STRUCTURAL STUDIES OF HIV-1 REVERSE TRANSCRIPTASE: RESISTANCE TO AZT VIA ATP-MEDIATED EXCISION

by XIONGYING TU

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemistry

written under the direction of

Professor Edward Arnold

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

[May, 2008]

ABSTRACT OF THE DISSERTATION

STRUCTURAL STUDIES OF HIV-1 REVERSE TRANSCRIPTASE: RESISTANCE TO AZT VIA ATP-MEDIATED EXCISION

BY XIONGYING TU

Dissertation Director:

EDDY ARNOLD

AZT, 3'-azido-2',3'-deoxythymidine, is a widely used anti-AIDS drug. Incorporation of AZTTP, the triphosphate form of AZT, into a growing DNA chain by the human immunodeficiency virus (HIV) reverse transcriptase (RT) blocks chain elongation. HIV develops resistance to AZT by acquiring mutations in reverse transcriptase (RT) that enhance ATP-mediated excision of AZTMP. This excision frees the end of the primer strand, making it possible for RT to continue viral DNA synthesis. The product of the excision reaction is AZT adenosine dinucleoside tetraphosphate (AZTppppA). In order to understand the structural bases of this resistance mechanism, five different HIV-1 RT complex structures involved in this ATP-mediated excision were determined using a cross-linking strategy and X-ray crystallography.

The structures of wild-type and AZT-resistant HIV-1 RT with a dsDNA and an AZTppppA (excision product complex) were determined to 3.15 and 3.2 Å resolution, respectively. Superposition of these two structures reveals that the ATP binds differently to wild-type RT and AZT-resistant RT. Primary resistance mutations do not improve a pre-existing site but create a new, higher-affinity binding site for ATP. The primary AZT-resistance mutations K70R and T215Y make significant contributions to ATP-binding, whereas, secondary resistance mutations assist the binding of ATP to RT. These structures also provide insights into antagonism between AZT-resistance mutations and other nucleoside-analog resistance mutations, including K65R and K70E.

The structure of AZT-resistant HIV-1 RT with a pre-translocation AZTMP-terminated DNA was determined to 3.7 Å. This structure has the fingers subdomain in a closed configuration instead of an open configuration found in the reported structure of the wild-type HIV-1 pre-translocation complex. The pre-translocation complex with the fingers in a closed configuration may be the preferred acceptor for ATP. This structure also implies polymerization/pyrophosphorolysis may occur without major opening of the fingers.

The structures of the AZT-resistant HIV-1 RT with a post-translocation AZTMP-terminated DNA and the unliganded AZT-resistant HIV-1 RT were determined to 2.9 and 2.65 Å resolution, respectively. Structural comparisons of these two complexes and excision product complexes suggest that the side chains of the two primary resistance mutations may undergo conformational changes that should be induced by the ATP.

Dedication

To

My parents, Sisters,

and Yunzhuo Zheng with love always

ACKNOWLEDEMENTS

To all the wonderful people that have contributed to my growth as a scientist and a person, you have my gratitude. I would first like to thank Dr. Eddy Arnold for taking me into his lab and his many-year support. I would also single out Dr. Sarafianos for all of his mentorship over the years. I am grateful for their advice and guidance.

I would also like to thank my committee members: Dr. Stefan Sarafianos, Dr. Wilma Olson, Dr. Roger Johns, Dr. Stephen Hughes, and Dr. Kalyan Das. I am grateful for all their support and help during my Ph.D. study.

I also want to thank both past and present members of the Arnold Lab for sharing with me their knowledge and their friendship. Dr. Kalyan Das has been a wonderful person who is willing to share his incredible knowledge of crystallography and research. I also wish to thank Dr. Steve Tuske, Art Clark, Chhaya Dharia, Daniel Himmel, Gail Ferstandig Arnold, and Deena Oren for all their friendship, support, encouragement, and knowledge. I would also like to thank Yulia Frenkel, Joe Bauman, Rajiv Bandwar, Mary Fitzgerald, Sergio Martinez, and Brian Hudson for their great help during my Ph.D study. Thanks must also go to Madeline Frances for making sure that I was registered every semester.

This dissertation has involved contributions from many scientists. Dr. Stefan Sarafianos crystallized, and processed the data from crystals of the AZT-resistant pre-translocation complex, post-translocation complex, and unliganded RT. Qianwei Han successfully synthesized the AZTppppA in time. Xiaorong Hou synthesized the crosslinkable primer.

Dr. Hughes' Lab generously provided AZT-resistant HIV-1 RT. Dr. Joe Bauman constructed the wild-type HIV-1 RT plasmid. Yulia Frenkel contributed to the structural determination of the AZT-resistant pre-translocation complex. For all these wonderful scientists, you have my great gratitude for your cooperation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

6XHis	hexahistidine
AIDS	acquired immunodeficiency syndrome
APS	Advanced Photon Source
AZT	3'-azido-3'-deoxythymidine
CHESS	Cornell high energy synchrotron source
CNS	Crystallography and NMR system
COOT	(Crystallographic Object)-Oriented Toolkit
dATP	2'-deoxyadenosine 5'-triphosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
DM	density modification
dNTP	deoxyribonucleotide triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
HIV-1	human immunodeficiency virus type 1
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
LB	Luria broth
MW	molecular weight
N site	dNTP binding site
NNRTI	non-nucleoside RT inhibitor
NRTI	nucleoside/nucleotide RT inhibitor
P site	priming site
PDB	protein data bank
RMSD	root mean square deviation
RNase H	ribonuclease H
RT	reverse transcriptase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

PART I. INTRODUCTION & LITERATURE REVIEW

1. AIDS and HIV

AIDS stands for acquired immunodeficiency syndrome. AIDS is caused by human immunodeficiency virus (HIV). HIV infects and destroys CD4-positive T-helper lymphocytes and macrophages. Once the CD4 cell-count decreases to a level at which the immune system is no longer effective against opportunistic pathogens, the patients die from multiple infections and tumors. AIDS is one of the most devastating pandemics of our time. According to estimates from the UNAIDS/WHO AIDS epidemic update (December 2006), an estimated 38.6 million people worldwide were living with HIV at the end of 2005. About 4.1 million people became newly infected with HIV and 2.8 million died from AIDS in 2005. AIDS has already orphaned 14 million children, and will orphan 25 million by 2010. Highly effective medicines and therapies are desperately needed to fight this terrible pandemic. The biggest challenge is drug resistance, since HIV undergoes very rapid genetic variation that eventually causes all current drugs to fail. Development of new medicines and therapies that can broadly inhibit resistant HIV variants will depend on a deeper understanding of the mechanisms of drug resistance.

There are two types of HIV: HIV-1 and HIV-2. Both types are transmitted by sexual contact, through blood and from mother to child. HIV-2 is only found in west Africa. Very few (under 100) people are infected by HIV-2 in the USA. HIV-2 weakens the

immune system more slowly. HIV-1 is responsible for the majority of global infections. Hence, when people talk about HIV, they usually mean HIV-1.

2. HIV life cycle and targeted processes for chemo-intervention of AIDS

Studies of the details of the HIV life cycle lead to the discovery of drug targets. The HIV life cycle can roughly be described in following steps (Fig. 1). The first step is entry into a $CD4^+$ cell. HIV binds to the CD4 receptor, followed by fusion of the viral envelope with the $CD4^+$ cell membrane, and then followed by release of an HIV capsid assembly of viral proteins and two copies of an RNA genome into the cell. The fusion process requires the assistance of co-receptors such as CCR5 and CXCR4. This step can be inhibited by fusion/entry inhibitors. The second step is reverse transcription. Once the viral capsid enters into the cell, an enzyme called reverse transcriptase (RT) copies the single-stranded genomic 10 kbase RNA into a complementary double-stranded DNA. This step is error-prone, and errors may occur at a frequency of about one per replication cycle. This step can be inhibited by RT inhibitors. The third step is integration. The replicated double-stranded viral DNA is transported into the cell nucleus, and integrated into the cell genome. Integration is carried out by another enzyme called integrase. This step can be blocked by integrase inhibitors. The fourth step is provirus transcription and viral protein synthesis. In this step, the provirus is copied to mRNA, and cellular machinery is used to synthesize viral proteins. The full-length RNA, which becomes the viral genome, binds to the Gag polyprotein and is packaged into new virus particles. The

final step of the HIV life cycle is the assembly, maturation and budding of new HIV virions. During the maturation process, HIV proteases cleave the Gag and Pol polyproteins into individual functional HIV proteins and enzymes. These various structural components are then assembled to form mature HIV virions. HIV protease inhibitors are designed to inhibit this cleavage process.

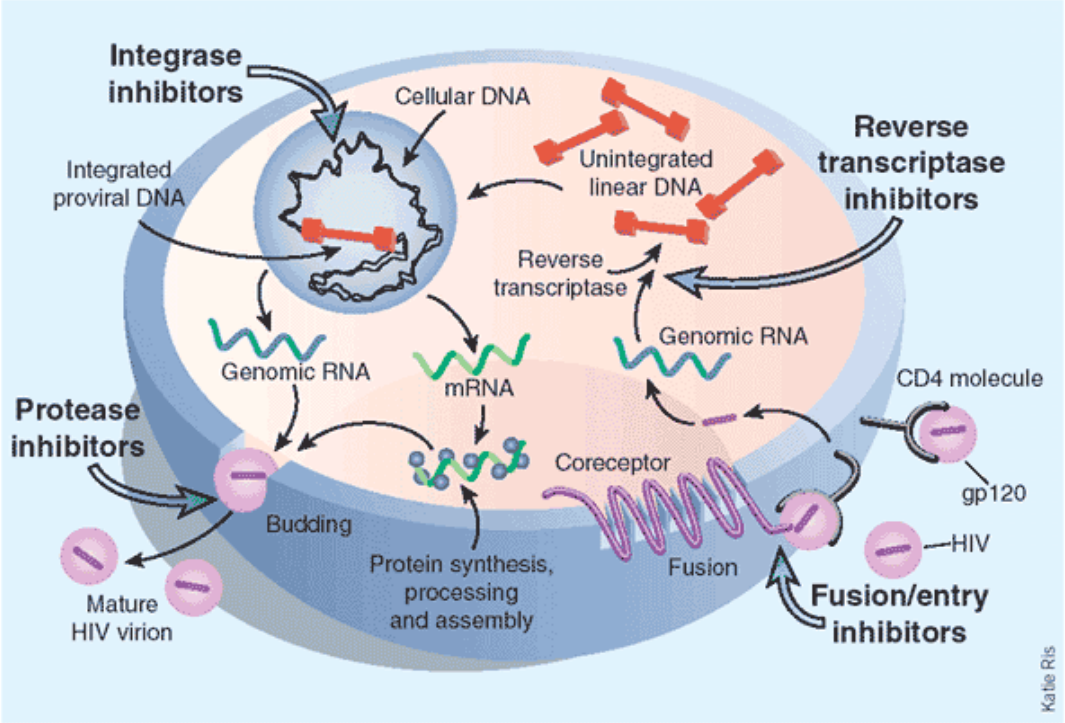


Fig. 1. HIV life cycle and drug-targeted processes

http://www.biology.arizona.edu/immunology/tutorials/AIDS/treatment.html

3. HIV-1 RT is one of the primary targets for drug design

HIV-1 RT plays a pivotal role in the virus replication cycle, since it is responsible for copying viral RNA into proviral DNA and is a primary source of viral mutations. Inhibition of its enzymatic activities will stop the virus life cycle. Hence, HIV-1 RT has been one of the primary targets for chemotherapeutic intervention in the control of AIDS (Jonckheere et al., 2002; Arnold et al., 1996). HIV-1 RT is a multifunctional enzyme with RNA/DNA-dependent DNA polymerase activities and ribonuclease H (RNase H) activity (Goff 1990; Whitcomb 1992; Le Grice 1993). The RNA/DNA-dependent DNA polymerase activities convert the single-stranded virus RNA genome into double-stranded DNA provirus capable of being integrated into the host cell genome. The RNase H has ribonuclease activity to cleave the RNA of the RNA/DNA hybrid. The functional HIV-1 RT is a heterodimer composed of two related polypeptide chains, a 66-kDa subunit (p66) and a 51-kDa subunit (p51) (Fig. 2). The p66 subunit contains two domains: the N-terminal polymerase domain (440 residues) and the C-terminal RNase H domain (120 residues). The N-terminal polymerase domain can be divided into four subdomains. The first three subdomains resemble a right hand, so they are referred to as fingers, palm, and thumb subdomains, respectively (Kohlstaedt, et al., 1992). The fourth subdomain connects the first three subdomains to the RNase H domain, which is therefore called the connection subdomain. The polymerase catalytic site is located in the p66 palm subdomain with three catalytic aspartic acid residues (Asp110, Asp185 and Asp 186) that are known to chelate two Mg^{2+} ions required for polymerase activity. The p51 subunit is

produced through proteolytic removal of the RNase H domain from a p66 chain. Of the two subunits, only p66 has functional polymerase and RNase H active sites.

Although p66 and p51 have identical primary amino acid sequences, the two subunits are structurally very different. The three catalytic aspartic acids that are exposed in the DNA-binding cleft in p66 are buried in p51. Another prominent difference between the two subunits is the orientation of the connection subdomain. The connection subdomain contacts only the RNase H domain and the thumb subdomain in p66 subunit; however it contacts all of the other subdomains and is tucked into a central position in p51 subunit (Jacobo-Molina et al., 1993).

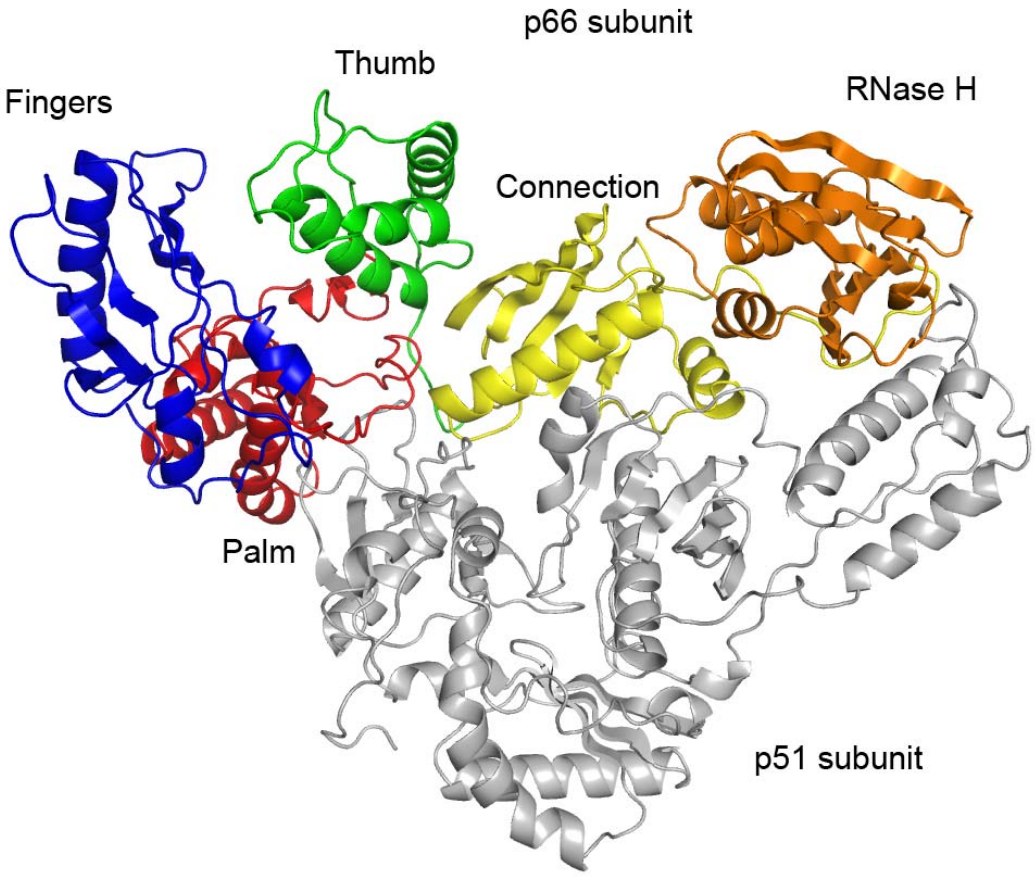


Fig. 2. Overall structure of HIV-1 RT p66/p51 heterodimer showing the relative locations of domains or subdomains. The p66 subunit is colored as follows: fingers in blue; palm in red; thumb in green; connection in yellow; and RNase H domain in orange. The p51 subunit is colored in grey.

4. DNA polymerization by HIV-1 RT

DNA polymerization by HIV-1 RT is a multiple-step event and involves many intermediates. During one polymerization cycle, RT must fit template-primer DNA to the catalytic site, recognize the incoming deoxynucleoside triphosphate (dNTP), line up the dNTP with the 3'-primer terminus, catalyze formation of the new phosphodiester bond, release the pyrophosphate product, and finally translocate to the next location (Lu et al., 2004).

4.1 DNA polymerization by HIV-1 RT is a multi-step reaction

According to previous studies, polymerization by HIV-1 RT can be described as follows (Fig. 3): 1). Binding of the DNA substrate to the free enzyme with primer 3'-end at the priming site (P site); 2). Binding of dNTP and Mg^{2+} (required for catalysis) to enzyme-DNA to form "open" catalytic complex; 3). Rate-limiting conformational change; 4). Formation of a phosphodiester bond between the 3'-OH primer terminus and the α-phosphate group of dNTP, followed by release of the pyrophosphate product (PPi); 5). Translocation of the elongated DNA primer from the dNTP-binding site (N site) to the priming site (P site) (processive DNA synthesis) or release of the nucleic acid (distributive DNA synthesis) (Sarafianos et al., 2002).

