World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Human Immunodeficiency Viral Reverse Transcriptases: Analyses of the Active Sites of the Polymerase Domain and Drug-Resistant Mutants of the Reverse Transcriptases

Peramachi Palanivelu *

Department of Molecular Microbiology, School of Biotechnology, Madurai Kamaraj University, Madurai – 625 021, India

World Journal of Advanced Research and Reviews, 2024, 22(03), 409–426

Publication history: Received on 26 April 2024; revised on 04 June 2024; accepted on 06 June 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.3.1699

Abstract

It is well-known that human immunodeficiency viruses (HIVs) cause the chronic, potentially life-threatening condition known as human Acquired Immuno-Deficiency Syndrome (AIDS). As the lifecycle of HIVs heavily depends on the crucial enzyme, the reverse transcriptase (RT), it has been used as a potential therapeutic target to treat and control the spread of AIDS. The active site amino acids of the polymerase domain of the RTs and their anti-HIV drug-binding sites are analyzed. The catalytic region of HIV RTs and the *E. coli* DNA polymerase I showed very similar active site amino acids, suggesting that HIV RTs would have possibly evolved from the bacterial DNA polymerase. The catalytic proton abstractor is identified as a K and the nucleotide selection amino acid as an N. However, the regular template-binding pair –YG- is slightly modified to a –YXG- in HIV RTs. Three completely conserved Ds in HIV RTs are involved in binding to the catalytic Mg^{2+}. The sensitive and resistant strains of HIV-1 for the HIV antiretroviral drug, azidothymidine (AZT), a nucleoside analogue of thymidine, show only a few non-isofunctional amino acid replacements and are located mainly in the palm and thumb subdomains of the RT polymerase, whereas the rest of the polymerase catalytic core and metal-binding sites are completely conserved in AZT-sensitive and -resistant HIV-1 strains. In the drug-resistant mutants of non-nucleoside RT inhibitors of HIV-2, the crucial mutations are located mainly near the -MDD- motif of the catalytic metal-binding region. Even though a large number of amino acid replacements are seen between the RTs of HIV-1 and HIV-2, the polymerase active sites are completely conserved in both. The HIV-1 and HIV-2 catalytic and metal-binding sites are completely conserved in simian immunodeficiency virus (SIV) as well. The absence of DEDD-superfamily of proofreading exonuclease domain in the HIV RTs, might cause the virus to evolve rapidly in patients. A possible mechanism of action for the HIV RTs is also proposed.

Keywords: Human Immunodeficiency Viruses; AIDS; Reverse Transcriptases; Reverse Transcriptase Inhibitors; Polymerase Domain; Active Sites; Mechanism of Action.

1. Introduction

HIV-1 was first isolated in 1983, and its association with AIDS was confirmed in 1984. HIV infection attacks mainly the body's immune system, and if not treated, it can lead to AIDS, which is the most advanced stage of the disease. Currently, there is no effective cure for AIDS. AIDS has caused at least 40 million deaths worldwide so far. In 2021 alone, there were 6,50,000 deaths and ~ 38 million people are living with HIV worldwide and ~ 3 million new infections are reported every year [1]. HIVs target mainly the body's white blood cells and weaken the immune system. It mainly destroys $CD4^+$ T cells, which are T helper (T_h) cells that express the surface protein CD4. These cells play crucial roles in the activation of the adaptive immune system and in achieving a regulated, and effective immune response to pathogens. Thus, the HIV infection results in the progressive destruction of $CD4^+$ T lymphocytes, leading to the inexorable collapse of the immune function.

* Corresponding author: Peramachi Palanivelu

HIV viruses belong to the group of retroviruses and possess single-stranded, positive-sense, linear RNA genomes. They belong to the family *Retroviridae,* which are now divided into two subfamilies: *Spumaretrovirinae* and *Orthoretrovirinae*. The spumaretroviruses are highly prevalent in diverse, non-primate mammalian families [e.g., Bovidae (cloven-hooved ruminants), Felidae (cats), Equidae (horses and relatives), and Rhinolophidae (horseshoe bats), as well as non-human primates including apes, Old- and New-world monkeys, and prosimians]. On the other hand, the orthoretroviruses are divided into six genera, viz. alpha, beta, gamma, delta, epsilon, and lentiviruses. Among them, the lentiviruses, also known as the slow viruses (lentus, in Latin for slow), include the deadly human immunodeficiency viruses (HIV-1 and HIV-2) that characteristically attack the immune system and cause AIDS in humans. This group also includes simian (apes and monkeys) and feline (cats) immunodeficiency viruses.

1.1. Lifecycle and Genome Structure of HIVs

The HIVs possess RNA genomes and are completely sequenced. The HIV-1 RNA genome is 9750 nucleotides long, and the virions measure ~120 nm in diameter. HIV-1 virions contain two copies of a single-stranded RNA genome (known as psuedodiploidy) within a conical-type capsid, surrounded by a plasma membrane of host-cell origin, embedded with the viral envelope proteins. During infection, the envelope glycoprotein gp120 of the virus interacts with the cell surface receptor, CD4 on the surface of the T cells. HIV-1 enters the host cells through interactions with the CD4 receptor and a chemokine co-receptor (CXCR4 or CCR5), resulting in the fusion of the virus with the host cell and insertion of the viral genome along with its enzymes (an RT, a protease and an integrase). After successful infection, the viral enzyme RT, an RNA-dependent DNA polymerase (EC 2.7.7.49) transcribes the RNA genome into a DNA copy that is followed by the integration into the host chromosome by the viral integrase to form a provirus. Then the infected CD4 cells start making new copies of the virus in large numbers and the virus is finally released from the CD4 cells by a process known as budding. The new HIV viruses infect other CD4 cells and repeat their lifecycle. The lifecycle of the virus is only 1-2 days. It is interesting to note that the provirus may remain latent for years, producing few or no new copies of HIV, which has hampered the treatment of individuals infected with HIV, as antiretrovirals can only target the replicating virus. Thus, in the absence of any treatment, HIV infection proceeds unchecked, killing more and more of the CD4 cells, and finally destroying the immune system. Although HIV-1 and HIV-2 belong to the same family of *Retroviridae* and subfamily *Orthoretrovirinae*, HIV-1 shows worldwide infection, whereas HIV-2 is reported predominately in West Africa.

The HIV genomic RNA possesses a 5’-cap, a 3’-poly(A) tail, and many open-reading frames (ORFs). There are two long terminal repeats (LTRs) of about 600 nt long at the 5'- and 3'- ends of the viral genome. The LTRs are the control centre of gene expression in retroviruses. That is, all the regulatory elements for gene expression are found in the LTRs, like the enhancer, promoter, transcription initiation (capping), transcription termination and polyadenylation signal. The 5’-LTR contains both the enhancer and promoter elements. The 3'-LTR, although it has exactly the same sequence arrangement as the 5' LTR, is not normally functional as a promoter. The viral gene expression, directed by the LTR signals is carried out entirely by host cell enzymes (RNA pol II, poly A synthetase, guanyl transferase). The LTRs are subdivided into three regions: U3 (a unique element), R (a repeat element), and U5 (a unique element that possesses a specific sequence required for efficient polyadenylation). The enhancer and other transcription regulatory signals are contained in the U3 region of the 5'-LTR, and the TATA box is located roughly 25 bp from the beginning of the R sequence. Viral structural proteins are encoded by the longer ORFs, whereas the shorter ORFs encode regulators of the viral lifecycle, like for attachment, membrane fusion, replication, and assembly. The entire lifecycle is completed in the CD4 cells.

The RT is encoded by the *pol* gene that is cleaved from a polyprotein, known as the ‘gag-pol’. The *gag* gene is one of the three ORFs of the retrovirus family. (Gag, a group-specific antigen, is the major structural protein of all retroviruses and comprises ~ 50% of the mass of a viral particle). The *gag* gene codes for the four core structural proteins, viz. p17 (matrix protein), p24 (viral capsid), p7 and p6 (nucleo-capsid proteins). The second ORF is the *pol* gene, which encodes another polyprotein containing the precursor to the viral enzymes, viz. a protease (p11), a RT (p66) and an integrase (p32). The RT converts the viral RNA genome into a double-stranded (ds) DNA. The DNA copy of the virus is integrated into the host's genome with the viral enzyme, integrase, and the viral DNA becomes part of the cellular DNA and replicates along with it. The third ORF, *env*, encodes two envelope glycoproteins, viz. gp120 (a spike protein) and gp41 (a transmembrane protein), which are displayed on the viral surface and used for attachment of the virus onto the host cell and to fuse with them. In addition to the above three longer ORFs, other regulatory genes like, *tat, rev, nef, vif, vpr, vpu*, encode for a single protein and are called by the same name [2].

1.2. Salient Features of RTs

RTs are a diverse group of enzymes and exhibit 2 different enzyme activities, viz. i) a DNA polymerase activity which uses either DNA or RNA as a template, and ii) an RNase H activity, which hydrolyses specifically the RNA strand within

an RNA/DNA hybrid. Both the DNA polymerase and RNase H activities are essential for the successful viral replication in the host cells. Therefore, the retroviral RT is a multifunctional enzyme that eventually converts the viral RNA genome into a dsDNA in the cytoplasm, shortly after entry of the virus into the host cells. Like DNA polymerases, RTs also require both a primer and a template. However, unlike DNA polymerases, the RTs are error-prone enzymes as they lack a proofreading exonuclease activity. Therefore, high mutation rates of RTs are the direct consequence of this characteristic. Interestingly, this property of the enzyme causes the virus to evolve rapidly in patients and particularly creates challenging problems for vaccine development and anti-RT drug therapy [3]. In addition to that, the RTs also display frequent template-switching, leading to high recombination rates. (Recombination mostly occurs between homologous regions of the two co-packaged HIV RNA genomes. If these two RNA molecules are derived from different viral strains, reverse transcription will give rise to highly recombinant proviral DNAs).

The HIV-1 RT is a heterodimer, comprised of a 66-kDa, (560 amino acid subunit p66), and a 51-kDa, (440 amino acid subunit p51) subunits. (HIV-2 RT is also a heterodimer of p68/p55 and corresponds to p66/p51 subunits of HIV-1 RT). The main subunit p66 contains the polymerase subdomains, viz. fingers, palm and thumb of the RT (these are the regular polymerase subdomains reported in other DNA polymerases too), a connector region, and the RNase H domain. The two independent, but functionally related enzymes of the RT suggest that the genes for the two enzyme activities would have fused to form a multifunctional enzyme during evolution, as suggested for *E. coli* DNA polymerase I, where the genes for three functionally related enzymes are fused to form the multifunctional polymerases and exhibits three different enzyme activities. The three RT polymerase subdomain regions are identified as: fingers (residues 1–85 and 118–155), palm (residues 86–117 and 156–236) and thumb (residues 237–318), and the connector region (residues 319–426) which connects the polymerase to the RNase H domain [4]. Interestingly, the p51 subunit has the identical sequence of p66, but lacks the RNase H domain. Interestingly, RTs find important applications in advanced diagnostic tools based on RNA analysis, where they are used for fast and direct 'One-Step RT-PCR' assays. (In these assays, the first-strand complementary DNA synthesized by the RT, is exponentially amplified in the end-point or real-time PCRs).

