain is real is that it occurs in both Chains A and C, to about the same extent, at the same pH transition. Further, the lattice environment is different for the fingers of Chains A and C, with no lattice contacts to the fingers in either case. This argues against lattice constraints generating the observed tilts of the fingers subdomains. This P2₁ crystal form allows for movement of the fingers subdomain and binding of dNTP to the N site in binary complex crystals, and has been used to study conformational changes induced by binding of nevirapine, an NNRTI.³ A second source of evidence is the similarity between the pH 9.5 structure and that from a pH 9.5 dataset from a different crystal. A plot of differences between C α pairs shows values $\sim$ 0.1 Å for most of both Chains A and C (Figs. S5 and S6). While extra datasets were not collected at the other pHs, these two show virtually identical tilts of the fingers relative to the palm, where the active site is located. An independent

confirmation of the pH-dependence of the fingers closure could be achieved by using electron microscopy, an increasingly successful tool for studying enzyme conformations. In general, this approach is unconstrained by any possible contribution of the lattice contacts seen in X-ray crystal structures. However, at this time, the resolution of the crystal structures presented here (2.5–2.7 Å) is not routinely achievable. Other methods could be applied such as NMR and FRET.

Although d4T itself is being phased out as a treatment for HIV, chemically modified versions of d4T may yet find a place in the future. A 4'-ethynyl derivative of d4T (BMS-986001) in Phase 2 trials showed similar efficacy as tenofovir disoproxil fumarate; however, development was discontinued due to a lower resistance barrier and gains in both peripheral and central fat accumulation.³¹ Such modifications could add specificity while retaining the flat ring that may allow d4TTP to retain the same catalytic efficiency as dTTP, despite the absence of the 2' hydroxyl.

Materials and Methods

RT expression, RT/DNA cross-linking, purification, and crystallization

A previously described wild-type RT construct, RT35A, was used to generate the RT139 construct used in this study;³² the I63C mutation in p66 subunit was introduced, and the resulting RT139A construct was expressed and purified following the described protocol.

A 27-mer template (5'-ATGAACGCGCGCCCGAA CAGGGACTGTG-3') was used for cross-linking to the template second base overhang; the cross-linking N⁶ dA is indicated by an underlined A. This oligomer was ordered from Midland Certified Reagent Company in Midland, TX, but with O⁶-phenyl-dI substituted for the cross-linking N⁶ dA. The O⁶ phenyl group was substituted with a cystamine group (H₂NCH₂CH₂SSCH₂CH₂NH₂), bonded at one terminal amino group to C6. This converted the inosine to adenine. Cystamine is commercially available as the hydrochloride and has to be converted to the free base amine. 2.5 g (0.011 mole) of cystamine dihydrochloride were dissolved in 3 mL water. 2 mL of 10 M NaOH (0.02 mole) were added to this solution. The solution was then placed in a 15 mL screw cap tube and shaken with 5 mL of chloroform to extract the free base amine. The extraction was repeated twice. The pooled chloroform extracts were placed in a beaker in the fume hood to evaporate. About 1–1.5 g of an oily liquid remained with a slight yellow tinge and no chloroform odor. A small amount of free base cystamine spun out as white particles and stored in an Eppendorf vial under nitrogen at -80°C (using

needle, rubber tube, and inflated balloon to flush out the air). After 4 weeks, the cystamine remained clear.

The oligo was reacted with free base cystamine as follows.¹¹ The oligo was dissolved in 1 M cystamine in water. The reaction was allowed to proceed at 68°C overnight after flushing out the air with N₂. The oligo was exchanged into 0.1 M TEAB (triethylammonium bicarbonate) pH 8.5, using a NAP-10 column. The reaction was loaded on a C18 Sep-Pak disposable cartridge (Waters Corporation). Before use, the cartridge was washed with 2 × 2 mL acetonitrile, then 2 × 2 mL 0.1 M TEAB pH 8.5. The oligo was loaded, washed with 3 × 1 mL 0.1 M TEAB pH 8.5, and eluted with 3 × 1 mL 30% v/v acetonitrile, 25 mM TEAB pH 8.5. Fractions with significant A₂₆₀ absorbance were lyophilized, suspended in nitrogenated water, and stored frozen under nitrogen.