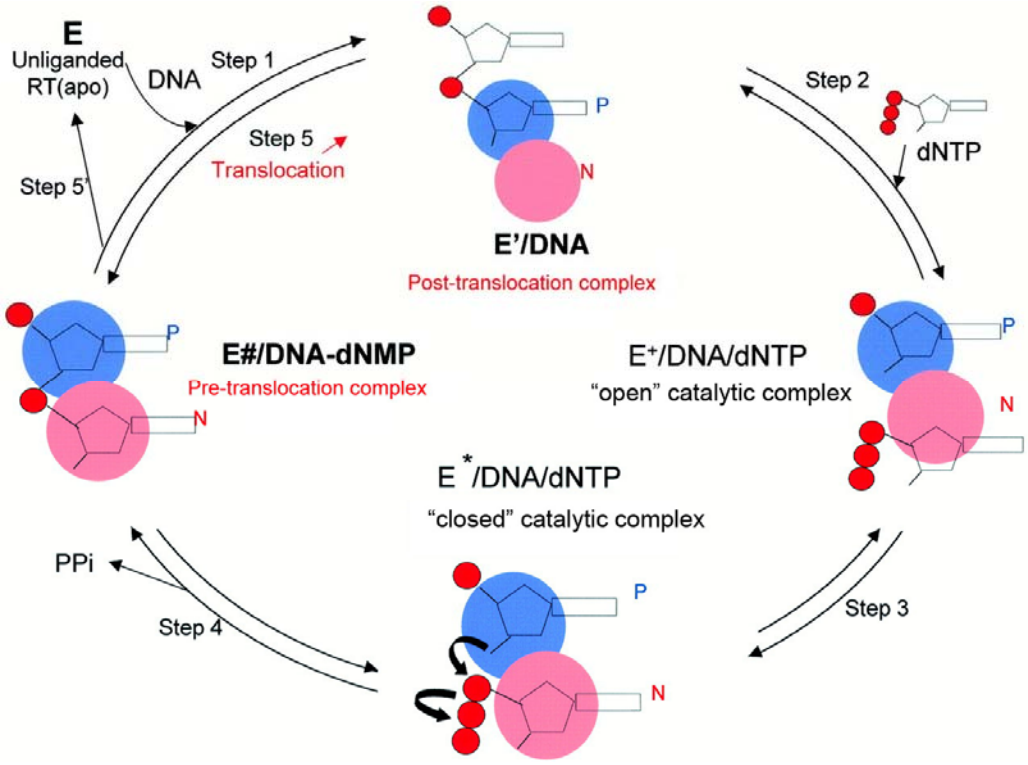


Fig. 3. Reaction steps and intermediates involved in RT-catalyzed polymerization (Sarafianos et al., 2002). P and N represent priming site (in blue) and nucleotide-binding site (in red), respectively.

4.2 Structures of intermediates involved DNA polymerization

The crystallographic structures of several polymerization reaction intermediates have already been solved (Sarafianos et al., 2002; Tuske et al., 2004; Huang, et al., 1998; Jacobo-Molina et al., 1993; Hsiou et al., 1996; Rodgers et al., 1996). All of the previously solved HIV-1 RT structures can be classified into one of the five categories: unliganded; in a complex with non-nucleoside inhibitors; in a complex with a Fab fragment and a double-stranded DNA with 3'-primer end at the P site (referred to as post-translocation complex or complex P); in a complex with a double-stranded DNA and a dNTP nucleotide (referred to as ternary complex); in complex with an Fab fragment and a double-stranded DNA with 3'-primer end at the N site (referred to as pre-translocation complex or complex N) (Sarafianos et al., 1999).

Comparisons of these structures have uncovered significant conformational movements or changes during polymerization, and also provided important insights into the structural mechanisms of drug resistance. Unliganded RTs have their thumb subdomains in a closed configuration. On the other hand, structural superposition shows that the RT/DNA complexes (pre-translocation or post-translocation complexes) have their thumb subdomains in an open configuration. This result indicates that binding of DNA to RT is accompanied by outward rotation of the thumb subdomain. This subdomain movement results in formation of a deep crevice between the thumb and the fingers, through which the DNA can bind to the polymerizase active site (Jacobo-Molina et al., 1993).

The solved structures of HIV-1 ternary complexes have their fingers subdomains in a closed configuration; however the solved structures of post-translocation and pre-translocation complexes have their fingers subdomains in an open configuration. Based on structural superposition of currently available ternary complexes, pre-translocation complexes and post-translocation complexes, The DNA polymerization by HIV-1 RT is accompanied by periodic movement of the fingers subdomain and a conserved $Y_{183}M_{184}D_{185}D_{186}$ (YMDD) motif. The movement of the fingers subdomain can be described as follows. The fingers subdomain begins in an open configuration in the post-translocation complex. Subsequently binding of an incoming dNTP to a post-translocation complex results in formation of an open ternary complex (E^{+}/DNA/dNTP). The fingers subdomain then closes down on the incoming dNTP, and aligns the dNTP with the 3'-OH of the primer. Hence, the open ternary complex (E^{+}/DNA/dNTP) is then converted to a closed catalytic ternary complex (E^{*}/DNA/dNTP). This has been proposed to be the rate-limiting step (Kati et al., 1992; Hsieh et al., 1993). The chemical reaction of nucleotide incorporation is an in-line nucleophilic attack of the 3'-OH of primer to the α-phosphate group of dNTP. After incorporation, the fingers subdomain opens to release PPi. It is still unclear whether the release of PPi and the opening of the fingers subdomain happen simultaneously or sequentially. Although the solved wild-type pre-translocation complex has the fingers subdomain in an open conformation, the closed state of catalytically active initiation complex of reovirus RNA

polymerase has suggested that translocation and insertion of a new NTP (or dNTP in RT) can occur without a major opening of the fingers (Tao et al., 2002). If we can obtain pre-translocation intermediate with the fingers subdomain in a closed conformation, then it will help clarify the potential contribution of fingers motion to the polymerization. The subsequent translocation moves the primer 3'-end from the N site to the P site, resulting in the formation of the post-translocation intermediate.

DNA polymerization is also accompanied by the periodic movement of the YMDD motif on a relatively rigid scaffold (Fig. 4 and Fig. 5). The YMDD ($Y_{183}M_{184}D_{185}D_{186}$) motif is a β-turn between β10 and β9, and contains catalytic residues (Asp185 and Asp186) for the polymerization. Structural superposition of trapped RT intermediates (unliganded RT, post-translocation complexes, ternary complexes and pre-translocation complexes) has shown that the overall RMSD for C_α atoms is less than 0.5 Å, but the YMDD motif is displaced about 2.8 Å away from its starting position that is defined by unliganded HIV-1 RT (superposition based on residues 107-112 and 155-215 in p66) (Ding et al., 1998; Sarafianos et al., 2002). The mobility of the YMDD motif combined with the rigidity of the scaffold in the active site has led to a proposal of a springboard motion of translocation (Sarafianos et al., 2002). Binding of DNA to the unliganded structure is accompanied by ~1.7 Å downward displacement of the YMDD motif from its starting position (Ding et al., 1998). Binding of an incoming dNTP to RT/DNA and the subsequent polymerization reaction result in further downward displacement of ~0.5 Å and ~1.0 Å, respectively

(Sarafianos et al., 2002). These processes have been compared to the loading of a springboard, and may store part of energy released from dNTP incorporation (Sarafianos et al., 2002). After dNTP incorporation, the 'bounce' of the YMDD springboard may supply some of the energy for translocation, and the YMDD motif moves upward ~1.5 Å or approximately halfway between its extreme positions. For the remainder of processive polymerization, the YMDD motif oscillates in cycle from the post-translocation position, to the catalytic position, then to the pre-translocation position and finally backs to the post-translocation position (Fig. 5) (Sarafianos et al., 2002).

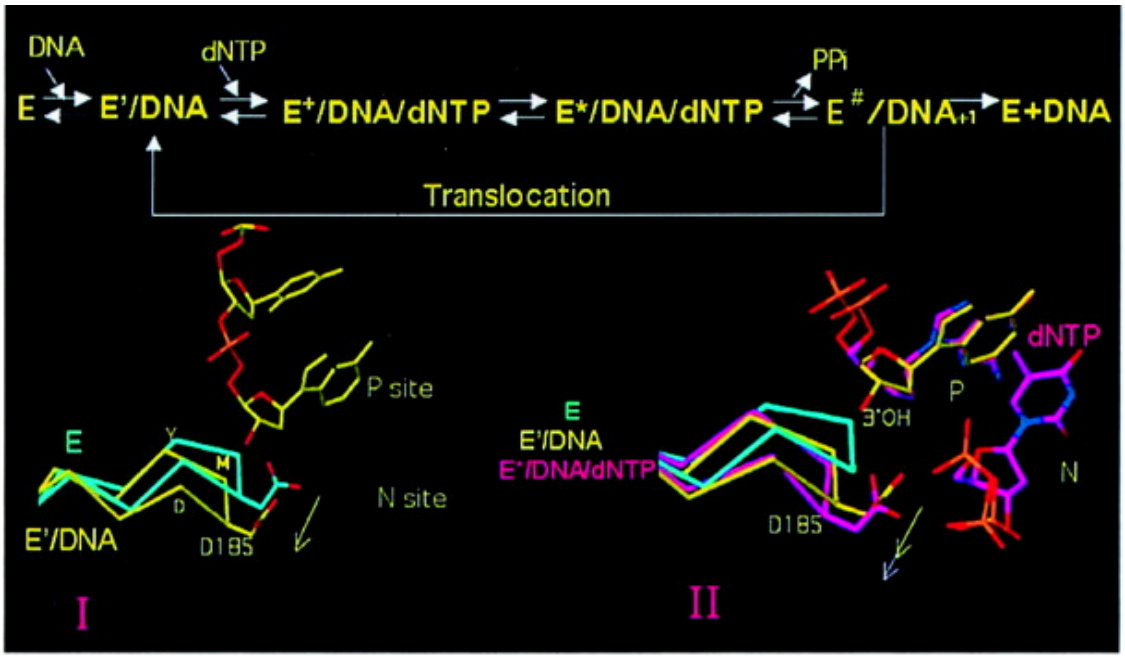


Fig. 4. The motion of the conserved YMDD motif of the polymerase active site during polymerization (Sarafianos et al., 2002). unliganded [RT (apo), cyan] (Hsiou *et al*., 1996), post-translocation complex [HIV-1 RTWT/$DNA_{AZTMP(P)}$, yellow] (Sarafianos et al., 2002), catalytic complex [HIV-1 RTWT/DNA_{ddAMP}/dTTP, white] (Huang *et al*., 1998) and pre-translocation complex [HIV-1 RT/$DNA_{AZTMP(N)}$, magenta] (Sarafianos et al., 2002). Superposition is based on p66 residues 107–112 and 155–215. In I, superposition of the unliganded RT (apo) and the post-translocation complex reveals the movement of the YMDD motif in the post-translocation complex relative to RT(apo). In II, superposition of the unliganded RT (apo), the pre-translocation complex and the catalytic complex shows further displacement of the YMDD motif in the catalytic complex.

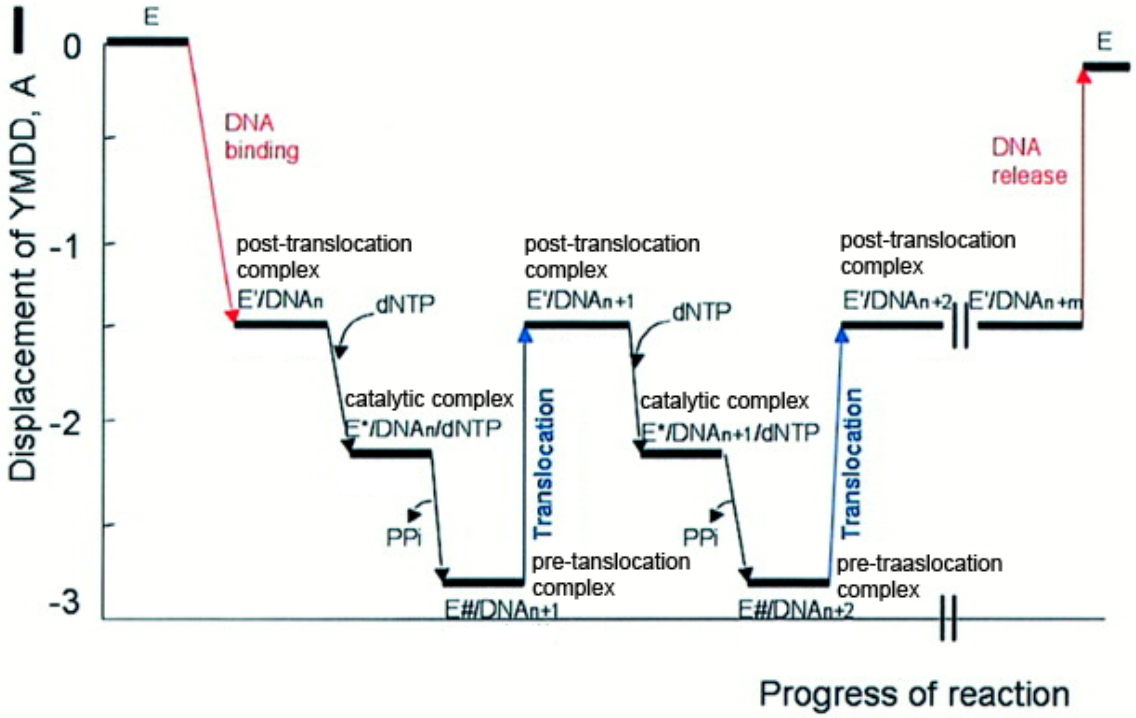


Fig. 5. Springboard motion of YMDD motif (Sarafianos et al., 2002)

Displacement of the C_α of D185 of the conserved YMDD motif relative to the unliganded structure of RT (apo) (PDB id: 1DLO) during polymerization. Negative values in Å correspond to compression. The C_α of D185 is displaced ~1.7 Å in the post-translocation complex, ~2.2 Å in the catalytic complex and ~2.7 Å in the pre-translocation, ~2.7 Å in the pre-translocation complex from its reference position in unliganded RT (apo, cyan).

5. HIV-1 RT inhibitors

The introduction of highly active antiretroviral treatments against HIV-1 has been successful in decreasing viral load and prolonging life-span of HIV-infected patients (Richman et al., 2001). All of the current HIV-1 RT inhibitors fall into four categories: nucleoside analog RT inhibitors (NRTIs), non-nucleoside RT inhibitors (NNRTIs), RNase H inhibitors, and PPi analog inhibitors. The nucleoside analogs are prodrugs. They are phosphorylated to their triphosphate forms in vivo, and then bound to the polymerization active site (Fig. 6), and then incorporated into growing DNA chains. Because NRTIs lack the 3'-OH group, they function as chain terminators (Parker et al., 1991). NNRTIs are a structurally diverse class of hydrophobic compounds. A number of RT/NNRTI complex structures have been solved, which show that NNRTI bind a pocket ~10 Å away from the polymerase active site (Kohlstaedt et al., 1992; Tantillo et al., 1994) (Fig.6). Steady-state kinetic studies suggest NNRTIs are noncompetitive with respect to the dNTP or primer/template. Binding of NNRTI to RT can cause conformational changes that can extend to the polymerase active site, through which they affect the chemical formation rate of phosphodiester bond (Spence et al., 1995; Rittinger 1995). RNase H inhibitors inhibit nuclease activity, and prevent cleavage of RNA from a DNA/RNA duplex. There are few reported examples of PPi analogs (Crumpacker 1992; Öberg 1989). PPi analogs can inhibit pyrophospholysis activity. No structure of an RT/PPi analog complex has been reported as yet, and the binding site of PPi is still unclear.

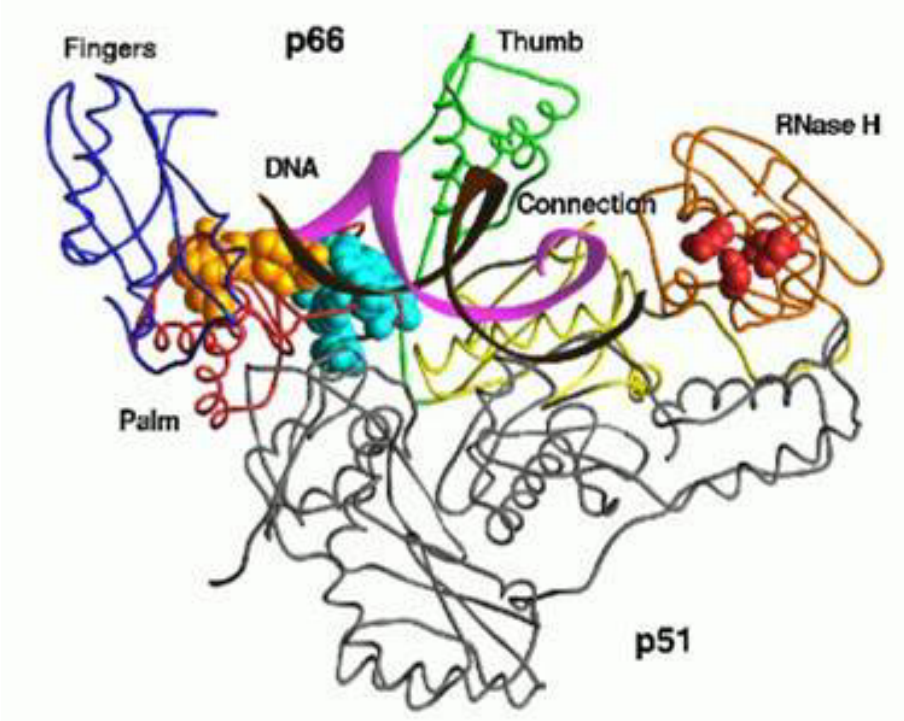


Fig. 6. Binding sites in HIV-1 RT (Made by Kalyan Das)

Binding sites for inhibitors are colored orange for NRTIs and cyan for NNRTIs; the RNase H active site is colored red.

The antiretroviral agents currently approved by the FDA include 8 nucleoside analogs (AZT, ddI, ddC, d4T, 3TC, abacavir, FTV, and tenofovir disoproxil fumarate), 4 NNRTIs (nevirapine, delavirdine, efavirenz, and etravirine), 8 viral protease inhibitors (saquinavir, indinavir, ritonavir, nelfinavir, amprenavir, lopinavir, atazanavir, and daruanivir), and 1 viral fusion inhibitor (T-20), a viral integrase inhibitor (Raltegravir), and a CCR5 inhibitor (maraviroc). Monotherapy treatment against HIV-1 infection can decrease viral load at the beginning, but is ultimately rendered ineffective by drug-selected mutations. Combination therapy (also called cocktail therapy), however, can greatly reduce the chance that resistance mutations will develop to render the treatment ineffective, since drugs have different mechanisms of action. The commonly used combinations include two or three nucleoside analogs with or without a protease inhibitor or an NNRTI.