1.3. Inhibitors of RTs: Nucleoside and Non-nucleoside inhibitors

Two distinct types of RT inhibitors, which block the polymerase activity of the RT are approved for the treatment of AIDS. Based on their mechanism of action, the HIV RT inhibitors are broadly classified into two groups: i) nucleoside analog RT inhibitors (NRTIs, a competitive type) and ii) non-nucleoside RT inhibitors (NNRTIs, a non-competitive type). NRTIs are prodrugs that are converted into their active form, i.e., to their 5'-triphosphates by cellular enzymes. Importantly, as they lack the essential 3'-hydroxyl groups required for chain extension their absence results in chain termination. The following NRTIs have been approved by the US Food and Drug Administration (FDA): AZT-zidovudine, 3TC-lamivudine, FTC-emtricitabine, ddI-didanosine, and ABC-abacavir. All the approved NRTIs lack the 3'-OH and act as chain terminators when added to the viral DNA by the RT. For NRTIs to be effective against HIV, they must be taken up by the host cell and then phosphorylated by cellular enzymes to convert them into their active forms, i.e., their triphosphates. The efficiency of this conversion to the active metabolite and the stability of NRTIs (and their triphosphates) in the presence of catabolic enzymes and their effective concentrations in the bloodstream are important considerations in this type of antiviral therapy.

On the other hand, the NNRTIs bind to HIV-1 reverse transcriptase at a hydrophobic site remote from the enzyme's active site to produce a conformational change on the enzyme that prevents substrate-binding and enzyme activity. Biochemical data have shown that NNRTIs are non-competitive inhibitors and do not directly interfere with the binding of either the dNTP or the nucleic acid substrates of the RT. There are five NNRTI drugs [nevirapine (approved in 1996), delavirdine (first-generation drug, approved in 1997), efavirenz (second-generation drug, approved in 1998), and etravirine (third-generation, approved in 2008), and rilpivirine, approved in 2011] that are currently approved by the FDA for treating HIV-1 infections. Pre-steady state kinetic analysis of single nucleotide addition in the presence of NNRTIs has shown that the binding of NNRTI interferes with the chemical step in DNA synthesis. However, the molecular details of NNRTI inhibition are not clearly understood [5]. Combination drugs have become more popular in the treatment of AIDS. Atripla, a combo-drug approved in 2006, contained three inhibitors for reverse transcriptase (efavirenz (NNRTI), emtricitabine (an NRTI, a cytidine analogue) and tenofovir disoproxil fumarate (an NRTI, an adenosine analogue) which reduced the number of pills to one and three times a day, instead of 15-20 pills. Many combination drugs are being developed to contain the virus and also to help ease treatment regimens for patients. Recently, one of the combination drugs, developed with three antivirals (emtricitabine; the integrase inhibitor, bictegravir; and tenofovir) reduced the treatment to a single pill per day. These antivirals are effective at knocking down the replication of HIVs and reducing the viral load, but they cannot cure. As soon as someone stops taking the treatment, the virus rebounds. Besides, HIVs develop resistance to all the available drugs, making it harder to treat them. Therefore, searching for new HIV drugs has become a continuous process.

As compared to the NNRTIs, the NRTIs often have several disadvantages, such as low bioavailability, high toxicity, efficiency of prodrug activation and, most notably, the tendency to develop drug resistance. Therefore, NNRTIs are designed to solve these problems. Inhibition by both types of inhibitors has been studied extensively by several groups. Interestingly, they found that the mutations that confer resistance to NRTIs and NNRTIs are located in the polymerase domain of HIV RTs [4, 6-10]. Mutations that have been shown to confer resistance to NNRTIs are found to cluster around the hydrophobic pocket filled by Y^{181} and Y^{188}, suggesting that most of these resistance mutations indirectly lead to distortion of the geometry of the polymerase active site and block polymerization. The main contribution to drug resistance to the first-generation NNRTIs is found to be due to Y^{181}→Cys and Y^{188}→Cys RT mutations (these two mutations are located on either side of the –M^{184}DD-, the catalytic metal-binding motif of the polymerase domain). Such modifications result in the loss of the aromatic ring stacking interactions at the catalytic metal-binding motif. Therefore, these two mutations at either Y^{181} and/or Y^{188} within the HIV-1 RT polymerase domain give a high level of resistance to many of the first-generation NNRTIs such as the main anti-AIDS drug, Nevirapine. By comparison, the second generation NNRTIs like the drug Efavirenz, show much greater efficacy.

Although tremendous progress has been made over the past 20 years in characterizing the structures of proteins and enzymes from HIVs, many unanswered questions still remain. The RT polymerase domain and the mutation sites that make the HIV drugs ineffective are analyzed and reported in this communication.

2. Materials and Methods

The protein sequence data of RTs from HIV-1, HIV-2 and simian viruses were obtained from the PUBMED and the SWISS-PROT databases. The advanced version of Clustal Omega was used for protein sequence analysis [11]. The polymerase and PR active sites are arrived at by sequence similarities, site-directed mutagenesis (SDM), chemical modification of active site amino acids and X-ray crystallographic data. The ExPASy tool was used to determine the pI values of HIV RTs.

3. Results and Discussion

3.1. Active sites of the HIV RTs

Figure 1 shows the MSA of RTs from different HIV-1 strains and a simian immunodeficiency virus. (Only the regions required for discussions are shown here). The H9BTT2-HIV-1 sequence is used as the reference and is highlighted in yellow. Different regions of the RT polymerase subdomains are highlighted in different colours (Fingers in yellow, Palm in green, Thumb in magenta, Connector region in grey and the RNase domain in red). All the polymerase subdomains and the RNase H domain are highly conserved among different strains of the HIV-1 and simian virus. Interestingly, among the polymerase subdomains, the finger subdomain contains a large number of conserved basic amino acids, K and R. The three catalytic metal-binding Ds (D^{110}, D^{185} and D^{186}) are located in the palm subdomain and are completely conserved in all HIV-1 strains (–^{184}MDD-) and simian virus (highlighted in dark green). The proposed polymerase catalytic core region, -$^{\mathbf{259}}$K^{-4}LVGKL^{1}NWASQIY^{8}A/P/Q/SG^{10}- is located in the thumb subdomain and is again completely conserved in all HIV-1 strains and the simian virus. The catalytic proton abstractor amino acid K, and the nucleotide discriminating amino acid, K at -4 from the catalytic K are completely conserved. Interestingly, the RT polymerase active-site, -$^{\mathbf{258}}$QK^{-4}LVGK$^{\mathbf{263}}$L^{1}NWASQIY^{8}A/P/Q/S/GI- is found to be similar to the confirmed active site of *E. coli* DNA pol I, -QR^{-4}RSAK^{758}A^{1}INFGLIY^{8}GM- [12] and in close agreement to the active sites of other DNA/RNA polymerases already reported (Table 1) [12, 13] The usual G residue in the template-binding -YG- pair is replaced with different amino acids in different HIV-1 strains, but followed by an invariant G. (The natural mutants in these subdomains are highlighted in red which play a major role in drug-resistant phenotypes). A -KV/IK- tirade of direct repeat is found in the N–terminal domain.

X-ray crystallographic analysis of the palm subdomain of p66 revealed a considerable structural similarity to the polymerase active site of the Klenow fragment of *Escherichia coli* DNA polymerase I [6]. The RT polymerase domains showed a large cleft analogous to that of the Klenow fragment of *E. coli* DNA polymerase I, suggesting that these polymerases would have diverged from a common ancestor. However, the subdomains that were likely to bind the template strand at the polymerase active site had a different structure in the two polymerases. Huang *et al*. [14] studied the ternary complex, RT-template:primer-dNTP, by disulfide trapping of the RT on an RNA:DNA heteroduplex template:primer. They found that Q^{258}-C in HIV-1 RT strongly prefers to cross-link a tethered G residue located on the template strand, six base pairs away from the nucleotide bound to the active site. The particular Q^{258} is very close to the proposed polymerase catalytic site as shown here, -Q^{258}K^{-4}LVGK^{263}L^{1}NWASQIY^{8}PG-.

Further, X-ray crystallographic structure of a ternary complex (RT-dsDNA-Fab) of the enzyme was reported by Jacobo-Monolina *et al*. [4]. Each subunit of the heterodimer consists of the three common polymerase subdomains, viz. fingers, palm, thumb, and a connector domain. Although the structures of the subdomains within p66 and p51 were found to be similar, the relative arrangement of the three subdomains and the connector domain within the two subunits was found to be different. In addition to the above domain and subdomains, the carboxyl terminus of the p66 subunit has a fifth domain with RNase H activity [4, 6].

Further, X-ray crystallographic studies have revealed that the polymerase subdomains of the p66 subunit formed a large nucleic acid binding cleft. The template-primer bound in the cleft formed by the fingers, palm, and thumb subdomains of p66 [4]. Also, it was found that the primer 3'-OH was positioned close to the polymerase active site, and numerous contacts between the enzyme and the DNA occurred in the palm, thumb, and fingers subdomains. As discussed elsewhere, the palm subdomain harbours the catalytic metal-binding site of the polymerase that is defined by a triad of Asp residues at positions D^{110}, D^{185} and D^{186} and the last two Ds are found in the invariant –MDD- motif. It was suggested these amino acids might bind the divalent cations (Mg^{2+}) that are required for the enzyme catalysis. The 3'-OH of the primer terminus was close to the catalytic triad and was appropriately positioned for nucleophilic attack on the α-phosphate of an incoming nucleoside triphosphate. The αH helix of the p66 thumb made contact with the sugar-phosphate backbone of the primer strand, whereas the adjacent antiparallel helix, αI, made contact with the sugar-phosphate backbone of the template strand. Therefore, it was suggested that these helices might function as tracks over which the template-primer moves during translocation. Besides, a bend observed at the template-primer terminus may have functional implications for RT catalysis, translocation, fidelity, and/or processivity. Moreover, the structure also showed two conserved Ys in the metal-binding pocket, viz. Y^{181} and Y^{188} (marked in red), where NNRTIs of the HIV-1 RT might bind (however, the Y^{181} is not conserved as it is replaced by a C in many strains of HIV-1 RT, whereas the Y^{183} is completely conserved in all) (Fig. 1). Of the p66/p52 heterodimer complex, the p5l subunit did not show any polymerase activity as it lacks any DNA-binding cleft [15, 16]. However, in the absence of p66, the p51 showed activity and DNA synthesis occurred on heteropolymeric RNA and DNA templates [15]. Thus, it was concluded that p51 *per se* was active, but the activity was masked in the presence of p66.

3.2. RT Inhibitors that Bind to the Polymerase Domain of HIV RTs

Both the nucleoside and non-nucleoside analogue inhibitors are shown to bind at or near active site amino acids of the polymerase domain. Insights into their binding sites are obtained by further analysis.

3.2.1. Nucleoside Analogue RT Inhibitors

HIV RTs are the target of many antiviral drugs including 3'-azido-3' deoxythymidine (AZT/Zidovudine), dideoxyinosine (ddI), and dideoxycytidine (ddC), and are the only nucleoside analogue drugs currently approved for treating HIV infections. Among them, AZT therapy is the first one that was approved in 1987. Although these drugs are widely used, each shows serious side effects, which include toxicity and rapid emergence of resistant HIV strains [17, 18]. In further analysis of the resistant strains, Lader and Kemp [17] identified three predicted amino acid substitutions common to all the HIV resistant strains, viz. (D^{67} →N, K^{70} →R, T^{215} →F/Y, highlighted in red in Fig. 1) and a fourth amino acid, viz. (K^{219} →Q) in three of the isolates. However, Kondo [19] found that in the 41 mutant clones of HIV-1 analyzed from 7 patients, the **T^{215}** mutation was the most predominant (97.6%). Thus, the single amino acid mutation in the **T^{215}** codon was found to be the most important factor in AZT resistance and it should be noted that the **T^{215}** is located in the palm subdomain whereas the other two are in the fingers subdomain (Fig. 1). It is interesting to note that an infectious molecular clone constructed with all these four mutations in HIV-1 RT yielded highly resistant HIV after transfection of T cells.