The 20-mer primer (5'-ACAGTCCCTGTTCGGGC GCC-3') was synthesized by Integrated DNA Technologies. To anneal the 20-mer primer and 27-mer template, a volume of 197.4 µL solution was prepared containing: 290.0 nmoles of cross-linking N⁶ dA DNA 27-mer template (1.47 mM), 323.0 nmoles DNA 20-mer primer (1.64 mM), 50 mM NaCl, 50 mM Tris-HCl, pH 8.0. The primer was in 10% stoichiometric excess to the template. The template-primer mix was placed in a sterile 1.5 mL Eppendorf vial and floated on a liter of water at 95°C. The beaker was placed in an ice bucket with a lid and left on the bench to cool overnight to room temperature.

For preparative scale, the cross-linking reaction to the template second base overhang used a 2.7-fold excess of template-primer to RT and was done using ddGTP (2',3'-dideoxy GTP) for primer extension. A volume of 5.183 mL solution was prepared containing: 108 nmoles (12.6 mg, 18.6 µM) RT139A in 75 mM NaCl and 10 mM Tris pH 8.0, 293 nmoles (51 µM) annealed template-primer in 50 mM NaCl and 50 mM Tris-HCl pH 8.0, 1150 nmoles ddGTP (200 µM), 50 mM Tris-HCl pH 8.0, 84 mM total NaCl, 5 mM MgCl₂, 1 mM βME. This was incubated at 37°C for 4 h, then placed on ice. In the cross-linking reaction, the total NaCl concentration in the cross-linking reaction was kept at ~70–80 mM, well below the ~150 mM NaCl that will dissociate RT/DNA complex. SDS-PAGE showed that about 70% of the p66 from the RT construct (RT139A) became cross-linked (data not shown).

Unreacted RT (~30%) was removed with a column that consisted of three, 1 mL Invitrogen cartridges in this series: Ni HisTrap (Ni²⁺), heparin HiTrap, heparin HiTrap. Three buffers were used. Buffer A was 75 mM NaCl, 10 mM Tris-HCl, pH 8.0. Buffer B was 75 mM NaCl, 10 mM Tris-HCl, 100 mM imidazole, pH 8.0. Buffer C was 1 M NaCl, 20 mM Tris-HCl pH 8.0. The 1 mL reaction volume was loaded directly into a 1 mL sample loop on an AKTA Explorer. The flow rate was set to 0.7 mL/min,

one column volume (CV) was set to 0.785 mL, and 1 mL fractions were collected. The column was equilibrated with 5 mL (6.36 CV) buffer A. The sample was injected onto the column with 2.25 mL buffer A. The column was washed with 7.1 mL (9.0 CV) buffer A, then eluted with 10 CV (7.9 mL) 100% buffer B. An NaCl gradient was then applied from 0 to 100% buffer C in 15 CV (11.8 mL). 100% buffer C was sustained for 6 CV (4.7 mL). The column was then equilibrated with 8 CV (6.3 mL) 100% buffer A. In the wash is a large peak with unreacted DNA and ddGTP (Fig. S7, Fractions A3–A7). In the imidazole elution step, a very small fraction (a few percent) of the cross-linked RT elutes (Fractions B2–B11). The vast majority of the cross-linked RT elutes as one peak (Fractions C2–C6) at approximately 200 mM NaCl in the NaCl gradient, followed closely (not complete baseline separation) by a second smaller peak (Fractions C8–C9). SDS-PAGE across these two peaks (Fig. S8) shows that the first peak is cross-linked RT139A, followed by the unreacted RT139A. An explanation for the binding of cross-linked RT139A, and differing affinity from unreacted RT139A, may originate from the template-primer being tethered at the single-stranded template overhang at the 5' end. The template-primer could swing in and out of the nucleic acid binding groove of RT. This reduces but does not eliminate the affinity for heparin. Therefore, the unreacted RT, with higher affinity for heparin, elutes at higher NaCl. The imidazole step is not essential, as nearly all the RT (cross-linked or unreacted) binds to the heparin afterwards. While Ni-NTA is needed to purify the His-tagged RT from *E. coli*, it is not needed to purify the cross-linked RT. Subsequent tests have shown that the heparin sepharose alone is sufficient to separate cross-linked from unreacted RT.

After this purification, the p51 N-terminal His tag was removed with His-tagged HRV14 3C protease added at a mass ratio of 1:10 protease:RT for 24 h at 4°C. A small but distinct shift downwards for p51, but not for p66, could be seen on SDS-PAGE with ~0.3 µg of protein loaded. The cleaved His tags and 3C protease were removed by adding a small amount of fresh Ni-NTA beads. The cross-linked protein was then concentrated and buffer-exchanged into 75 mM NaCl, 10 mM Tris-HCl pH 8.0.