6. AZT is a widely used anti-AIDS drug

AZT, 3'-azido-2',3'-deoxythymidine, was the first drug approved for the treatment of AIDS. AZT was approved by the Food and Drug Administration (FDA) for use against HIV in 1987, and as a preventive treatment in 1990. It has become one of the most commonly prescribed anti-HIV drugs nowadays. Combination therapy with AZT has been successful in decreasing viral load and prolonging lifespan of HIV-infected patients. AZT belongs to the category of NRTIs, and is an analog of thymidine. AZT differs from thymidine in that the 3'-substituent group is 3'-azido instead of 3'-OH. AZT is activated to its triphosphate form (AZTTP) *in vivo*, and then incorporated into DNA primer 3'-end through DNA polymerization by HIV-1 RT. After incorporation into a DNA chain, AZT

will terminate the elongation of the DNA strand. The triphosphate form of AZT also has some ability to inhibit cellular DNA polymerases. However, AZT has a 100- to 300-fold greater affinity for the HIV-1 RT than the human DNA polymerase, accounting for its antiviral selectivity (Mitsuya et al., 1985). The DNA polymerase in mitochondria is relatively more sensitive to AZT, which causes certain toxicities such as damage to cardiac and other muscles (also called myositis) (Yarchoan et al., 1989; Balzarini et al., 1999; Rang et al., 1995; parker et al., 1991; Collins et al., 2004).

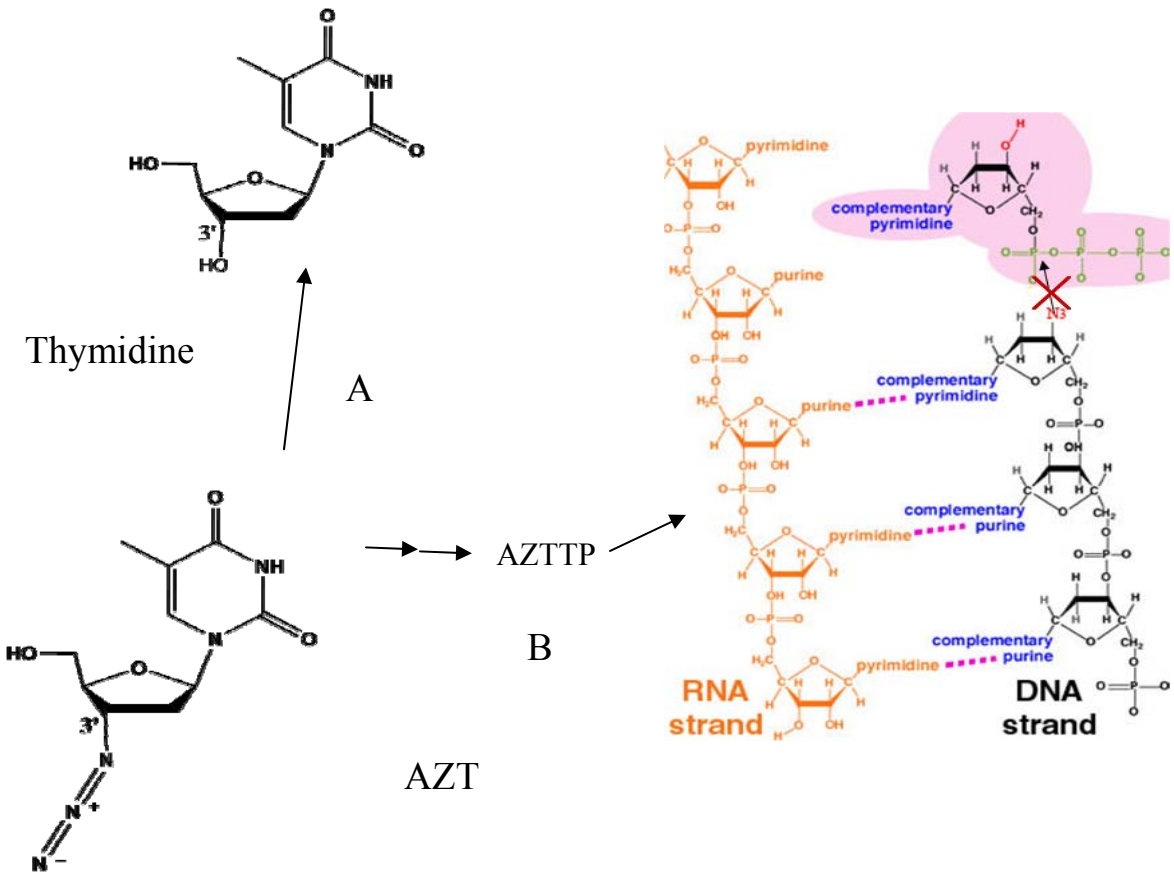


Fig. 7. AZT is a chain terminator

A. AZT is an analog of thymidine

B. AZT terminates DNA chain elongation.

7. AZT resistance mutations

Resistance to AZT is often caused by cumulative HIV-1 RT mutations that usually include M41L, D67N, K70R, L210W, T215Y/F, and K219Q/E. These AZT-resistance mutations are also referred to as thymidine analog mutations (TAMs), because another thymidine analog drug (stavudine) also selects for them (Pellegrin et al., 1999; Coakley et al., 2000). AZT-resistance mutations emerge in a stepwise manner (Fig. 8). Although the order of appearance of these mutations can vary among individual patients, some general trends can be recognized (Boucher et al., 1992; Kellam et al., 1992; Cleland et al., 1996). Some of the resistance-inducing mutations appear relatively early, e.g., K70R and M41L. Other mutations are observed almost exclusively at relatively later stages such as T215Y/F, which dramatically enhance the resistance to AZT in the background of early mutations (Larder 1994). One reason for the differences in time of appearance of AZT-resistance mutations is the number of codon nucleotides changed (Keulen et al., 1996). For instance, a relatively simple transition or transversion is required for the AZT-resistance mutations at codon 41 and 70, respectively, and a more difficult two-nucleotide transversion is required at codon 215.

The selection of resistance mutations during antiretroviral therapy is associated with reduction in drug susceptibility and viral fitness. Some resistance-inducing mutations might cause a significant polymerization activity defect, but other mutations might compensate for this defect. Hence, a complex interplay of parameters including drug

resistance and catalytic properties of RT variants direct the evolution of drug resistance (Berkhout 1999). Resistance-related mutations have been conventionally classified as primary or secondary based on drug susceptibility. Secondary mutations do not confer high level of resistance on their own, but can improve the replication fitness and enhance the resistance level conferred by primary mutations (Larder 1994; Jeeninga et al., 2001). Of the mutations selected by AZT, T215Y/F and K70R are generally considered primary, and D67N, L210W and K219Q/E are considered secondary. The primary mutations are selected at relatively early stage to confer resistance to drugs; however the secondary mutations are always selected at relatively late stages in some cases to improve the replication fitness harmed by the primary resistance mutations.

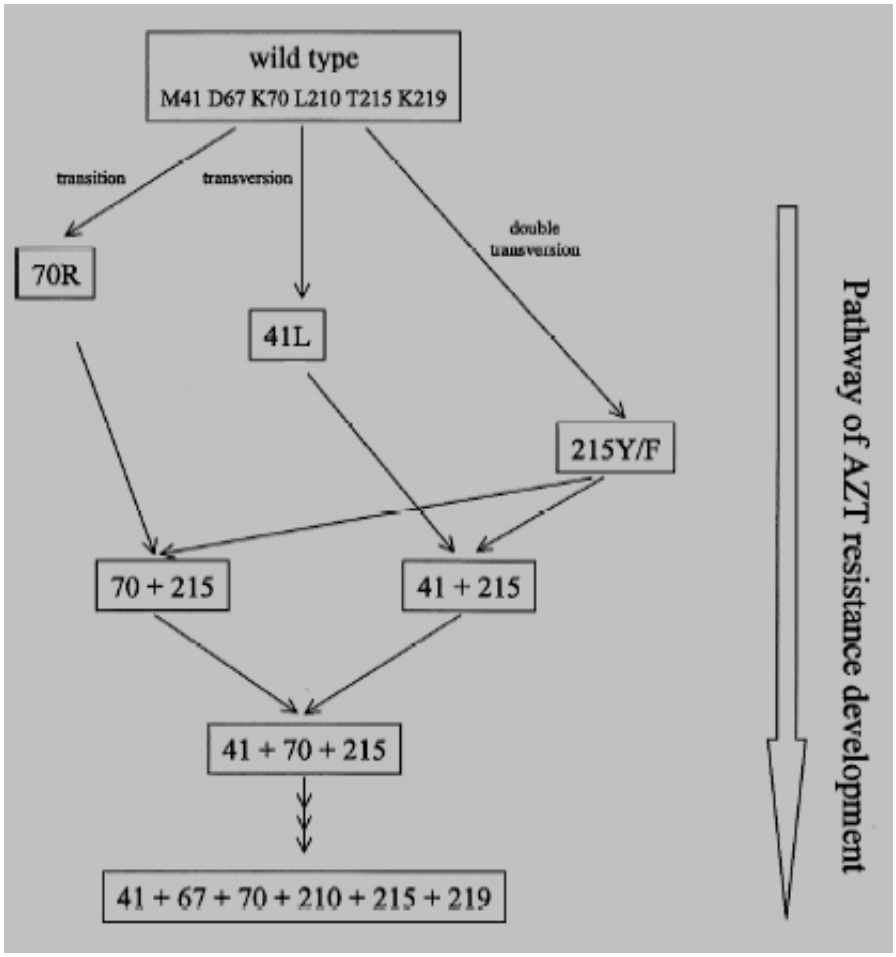


Fig. 8. Pathway of AZT resistance development (Jeeninga et al., 2001)

Most AZT-resistant HIV-1 RT isolates contain only a subset of the AZT-resistance mutations (Hooker et al., 1996; Kellam et al., 1992; Larder & Kemp, 1989; Richman et al., 1991). D67N/K70R/K219Q and M41L/L210W/T215Y are frequent AZT-resistance mutation clusters selected by AZT (Hanna et al., 2000; Yahi et al., 1999). Both of them have only one of the two primary mutations, which might reflect different evolutionary pathways of AZT resistance.

8. The chemical mechanism of HIV-1 RT resistance to AZT

8.1 The long-standing mystery of AZT resistance mechanism

While clinical data clearly had demonstrated that HIV-1 RT with M41L/D67N/K70R/L210W/T215Y/K219Q mutations confer more than 100-fold resistance to AZT relative to wild-type RT, the resistance mechanism had remained obscure for more than 10 years, because the resistance couldn't be demonstrated *in vitro*. The reason is that the polymerization assay originally used for studying resistance didn't contain reagents necessary for AZT excision.

8.2 ATP-mediated pyrophosphorolysis

A solution to this mystery didn't begin to emerge until 1998. HIV-1 RT doesn't have 3'-exonuclease activity (Roberts et al., 1998), but it has pyrophosphorolysis activity (the reversal of polymerization). In 1998, it was found that the pyrophosphorolysis could be mediated by nucleoside triphosphates and nucleoside diphosphates in *vivo* (Meyer et al.,

1998). The nucleotide-mediated pyrophosphorolysis activity is too low for wild-type HIV-1 RT to be significantly resistant to AZT, but it can be dramatically enhanced by AZT-resistance mutations. Biochemical studies have shown that nucleoside triphosphates, nucleoside diphosphates, and pyrophosphate (PPi) itself can act as PPi donors to initiate AZT excision (Arion et al., 1998; Meyer et al., 1998); however ATP is the preferred candidate *in vivo*, considering its efficiency and physiological concentrations (Meyer et al., 1999; Boyer et al., 2001). The ATP-mediated pyrophosphorolysis is a nucleophilic substitution reaction (Fig. 9). ATP binds to HIV-1 RT, and its γ-phosphate is poised toward the scissile phosphate of AZTMP at the primer 3'-end. A nucleophilic attack by the γ-phosphate excises AZTMP from the primer 3'-end, and forms an excision product, AZT adenosine dinucleoside tetraphosphate (AZTppppA): One end is referred to as the ATP moiety and the other end as the AZTMP moiety (Fig. 9).

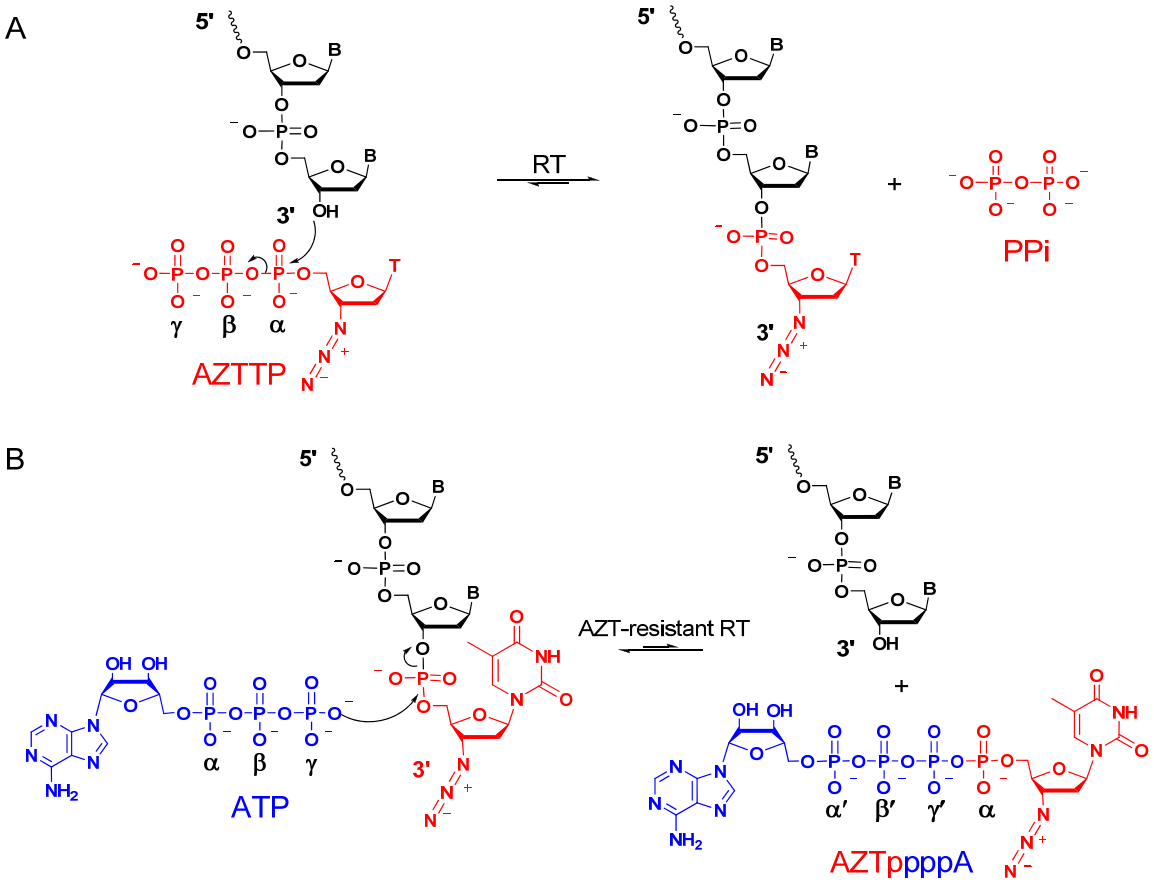


Fig. 9. Resistance mechanism of HIV-1 RT to AZT

(A) Polymerization of AZTTP

(B) AZT-resistant RT has enhanced ATP-mediated pyrophosphorolysis activity

8.3 AZT excision model

The first structure of an HIV-1 ternary complex, wild-type HIV-1 RT-DNA/dTTP complex, was published by Dr. Harrison's group in 1998 (Huang et al., 1998). Another key structure, wild-type HIV-1 RT/DNA/Fab binary complex, had been solved by Dr. Arnold's group in 1993 (Jacobo-Molina et al., 1993). Based on these two structures, Dr. Boyer and Dr. Sarafianos performed two types of modeling studies. (i) Starting with the ternary structure, a model was prepared in which the 3'-end is terminated with AZTMP and located at the N site. ATP was then modeled into the proposed ATP-binding site (Fig.10). (ii) Starting with the binary complex, a model was developed with AZTMP-terminated primer end at the P site and an incoming dNTP at the N site (Fig. 11). The first modeling studies showed that the scissile phosphate of primer 3'-end can be accessible by the γ-phosphate group of ATP only when it is at the N site. In other words, the AZTMP can be excised from 3'-end of the primer only when it is at the N site; the occupation of a dNTP at the N site will block excision (Boyer et al., 2001). The second modeling study showed that the azido group of the ribose ring clashes with D185 and/or incoming dNTP, if AZTMP-terminated 3'-end of primer is at the P site (Fig. 11). It tells us that the AZTMP-terminated 3'-end of primer is more likely to occupy the N site than the dideoxynucleotide-terminated 3'-end of primer, since the dideoxynucleotides don't have a 3'-azido group. These ideas have been confirmed by biochemical data that show the presence of complementary incoming dNTP interferes with the excision of a dideoxy nucleoside instead of AZT (Boyer et al., 2001). In 2002, two key structures were

published, which are wild-type HIV-1 RT with pre- and post-translocation AZTMP-terminated DNA (Sarafianos et al., 2002). Modeling studies based on these two structures provided further support of the previous modeling results (Sarafianos et al., 2002).