Furthermore, the **K^{65}** mutation in the –IK**K^{65}**K- motif of the fingers subdomain was found to be responsible for virus resistance to ddC and 2',3'-dideoxy-3'-thiacytidine [20] (Fig. 1). Interestingly, among the nucleoside analogue mutants, the most resistant to clinical treatment for HIV-1 infection, is the M^{184}→V in the –M^{184}DD- motif of the catalytic metal-binding site. (highlighted in red in Fig. 1) suggesting direct inhibition of the metal-binding leads to HIV-resistant strains [21]. The phenoxyl side-chain of Y^{183}, which is part of the conserved -^{183}YMDD- motif, has hydrogen-bonding interactions with nucleotide bases of the second duplex base-pair and is predicted to have at least one hydrogen bond with all Watson-Crick base-pairs at this position [8]. Analysis of HIV-1 variants confirmed that the ddI resistant mutation conferred both ddI as well as ddC resistance, but suppressed the effect of the AZT resistance mutation. Therefore, it is suggested that the use of a combination therapy for HIV-1 may prevent the emergence of drug-resistant strains [18]. Some mutations that cause resistance to the nucleoside analogues, such as AZT, ddI, and ddC, are located in close proximity to the dNTP-binding site sufficient to directly interfere with the binding of nucleoside analogues, while many are located away from the dNTP-binding site [22].

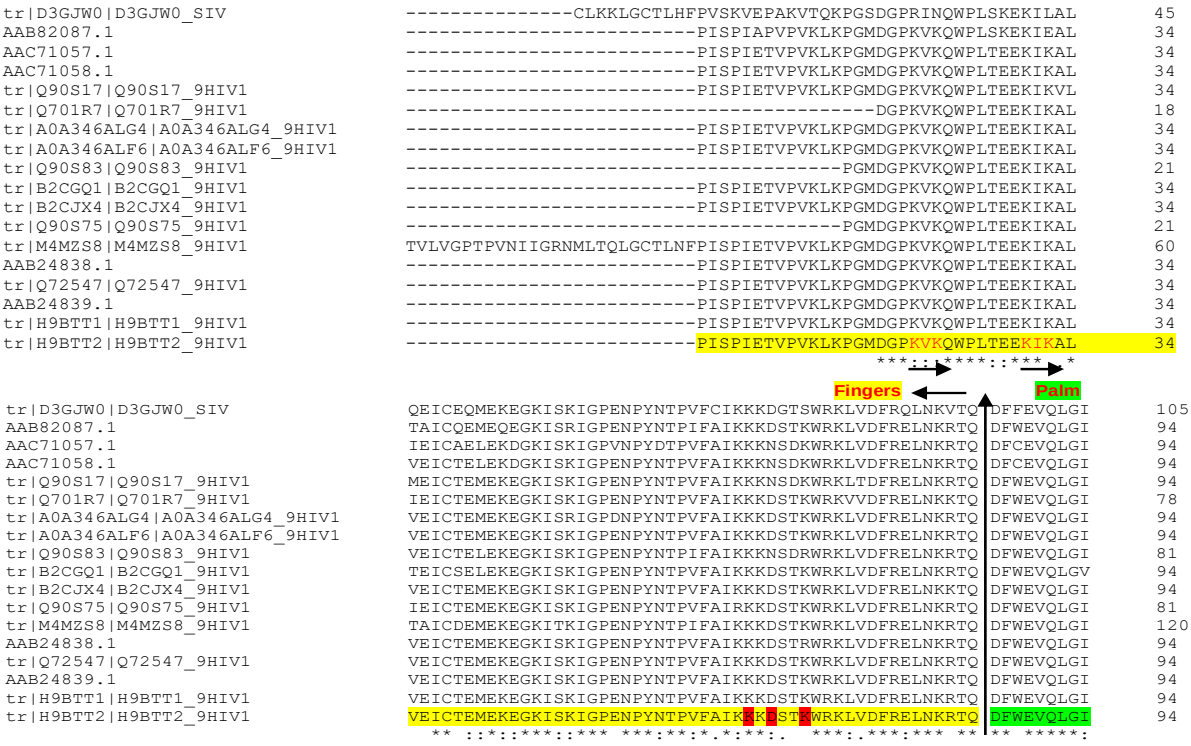
3.2.2. Non-nucleoside RT Inhibitors

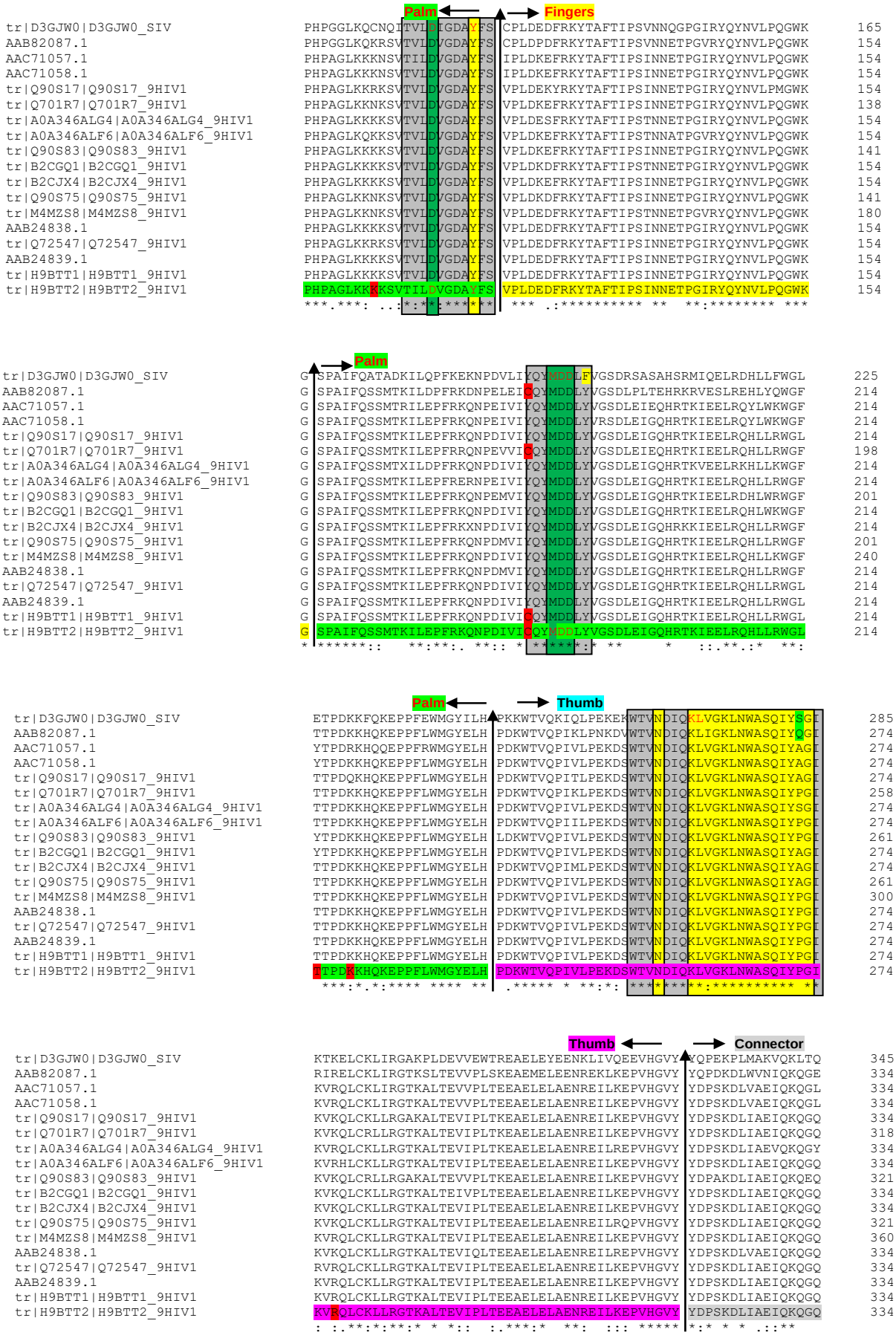
In addition to the nucleoside analogue HIV drugs, numerous non-nucleoside compounds like, the TIBO compounds tetrahydro-imidazo[4,5,1-jk][l,4]-benzodiazepin-2(1H)-one and –thione, TIBO derivatives [23] and Nevirapine [24] are also found to be effective, especially, against HIV-1 RT. In contrast to nucleoside analogue inhibitors which need to be converted to triphosphates by the host cells, these NNRTIs directly inhibit HIV-1 RT by a non-competitive mechanism. TIBO and Nevirapine are found to be effective against HIV-1, but not against HIV-2. This was further confirmed by Pauwels *et al*. [23] where they found that the TIBO derivatives inhibited the replication of HIV-1, but not of HIV-2, or of any other DNA or RNA viruses, whereas the NRTIs inhibited both the HIV RTs, suggesting their competition with the nucleotide binding site.

HIV-1 has been divided into three major groups: M, O and N, Main, Outlier and Non-M/Non-O, respectively. Group M viruses are found globally and are largely responsible for the AIDS pandemic, while group O and N viruses are restricted to West and Central Africa. The low spread of the newest group P, indicates it most likely emerged very recently. These O group isolates showed high level of resistance to the NNRTIs, viz. nevirapine and loviride, suggesting that the NNRTIs bind not at the active site(s), but close to the active site(s). In the O group, the V^{179} and Y^{181} are mutated to E^{179} and C^{181}, (i.e.), near the catalytic metal-binding motif. HIV-2 isolates are resistant to NNRTIs, an effect that appears to be mediated by the presence of I^{181} and L^{188} at positions Y^{181} and Y^{188}, respectively of the RT sequence. (Fig. 3).

Nunberg *et al*. [25] found that the TIBO-resistant RT genes encoded two amino acid changes, K^{103} →N and/or Y^{181} →C, each of which contributed partial resistance (present in the palm subdomain and highlighted in red). The mutation at amino acid Y^{181} lies adjacent to the conserved metal-binding motif -Y^{181}QYM^{184}DD- motif (in the palm subdomain). The second mutation at amino acid K^{103} lies within the polybasic -KKK^{103}K- motif in the fingers domain (Fig. 1). Thus, K^{103}→N and Y^{181}→C mutations contributed partial resistance. However, the combination of the two mutations was more than additive, resulting in ~2,000-fold resistance to NNRTIs, but sensitive to AZT. However, it was found that the strains of the virus, resistant to these compounds, also arise rapidly [25]. These results suggest that the virus resistance for NNRTIs is mainly due to substitutions of amino acids in fingers and palm subdomains, whereas for the NRTIs it is mainly in the palm subdomain. Usually, the DEDD-superfamily of PR exonuclease domain that precedes the polymerase domain in other polymerases [12] is not found in HIV RTs and confirming that the RTs are PR exonuclease deficient. The polymerase domain is followed by a connector region of ~ 200 amino acid residues, which is followed by the RNase H domain. Both the connector and the RNase H are also highly conserved in all, including the SIV (Fig. 3).