Sitting drops of 0.5 µL of protein plus 0.3 µL of well solution were set up at 4°C on CombiClover Junior plates from Emerald BioStructures. Each well contained 0.3 mL of 11.5% PEG 8000 (w/v), 0.1 M N-cyclohexyl-2-aminoethanesulfonic acid (CHES)-NaOH pH 9.5, 0.2 M NaCl. The protein solution was 10 mg/mL cross-linked RT/DNA, in 75 mM NaCl, 10 mM Tris pH 8, 20 mM MgCl₂, and 1 mM d4TTP. The tray was incubated at 4°C. Crystals began appearing in 2 days and were done growing by 2 weeks. X-ray diffraction datasets were collected

from crystals approximately $200 \times 100 \times 60 \mu\text{m}$ in size.

The RT139A sample was also cross-linked at the first base overhang. This was prepared and crystallized similarly, except that the cross-linking adenine in the template sequence was the fifth base (A) from the 5' end, not the fourth (A). The cross-linking reaction is very sensitive to the exact concentration of beta-mercaptoethanol (βME). About 0–5 mM concentrations of βME were tested in a small-scale cross-linking to the first base overhang. Reactions (25 μL) were set up with 5 μM RT139A, 15 μM of the 27/20 template-primer DNA/DNA, 5 mM MgCl_2 , 80 mM NaCl , 50 mM Tris-HCl pH 8.0, 200 μM ddGTP, and one of 0, 1, 2, 3, 4, or 5 mM βME . The solutions were incubated for 15 h at 37°C. SDS sample buffer without any reducing agent was added to 4.2 μL (2.5 μg) of each reaction. Samples were run on non-reducing SDS-PAGE (4–20%) without boiling, and the gel was stained with Coomassie Blue G-250. The maximum conversion of p66 to cross-linked p66 is about 70% at 1 mM βME (Fig. S9); at 0 or 5 mM βME only about 10% or less was converted. The p51 band intensity stays constant across βME concentration because it was not cross-linked, providing a useful check. 1 mM βME was also used for cross-linking to the second base template overhang. On SDS-PAGE, p66 cross-linked to the first-based overhang returns to the same position as unreacted p66, after reduction with sample buffer containing βME (data not shown).

Crystal soaking and freezing

For datasets collected at pH 9.0 or below, the crystals grown at pH 9.5 were pre-soaked for 1 minute in 50 μL 10% glycerol (v/v), 11.5% PEG 8 K (w/v), 0.1 M CHES-NaOH pH 9.5, 0.2 M NaCl , 20 mM MgCl_2 , 1 mM d4TTP. Then crystals were transferred for 1 min each through a descending pH series starting at pH 9.0, in steps of 0.5 pH unit. These solutions were each 50 μL in volume however, with glycerol raised to 20% (v/v) and the CHES-NaOH pH 9.5 buffer substituted with one of: CHES-NaOH pH 9.0, Tricine-NaOH pH 8.5 or 8.0, HEPES-NaOH pH 7.5 or 7.0. After the crystal arrived at the desired endpoint pH, it was then flash frozen in liquid N_2 . For datasets collected at pH 9.5, the crystals grown at pH 9.5 were transferred to 10% and 20% glycerol (v/v) solutions that were both 10% PEG 8 K (w/v), and both pH 9.5 CHES-NaOH.

X-ray crystallography

All X-ray diffraction datasets were collected from the crystals of ternary RT complexes using the F1 beamline at Cornell High Energy Synchrotron Source (CHESS). The data were processed and scaled using HKL2000³³ as summarized in Table S1. The pH 7.0 structure was solved by molecular replacement, using protein atoms from one RT heterodimer (Chains A

and B, p66 and p51, respectively) in the crystal structure of the RT/DNA/RNA/dATP (PDB ID 4PQU) complex.³ All B factors were set to 52 $\text{\AA}^2$. The subdomains were positioned by rigid-body refinements in which each RT heterodimer was divided into 13 segments. The template-primer models were taken from PDB 3V4I and fitted manually to a different map (Fo–Fc) density. For the remaining (higher) pHs, the final, refined pH 7.0 structure was used as the starting model and the molecular replacement and refinement processes were carried out in the same manner as above. Each crystal structure has two RT p66/p51 heterodimers per asymmetric unit, Chains A (p66) and B (p51), and C (p66) and D (p51). Individual structures were refined using PHENIX,³⁴ and the model building was carried out using COOT.³⁵ Structural figures were generated using PyMOL.³⁶

Significance of work. Stavudine (d4T), a drug used to treat HIV, acts as chain terminator following incorporation into viral DNA by the enzyme reverse transcriptase (RT). We present the first atomic structure of RT/DNA in a replication-competent state that binds d4T-triphosphate (d4T-TP) as a dTTP mimic. In structures obtained at different pH values, RT conformational changes were observed that are correlated with pH dependence of polymerase activity. A new, economical method for preparing cross-linked RT-DNA complexes is described.