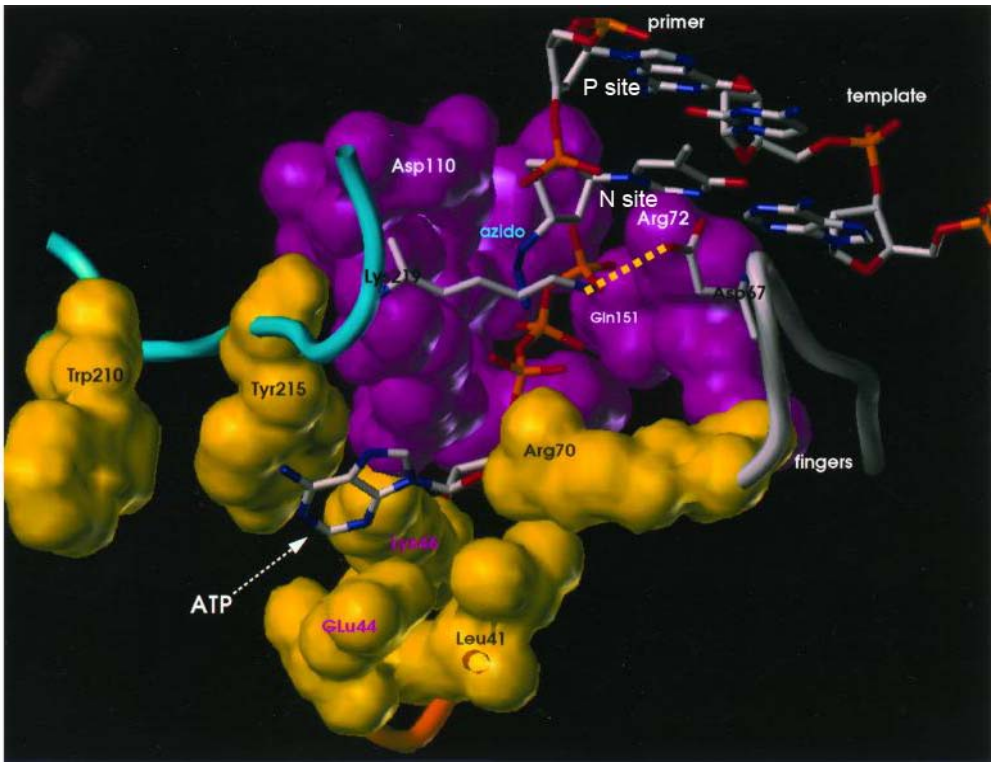


Fig. 10. Models showing the binding of ATP to AZT-resistant HIV-1 RT

Modeling shows that AZT-resistance mutations are actually around the ATP-binding site. The scissile phosphate of AZTMP-terminated 3'-end of primer can be accessible by the γ-phosphate group of ATP only when it is at the N site (Boyer et al 2001).

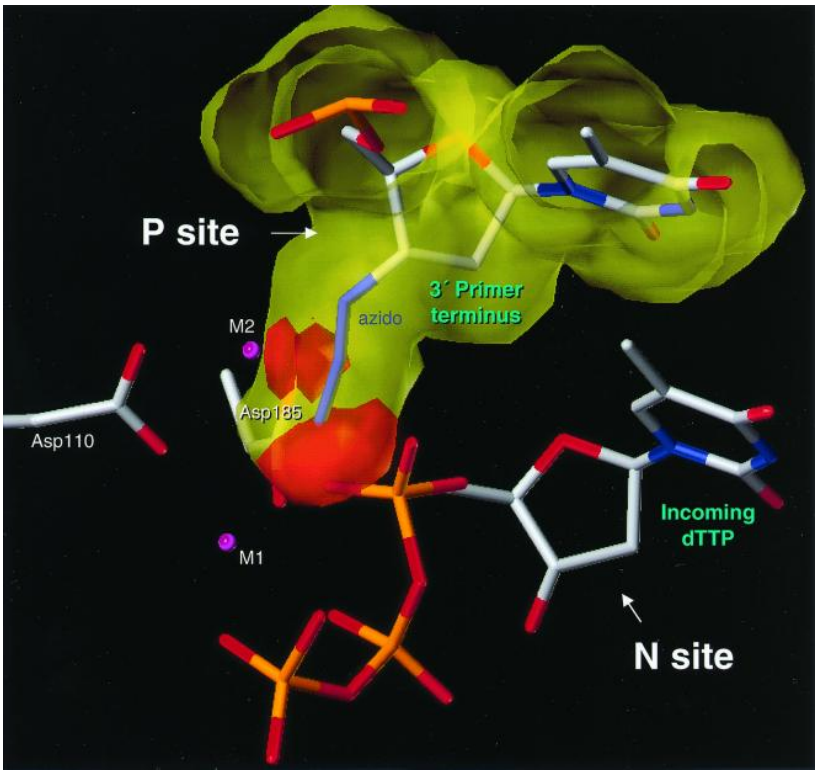


Fig.11. Steric clash when an AZTMP-terminated primer binds to RT at the P site. Modeling based on the wild-type HIV-1 RT-DNA/dTTP and wild-type HIV-1 RT/DNA/Fab shows that AZT clashes with D185 and incoming dNTP (Boyer et al., 2001).

Based on the above modeling and biochemical results, an excision model has been proposed (Boyer et al., 2001; Sarafianos et al., 2002). This model has been widely adopted by research groups working in the AZT resistance fields. In this model, excision occurs only when the AZTMP is positioned at the N site. The excision rate is affected by the accessibility of ATP and precise positioning of the reactive components including the 3'-primer end and the β,γ-phosphate moiety of ATP.

The events and intermediates involved in ATP-mediated excision process can be briefly described as follows (Fig. 12). After a dNTP is incorporated into the growing chain, translocation can move the 3'-primer terminus from the N site to the P site in a reversible process. The 3'-primer terminus occupies the N site or the P site in a dynamic manner, forming an equilibrium between pre-translocation and post-translocation intermediates. If the 3'-primer terminus is positioned at the N site (this intermediate is called pre-translocation intermediate), it creates the opportunity for excising the incorporated nucleotide from the 3'-primer end. In the normal case, this excision is inhibited, because the next incoming dNTP occupies the N site and halts the 3'-primer terminus at the P site. However, if the 3'-primer is terminated by AZTMP, the next incoming dNTP is prevented from occupying the N site because of steric clashes between its α-phosphate and the azido group of AZTMP. Thus, the AZTMP-terminated primer 3'-end has a better chance to occupy the N site than other NNRTIs, and it will favor AZT excision. Once the ATP binds to RT, its γ-phosphate will be positioned toward the scissile phosphate of the

AZTMP-terminated primer, and then elicit a nucleophilic attack. Immediately afterward, the excision product intermediate (AZT-resistant RT/DNA_{ddNMP}/AZTppppA) is formed. The followed event is the release of AZTppppA, which will leave a free 3'-OH of primer for polymerization.

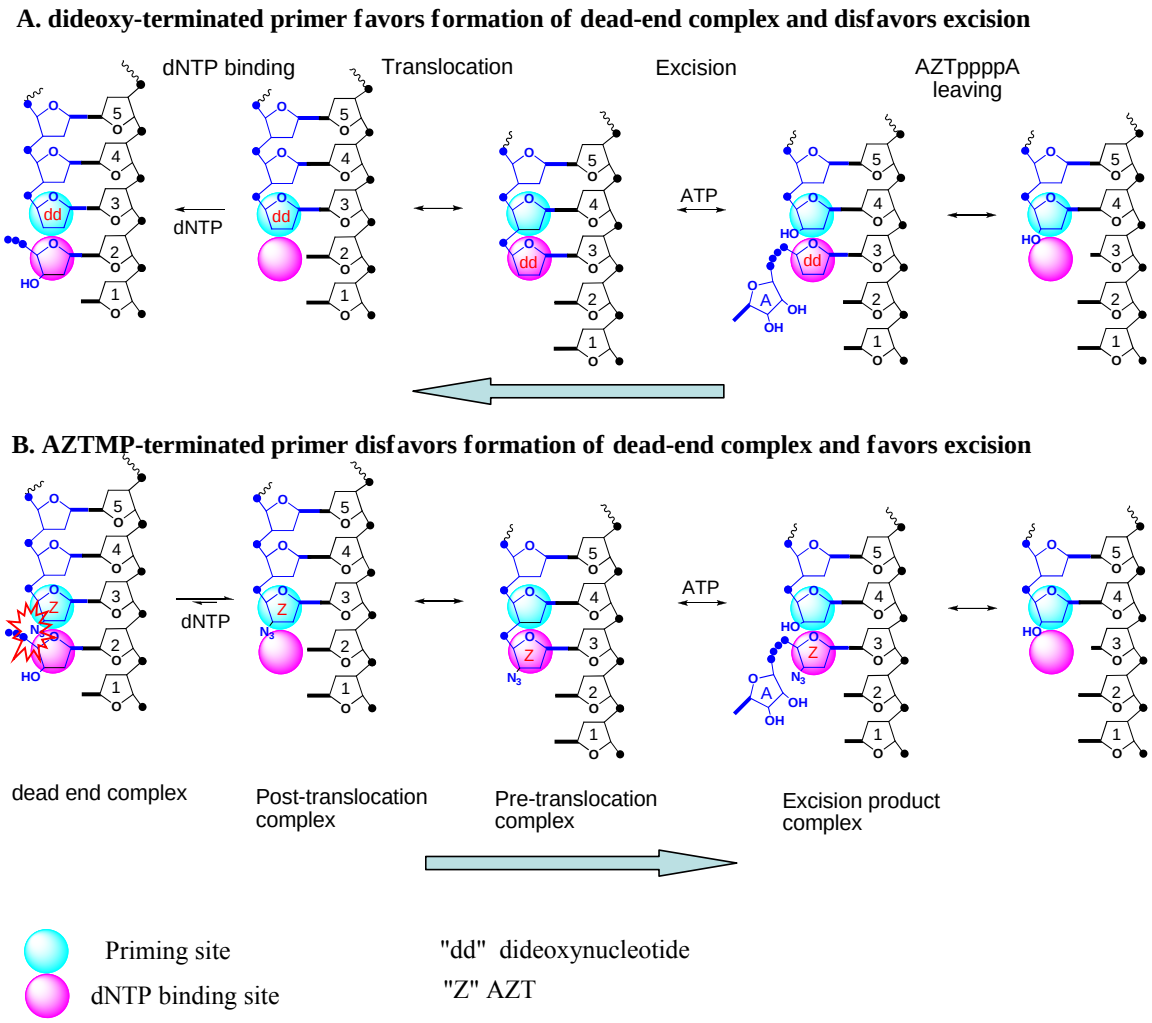


Fig. 12. AZT excision model

A. If the 3'-primer is terminated by a dideoxynucleotide, the catalytic dead-end complex (left) is a stable species that shifts the overall equilibrium to the left (green arrow to the left).

B. If the 3'-primer is terminated by an AZT, the catalytic dead-end complex (left) is unstable because of the azido group. It will shift the overall equilibrium to the right (green arrow to the right), therefore favoring the excision step (right).

8.4 The conjectured roles played by AZT-resistance mutations

The previous hypothesis about the roles of AZT-resistance mutations were mostly derived from the structural studies or modeling studies of wild-type HIV-1 RT complexes especially the wild-type catalytic ternary complex (HIV-1 RT/DNA/dTTP, Fig. 13) (Huang et al., 1998) and AZT-terminated pre- and post-translocation wild-type HIV-1 RT/DNA complexes (Sarafianos et al., 2002). These hypotheses were not confirmed by direct structural data so far. The previous hypothesis about the roles of AZT-resistance mutations are listed in the Table 1, and described as follows. The T215Y/F can possibly form π-π interactions with the aromatic adenine group of ATP, since it has a bulky phenyl side chain (Boyer et al., 2001, Sarafianos et al., 2002). This π-π stack could be enhanced by another mutation L210W, since it is proximal to the T215Y/F. K70R may form a hydrogen bond with the Asp113 and the γ-phosphate of the incoming dNTP. M41L mutation may reshape part of the 3'-OH pocket wall through interacting with Phe116 (Huang et al., 1998; Sarafianos et al., 1999). In the wild-type ternary structure, D67 interacts with K219 through a salt link across the cleft between palm and fingers subdomain. Their corresponding mutations (D67N and K219Q/E) may disrupt this salt link, through which they may destabilize the closed configuration of the fingers subdomain or enhance the accessibility of ATP to the binding cleft located between the fingers and palm subdomains. Although these studies has shed some insights into the roles of

AZT-resistance mutations, the knowledge about their roles has been limited by the lack of structures of HIV-1 RT excision intermediates containing the AZT-resistance mutations.

Table 1. The conjectured roles played by AZT-resistance mutations		
Resistant Mutations	Mechanism	Comments
T215Y/F	π-π stacks with aromatic adenine group of ATP	Primary resistance mutation
K70R	Mutant R70 may make a new hydrogen bond with γ-phosphate group of dNTP or D113	Primary resistance mutation
M41L	Causes indirect rearrangement of dNTP-binding pocket through residue 116	Associated with T215Y
L210W	π-π stacks with T215Y, and enhances the π-π stack between T215Y and adenine group of ATP	Secondary mutation associated with T215Y
D67N	Changes interaction with residue 215 and formation of dNTP-binding pocket	Secondary mutation
K219Q	Affects interactions of residue 219 with residue 67 across the cleft, as well as formation of the 3' pocket	Secondary mutation, associated with mutations at position 67, 70 and 215

9. Puzzles in this enhanced ATP-mediated pyrophosphorolytic excision

Although the AZT excision model has provided some insights into the AZT excision mechanism, the structural bases for this enhanced ATP-mediated pyrophosphorolysis are still unclear. The reason is that none of the excision intermediates containing AZT–resistance mutations has been solved as yet. This missing structural information causes a lot of questions to remain unanswered. In particular, How is the excision enhanced? What roles does each AZT-resistant mutation play? How is ATP bound in the AZT-resistant RT or wild-type RT? Are there any conformational changes associated to the AZT-resistance mutations during the ATP-mediated pyrophosphorolysis?

10. ATP-mediated pyrophosphorolytic excision is a general mechanism for drug resistance to NRTIs

NRTIs are nucleoside analogs. HIV-1 RT has to maintain its ability to incorporate normal nucleotides, no matter what NRTI-resistance mutations it bears. Thus HIV-1 RT resistance mechanism always involves enhanced discrimination of normal nucleotides and their analogs. This enhanced discrimination occurs either after incorporation or before incorporation. The discrimination after incorporation relies on selective excision. According to the excision model, any NRTI can theoretically be excised only if the NRTI-terminated primer is positioned at the N site. It is true when the RT bears 69-insertion or Δ67 deletion mutation in the background of AZT resistance mutations. Both mutation clusters can confer resistance to multiple-NRTIs through this

ATP-mediated pyrophosphorolytic excision (Meyer et al., 2003; White et al., 2004; Mas et al., 2002; Boyer et al., 2002). It was reported that the 69-insertion mutation cluster can confer resistance to all FDA-approved NRTIs (Larder et al., 1999). The 69-insertion mutation cluster contain a dipeptide insertion (SS, SA, or SG) between codons 69 and 70 and the AZT-resistance mutations. The 69-insertion mutation cluster has been observed in about 1% of patients receiving long-term NRTI-based therapies (Larder et al., 1999). The Δ67 mutation cluster has a deletion mutation at codon 67 and contain a set of other mutations including the AZT-resistance mutations (M41L/T69G/K70R/L74I/K103N/T215Y(F)/K219Q(E)). The Δ67 mutation cluster can cause resistance to AZT, tenofovir and, to a lesser extent, d4TMP (Imamichi et al., 2000; Boyer et al., 2004). In this sense, drug design against ATP-mediated excision will have general significance of re-sensitizing NRTIs-resistant variants to almost all of the NRTIs.

Discrimination before incorporation relies on interference with the incorporation of NRTIs. One simple example is the resistance mutation M184V/I that confers high level resistance to 3TC and FTC (Schuurman et al., 1995; Shafer et al., 2000). Residue 184 is part of the YMDD motif. Structural studies have shown that the β-branched side chain of mutant V184 or I184 can cause steric conflict with the L-oxathiolane ring of 3TCTP or FTC. This clash decreases the incorporation rate of 3TCTP or FTC, because there is an improper alignment of the substrates at the reactive center (Sarafianos et al., 2004, 1999; Gao et al., 2000).

11. Drug interactions

Over the last twenty years, the antiretroviral therapies have evolved from monotherapy to combination therapy, as the drug-resistance mutations evolve. Thus, understanding drug synergy will be very helpful in designing antiretroviral regimen and therapy strategy based on HIV-1 RT genotypes of patients (Sluis-Cremer et al., 2006). Some drug synergies are resulted from interplay of different classes of drugs. NNRTIs can re-sensitize the AZT-resistant HIV-1 RT to AZT by indirectly inhibiting the unblocking activity through modulating the binding of ATP (Meyer et al., 1998; Basavapathruni et al., 2004; Odriozola et al., 2003). Some drug synergies are related to specific patterns of resistance mutations selected by different drugs. K70E is a resistance mutation after treatments of ddI, abacavir, but this mutation re-sensitize AZT-resistance RT to AZT. K65R mutation was originally isolated from patients treated with ddC or ddI, and subsequently shown to confer resistance to TDF and foscarnet (Gu et al., 1994; Gallant et al., 2003; Meyer et al., 2003). Biochemical studies have shown that K70E and K65R are antagonistic to ATP-mediated pyrophosphorolysis and can re-sensitize AZT to AZT-resistance mutations (Miller et al., 2005). Mutations selected by foscarnet, a PPi analog, can also re-sensitize AZT to AZT-resistant variants.

12. Excision inhibitors

The ATP-binding pocket can possibly be a new target for designing new drugs to inhibit HIV-1 RT variants containing AZT-resistance mutations. These drugs inhibit ATP-mediated excision by preventing the binding of ATP to RT. A new class of compounds called excision inhibitors have already been proposed (Sarafianos et al., 2004). This class of inhibitors can be competitive with ATP, but non-competitive with AZTTP. The combinatorial therapy with AZT and excision inhibitor would be able to re-sensitize AZT-resistant HIV-1 RT variants to AZT. In more general, excision inhibitors would also be synergistic when combined with some other chain-terminating nucleotides, and might be useful in overcoming resistances to other NRTI drugs, such as tenofovir and d4T. By our knowledge, PPi analogs can also be regarded as one class of excision inhibitors.