CLUSTAL O (1.2.4) MSA of RTs from HIV-1 and SIV




Palm
Fingers
Palm
Palm
Thumb
Thumb
Connector

```
tr|D3GJW0|D3GJW0_SIV              GQWSYQIEQEENKPLKVGKYARTKNAHTNELRVLAGLVQKIAKEALVIWGQLPRFYLPIE  405
AAB82087.1                        GQWTYQIYQDEHKDLKTGKYTRQKASHTNDIRQLAEVLQKVSQESIVIWGKLPKFKLPVT  394
AAC71057.1                        GQWTYQIYQEPFKNLKTGKYAKMKGAHTNDVKQLTEVVQKVATESIVIWGKTPKFRLPIQ  394
AAC71058.1                        GQWTYQIYQEPFKNLKTGKYAKMKGAHTNDVKQLTEVVQKVATESIVIWGKTPKFRLPIQ  394
tr|Q90S17|Q90S17_9HIV1           GQWTYQIYQEPFKNLKTGKYAKMKGAHTNDVRQLTEAVQKITTESIVIWGKTPKFKLPIQ  394
tr|Q701R7|Q701R7_9HIV1           GQWTYQIYQEPFKNLKTGKYARTRGAHTNDVKQLTEAVQKIATEGIVIWGKTPKFKLPIQ  378
tr|A0A346ALG4|A0A346ALG4_9HIV1   GQWSYQIYQEPFKNLKTGKYARMRGXHTNDVKQLTEAVQKITTESIVIWGKIPKFRLPIQ  394
tr|A0A346ALF6|A0A346ALF6_9HIV1   GQWTYQIYQEPFKNLKTGKYAKMRSTHTNDVKQLTEAVQKIATEGIVIWGKIPKFRLPIQ  394
tr|Q90S83|Q90S83_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  381
tr|B2CGQ1|B2CGQ1_9HIV1           GQWTYQIYQEPFKNLKTGKYAKMKGAHTNDVRQLTEAVQKITTESIVIWGKIPKFKLPIQ  394
tr|B2CJX4|B2CJX4_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTETVQKIXTESIVIWGKTPKFKLPIQ  394
tr|Q90S75|Q90S75_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  381
tr|M4MZS8|M4MZS8_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  420
AAB24838.1                        GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKIPRFKLPIQ  394
tr|Q72547|Q72547_9HIV1           GQWTYQIYQEPFKNLRTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  394
AAB24839.1                        GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPRFKLPIQ  394
tr|H9BTT1|H9BTT1_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  394
tr|H9BTT2|H9BTT2_9HIV1           GQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQ  394
                                  ***:*** *:  * *:.***:: :  ***::: *:  :**:  *.:****: *:* **:
```

```
                                                  Connector ◄──┃──► RNase H
tr|D3GJW0|D3GJW0_SIV              REVWDQWWPEYWQVTWIPDWEFISTPPLIRLWYNLLKEPIPGEDVY┃YVDGAANRTSKLGK  465
AAB82087.1                        RETWETWWADYWQATWIPEWEYVSTPPLIKLWYRLESEPIMGAETY┃YVDGAANRDTKLGK  454
AAC71057.1                        RETWEAWWMEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
AAC71058.1                        KETWEAWWMEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|Q90S17|Q90S17_9HIV1           KETWGAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|Q701R7|Q701R7_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIIGAETF┃YVDGAANRETKLGK  438
tr|A0A346ALG4|A0A346ALG4_9HIV1   KETWEAWWMEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|A0A346ALF6|A0A346ALF6_9HIV1   KETWEAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKIGK  454
tr|Q90S83|Q90S83_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIIGAETF┃YVDGAANRETKLGK  441
tr|B2CGQ1|B2CGQ1_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPILGAETF┃YVDGAANRETKLGK  454
tr|B2CJX4|B2CJX4_9HIV1           KETWEAWWTEYWQATWIPEWEFXNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|Q90S75|Q90S75_9HIV1           KETWDTWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIAGAETF┃YVDGAANRETKLGK  441
tr|M4MZS8|M4MZS8_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  480
AAB24838.1                        KETWEAWWIEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|Q72547|Q72547_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
AAB24839.1                        KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETKLGK  454
tr|H9BTT1|H9BTT1_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETRLGK  454
tr|H9BTT2|H9BTT2_9HIV1           KETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETF┃YVDGAANRETRLGK  454
                                  :*.*  ** :***.****:**: .****::***.* .*** * :.: ******** :::**
```

```
tr|D3GJW0|D3GJW0_SIV              AGYITARNKSKVVALEETTNQKAELEAIKLALQDSGPRVNIVTDSQYALGILTASPDQSD  525
AAB82087.1                        AGYVTEQGKQKIIKLNETTNQKAELMAVLLALQDSKEKVNIVTDSQYVLGIISSQPTQSE  514
AAC71057.1                        AGYVTDRGRQKVVTLTDTTNQKTELQAIHLALQDAGLEVNIVTDSQYALGIIQAQPDKSE  514
AAC71058.1                        AGYVTDRGRQKVVTLTDTTNQKTELEAIHLALQDAGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|Q90S17|Q90S17_9HIV1           AGYVTDKGKQKVVTLTDTTNQQTELQAIYLALQDSGLEVNIVSDSQYALGIIQAQPDKSE  514
tr|Q701R7|Q701R7_9HIV1           AGYVTDKGRQKVVSLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDRSE  498
tr|A0A346ALG4|A0A346ALG4_9HIV1   AGYVTDRGRQKVVPLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|A0A346ALF6|A0A346ALF6_9HIV1   AGYVTDRGRQKVVSITDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|Q90S83|Q90S83_9HIV1           AGYVTNKGRQKVVSLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  501
tr|B2CGQ1|B2CGQ1_9HIV1           AGYITNKGRQKVVSLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|B2CJX4|B2CJX4_9HIV1           AGYVTNRGRQKVVSLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAXPDKSE  514
tr|Q90S75|Q90S75_9HIV1           AGYVTNKGRQKVVSLTDTTNQRTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  501
tr|M4MZS8|M4MZS8_9HIV1           AGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSE  540
AAB24838.1                        AGYVTNKGRQKVVSLTDTTNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|Q72547|Q72547_9HIV1           AGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSE  514
AAB24839.1                        AGYVTNKGRQKVVPLTNTTNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSE  514
tr|H9BTT1|H9BTT1_9HIV1           AGYVTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSE  514
tr|H9BTT2|H9BTT2_9HIV1           AGYVTNKGRQKVVPLTNTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSE  514
                                  ***:* :.:.*:: : :****::** *: *****:  .****:****.***: : * :*:
```

```
tr|D3GJW0|D3GJW0_SIV              NPIVREIIELMIGKEGVYLGWVPAHKGIGGNEQVDKLVSQGIRQVLFLEGIDKAQEEHDK  585
AAB82087.1                        SPIVQQIIEELTKKEQVYLTWVPAHKGIGGNEKIDKLVSKDIRRVL--------------  560
AAC71057.1                        SEIVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  560
AAC71058.1                        SEIVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  560
tr|Q90S17|Q90S17_9HIV1           SELVSQIIEQLIKKEKVYLXWVPAHKGIGGNE-----------------------------  546
tr|Q701R7|Q701R7_9HIV1           SELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSNGIRRVLFLDGIDKAQEEHEK  558
tr|A0A346ALG4|A0A346ALG4_9HIV1   SELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQIDKLVSDGIRKVL--------------  560
tr|A0A346ALF6|A0A346ALF6_9HIV1   SEQVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSNGIRKVL--------------  560
tr|Q90S83|Q90S83_9HIV1           SELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSTGIRKVL--------------  547
tr|B2CGQ1|B2CGQ1_9HIV1           SELVSQIIEELIKKEKVYLAWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEK  574
tr|B2CJX4|B2CJX4_9HIV1           SEXVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  560
tr|Q90S75|Q90S75_9HIV1           SELVNKIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  547
tr|M4MZS8|M4MZS8_9HIV1           SELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  586
AAB24838.1                        SELVSQIIEELIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKV---------------  559
tr|Q72547|Q72547_9HIV1           SELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVLFLDGID--------  566
AAB24839.1                        SELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKV---------------  559
tr|H9BTT1|H9BTT1_9HIV1           SELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  560
tr|H9BTT2|H9BTT2_9HIV1           SELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL--------------  560
                                  .  * :*** :  ** *** ************
```

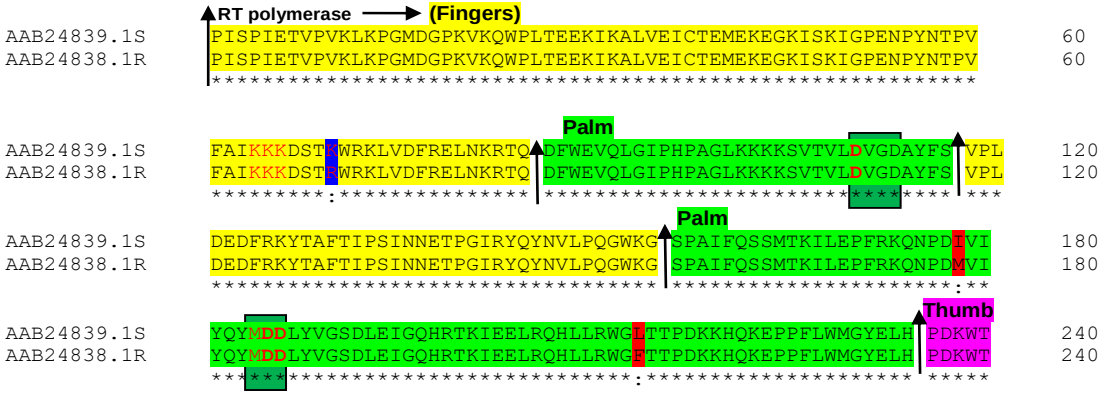
```
//End of HIV-1 & SIV RT sequences
tr|D3GJW0|D3GJW0_SIV              YHNNWRALAQDFSIPNIVAKEIVAQCPKCQTKGEPVHGQVDASPGTWQMDCTHLEGKVII      645
AAB82087.1                        ------------------------------------------------------------      560
AAC71057.1                        ------------------------------------------------------------      560
AAC71058.1                        ------------------------------------------------------------      560
tr|Q90S17|Q90S17_9HIV1            ------------------------------------------------------------      546
tr|Q701R7|Q701R7_9HIV1            YHSNWRAMA---------------------------------------------------      567
tr|A0A346ALG4|A0A346ALG4_9HIV1    ------------------------------------------------------------      560
tr|A0A346ALF6|A0A346ALF6_9HIV1    ------------------------------------------------------------      560
tr|Q90S83|Q90S83_9HIV1            ------------------------------------------------------------      547
tr|B2CGQ1|B2CGQ1_9HIV1            YHNNWRAMAS--------------------------------------------------      584
tr|B2CJX4|B2CJX4_9HIV1            ------------------------------------------------------------      560
tr|Q90S75|Q90S75_9HIV1            ------------------------------------------------------------      547
tr|M4MZS8|M4MZS8_9HIV1            ------------------------------------------------------------      586
AAB24838.1                        ------------------------------------------------------------      559
tr|Q72547|Q72547_9HIV1            ------------------------------------------------------------      566
AAB24839.1                        ------------------------------------------------------------      559
tr|H9BTT1|H9BTT1_9HIV1            ------------------------------------------------------------      560
tr|H9BTT2|H9BTT2_9HIV1            ------------------------------------------------------------      560
```