Acknowledgments

We are grateful to CHESS staff for support and access to the F1 beamline for data collection, and to Steve Tuske for helpful discussions. This work was supported by the National Institutes of Health MERIT Award R37 AI027690 to E.A.. Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6AMO, 6AN2, 6AN8, 6ANQ, 6ASW, and 6AVM for wild-type RT (I63C)/dsDNA/d4TTP ternary complexes at pH 7.0, 7.5, 8.0, 8.5, 9.0, and 9.5, respectively, with cross-linking at the template second base overhang. The ternary complex with cross-linking at the first base overhang, at pH 9.5, has accession number 6AVT. The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

1. Sierra-Aragon S, Walter H (2012) Targets for inhibition of HIV replication: entry, enzyme action, release and maturation. *Intervirology* 55:84–97.
2. Sarafianos SG, Marchand B, Das K, Himmel DM, Parniak MA, Hughes SH, Arnold E (2009) Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* 385: 693–713.

3. Das K, Martinez SE, Bauman JD, Arnold E (2012) HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. *Nat Struct Mol Biol* 19:253–259.
4. Antiretroviral Therapy for HIV Infection in Adults and Adolescents. *Recommendations for a public health approach: 2010 revision*. Geneva: World Health Organization, 2010.
5. Brennan AT, Davies MA, Bor J, Wandeler G, Stinson K, Wood R, Prozesky H, Tanser F, Fatti G, Boule A, Sikazwe I, Wool-Kaloustian K, Yuannoutsos C, Leroy V, de Rekeneire N, Fox MP (2017) Has the phasing out of stavudine in accordance with changes in WHO guidelines led to a decrease in single-drug substitutions in first-line antiretroviral therapy for HIV in Sub-Saharan Africa? *AIDS* 31:147–157.
6. Magula N, Dedicoat M (2015) Low dose versus high dose stavudine for treating people with HIV infection. *Cochrane Database Syst Rev* 1:CD007497.
7. Gray LR, Tachedjian G, Ellett AM, Roche MJ, Cheng WJ, Guillemin GJ, Brew BJ, Turville SG, Wesselingh SL, Gorry PR, Churchill MJ (2013) The NRTIs lamivudine, stavudine and zidovudine have reduced HIV-1 inhibitory activity in astrocytes. *PLoS One* 8:e62196.
8. Vaccaro JA, Parnell KM, Terezakis SA, Anderson KS (2000) Mechanism of inhibition of the human immunodeficiency virus type 1 reverse transcriptase by d4TTP: an equivalent incorporation efficiency relative to the natural substrate dTTP. *Antimicrob Agents Chemother* 44:217–221.
9. Huang H, Chopra R, Verdine GL, Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 282:1669–1675.
10. Huang H, Harrison SC, Verdine GL (2000) Trapping of a catalytic HIV reverse transcriptase/template:primer complex through a disulfide bond. *Chem Biol* 7:355–364.
11. Corn JE, Berger JM (2007) FASTDXL: a generalized screen to trap disulfide-stabilized complexes for use in structural studies. *Structure* 15:773–780.
12. Corn JE, Pelton JG, Berger JM (2008) Identification of a DNA primase template tracking site redefines the geometry of primer synthesis. *Nat Struct Mol Biol* 15: 163–169.
13. Hou X, Wang G, Gaffney BL, Jones RA (2009) Synthesis of guanosine and deoxyguanosine phosphoramidites with cross-linkable thioalkyl tethers for direct incorporation into RNA and DNA. *Nucleosides Nucleotides Nucleic Acids Res* 28:1076–1094.
14. Das K, Martinez SE, Bandwar RP, Arnold E (2014) Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. *Nucleic Acids Res* 42:8125–8137.
15. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera – a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.
16. Wallden K, Ruzzenente B, Rinaldo-Matthis A, Bianchi V, Nordlund P (2005) Structural basis for substrate specificity of the human mitochondrial deoxyribonucleotidase. *Structure* 13:1081–1088.