13. The aims of this dissertation

The current challenge for understanding this enhanced ATP-mediated pyrophosphorolytic excision is that the ATP-binding site of RT has only been modeled but not defined experimentally. Thus the structural basis for this excision is unknown. This challenge can be approached through determining the structure of the excision reaction intermediate (AZT-resistant HIV-1 RT/$DNA_{AZTMP(N)}$/ATP) or the excision product intermediate (AZT-resistant HIV-1 RT/$DNA_{ddAMP(P)}$/AZTppppA). It is difficult to trap the excision reaction intermediate (AZT-resistant RT/$DNA_{AZTMP(N)}$/ATP) in a stable form.

One reason is that ATP has relatively weak binding affinity (mM range); another reason is that ATP can initiate pyrophosphorolytic excision. However, the excision product intermediate can be easily trapped in a stable form, if the primer strand is terminated with a dideoxynucleotide to prevent the polymerization reaction. Thus we attempted to solve the structure of excision product complex (AZT-resistant HIV-1 RT /DNA_{ddAMP}/AZTppppA) instead of the excision reaction complex (AZT-resistant HIV-1 RT/$DNA_{AZTMP(N)}$/ATP). In addition to the AZT-resistant HIV-1 RT excision product complex, we also attempted to solve wild-type HIV-1 RT excision product complex. These two structures together will help us to elucidate how AZT-resistance mutations enhance the ATP-mediated pyrophosphorolytic excision, and solve some long-standing puzzles about the AZT resistance mechanism.

We also attempted to solve structures of other HIV-1 RT intermediates containing AZT-resistance mutations. These structures include AZT-resistant HIV-1 pre-translocation complex, AZT-resistant HIV-1 post-translocation complex and unliganded AZT-resistant HIV-1 RT. These structures can be compared with each other or their previously solved wild-type correspondents. All the structures together will cover almost every step of this ATP-mediated pyrophosphorolytic excision by HIV-1 RT (Fig 10), and will be able to give a relatively complete image about the structural basis of this AZT excision mechanism.

PART II. MATERIALS & METHODS

1. Strategy for trapping HIV-1 RT intermediates in stable forms

The biggest challenge in crystallographic studies of the HIV-1 RT-catalyzed polymerization and pyrophosphorolytic excision is to trap intermediates in stable forms, because RT binds to polymerizable versus non-polymerizable sites of DNA in relatively modest specificity even in the presence of a dNTP (Huang et al., 1998). In order to trap intermediates in their stable forms, a disulfide cross-linking methodology was originally developed by Harrison and Verdine (Huang et al., 1998 & 2002). Disulfide cross-linking was chosen because of the following reasons: (1) Thiol-disulfide interchange easily under physiological conditions; (2) Formation of a disulfide bond is reversible and selective, therefore unlikely to yield an irrelevant cross-linked species through kinetic trapping, especially in the presence of a thiol-reducing agent; and (3) Efficient methods available to tether a thiol to a base of DNA, and mutate a protein residue to cysteine (Huang et al., 1998).

The RT was covalently linked to DNA by disulfide interchange between an engineered sulfhydryl group of RT and a disulfide-containing tether attached to the N^2 amino group of a deoxyguanosine in the template strand (Fig. 13). In order to achieve the best specificity, three potential mutation sites (Q258, G262, W266) of RT were evaluated (Huang et al., 1998). These sites were located on one face of α-helix H in p66. Modeling studies

showed their side chains could extend to the minor groove of the DNA template/primer duplex (Huang et al., 1998). If a cysteine residue is introduced at one of these three positions and a thiol-bearing tether is protruded from the minor groove, then a disulfide bond can be formed between protein and DNA, hence the RT/DNA complex will be stalled in a stable form (Huang et al., 1998). Of the three cysteine-engineered mutants, Q258C exhibited the greatest degree of positional selectivity in the cross-linking reaction of DNA to RT.

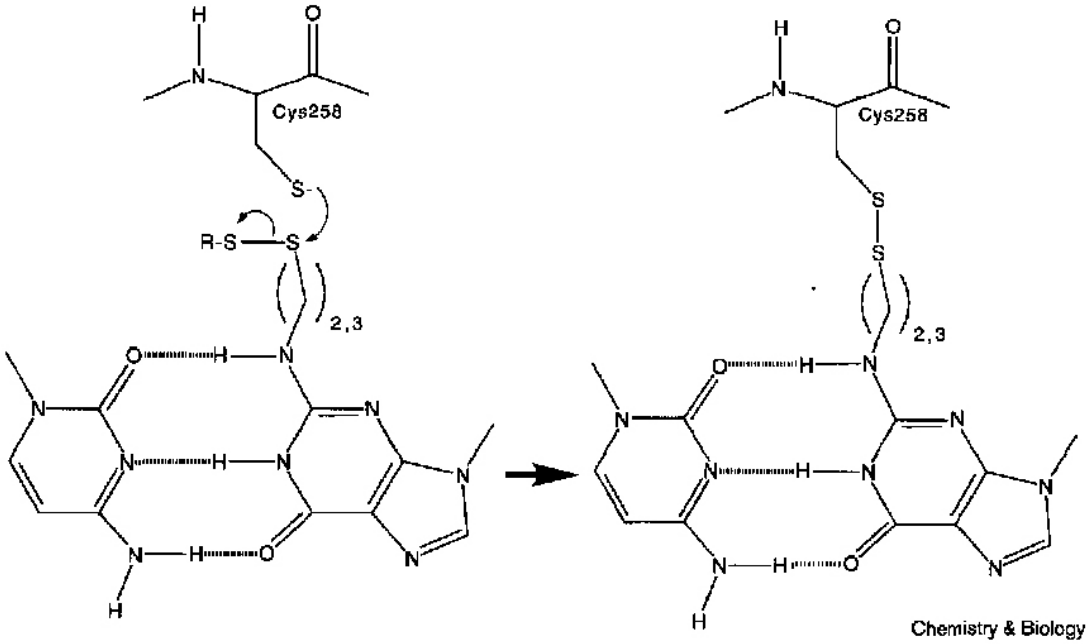


Fig. 13. RT-DNA cross-linking reaction. Disulfide bond formed between the side chain of engineered Cys258 in a p66 subunit and a thiol group tethered to N^2 of a guanine base in a primer strand (Sarafianos et al., 1999)

This strategy was modified by our laboratory (Sarafianos et al., 2002). The mutated residue was still Q258, but the thiol-bearing tether was attached to a guanosine of the primer other than the template strand. This modification brings great convenience and economic benefit. In the original strategy, we needed to chemically synthesize different modified templates with sequences complementary to the incoming dNTP or dNTP analogs. On the other hand, the un-modified primer remains the same. In the new strategy, the modified primer sequence remains the same, and the un-modified template is instead designed to have different sequences complementary to the incoming dNTP or dNTP analogs. Un-modified DNA oligonucleotides are more readily available and less costly, and sequences can be more easily changed according to the type of incoming dNTP or dNTP analogs such as AZT, 3TC, tenofovir and so on. Therefore, this new strategy allows us to more conveniently trap different intermediates involved in polymerization or pyrophosphorolysis.

Studies were also conducted to investigate the effect of linker length on the cross-linking efficiency (Sarafianos et al., 2002). It was found that either C2 or C3 tethers had the same efficiency, which suggested that there were no substantial changes in the binding of nucleic acid. The crystal structures of HIV-1 RT-DNA$_{ddA(P)}$/Fab with C2 tether (primer linked to RT with a C2 tether) and HIV-1 RT-DNA$_{ddA(P)}$/Fab with C3 tether (primer linked to RT with a C3 tether) showed the binding of nucleic acid are the same (Sarafianos et al., 2002). These data together indicated that neither the C2 nor C3 tethers constrained the

binding of DNA in a biologically irrelevant conformation. In other word, the cross-linking was catalytically relevant.

The optimal distance was also determined (Sarafianos et al., 2002). The modified G was placed at the fourth, fifth, or sixth base pair relative to the P site, and the cross-linking efficiencies were assessed. Experiments showed that the optimal distance was six base pairs away from the P site. About 50% of RT was cross-linked to DNA, which was much greater than the amount cross-linked to the other sites. This indicated positional specificity for the modified G at the primer strand. The optimal distance is the same as that when the modified G is in the template

Based on the above data, a 20-mer primer was designed. This primer has a modified G at the fifth position from the 3'-end of the primer, which allow preparation of P site complexes or excision product complexes by a single cycle of polymerization reaction and N site complexes by two cycles of polymerization reaction.

2. Expression and purification of cross-linkable HIV-1 RT

2.1 Cross-linkable AZT-resistant HIV-1 RT

The purified cross-linkable AZT-resistant RT was kindly provided by Dr. Stephen Hughes' lab. This AZT-resistant RT was produced by co-expressing the p66 and the p51 subunits in E. coli strain BL21(DE3)(pLysE), and purified by Ni-affinity and anion

exchange chromatography (Boyer et al., 2001). The p66 subunit of AZT-resistant RT contained five AZT resistance mutations (M41L/D67N/K70R/T215Y/K219Q), along with C280S and Q258C. The p51 subunit of AZT-resistant RT contained C280S mutation and a deletion at 428. Two Gly and seven His were added after the Δ428.

2.2 Cross-linkable wild-type HIV-1 RT

The expression plasmid of cross-linkable wild-type RT was engineered by Dr. Joseph Bauman in the Arnold lab. The p66 subunit of wild-type RT contained C280S and Q258C mutations, and was truncated at the residue 555. Six histidines and a HIV-1 C2 protease cleavage site were added to the N-termini of p51 subunit. The p51 subunit of wild-type RT contained C280S mutation, and was truncated at Q428. The expression plasmid was transformed into BL21(DE3)(pLysE) cells and grown overnight at 37˚C on LB-agar plates containing 35 mg/liter streptomycin and 0.1% glucose. A single colony was then inoculated into 50 ml LB containing 35 mg/liter streptomycin and 0.5% glucose, and grown overnight at 37°C with 2000 rpm shaking. The overnight culture was diluted to same fresh medium in a ratio of 1:100, and grown at 37°C with 2000 rpm shaking. When the OD600 absorption reached 0.9, the cells were induced with 1mM IPTG. After three hours of induction, the cells were harvested by centrifuge at 6000 rpm, and stored at -80°C. The cells were re-suspended to a lysate buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM imidazole), and sonicated in a ice bath until the solution became clear. The solution was then ultracentrifuged at 12000 rpm and 4°C for 20 minutes. The supernatant

were pooled together and filtered using a 0.45 μm syringe filter.

Ni-affinity purification was performed according to the manufacturer's recommendations (Qiagen). The filtered lysis supernatant were incubated with a Ni^{2+} column pre-equilibrated with the lysate buffer on a shaker at 4°C for 1 hr. The column was then washed by 10X column volume of a wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10mM imidazole). After the column was drained, the column was eluted with 3X column volume of elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 250 mM imidazole). The yield of RT was checked by OD280 (OD280/3.1 X dilution factor).

Further purification was performed using a Mono Q column. The buffer of eluted RT was exchanged into buffer A (50 mM diethanolamine pH 8.9) using an Amicon Ultra-15 Centrifugal unit with an Ultracell-30 membrane (Millipore). The RT solution was filtered with a 0.22 μm filter, and about 20 mg of RT was loaded onto a Mono Q column (Amersham Biosciences) equilibrated with buffer A. The sample was eluted during a one-hour gradient from 0 to 25% buffer B (buffer A + 1 M NaCl) with a flow rate of 4 ml/minute. The peaks containing RT were pooled together and buffer exchanged into 10 mM Tris pH 8.0 and 75 mM NaCl. The RT solution was then concentrated to 3.6 mg/ml using an Amicon Ultra-15 Centrifugal unit with an Ultracell-30 membrane (Millipore). The concentrated RT was aliquoted and stored at -80°C or 4°C for immediate use.

3. Synthesis of cross-linkable DNA primer and AZTppppA

The cross-linkable 20 mer DNA primer and AZTppppA were synthesized by Dr. Roger Jones' lab. The sequence of primer is 5'-ACAGTCCCTGTTCGGG*CGCC-3' (G* has a thioalkyl tether covalently attached to the N^2 group of the guanosine residue). The cross-linkable primer was synthesized according to a procedure as described (Hou 2006; Wang 2004). AZTppppA was synthesized by reaction of adenosine monophosphate with a trimetaphosphate derivative of AZT as described (Han et al., 2006).

4. Synthesis of DNA template

27 mer DNA templates were purchased from the DNA synthesis laboratory at the University of Medicine and Dentistry of New Jersey. Sequences of templates for HIV-1 RT excision complexes, pre-translocation complex, and post-translocation complex are shown in Table 2.

Table 2. Sequences of templates for HIV-1 RT excision complexes, pre-translocation complex and post-translocation complex	
Excision product complexes	27 mer 5'-ATGCATGGCGCCCGAACAGGGACTGTG -3'
Pre-translocation complex	27 mer 5'-ATGCATGGCGCCCGAACAGGGACTGTG -3'
Post-translocation complex	27 mer 5'-ATGCTAGGCGCCCGAACAGGGACTGTG -3'

5. Preparation of RT-DNA complex

5.1 Template-primer annealing reaction

Both primer and template were white dry solid. Same procedures were used to prepare template and primer solution at the same concentration. 2.7 mg of template or primer was dissolved in 200 μl ddH_2O. Absorptions at wavelength of 260 nm were measured using a UV spectrometer. Concentrations were calculated using the calculated extinction coefficients and adjusted to 1.6 mM. Template or primer solutions were aliquoted into 50 μl/tube (equal to 80 nmol/tube) and stored at –80 ºC.

The template-primer annealing solution was prepared as described in Table 3. 0.1 mM oligonucleotide solution containing 10 mM Tris-HCl pH8.0, 50 mM NaCl, 0.1 mM template, and 0.1 mM primer was annealed by heating in a water bath for 10 min at 85°C and then cooled overnight to ambient temperature.

Table 3. Template-primer annealing reaction	
20 mer Primer (1 mM)	80 µl
27 mer Template (1mM)	80 µl
NaCl (4 M)	20 µl
1M Tris-HCl pH 8.0	40 µl
ddH_2O	580 µl
Total Volume	800 µl

5.2 Cross-linking reaction of HIV-1 RT excision product complexes (HIV-1 RT-DNA$_{ddAMP}$/AZTppppA), pre-translocation complex (HIV-1 RT-DNA$_{AZTMP(N)}$), and post-translocation complex (HIV-1 RT-DNA$_{AZTMP(P)}$/Fab)

For the excision product complexes, the annealed substrates were reacted with RT in 3:1 molar ratios in the presence of 150 mM NaCl, 50 μM β-mercaptoethanol, 100 μM ddATP, and 50 mM Tris-HCl, pH 8.0. Polymerization was initiated by addition of 5 mM $MgCl_2$, and the samples were allowed to react two hours at room temperature. The reaction mixtures were then incubated for 3 hours or overnight at 4˚C to achieve the highest cross-linking efficiency. When needed, 2 mM AZTppppA could be added into the reaction mixture, but no matter whether the AZTppppA was added or not, the overnight cross-linking efficiencies were the same. The standard protocol of the cross-linking reaction for excision product complexes was shown in Table 4.

For the pre-translocation and post-translocation complexes, the cross-linking reaction was conducted by Dr. Sarafianos. The annealed substrates were reacted with RT in 3:1 molar ratios in the presence of 150 mM NaCl, 5 mM MgCl2 and 50 mM Tris-HCl, pH 8.0. The pre-translocation complex was prepared by two cycles of polymerization with 1 mM dATP and 1 mM AZTTP added to the cross-linking solution. The post-translocation complex was prepared by one cycle of polymerization reaction with only 1 mM AZTTP added in the cross-linking reaction. The reaction mixtures were then incubated overnight at 4˚C to achieve the highest cross-linking efficiency. The standard protocols of the

cross-linking reaction for HIV-1 RT pre-translocation complex and post-translocation complex are shown in Table 4.

Table 4. Cross-linking reaction for RT intermediates				
Components	**Excision product complex**	**Pre-translocation Complex**	**Post-translocation complex**	**Final Concentration**
HIV-1 RT (3 mg/ml)	500 μl	500 μl	500 μl	2.8 μM
27mer/20mer (0.1 mM)	400 μl	400 μl	400 μl	8 μM
$MgCl_2$ (1 M)	25 μl	25 μl	25 μl	5 mM
P site nucleotide (100 mM)	ddATP 1 μl	dATP 1 μl	AZTTP 1 μl	20 μM
N site nucleotide (100 mM)	AZTppppA or ddH2O 1 μl	AZTTP 1 μl	N/A	20 μM
pH 8.0 Tris-HCl (1 M)	250 μl	250 μl	250 μl	50 mM
NaCl (4 M)	125 μl	125 ul	125 μl	100 mM
β-mercaptoethanol (2 mM)	50 μl	N/A	N/A	2 mM
ddH_2O	3648	3698 μl	3699 μl	2 mM
Total Volume	5000 μl	5000 μl	5000 μl	

The final reaction mixture contained the desired product, RT/template-primer (HIV-1 RT cross-linked to dsDNA template-primer), together with uncross-linked RT, free template-primer, un-reacted nucleosides and $MgCl_2$. The results were analyzed using non-reducing SDS-PAGE (Fig. 14). Three bands were observed, which represented p66-template-primer, p66 and p51, respectively. RT-primer/template is at the highest molecular weight position. If samples loaded for electrophoresis were heated, there would be an extra band corresponding to p66-primer at a position between the p66-template-primer and the p66, because heating caused disruption of template-primer duplex and a portion of template was dissociated from the cross-linked protein-DNA complex. Successful cross-linking was assessed by the appearance of a band corresponding to the p66-template-primer with concomitant weakening of the band corresponding to p66. The ratio of band intensities corresponding to cross-linked *versus* uncross-linked p66 subunits was used to calculate the percent yield of the cross-linking reaction (Sarafianos et al., 2002).