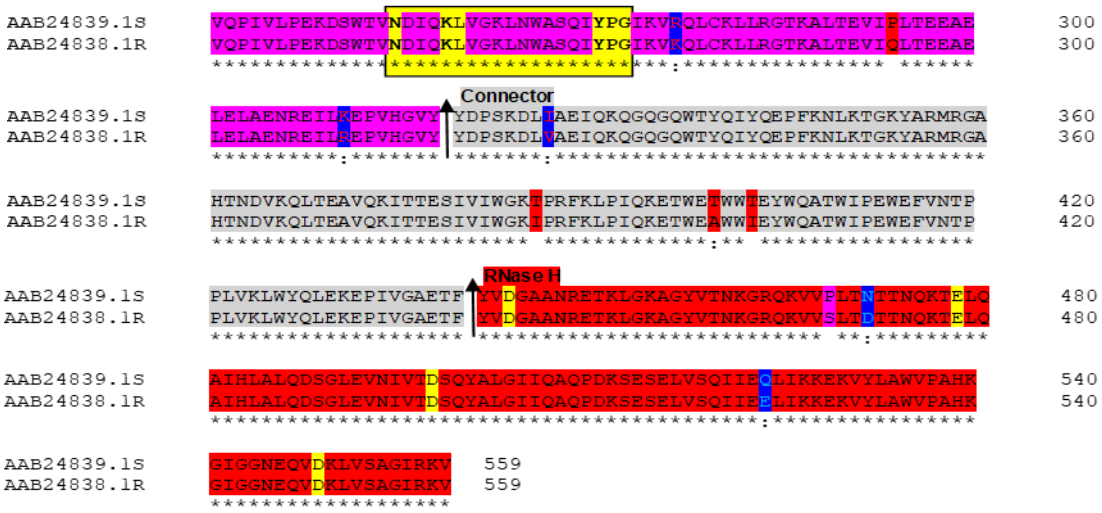
Figure 1 MSA of RTs from different HIV-1 strains and SIV

D3GJW0_SIV Pol protein, SIV
AAB82087.1 Reverse transcriptase, HIV-1
AAC71057.1 Reverse transcriptase, HIV-1
AAC71058.1 Reverse transcriptase, HIV-1
Q90S17_9HIV1 Reverse transcriptase, HIV-1
Q701R7_9HIV1 Reverse transcriptase, HIV-1
A0A346ALG4_9HIV1 Pol protein, HIV-1
A0A346ALF6_9HIV1 Pol protein, HIV-1
Q90S83_9HIV1 Pol, polyprotein, HIV-1
B2CGQ1_9HIV1 Pol protein, HIV-1
B2CJX4_9HIV1 Reverse transcriptase, HIV-1
Q90S75_9HIV1 Pol, polyprotein, HIV-1
M4MZS8_9HIV1 Pol protein, HIV-1
AAB24838.1 Reverse transcriptase, HIV-1, (AZT-resistant)
Q72547_9HIV1 Reverse transcriptase, HIV-1
AAB24839.1 Reverse transcriptase, HIV-1, (AZT-sensitive)
H9BTT1_9HIV1 Pol protein, HIV-1
H9BTT2_9HIV1 Pol protein, HIV-1

To find out the important changes made in the RTs of AZT-sensitive and -resistant strains of HIV-1, a 'mix and match' analysis was performed. Fig. 2 shows the 'Mix and Match' MSA of RTs from an AZT-sensitive and an AZT-resistant strain of the HIV-1. Interestingly, the fingers subdomain is almost completely conserved. In the palm and thumb subdomains, only 3 non-isofunctional amino acid substitutions are observed (two in the palm and 1 in the thumb), but they are located away from the proposed catalytic amino acids (highlighted in red). Thus, the change of only just three amino acids, viz. ^{178}I(s)→M(r); ^{214}L(s)→F(r); ^{294}P(s)→Q(r) in polymerase palm and thumb subdomains make one sensitive and the other one resistant to AZT. Similar amino acid replacements are highlighted in blue and non-similar replacements are shown in red (Fig. 2). These results suggest that without making any change on the catalytic metal-binding motifs and the polymerase catalytic core, but making only 3 changes in the thumb and palm subdomains, the AZT-resistant HIV-I could successfully exclude the AZT from its active site and keeps multiplying in human cells. (It should be noted that in the AZT, only the 3'-OH group of thymidine is replaced by an azido group, $-N_3$ ($N^-=N^+=N$) at the 3'of ribose). Therefore, in the absence of the -OH group, the incorporation of AZT at the polymerization site stops the addition of the next nucleotide and as a consequence, stops the multiplication of the HIVs in human cells. However, a few amino acid substitutions in the AZT-resistant HIV-1 do not allow AZT to bind onto the enzyme and continue the replication of the virus.

CLUSTAL O (1.2.4) 'Mix and Match' MSA of RTs from AZT-sensitive (S) and AZT-resistant (R) strains of the HIV-1.



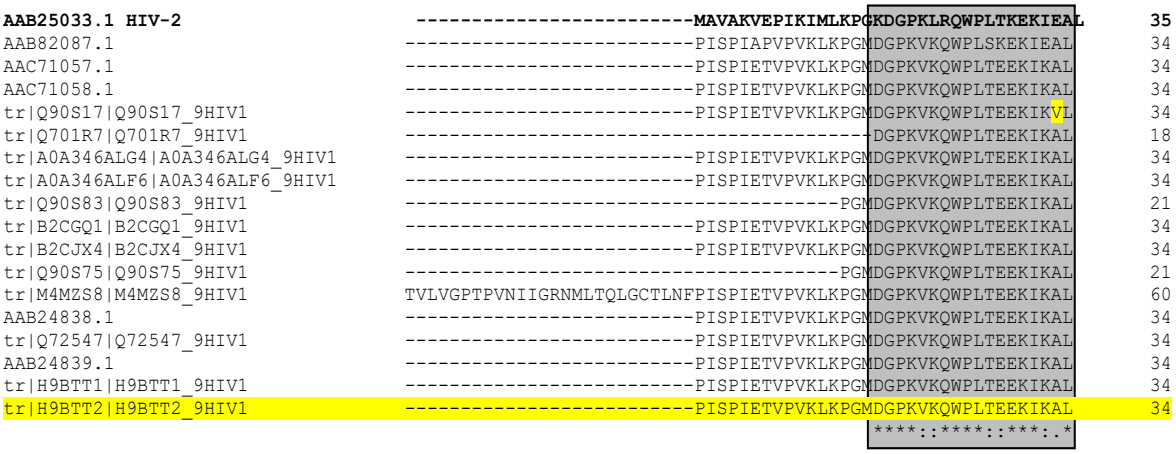


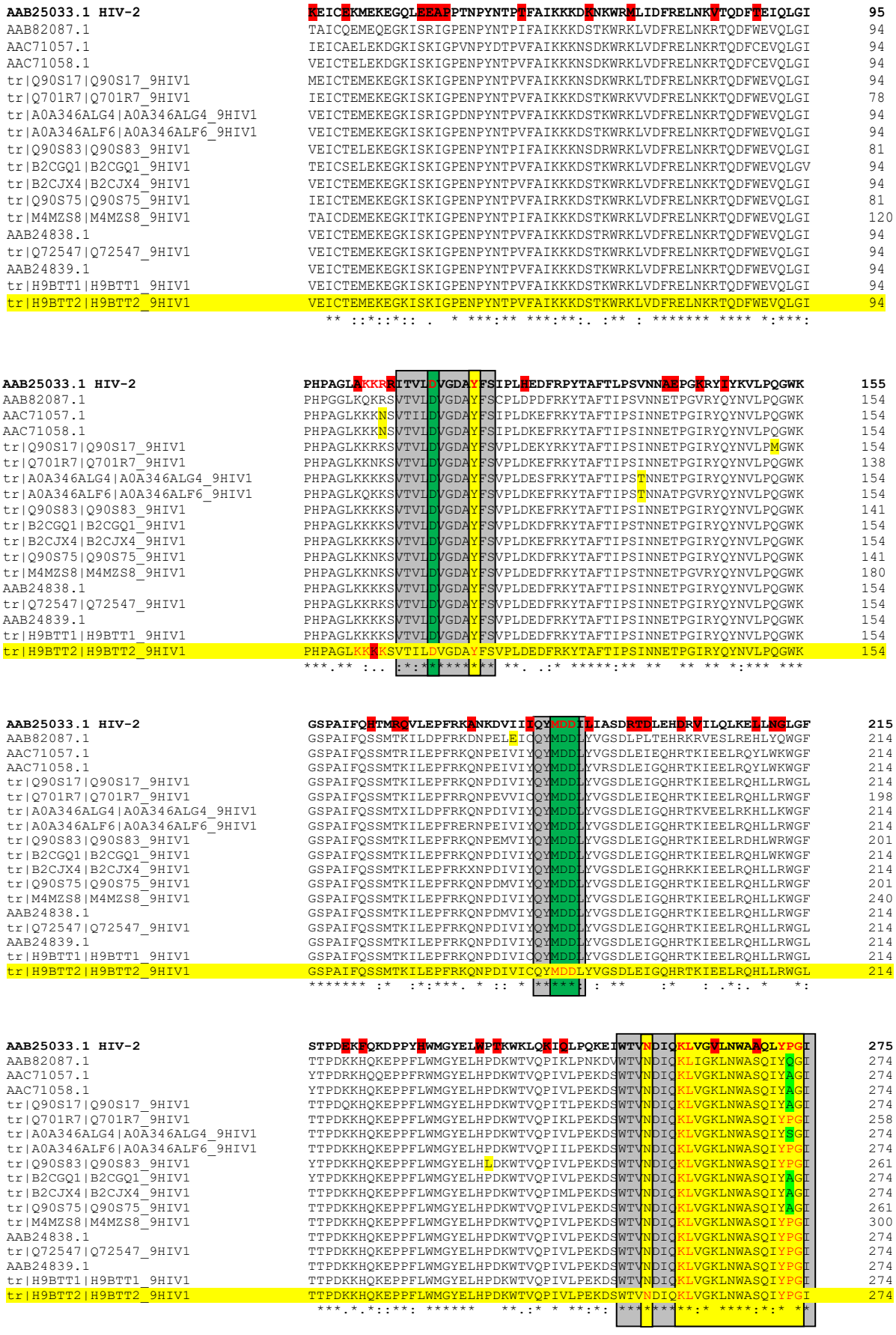
Similar amino acid changes are highlighted in blues (6); non-similar amino acid changes are in red/magenta (3) in the polymerase domain and only one in the RNase H; Catalytic region amino acids are in bold and highlighted in yellow.