17. Meyer P, Schneider B, Sarfati S, Deville-Bonne D, Guerreiro C, Boretto J, Janin J, Veron M, Canard B (2000) Structural basis for activation of alpha-boranophosphate nucleotide analogues targeting drug-resistant reverse transcriptase. *EMBO J* 19:3520–3529.
18. Viterbo D, Milanesio M, Pomez Hernandez R, Rodriguez Tanty C, Colas Gonzalez I, Sablon Carrazana M, Duque Rodriguez J (2000) 2',3'-Didehydro-3'-deoxythymidine N-methyl-2-pyrrolidone solvate (d4T.NMPO). *Acta Cryst C* 56:580–581.
19. Ueno T, Mitsuya H (1997) Comparative enzymatic study of HIV-1 reverse transcriptase resistant to 2',3'-dideoxynucleotide analogs using the single-nucleotide incorporation assay. *Biochemistry* 36:1092–1099.
20. Das K, Martinez SE, Arnold E (2017) Structural insights into HIV reverse transcriptase mutations Q151M and Q151M complex that confer multinucleoside drug resistance. *Antimicrob Agents Chemother* 61:1–11.
21. Arion D, Kaushik N, McCormick S, Borkow G, Parniak MA (1998) Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry* 37:15908–15917.
22. Meyer PR, Matsuura SE, So AG, Scott WA (1998) Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc Natl Acad Sci U S A* 95:13471–13476.
23. Boyer PL, Sarafianos SG, Arnold E, Hughes SH (2001) Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. *J Virol* 75:4832–4842.
24. Tu X, Das K, Han Q, Bauman JD, Clark AD Jr, Hou X, Frenkel YV, Gaffney BL, Jones RA, Boyer PL, Hughes SH, Sarafianos SG, Arnold E (2010) Structural basis of HIV-1 resistance to AZT by excision. *Nat Struct Mol Biol* 17:1202–1209.
25. Rubinek T, Loya S, Shaharabany M, Hughes SH, Clark PK, Hizi A (1994) The catalytic properties of the reverse transcriptase of the lentivirus equine infectious anemia virus. *Eur J Biochem* 219:977–983.
26. Das K, Balzarini J, Miller MT, Maguire AR, Destefano JJ, Arnold E (2016) Conformational states of HIV-1 reverse transcriptase for nucleotide incorporation vs. pyrophosphorolysis-binding of foscarnet. *ACS Chem Biol* 11:2158–2164.
27. Radzio J, Sluis-Cremer N (2005) Stereo-selectivity of HIV-1 reverse transcriptase toward isomers of thymidine-5'-O-1-thiotriphosphate. *Protein Sci* 14: 1929–1933.
28. Reardon JE (1992) Human immunodeficiency virus reverse transcriptase: steady-state and pre-steady-state kinetics of nucleotide incorporation. *Biochemistry* 31: 4473–4479.
29. Hsieh JC, Zinnen S, Modrich P (1993) Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase. *J Biol Chem* 268:24607–24613.
30. Benkovic SJ, Schray KJ. In: Boyer PD, Ed., (1973) *The Enzymes: Group Transfer Part A*. New York: Academic Press; p. 201–238.
31. Gupta SK, McComsey GA, Lombaard J, Echevarria J, Orrell C, Avihingsanon A, Osiyemi O, Santoscoy M, Ray N, Stock DA, Joshi SR, Hanna GJ, Lataillade M (2016) Efficacy, safety, bone and metabolic effects of hiv nucleoside reverse transcriptase inhibitor BMS-986001 (a1467003): a phase 2b randomised, controlled, partly blinded trial. *Lancet HIV* 3:e13–e22.
32. Bauman JD, Das K, Ho WC, Baweja M, Himmel DM, Clark AD Jr, Oren DA, Boyer PL, Hughes SH, Shatkin AJ, Arnold E (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucl Acids Res* 36:5083–5092.

33. Otwinowski Z, Minor W. *Denzo and scalepack*. Boston: Kluwer Academic Publishers, 2001.
34. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) Phenix: a comprehensive PYTHON-based system for macromolecular structure solution. *Acta Cryst D*66:213–221.
35. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D*60: 2126–2132.
36. The PyMOL molecular graphics system, version 1.8 Schrödinger, llc.