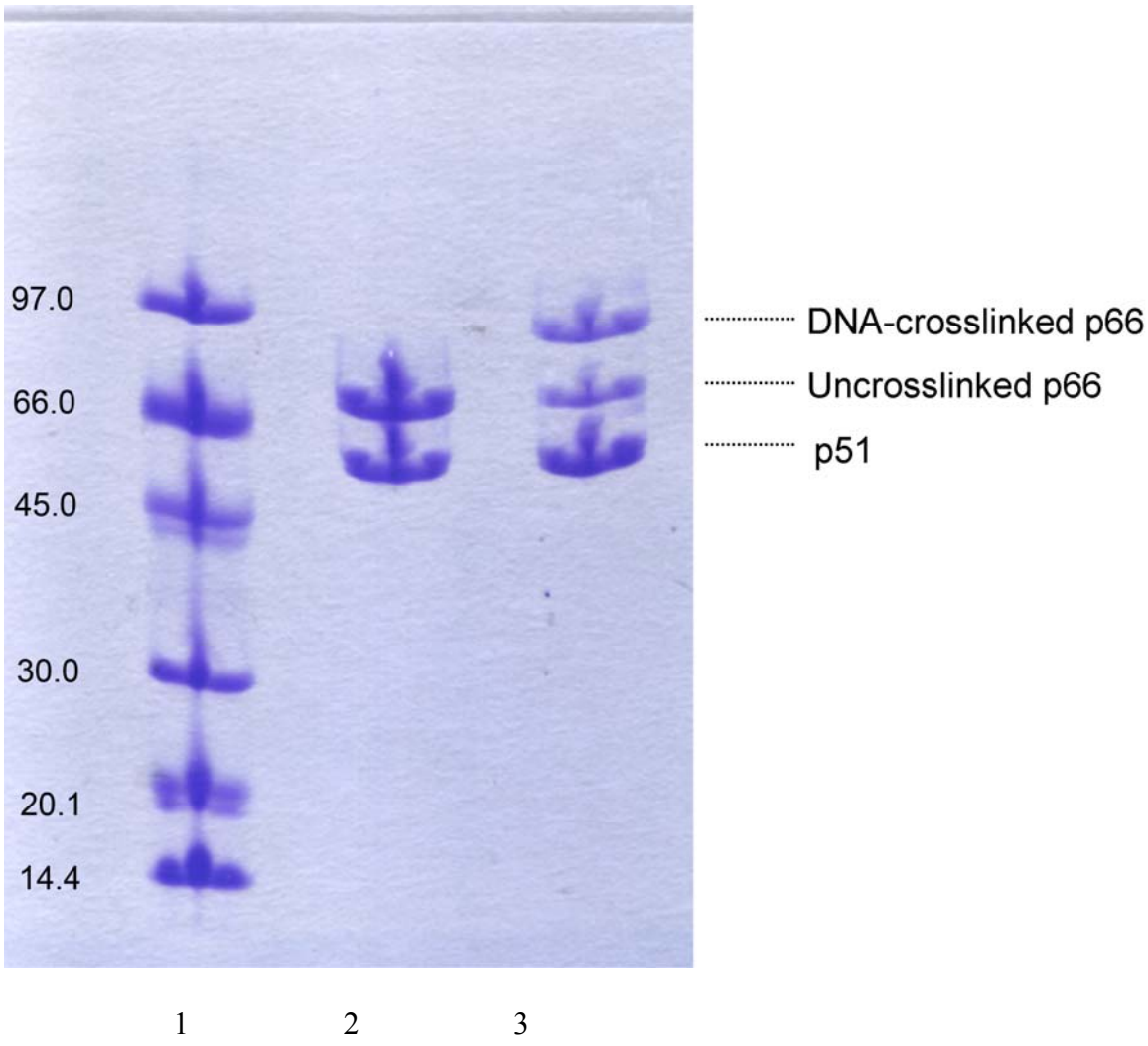


Fig. 14. Non-reducing SDS-PAGE of cross-linking reaction mixture

Lane 1: Molecular weight markers

Lane 2: HIV-1 RT

Lane 3: Cross-linking reaction mixture

6. Purification of cross-linking reaction products

The purification was performed with Ni^{2+} affinity chromatography (Fig. 15). Briefly, 5 ml of the overnight reaction mixture was loaded into a Hi-Trap IMAC column in tandem with two heparin columns. The un-reacted nucleotides or free DNA were first washed out by a buffer including 10 mM Tris-HCl, pH 8.0, 75 mM NaCl, while His-tagged free RT and DNA-crosslinked RT were bound to the IMAC column. His-tagged RTs including free and DNA-crosslinked species were then eluted by the second buffer containing 10 mM Tris-HCl, pH 8.0, 75 mM NaCl and 100 mM imidazole. The tandem-coupled heparin columns can absorb free RT, but do not absorb DNA-crosslinked RT, so the DNA-crosslinked RT was separated from the RT. The heparin-bound free RT was recovered with another buffer including 20 mM Tris-HCl pH 8.0 and 1 M NaCl. The peak fractions containing the DNA-crosslinked RT was pooled and analyzed by non-reducing 12% (w/v) PAGE (data not shown). The pooled DNA-crosslinked RT was then concentrated, and its buffer was simultaneously exchanged to a new buffer containing10 mM Tris-HCl, 0.75 mM NaCl using Centricon-30 spin columns. Protein concentration is determined by measuring absorbance at 280 nm and by interpolation of a standard curve of known cross-linked RT concentrations versus OD280.

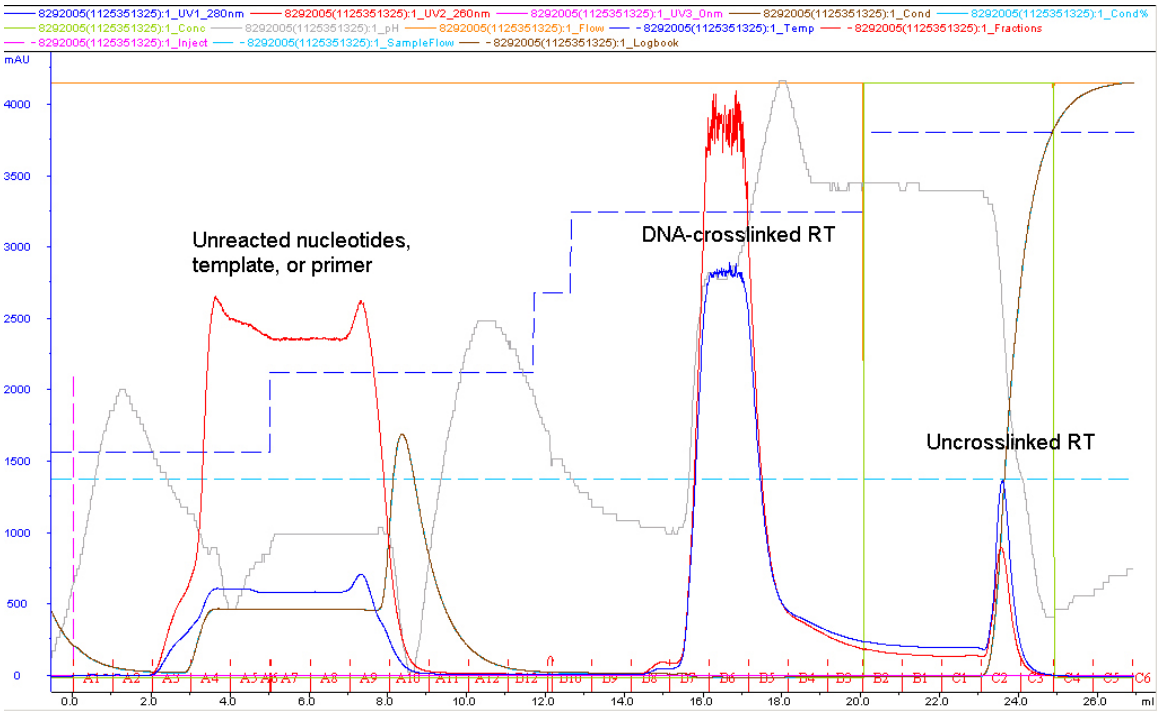


Fig. 15. Purification of HIV-1 RT-DNA

7. Crystallization of HIV-1 RT intermediates

Protein crystallization is an artificial process of formation of solid crystals from a homogenous protein solution. Protein solubility in aqueous solution depends on a combination of temperature, pH, amino acid composition, and other solution components. When the concentration of a protein solution is brought above its solubility limit, the solution becomes supersaturated (Fig. 16). The supersaturated protein can be in meta-stable, nucleation, or precipitate region, depending on its protein concentration from low to high. In the meta-stable region, no microscopic crystallite or amorphous precipitates can be formed. In the nucleation region, protein molecules can associate into microscopic crystallites. In the precipitate region, the protein molecules aggregate to amorphous precipitates. Crystallization is characterized by two steps: nucleation of microscopic crystallites (possibly having only 100 molecules), and growth of crystallites to sizes suitable for diffraction. If the protein concentration is brought to nucleation region too quickly, a shower of small crystallites will form in the droplet. If the protein concentration is brought to nucleation region slowly, only a few nuclei are formed, and they will continue to grow without spontaneous formation of new nuclei. Thus, the goal of crystallization condition screening is to identify solution conditions that can bring protein concentration to a point only slightly above its saturation point as slowly as possible.

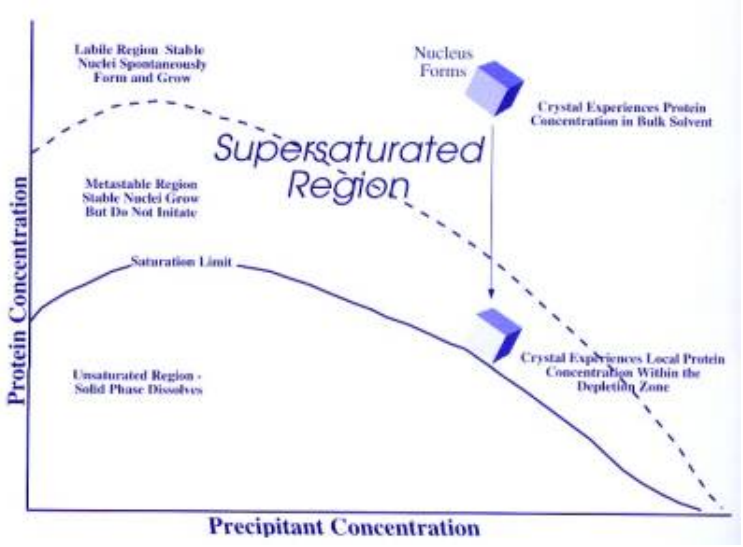


Fig. 16. The phase diagram for protein crystallization (from McPherson, 1999)

7.1 Crystallization of wild-type and AZT-resistant HIV-1 RT excision product complexes

The initial crystallization samples contained 8-10 mg/ml RT-DNA, 75 mM NaCl, 5 mM dinucleoside tetraphosphate, and 10 mM Tris-HCl pH 8.0. Crystals were grown at 4˚C by hanging-drop vapor diffusion method. Hanging drops were prepared by mixing 2 μl of a crystallization sample with 2 μl of crystallization solution. Based on our previous results, two crystallization conditions were selected as starting conditions. The first condition was the one used for crystallizing the HIV-1 RT-DNA/tenofovir ternary complex. This condition contained 50 mM Bis-Tris propane pH5.6, 10-11% PEG8000, 0.3 M $(NH4)_2SO_4$, 5% glycerol, 5% sucrose, and 20 mM Mg^{2+}. The second condition was modified from the one used for crystallizing HIV-1 RT-DNA/ddTTP ternary complex, and contained 50 mM Bis-Tris propane pH 6.8, 13% PEG6000, 1 M NaCl, 5% glycerol, 5% sucrose, and 20 mM Mg^{2+}. Two-dimensional grid searches (PEG concentration versus pH) were designed to optimize these two conditions. The PEG concentrations ranged from 8% to 14% with 1% increments. The pH values ranged from 5.6 to 7.2 with 0.2 increments. The results showed that pH could affect crystal size and number of crystals per drop. It was found that pH 6.8 resulted in relatively bigger crystals and fewer crystals per drop in both conditions. Furthermore, we also found that the crystallization behavior of this complex was affected by the PEG8000 concentration. In particular, when the concentration of PEG8000 was above 13%, sample solutions precipitate immediately after mixing with precipitant solution; however, when the concentration of PEG8000 was

below 9%, the drops were clear even after a few months. Both crystallization conditions had low rates of crystal formation, and the second condition had even lower than the first one. Only a low proportion of drops had crystals, and most of them were still clear even after one month. These problems were solved by microseeding and additives screening. Microseeding can improve crystal formation rate, but results in a crystallite shower in the seeded drop. Additives screening showed that spermine and spermidine could significantly improve the rate of crystal formation. Since addition of additives is handier than microseeding, we prefer adding spermine or spemidine into the crystallization condition.

Crystals of both complexes appear as thin-plate clusters (Fig. 17 and Fig. 18). Because crystals are thin and fragile, they are sensitive to handling. Two dimensions can be more than 600 μm, but a third dimension could only reach 10 μm. Crystals usually appeared after two weeks. It was good to harvest them within one month, once they grew to suitable sizes. Older crystal plates tended to stack on each other. Crystals used to collect full datasets were grown in a solution containing 50 mM Bis-Tris propane pH5.6, 10-11% PEG8000, 0.3 M $(NH4)_2SO_4$, 5% glycerol, 5% sucrose, 20 mM Mg^{2+} and 5 mM spermidine. Crystals were very carefully separated, and temporarily stabilized in a holding buffer for 5 min (the reservoir solution with PEG8000 concentration elevated to 12%). Crystals were then soaked with cryo-protecting solution (holding buffer containing 20% (v/v) glycerol) for about 30 s, and flash-cooled in liquid nitrogen. The

purpose of flash-cooling crystals is to prevent radiation damage, since protein crystals are exposed to synchrotron X-rays during data collection.

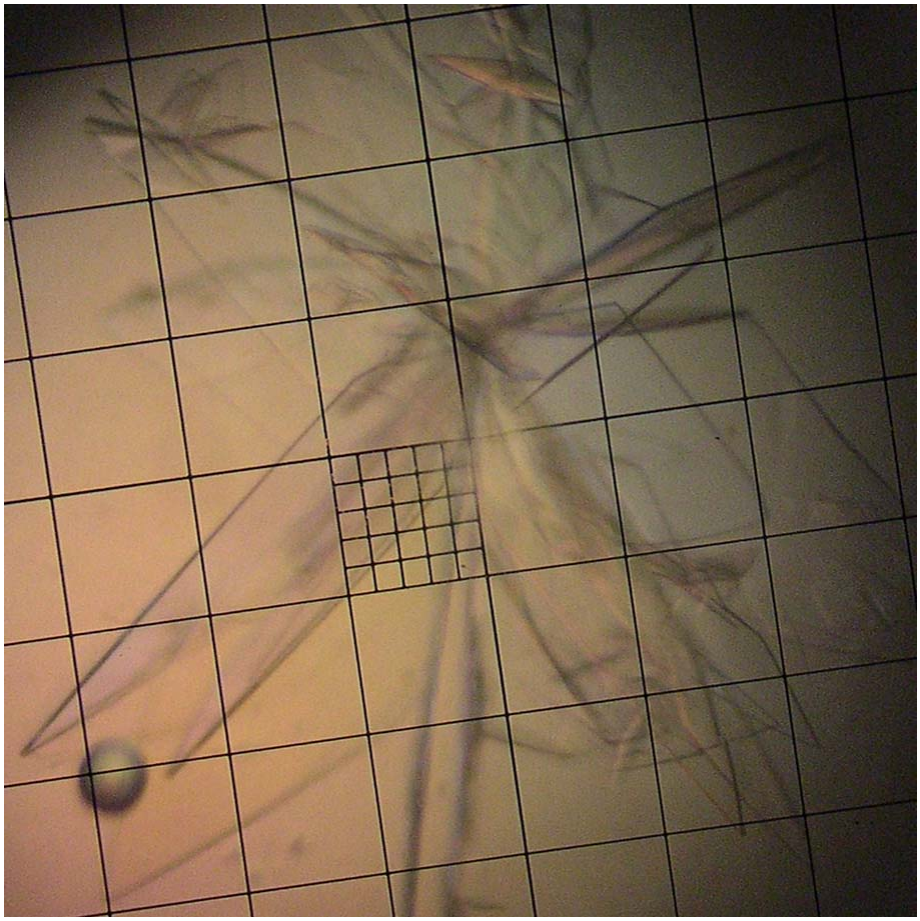


Fig. 17. Crystals of wild-type HIV-1 RT excision product complex

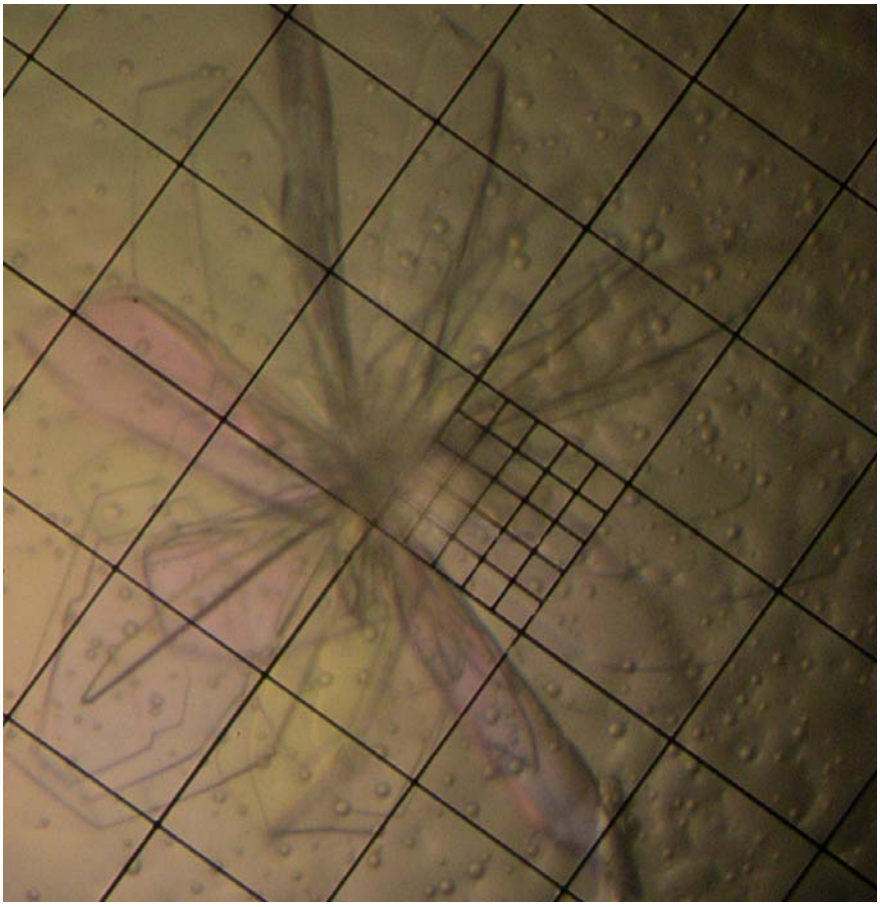


Fig. 18. Crystals of AZT-resistant HIV-1 RT excision product complex

7.2 Crystallization of AZT-resistant HIV-1 pre-translocation and post-translocation complexes

Crystallization of AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes were performed by Dr. Sarafianos and Yulia Frenkel in the Arnold lab. Unlike the wild-type HIV-1 RT pre-translocation complex (Sarafianos et al., 2002), this AZT-resistant HIV-1 pre-translocation complex was crystallized without an Fab antibody fragment. The initial crystallization sample contained ~8 mg/ml RT-DNA, 75 mM NaCl, and 10 mM Tris-HCl pH 8.0. Crystals were grown at 4°C using hanging-drop diffusion method. The hanging drops were prepared by mixing 2 μl crystallization sample and 2 μl crystallization solution (50 mM Bis-Tris propane pH 5.6, 10% PEG8000, 0.3 M $(NH4)_2SO_4$, 5% glycerol, 5% sucrose, and 20 mM Mg^{2+}). Before being flash-cooled in liquid nitrogen, crystals were first stabilized in a holding buffer (the reservoir solution with PEG8000 concentration elevated to 12%). Crystals were then soaked with cryo-protecting solution (holding buffer containing 20% (v/v) glycerol) for about 30 s.