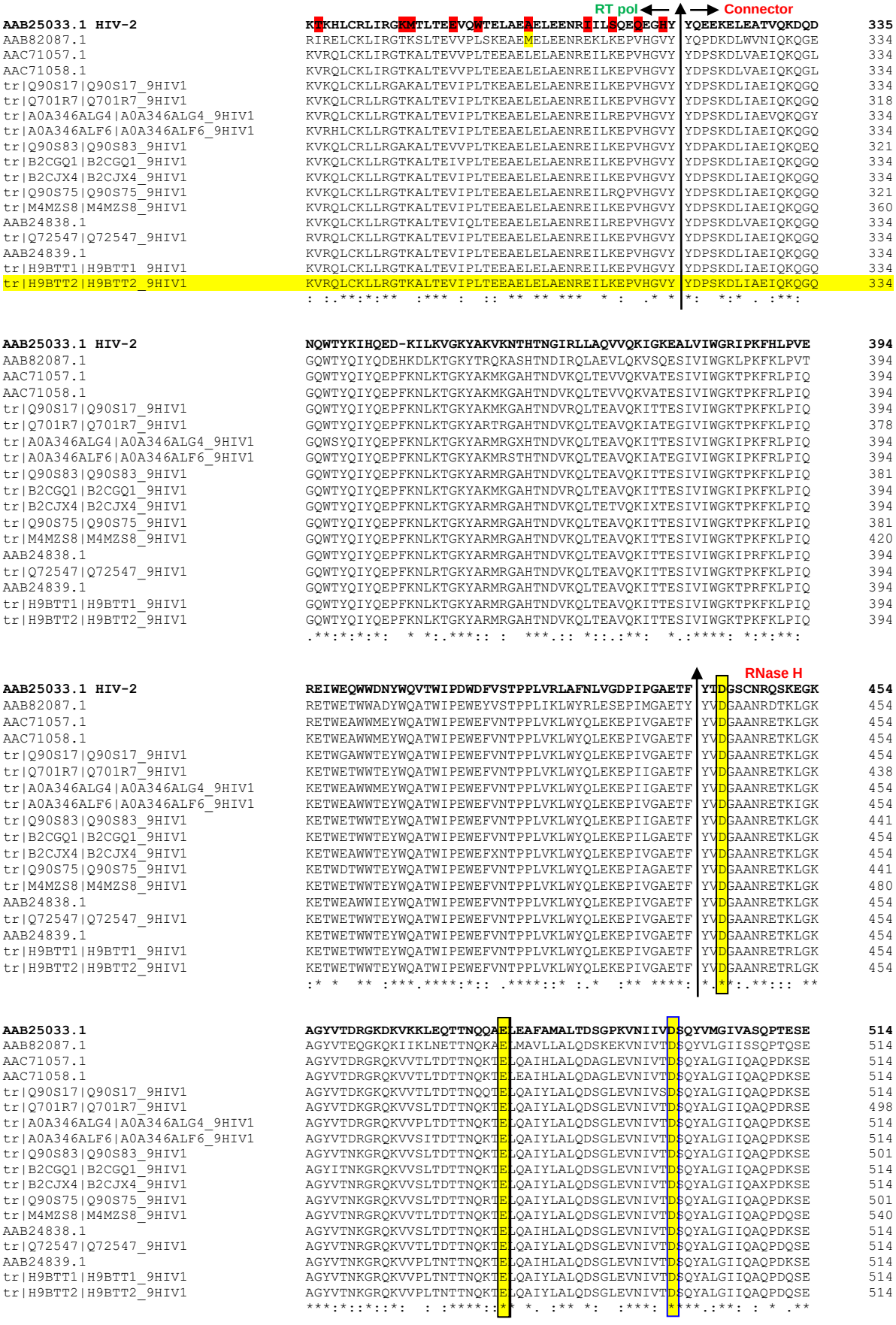
Figure 2 'Mix and Match' MSA of RTs from AZT sensitive and resistant strains of the HIV-1.

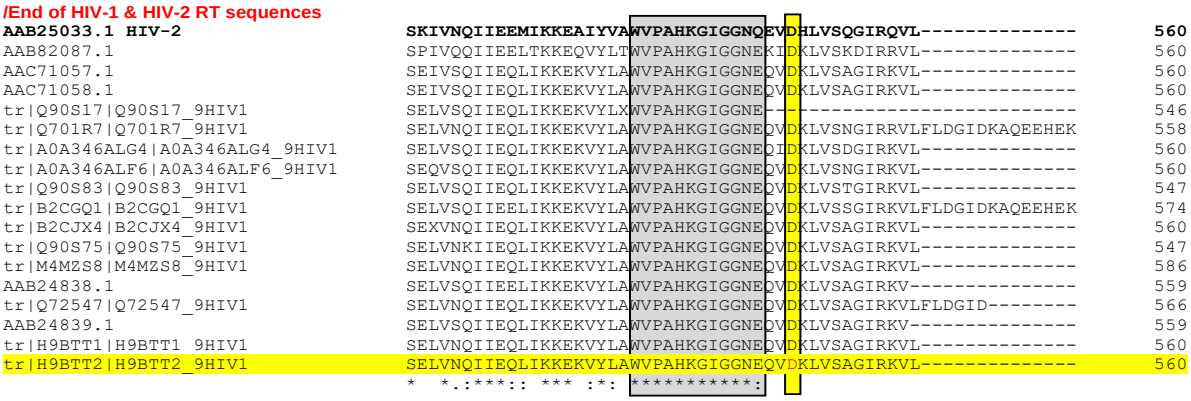
HIV-2 RT is intrinsically resistant to NNRTIs, but the HIV- 1 RT is susceptible to them. To find out the important changes made in the RTs of HIV-1 and HIV-2, a 'mix and match' MSA analysis was performed. Fig. 3 shows the 'Mix and Match' MSA of RTs from HIV-1 and HIV-2. In the HIV-2 standard strain (in bold) a large number of amino acids are modified in all three polymerase subdomains of the RT (highlighted in red) including some tri- and tetra-peptides, but interestingly keeping the active sites more or less intact. The BLASTp analysis has shown only ~62% identity between the two HIV RTs. However, the pIs vary significantly between the RTs of HIV-1 and HIV-2, and are 8.63 and 7.2, respectively (i.e., the HIV-1 RT is highly basic and the HIV-2 RT is near neutral). Furthermore, it is interesting to note that the Ys on both sides of the metal-binding –MDD- motif are replaced by I/L in HIV-2 RT. The marked changes in their pIs and the amino acid substitutions at Y^{181} and Y^{188} (both are replaced with branched-chain amino acids, I and L, respectively, in HIV-2) in the catalytic metal-binding region -MDD- are implicated in the selective inhibition of HIV-1 RT by NNRTIs, but not the HIV-2 RT [23]. However, the catalytic core region and the NTP selection amino acid (N) are highly conserved in both HIV-1 and HIV-2, and thus, confirming their susceptibility to NRTIs. The RNase H active site amino acids are highlighted in yellow.

CLUSTAL O (1.2.4) 'Mix and Match' MSA of reverse transcriptases from HIV-1 and HIV-2.




AAB25033.1 HIV-2
AAB82087.1
AAC71057.1
AAC71058.1
tr|Q90S17|Q90S17_9HIV1
tr|Q701R7|Q701R7_9HIV1
tr|A0A346ALG4|A0A346ALG4_9HIV1
tr|A0A346ALF6|A0A346ALF6_9HIV1
tr|Q90S83|Q90S83_9HIV1
tr|B2CGQ1|B2CGQ1_9HIV1
tr|B2CJX4|B2CJX4_9HIV1
tr|Q90S75|Q90S75_9HIV1
tr|M4MZS8|M4MZS8_9HIV1
AAB24838.1
tr|Q72547|Q72547_9HIV1
AAB24839.1
tr|H9BTT1|H9BTT1_9HIV1
tr|H9BTT2|H9BTT2_9HIV1


RT pol
Connector
RNase H



AAB25033.1 Reverse transcriptase, HIV-2. The rest of the legends are as in Fig. 2. The RNase H active site amino acids are highlighted in yellow.

Figure 3 'Mix and Match' MSA of Reverse transcriptases from HIV-1 and HIV-2.

4. Analysis of the Active sites of the RTs of HIVs

4.1. Catalytic site amino acids at the polymerase domain in HIV RTs

The active site amino acids of HIV-1 RT are arrived at from the data derived from three different sources. Firstly, from the sequence similarity of the active site of HIV-1 RT (-^{258}Q**K**$^{-4}$LVGK^{263}L^{1}NWASQI**Y**^{8}PGI-) and the *E. coli* DNA pol I (-^{753}Q**R**$^{-4}$RSA**K**^{758}A^{1}INFGLI**Y**^{8}G^{767}M-) (Fig. 1). The *E. coli* DNA polymerase I active site amino acids were elucidated both by chemical modification of the active site amino acids and SDM experiments. Pyridoxal-5'-phosphate (PLP), a competitive inhibitor of dNTP substrates of DNA polymerase, was shown to form an adduct with **K**^{758} resulting in the loss of substrate binding and polymerase activity [26]. This finding was further confirmed by SDM experiments by Pandey *et al*. [27]. They modified the active site amino acid -K^{758} to K^{758}→A and K^{758}→R. The catalytic activity of the purified, mutant enzymes of K^{758}→A and K^{758}→R, showed a drastic reduction in the polymerase activity but little difference in the 3'→5' PR exonuclease activity. Their experiments suggested a dual role for the K^{758} in catalysis: facilitating i) dNTP binding at the polymerization site and (ii) translocation along the template DNA [27]. Further photoaffinity labelling of the *E. coli* enzyme, using 8-azido-dATP, resulted in the covalent modification of **Y**^{766} and loss of enzyme activity [28]. The Y^{766} forms the template-binding -**Y**^{766}**G**- pair. It should be noted that the Y^{766} is 8 amino acids downstream from the catalytic K^{758}. Thus, the active site amino acids at the polymerization sites are similar in HIV-1 RT and *E. coli* DNA pol I.

Secondly, from the X-ray crystallographic data of the HIV-1 RT. Kohlstaedt *et al*. [6] found that the palm region of p66 has considerable structural similarity to the polymerase active site of the Klenow fragment of *E. coli* DNA pol I. The RT polymerase domains showed a large cleft analogous to that of the Klenow fragment of *E. coli* DNA pol I and suggested that these polymerases would have diverged from a common ancestor.

Thirdly, from biochemical modification of the active site amino acid of HIV-1 RT with the active site affinity reagent PLP. Basu *et al*. [29] used PLP for analysis of the substrate-binding site of HIV-1 RT. The HIV-1 RT reacted with PLP and formed an enzyme-PLP adduct which led to the irreversible inactivation of RT polymerase activity, whereas the RNase H activity was minimally affected. Furthermore, the reactivity of this site was also blocked by the inclusion of substrate dNTP and with an appropriate template-primer. The amino acid composition and sequence analysis of the resulting PLP-cross-linked peptide showed that K^{263} as the site of PLP reactivity. Therefore, they concluded that K^{263} could serve as an important part of the dNTP-binding domain in HIV-1 RT.

Martin *et al*. [30] analyzed the HIV-1 RT active site by SDM experiments. They replaced the particular K^{263} with a non-isofunctional amino acid S (K^{263}→S) and analyzed the activity of the mutant enzyme. The mutant enzyme, where the K^{263} is replaced by S^{263}, is bound to the natural dNTP substrates and primed polynucleic acid substrates with equal affinity when compared to the wild-type enzyme. Furthermore, they also found that the S^{263} substitution had no effect on the RNase H activity of the enzyme. Their results indicated that the K^{263} is not essential in the binding of dNTP and priming polymerization by the HIV-1 RT. Based on these results, a downstream K with an N at -4 is proposed as the polymerase proton abstractor in the catalytic site. Therefore, the proposed polymerase catalytic core is, -VN^{-4}DIQ**K**^{259}L^{1}VGK^{263}LNWASQI**Y**^{12}**PG**^{273}- -VN^{-4}DIQ**K**^{260}L^{1}VGV^{264}LNWAAQL**Y**^{12}**PG**^{274}- for HIV-1 and HIV-2 RTs, respectively. Furthermore, the catalytic site amino acid for proton abstraction is usually a basic amino acid, K/R and it is not conserved in HIV-2 (highlighted in red) and is replaced with a neutral amino acid V (Fig. 3). The crystallographic data

of HIV-2 RT presented evidence that the conformation of I^{181} compared with the Y^{181} of HIV-1 could be a significant contributory factor to the inherent drug resistance of HIV-2 to NNRTIs [31].

4.2. Metal-binding site amino acids of the polymerase domain in HIV-RTs

The metal-binding sites on HIV-1 RTs were analyzed by several investigators [32-34]. Genetic substitution experiments have shown that single amino acid alterations, viz. D^{110}→Q, D^{185}→N/E, or D^{186}→N/E in the p66 subunit produced an inactive HIV-1 RT (showed only <0.01% of the activity) and M^{184}→L substitution showed only 8% of the activity [32, 33]. This was further confirmed by Le Grice *et al*. [34] by SDM analysis of the two amino acids -$D^{185}D^{186}$- of the -MDD- motif. The active site mutants, in which these two amino acids were altered to N, were virtually devoid of any RT activity, indicating their direct involvement in metal-binding and in catalysis. Interestingly, these mutations did not affect the RNase H activity of the enzyme. Furthermore, the reconstituted heterodimer in which the p51 subunit was mutated in either D^{185} or D^{186} retained high levels of RT polymerase activity, suggesting that these residues are dispensable in the heterodimer associated with p51. However, the reciprocal reconstitutions (i.e., those in which the p66 is mutated) showed only ~2.5% of the RT activity as that of the wild-type enzyme, indicating that a wild-type p51 cannot compensate for mutations introduced into the active site of p66. Therefore, all the three invariant Ds are involved in the catalytic metal-binding and completely conserved in both the HIV RTs and also in SIV RT (Figs. 1-3). Table 1 shows the catalytic regions of the polymerase from different RNA/DNA polymerases and RTs.

Table 1 Catalytic core regions of various RNA/DNA polymerases and HIV RTs

Polymerase type	Catalytic core
SSU RNA/DNA pols	
T7 Viral SSU RNA pol	-^{620}WLAY^{-8}GVTR^{-4}SVTKR^{1}SVMTLAY^{8}GS-
SP6 Viral SSU RNA Pol	-^{612}WDSI^{-8}GITR^{-4}SLTKK^{1}PVMTLPY^{8}GS-
Mitochondrial SSU RNA pol (*Sc*)	-^{1009}TR^{-4}KVVKQ^{1}TVMTNVY^{8}GV-
Mitochondrial SSU RNA pol (*Hs*)	-^{986}TR^{-4}KVVKQ^{1}TVMTVVY^{8}GV-
E. coli DNA pol I (SSU)	-^{753}QR^{-4}RSAK^{758}A^{1}INFGLIY^{8}GM-
Chloroplast SSU DNA pol IA (*ARATH*)	-^{873}ER^{-4}RKAK^{878}M^{1}LNFSIAY^{8}GK-
Chloroplast SSU DNA pol IB (*ARATH*)	-^{857}ER^{-4}RKAK^{862}M^{1}LNFSIAY^{8}GK-
Chloroplast SSU RNA pol (NEP) (*ARATH*)	-^{765}DR^{-4}KLVK^{770}Q^{1}TVMTSVY^{8}GV-
Mitochondrial SSU RNA pol (NEP) (*ARATH*)	-^{748}DR^{-4}KLVK^{753}Q^{1}TVMTSVY^{8}GV-
Human DNA pol α (*Hs*)	-^{946}Q^{-4}KALK^{950}L^{1}TANSMY^{7}GCL-
Human DNA pol δ (*Hs*)	-^{690}Q^{-4}LALK^{694}V^{1}SANSV Y^{7}GFT-
Human DNA pol ε (*Hs*)	-^{805}Q^{-4}LAHK^{809}C^{1}ILNSFY^{7}GYV-
HIV-1 Reverse transcriptase	-^{255}N^{-4}DIQK^{259}L^{1}VGKLNWASQIY^{12}PG^{273}-*
HIV-2 Reverse transcriptase	-^{256}N^{-4}DIQK^{260}L^{1}VGVLNWASQIY^{12}PG^{274}-*
SIV Reverse transcriptase	-^{266}N^{-4}DIQK^{270}L^{1}VGKLNWASQIY^{12}SG^{284}-*
Human Influenza Virus C	-^{468}N^{-4}AVCK^{472}L^{1}IGINMSLEKSY^{12}G- [35]
Respiratory Syncytial Viruses	-^{293}N^{-4}TLNK^{297}S^{1}LGLRCGFNNVILTQLFLY^{18}G- [36]

Adapted from Palanivelu [37, 38];*Present work; *Sc, Saccharomyces cerevisiae*; *Hs, Homo sapiens*; *ARATH, Arabidopsis thaliana*; The active site amino acids, highlighted in dark blue, are confirmed by SDM and other techniques.

Generally, the catalytic core region of DNA polymerases essentially contains three components, viz. a template-binding pair –YG-, a basic catalytic amino acid -K/R as proton abstractor to initiate catalysis and a nucleotide discriminating amino acid –K/R-, placed at -4 from the catalytic K/R. The nucleotide discriminating amino acid –K/R-, placed at -4 from the catalytic K/R is replaced by an N in HIV-RTs, human influenza and respiratory syncytial viruses also (Table 1). These three highly conserved components of the polymerase catalytic core are also found in the RTs of HIV-1, HIV-2 and SIV (Figs. 1-3). Thus, the catalytic amino acids in the active site regions of the HIV-RTs are in close agreement with those already reported from other DNA/RNA polymerases (Table 1).

4.3. Proposed Mechanism of Action of HIV RTs

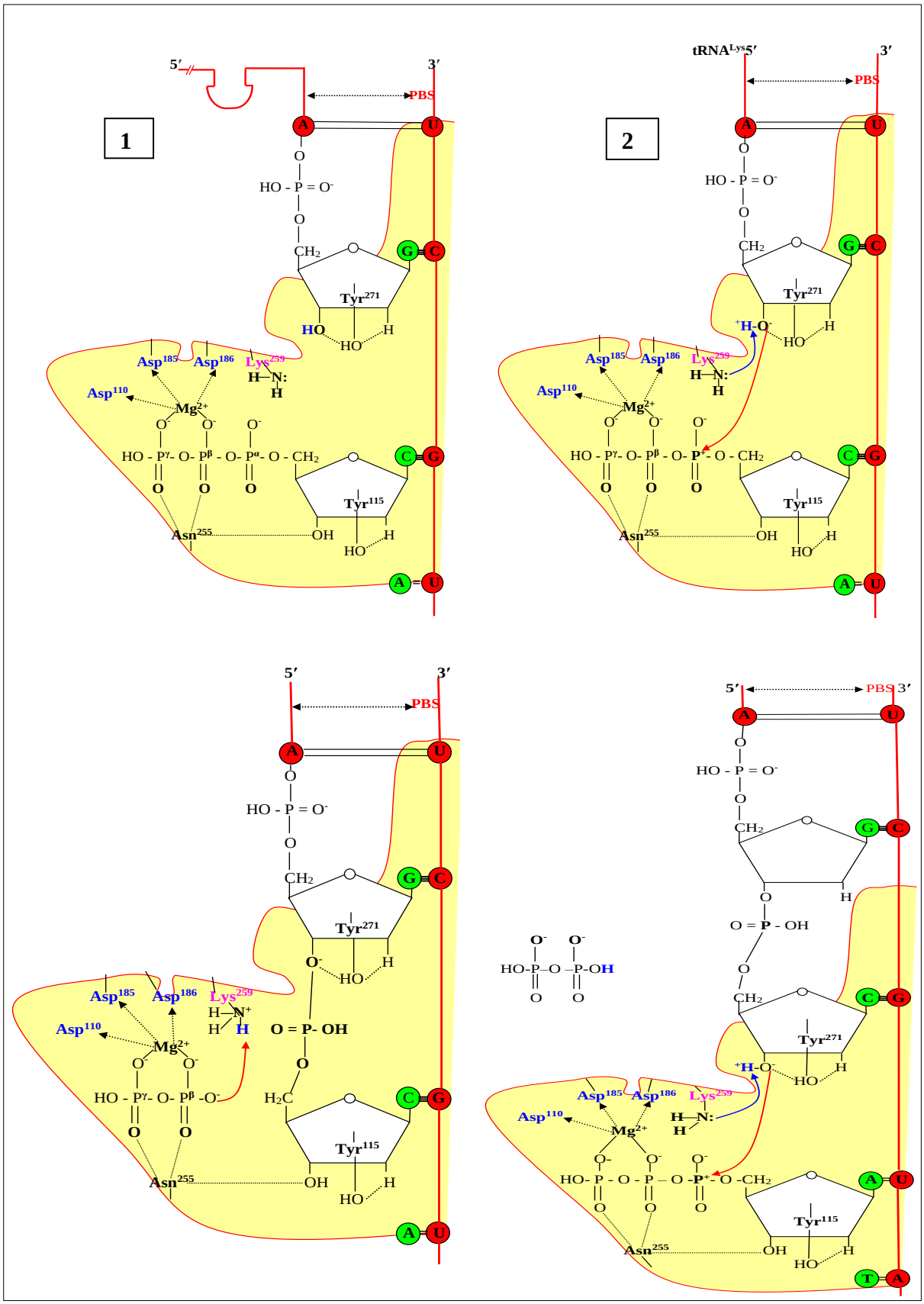


Figure 4 A schematic diagramme showing the proposed steps in the reverse transcription by HIV RTs

The crystallographic data have shown that the 3D structure of the RTs is very similar to DNA polymerases with fingers, palm and thumb subdomains. The MSA analysis has also shown that the catalytic metal-binding and catalytic proton abstractor and nucleotide selection amino acids are the same as in other DNA/RNA polymerases. Hence, they can also be considered as DNA polymerases, which make a double-stranded DNA from RNA/DNA templates through a complex series of steps [39]. Tyr^{115}, which is located in the vicinity of the polymerase catalytic site of HIV-1 RT is implicated in the dNTP-binding and misinsertion fidelity of DNA synthesis and is completely conserved in HIV-1, HIV-2 and SIV RTs. This is located very close to the catalytic Mg^{2+}-binding D^{110} and is highlighted in yellow [40] (Fig. 4). A distinct difference was observed in the template-binding amino acids between regular DNA/RNA polymerases and RTs, viz. the DNA/RNA polymerases use mostly a –YG- pair as the template-binding pair, but the RTs use a triad as -YXG-. Therefore, the RTs should also follow a very similar mechanism as proposed for other DNA polymerases [19].

HIVs and other lentiviruses use the host $tRNA^{lys}$ as the primer for the synthesis of the cDNA (the minus strand). The 18-nucleotide (nt) 3'-end of the $tRNA^{lys}$ is strictly complementary to the 18 nt of the 5' primer-binding site of the viral genome, designated as PBS. Interestingly, the same PBS is also found at the 3'-end of the viral genomic RNA. The DNA synthesis from the viral RNA is a multistep process. After the first priming, the minus DNA strand is copied from the viral RNA only up to a position of ~ 150 nucleotides at the 5'-end which carries the viral 5'-LTR. The plus strand RNA genome is degraded by the associated RNase H, leaving the nascent minus-strand cDNA as single-stranded. Now the minus-strand cDNA along with its covalently linked $tRNA^{lys}$ jumps and binds onto the 3'-LTR and continues and completes the transcription of the minus-strand synthesis all along the viral RNA up to the PBS region at 5'. As DNA synthesis proceeds, the RNase H degrades the RNA strand. However, the polypurine tract (PPT), 5'-AAAAGAAAAGGGGGG-3' located just 5' to the U3 sequence of the viral RNA genome is resistant to RNase H cleavage. RNase H makes a specific cut following the 6th "G" residue to define the 3'-end of the PPT primer, and the new PPT serves as the primer for second (plus) strand DNA synthesis which covers U3, R, U5 and PBS. Now, the RNase H removes the entire $tRNA^{lys}$ including the covalently linked PBS. In the following step, the partially synthesized second strand jumps and binds to the PBS at the 3'-end and completes the second strand synthesis.