Similar to the previously solved crystallographic structure of wild-type HIV-1 RT post-translocation complex, this AZT-resistant HIV-1 RT post-translocation complex was crystallized with Fab antibody fragment. The crystallization procedure was as described (Sarafianos et al., 2002). The initial crystallization sample (~20 mg/ml RT-DNA, 75 mM NaCl, and 10 mM Tris-HCl pH 8) was mixed with Fab28 antibody fragment in 1:0.8 mass ratios. Crystals were grown at 4°C using hanging-drop diffusion method. The hanging

drops were prepared by mixing 2 μl sample and 2 μl crystallization solutions (100 mM cacodylate pH 6.0, 31-34% saturated ammonium sulfate). Before being flash-cooled in liquid nitrogen, crystals were soaked in solutions containing 37% saturated ammonium sulfate, 0-10% glucose (w/v) and 0-20% glycerol (v/v), reaching the final concentration in three steps of 5 minutes each.

8. Crystallographic data collection and processing

8.1 Crystallographic data collection and processing of wild-type and AZT-resistant HIV-1 RT excision product complexes

Diffraction data of wild-type HIV-1 RT excision product complex was collected at the Cornell High Energy Synchrotron Source (CHESS) beam line A1 with a wavelength of 0.976 Å (Fig. 19). The exposure time for this wild-type excision product complex crystal was 45 sec. The oscillation angle range for this complex is 1.0˚ around the phi axis. The crystal-to-detector distance was 310.0 mm.

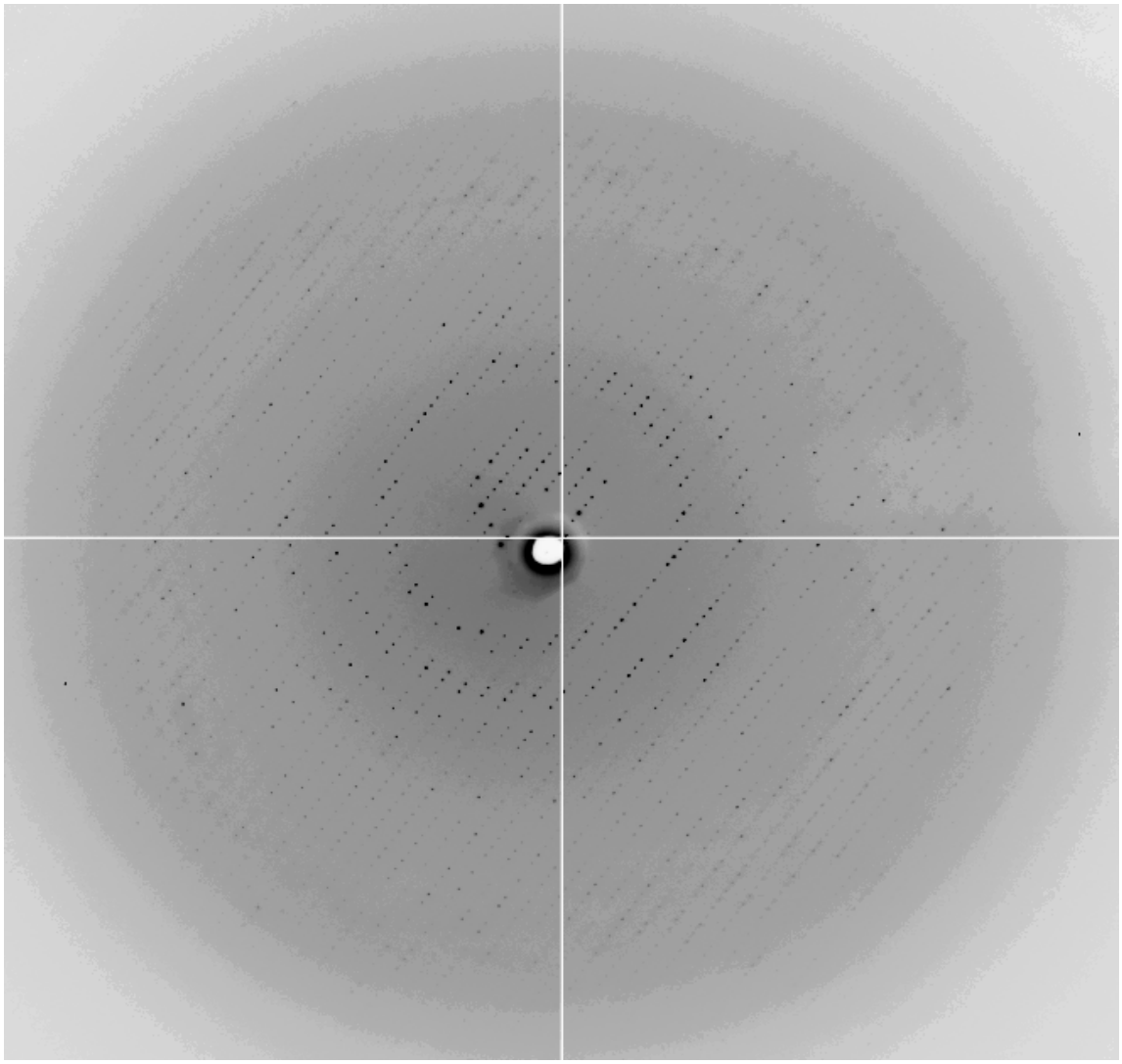


Fig.19. Diffraction of wild-type HIV-1 RT excision product complex

Diffraction data of AZT-resistant RT excision product complex was collected at the Advanced Photon Source (APS) beam line 19ID with a wavelength of 0.979 Å (Fig. 20). The exposure time for this AZT-resistant excision product complex crystal was 30 sec. The oscillation angle range for this complex was 1.0˚ around the omega axis. The crystal-to-detector distance is 451.0 mm.

The data indexing, integration and scale were performed using the HKL 2000 suite (Otwinowski et al., 1997). The analyses of systematic absence of axial reflections indicated that the crystallographic 2-fold axis is actually a 2-fold screw axis. The sigma cutoff used in scaling is -0.5. The final data processing results are shown in Table 5 and Table 6. Both crystals share the same space group of $P2_1$ and very similar corresponding cell parameters. Intensities are converted to magnitude by the single square procedure in CCP4 suite.

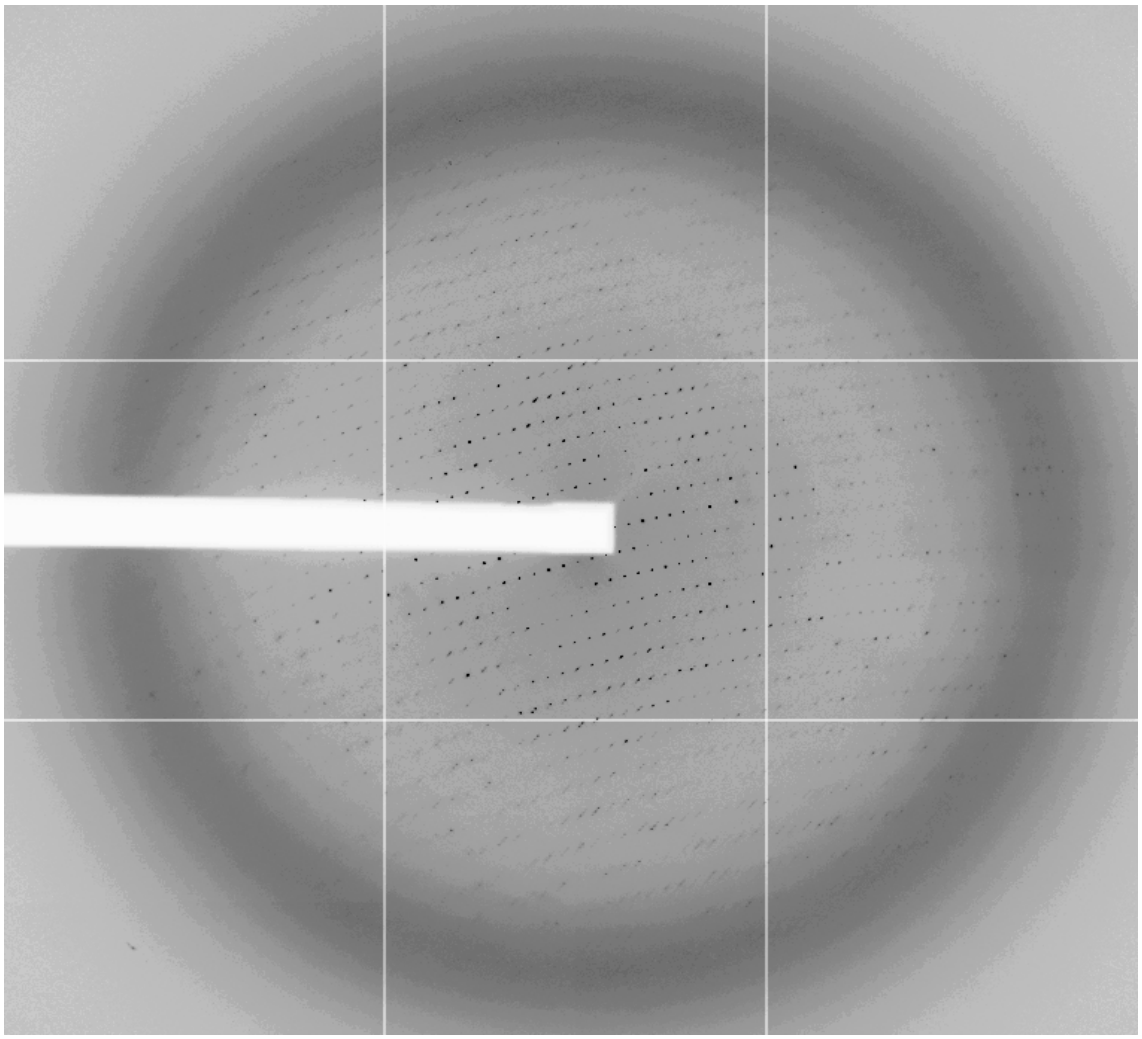


Fig.20. Diffraction of AZT-resistant HIV-1 RT excision product complex

Table 5 Summary of wild-type excision product complex data statistics

Resolution Shell (Å)	Average I/σ	Chi^2	Linear R-factor	Square R-factor	Completeness (%)	Average Redundancy
25.00-6.64	27.5	1.039	0.06	0.064	99.8	6.7
6.64-5.29	15.2	1.096	0.11	0.115	99.0	6.0
5.29-4.62	13.1	1.076	0.12	0.126	97.7	5.5
4.62-4.20	10.7	1.093	0.14	0.150	94.8	5.2
4.20-3.90	8.9	1.060	0.18	0.198	90.0	5.2
3.90-3.67	7.1	1.055	0.24	0.268	85.7	5.2
3.67-3.49	5.1	1.058	0.32	0.406	84.5	5.1
3.49-3.34	3.6	0.844	0.38	0.514	84.4	4.9
3.34-3.21	2.7	0.716	0.44	0.679	83.8	4.8
3.21-3.10	2.0	0.637	0.54	0.961	83.5	4.6
25.00-3.10	11.1	0.988	0.13	0.101	90.3	5.3

R linear = $\sum|I - \langle I \rangle| / \sum(I)$

R square = $\sum(I - \langle I \rangle)2 / \sum(I2)$

Chi2 = $\sum((I - \langle I \rangle)2 / (\sigma 2 * N / (N-1)))$

In all sums single measurements are excluded

Table 6 Summary of AZT-resistant excision product complex data statistics

Resolution Shell (Å)	Average I/σ	Chi^2	Linear R-factor	Square R-factor	Completeness (%)	Average Redundancy
25.00- 6.85	31.52	1.636	0.052	0.056	99.6	4.1
6.85-5.46	17.18	1.482	0.094	0.098	99.5	3.8
5.46-4.77	14.53	1.414	0.105	0.108	99.2	3.7
4.77-4.34	11.08	1.320	0.127	0.132	98.9	3.5
4.34-4.03	7.57	1.149	0.161	0.168	98.3	3.4
4.03-3.79	4.76	1.068	0.228	0.248	98.2	3.2
3.79-3.60	3.06	0.954	0.327	0.416	97.9	3.1
3.60-3.45	2.39	0.904	0.395	0.476	97.6	2.9
3.45-3.31	1.91	0.831	0.451	0.622	96.7	2.8
3.31-3.20	1.53	0.780	0.517	0.769	95.1	2.6
25.00-3.20	8.71	1.213	0.122	0.088	98.1	3.3

R linear = $\sum|I - \langle I \rangle| / \sum(I)$

R square = $\sum(I - \langle I \rangle)2 / \sum(I2)$

Chi2 = $\sum((I - \langle I \rangle)2 / (\sigma 2 * N / (N-1)))$

In all sums single measurements are excluded

8.2 Crystallographic data collection and processing of AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes

Diffraction datasets of AZT-resistant HIV-1 RT pre-translocation (Fig. 21) and post-translocation complex (Fig. 22) were collected at the CHESS beam line F1 with wavelengths of 1.000 Å and 0.950 Å, respectively. The exposure times for the pre-translocation and post-translocation complexes were 120 sec and 60 sec, respectively. The oscillation angle ranges for the two complexes are 1.8˚ around phi axis and 1.0˚ around omega axis, respectively. The distances of crystal-to-detector are 300.0 mm and 220.0 mm, respectively.

The data indexing, integration and scale were performed using the HKL 2000 suite (Otwinowski et al., 1997). Summaries of the data statistics for the AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes are given in Table 7 and Table 8, respectively.