As discussed elsewhere, the catalytic Mg^{2+}-binding site which is composed of three invariant Ds, viz. D^{110}, D^{185} and D^{186} position the incoming dNTP to the already Watson–Crick base-paired base, complementary to the template, at the polymerization site (Fig. 4). The catalysis is initiated by proton abstraction by the catalytic amino acid from the 3'-OH of the primer, which is followed by an electrophilic-nucleophilic attack between the α-phosphate of the Watson-Crick base-paired dNTP and the 3′-Oxyanion of the growing primer resulting in the phosphodiester bond formation. The pyrophosphate generated in the last step of the cycle is used for the translocation of the enzyme to the next nucleotide by the hydrolysis of the pyrophosphate by a pyrophosphatase, which generates the required energy for the translocation. For a description of the steps in the mechanism, see Palanivelu [35]. Unlike RNA transcriptases, the reverse transcriptases invariably use a primer and have an associated RNase H.

5. Conclusions

To control the AIDS, a protective vaccine or effective retroviral drugs that are able to block the binding of the virus to the host cell and/or the lifecycle of the virus are crucial. In this respect, the HIV RTs continue to be one of the prime targets for the current AIDS therapy and new drug development efforts. The present work sheds light on the active site amino acids present on the polymerase domain of HIV-RTs. The active site amino acids of the HIV RTs are found to be very similar to other DNA/RNA polymerases, suggesting their common evolutionary origin. The conspicuous absence of a regular DEDD-superfamily of the proofreading exonuclease domain on the HIV-RTs suggests mutations may not be corrected as in other RNA/DNA polymerases, leading to the rapid emergence of viral variants, which help them escape the antiviral drugs. Despite extensive studies on HIV-RTs and HIV drug-resistant strains, further understanding and characterization of HIV RTs and their mechanism(s) of drug resistance is necessary for the design of more effective drugs in the future.

Compliance with ethical standards

Acknowledgments

The author wishes to thank Dr. N. Srinivasan, former Professor, Department of Endocrinology, Post Graduate Institute of Basic Medical Sciences, University of Madras, Chennai, for the corrections and suggestions on the manuscript.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Global HIV and AIDS statistics - 2022 fact sheet. www.unaids.org. UNAIDS.

[2] Rossi E, Meuser ME, Cunanan CJ, Cocklin S. Structure, Function, and Interactions of the HIV-1 Capsid Protein. *Life.* 2021; *11*(2): 100; https://doi.org/10.3390/life11020100.

[3] Beard WA, Stahl SJ, Kim H, Bebenek K, Kumar A, Strub M, Becerra P, *et al*. Structure/ Function Studies of Human Immunodeficiency Virus Type 1 Reverse Transcriptase: Alanine scanning mutagenesis of an α-Helix in the thumb subdomain. J Biol Chem. 1994; 269: 28091-28097.

[4] Jacobo-Molina A, Ding J, Nanni RG, Clark Jr AD, Lu X, Tantillo C, Williams RL, *et al*. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc Natl Acad Sci (USA). 1993; 90:6320-6324.

[5] Reardon JE. Human immunodeficiency virus reverse transcriptase: steady-state and pre-steady-state kinetics of nucleotide incorporation. Biochemistry. 1992; 31:4473–4479.

[6] Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. Crystal structure at 3.5 Å resolution of HIV-1reverse transcriptase complexed with an inhibitor. Science. 1992; 256:1783–1790.

[7] Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Stuart D, *et al.* High resolution structures of HIV-1 RT from four RT-inhibitor complexes. Nat Struct Biol. 1995; 2: 293-302.

[8] Ding J, Das K, Tantillo C, Zhang W, Clark AD Jr, Jessen S, Lu X, *et al*. Structure of HIV-1 reverse transcriptase in a complex with the non-nucleoside inhibitor alpha-APA R 95845 at 2.8 Å resolution. Structure. 1995; 3:365-379.

[9] Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science. 1998; 282:1669-1675.

[10] Tuske S, Sarafianos SG, Clark AD Jr, Ding J, Naeger LK, White KL, Miller MD, *et al*. Structures of HIV-1RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. Nat Struct Mol Biol. 2004; 5:469–474.

[11] Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, *et al*, Search and sequence analysis tools services from EMBL-EBI in 2022. Nuc Acids Res. 2022; 50: W276-W279. https://doi.org/10.1093/nar/gkac240.

[12] Palanivelu P. DNA polymerases – An insight into their active sites and mechanism of action, In: Recent Advances in Biological Research, Vol 1, Chapter 2, pp 1-39, SCIENCEDOMAIN International Book Publishers, UK. ISBN: 9788193422441, DOI: 10.9734/bpi/rabr/v1; 2019.

[13] Palanivelu P. Single Subunit RNA Polymerases: An Insight into their Active Sites and Catalytic Mechanism, In: Advances and Trends in Biotechnology and Genetics. Vol 1, Chapter 1, pp 1-38, SCIENCEDOMAIN International Book Publishers, UK. ISBN: 978-93-89246-59-9, DOI:10.9734/bpi/atbg/v1; 2019.

[14] Huang H, Harrison, SC Verdi GL. Trapping of a catalytic HIV reverse transcriptase-template:primer complex through a disulfide bond. Chem Biol. 2000; 7:355–364.

[15] Hostomsky Z, Hostomska Z, Fu TB, Taylor J. Reverse Transcriptase of Human Immunodeficiency Virus Type 1: Functionality of Subunits of the Heterodimer in DNA Synthesis. J Virol. 1992; 66:3179-3182.

[16] Le Grice SFJ, Naas T, Wohlgensinger B, Schatz O. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1reverse transcriptase. The EMBO J. 1991;10: 3905-3911.

[17] Larder BA, Purifoy DJM, Powell KL, Darby G. Site-specific mutagenesis of AIDS virus reverse transcriptase. Nature. 1987; 327: 716–717,

[18] St. Clair MH, Martin JL, Tudor-Williams G, Bach MC, Vavro CL, King DM, Kellam P, Kemp S D, Larder BA. Resistance to ddI and Sensitivity to AZT Induced by a Mutation in HIV-1 Reverse Transcriptase. Science. 1991; 253: 1557-1559.

[19] Kondo M. Predominance of codon 215 mutation in reverse transcriptase-coding region of 3'-azido-3'-deoxythymidine (AZT)-resistant HIV-1 isolates after long-term AZT therapy. Kansenshogaku Zasshi. 1995; 69:1278-1285.

[20] Gu Z, Gao Q, Fang H, Salomon H, Parniak MA, Goldberg E, Cameron J, Wainberg MA. Identification of a mutation at codon 65 in the IKKK motif of reverse transcriptase that encodes human immunodeficiency virus resistance

to 2',3'-dideoxycytidine and 2',3'-dideoxy-3'-thiacytidine. Antimicrob Agents Chemother. 1994; 38:275-281. doi: 10.1128/AAC.38.2.275.

[21] Boyer PL, Julias JG, Ambrose Z, Siddiqui MA, Marquez VE, Hughes SH. The nucleoside analogs 4′C-methyl thymidine and 4′C-ethyl thymidine block DNA synthesis by wild-type HIV-1 RT and excision proficient NRTI resistant RT variants. J Mol Biol. 2007; 371:873–882.

[22] Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, Pauwels R, *et al*. Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. Implications for mechanisms of drug inhibition and resistance. J Mol Biol. 1994; 243:369-387.

[23] Pauwels R, Andries K, Desmyter J, Schols D, Kukla MJ, Breslin HJ, Raeymaeckers A, *et al*. Potent and selective inhibition of HIV-1 replication *in vitro* by a novel series of TIBO derivatives. Nature. 1990; 343: 470-474.

[24] Merluzzi, VJ, Hargrave, KD, Labadia, M, Grozinger, K, Skoog, M, Wu, JC, Shih, C, *et al*. Inhibition of HIV-1 Replication by a Nonnucleoside Reverse Transcriptase Inhibitor. Science.1990; 250: 1411-1413.

[25] Nunberg JH, Schleif WA, Boots EJ, O'Brien JA, Quintero JC, Hoffman, JM, Emini EA, Goldman ME. Viral resistance to human immunodeficiency virus type 1-specific pyridinone reverse transcriptase inhibitors. J Virol. 1991; 65: 4887-4892.

[26] Basu A, Modak MJ. Identification and amino acid sequence of the deoxynucleoside triphosphate binding site in Escherichia coli DNA polymerase I. Biochemistry. 1987; 26:1704-1709.

[27] Pandey VN, Kaushik N, Modak MJ. Role of Lysine 758 of *Escherichia coli* DNA Polymerase I as Assessed by Site-directed Mutagenesis. J Biol Chem. 1994; 269:13259-13265.

[28] Rush J, Konigsberg WH. Photoaffinity labeling of the Klenow fragment with 8-azido-dATP. J Biol Chem. 1990; 265:4821482.

[29] Basu A, Tirumalai RS, Modak MJ. Substrate Binding in Human Immunodeficiency Virus Reverse Transcriptase: An Analysis of Pyridoxal-5'-phosphate Sensitivity and Identification of Lysine 263 In The Substrate-Binding Domain*. J. Biol. Chem. 1989; 264:8746-8752.

[30] Martin JL, Jeanne E. Wilson, Eric S. Furfine, Samuel. E Hopkins, and Phillip A. Furman. Biochemical Analysis of Human Immunodeficiency Virus-1 Reverse Transcriptase Containing a Mutation at Position Lysine 263*. J Biol Chem. 1993; 268:2565-2570.

[31] Ren J, Bird LE, Chamberlain PP, Stewart-Jones GB, Stuart DI, Stammers DK. Structure of HIV-2 reverse transcriptase at 2.35-Å resolution and the mechanism of resistance to non-nucleoside inhibitors. Proc Natl Acad Sci, USA. 2002;99: 14410–14415.

[32] Lowe DM, Parmar V, Kemp SD, Larder BA. Mutational analysis of two conserved sequence motifs in HIV-I reverse transcriptase. FEBS. 1991; 282: 231-234.

[33] Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT) Science. 1989; 246: 1155-1158.

[34] Le Grice SFJ, Naas T, Wohigensinger B, Schatz O. Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. The EMBO J. 1991; 10:3905-3911.

[35] Palanivelu P. An Insight into the Active Sites of the Catalytic Basic Protein Subunit PB1 of the RNA Polymerase of Human Influenza Viruses, World J Adv Res Rev. 2022; 17:625–565.

[36] Palanivelu P. An Insight into the Active sites of the RNA Polymerase and Proofreading Exonuclease of the Human Respiratory Syncytial Virus. World J Adv Res Rev. 2023; 18:842–858.

[37] Palanivelu P. Identification of Polymerase and Proofreading Exonuclease Domains in the DNA Polymerases IA, IB and Nuclear-Encoded RNA Polymerase of the Plant Chloroplasts. World J Adv Res Rev. 2023;17, 706–727.

[38] Palanivelu P. An Insight into the Mechanism of Genome Duplication in Eukaryotes: Polymerase and Proofreading Functions by Eukaryotic DNA Replicases. Int J of Rec Sci Res. 2022; 13:2076-2116.

[39] Gilboa E, Mitra SW, Goff S, Baltimore D. A detailed model of reverse transcription and tests of crucial aspects. Cell. 1979; 18:93-100. doi: 10.1016/0092-8674(79)90357.

[40] Martin-Hernandez AM, Domingo E, Menendez-Arias L. Human immunodeficiency virus type 1 reverse transcriptase: role of Tyr115 in deoxynucleotide binding and misinsertion fidelity of DNA synthesis. The EMBO Journal. 1996;.15:4434-4442.