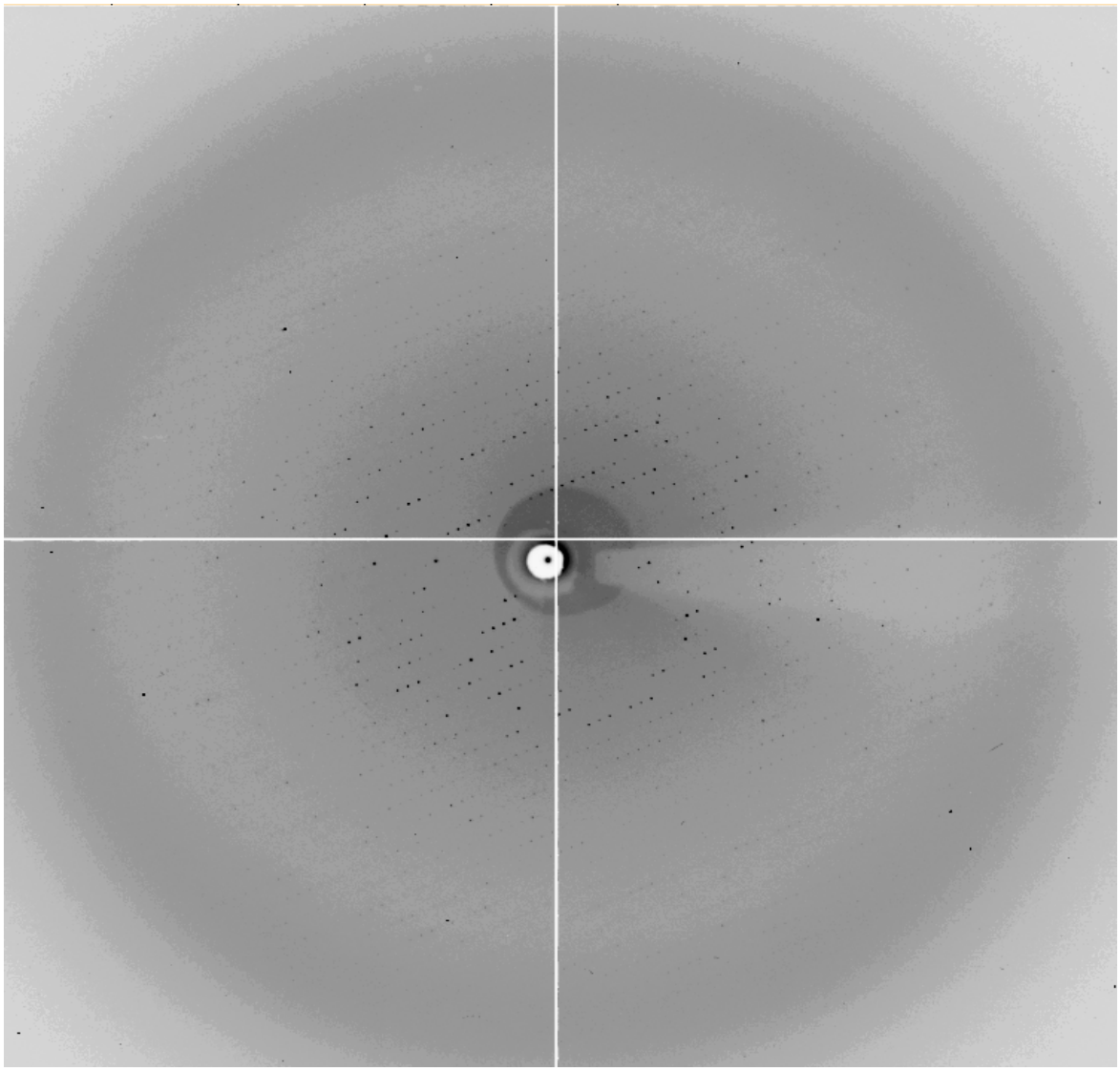


Fig. 21. Diffraction of AZT-resistant HIV-1 RT pre-translocation complex

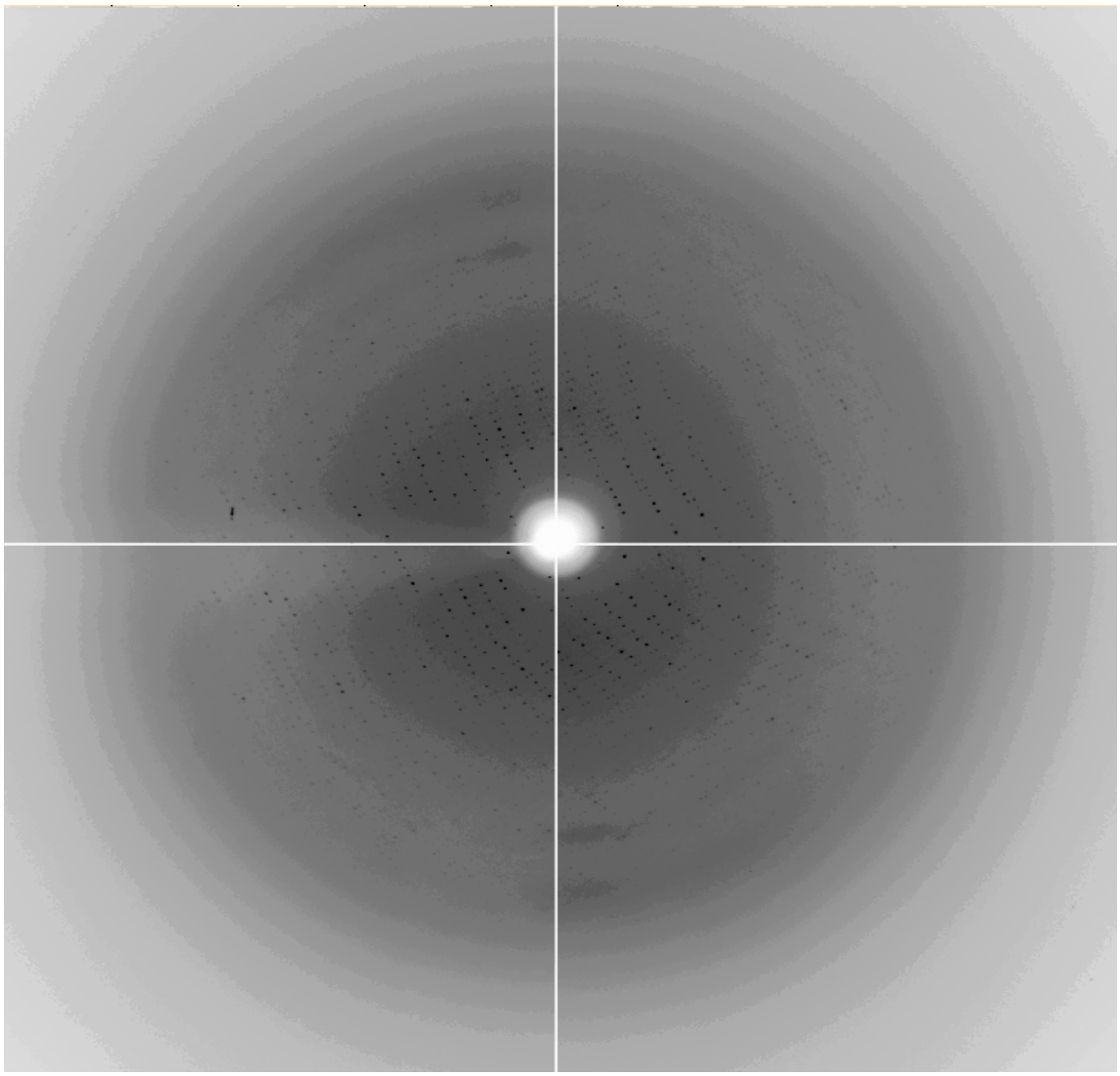


Fig. 22. Diffraction of AZT-resistant HIV-1 RT post-translocation complex

Autoindexing the images of pre-translocation complex gave a clear solution for the Bravais lattice type: primitive orthorhombic. The degree of distortion required for the cell parameters to comply with the primitive orthorhombic Bravais lattice is 0.3 %, which is reasonably low. The crystallographic space group was assigned to $P2_12_12_1$, based on the following two factors. One is the axis reflection presence conditions along h00, 0k0, and 00l, where h,k,l=2n, which is characteristic of a 2_1 screw axis. The other is the knowledge from the previously solved HIV-1 RT-DNA/dTTP catalytic ternary complex structure that is crystallized in the space group of $P2_12_12_1$ with similar cell parameters to the pre-translocation complex. The post-refined unit cell parameters are a= 79.1, b=154.4, c=277.9, α=90.0, β=90.0, and γ=90.0. According to the cell parameters and the molecular weight of the complex (~125 kDa), the asymmetric unit contains two complexes. The Matthews coefficient (V_M) is 3.39 $Å^3$/Dalton (Matthews, 1989). Assuming protein density is 1.34, the solvent content is 63%.

Autoindexing the images of the post-translocation complex gave a clear solution for the Bravais lattice type: primitive hexagonal. The degree of distortion required for the cell parameters to comply with the primitive orthorhombic Bravais lattice is 0.31 %. The crystallographic space group was assigned to $P3_212$, based on the axis reflection condition along 00l, where l=3n, which is characteristic of a 3_2 or 3_1 screw axis, and previous knowledge about this type of HIV-1 RT-DNA/Fab or HIV-1/DNA/Fab crystal which had nearly the same cell parameters. The post-refined unit cell parameters are a=166.5,

b=166.5, c=220.8, α=90.0˚, β=90.0˚, and γ=120.0˚. The asymmetric unit contains one complex with a molecular mass of ~180 kDa. The Matthew's coefficient (V_M) is 4.91 $Å^3$/Dalton (Matthews, 1989). Assuming protein density is 1.34, the solvent is 75%.

Table 7. Summary of AZT-resistant HIV-1 RT post-translocation complex data statistics

Resolution Shell (Å)	Average I	Average σ	Average I/σ	Chi^2	Linear R-factor	Square R-factor	Completeness (%)
40.00 7.78	840.1	35.6	23.6	0.892	0.062	0.067	99.8
7.78 6.18	205.8	13.9	14.8	1.075	0.121	0.116	99.8
6.18 5.40	114.5	13.2	8.7	1.153	0.215	0.202	97.5
5.40 4.91	111.3	14.7	7.6	1.203	0.255	0.239	93.8
4.91 4.56	95.2	16.0	6.0	1.191	0.330	0.299	89.7
4.56 4.29	83.3	18.7	4.5	1.150	0.380	0.352	87.9
4.29 4.08	101.2	25.2	4.0	1.088	0.484	0.445	82.1
4.08 3.90	46.9	23.8	2.0	1.039	0.625	0.611	74.4
3.90 3.75	38.1	26.0	1.5	1.027	0.671	0.654	62.9
3.75 3.62	27.3	27.8	1.0	0.993	0.786	0.836	49.3
All reflections	194.0	21.0	9.2	1.085	0.137	0.089	83.9

Table 8. Summary of AZT-resistant HIV-1 RT post-translocation complex data statistics

Resolution Shell (Å)	Average I	Average σ	Average I/σ	Chi^2	Linear R-factor	Square R-factor	Completeness (%)
40.00-6.24	3471.3	130.8	26.5	0.633	0.040	0.042	95.0
6.24-4.96	1476.7	91.2	16.2	0.889	0.090	0.089	96.5
4.96-4.33	1599.0	112.1	14.3	1.042	0.107	0.106	96.6
4.33-3.94	1180.2	119.2	9.9	1.114	0.150	0.146	96.2
3.94-3.65	812.4	127.5	6.4	1.173	0.231	0.242	96.0
3.65-3.44	594.6	144.2	4.1	1.135	0.315	0.350	96.0
3.44-3.27	468.1	164.6	2.8	1.054	0.395	0.451	95.2
3.27-3.12	380.2	184.4	2.1	0.938	0.462	0.574	93.1
3.12-3.00	342.1	203.0	1.7	0.884	0.520	0.650	90.5
3.00-2.90	316.1	215.1	1.5	0.837	0.558	0.711	87.1
All reflections	1085.1	147.9	7.3	0.966	0.127	0.096	94.2

8.3 Crystallographic data collection and processing of unliganded AZT-resistant RT

Diffraction datasets of the unliganded AZT-resistant HIV-1 RT was collected at the CHESS beam line F1 (Fig. 23) with a wavelength of 1.000 Å. The exposure time was 60 sec. The oscillation angle range is 1.5˚ around omega axis. The crystal-to-detector distance is 200.0 mm.

The data indexing, integration, and scaling were performed using the HKL 2000 suite (Otwinowski, et al., 1997). A summary of the data statistics is given in Table 9. Autoindexing the images of pre-translocation complex gave a clear solution for the Bravais lattice type: primitive monoclinic. The degree of distortion required for the cell parameters to comply with the C-centered monoclinic Bravais lattice is 0.5 %. The crystallographic space group was assigned to *C*2. The post-refined unit cell dimensions are a= 237.1, b=71.2, c=94.6, α=90.0, β=106.0, and γ=90.0. The asymmetric unit contains one complex with a molecular mass of ~114 kDa. The Matthews coefficient (V_M) is 3.36 $Å^3$/Dalton (Matthews, 1989). Assuming protein density is 1.34, the solvent content is 63%.

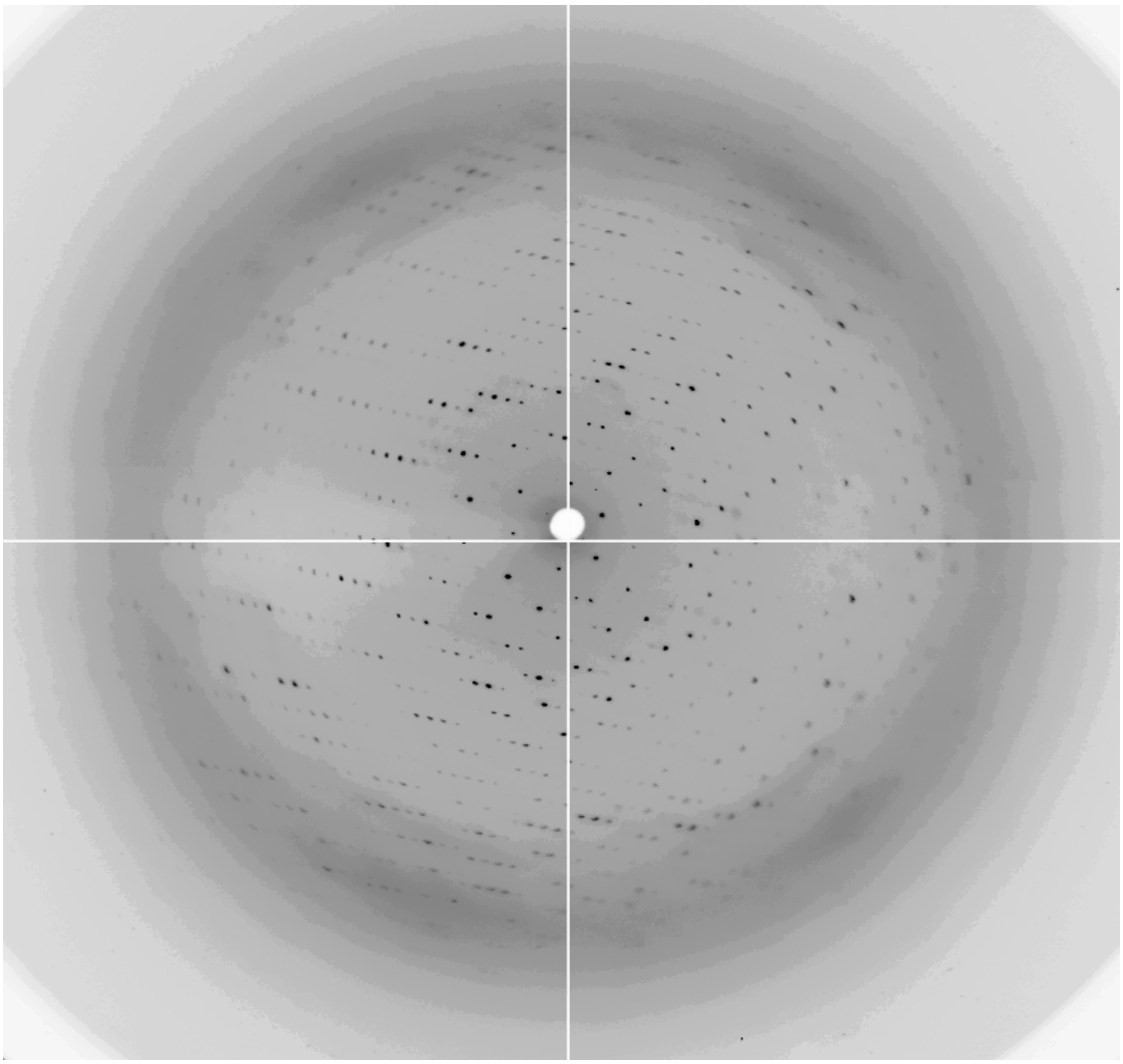


Fig.23. Diffraction of unliganded AZT-resistant HIV-1 RT

Table 9 Summary of unliganded AZT-resistant HIV-1 RT data statistics

Resolution Shell (Å)	Average I	Average σ	Average I/σ	Chi^2	Linear R-factor	Square R-factor	Completeness (%)
30.00-5.59	2146.7	63.4	33.9	1.076	0.033	0.039	93.7
5.59-4.44	1156.9	37.6	30.8	1.142	0.041	0.045	97.7
4.44-3.88	782.1	35.3	22.2	1.044	0.051	0.052	96.7
3.88-3.53	377.8	23.8	15.9	1.086	0.074	0.069	98.8
3.53-3.28	189.3	16.3	11.6	1.162	0.104	0.096	97.6
3.28-3.08	94.2	13.3	7.1	1.178	0.156	0.141	96.1
3.08-2.93	53.7	12.0	4.5	1.208	0.228	0.209	94.5
2.93-2.80	37.4	12.0	3.1	1.165	0.296	0.305	93.7
2.80-2.69	28.9	13.0	2.2	1.211	0.367	0.394	88.8
2.69-2.60	28.5	14.9	1.9	1.301	0.378	0.420	73.4
All reflections	509.1	24.6	20.7	1.134	0.051	0.044	93.1

9. Ellipsoidal truncation and anisotropic scaling of reflections

Diffraction anisotropy usually occurs when one cell direction has fewer lattice contacts than another. The diffraction anisotropy can severely stall the R-factor improvement by two means (Strong et al., 2006). First, a considerable proportion of poorly-measured weak reflections are included in the refinement process, which results in a high R factor. Inclusion of these weak reflections is a consequence of the shortcomings of currently available data processing programs such as Denzo and Mosflm, since they only implement spherical resolution cutoff. If one wishes to include the strong reflections in the high resolution shell, one is inevitably compelled to accept the poorly measured weak reflections bounded by the same spherical shell as well. If one discard the high resolution data, then one lose the details of the electron density map. Since the intensity contours of the anisotropic diffraction pattern is ellipsoidal, one would clearly like to impose an ellipsoidal resolution boundary on the data, rather than a conventional spherical boundary. Reflections bounded by ellipsoidal shells have the similar I/σ and R_{sym}, and the weak reflections falling outside the ellipsoidal boundary are removed from the dataset and from the refinement process. This procedure is called ellipsoidal truncation.

Second, diffraction anisotropy causes poor agreement between F_{obs} and F_{calc}, and degrades the appearance of electron density maps. $|F_{obs}|$ is directionally dependent; however the $|F_{calc}|$ is not. It results in poor agreement between F_{obs} and F_{calc}. In order to make them comparable, the anisotropic scaling algorithms in REFMAC and CNS

eliminate anisotropy by applying an appropriate scale factor to $|F_{obs}|$. The anisotropic scale factor has three principal components, β_{11}, β_{12}, and β_{13}, acting as B-factors along a*, b* and c* directions, respectively. The magnitudes of the reflections in the weakly diffracting direction are scaled up, while the magnitudes of the reflections in the strongly diffracting direction are scaled down. The adverse effect of this procedure is that the high resolution reflections in the strongly diffracting direction of the crystal are diminished to the point that they contribute very little to the electron density map, resulting in a map that appears to be low resolution. The solution is to apply a negative isotropic B-factor to the dataset to restore the magnitude of the high resolution reflections to their original values. As a result, the electron density maps can be greatly improved in details.

9.1 Ellipsoidal truncation and anisotropic scaling of diffraction datasets of wild-type and AZT-resistant HIV-1 RT excision product complexes

Both crystals were observed to have diffraction anisotropy, so their reduced datasets were subjected to ellipsoidal truncation and anisotropic scaling (Strong et al., 2006). Diffraction anisotropy was evidenced as a directional dependence in diffraction quality.

The dataset of wild-type excision product complex was mildly anisotropic (Fig. 24). The recommended resolution limits are 3.1 Å along all the reciprocal cell directions a*, b*, c*, so no reflections were truncated and all the reflections are remains after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the

dataset. Lastly, an isotropic B of -4.38 Å 2 was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling. The pseudo precession images (Fig. 24) illustrate the individual steps.

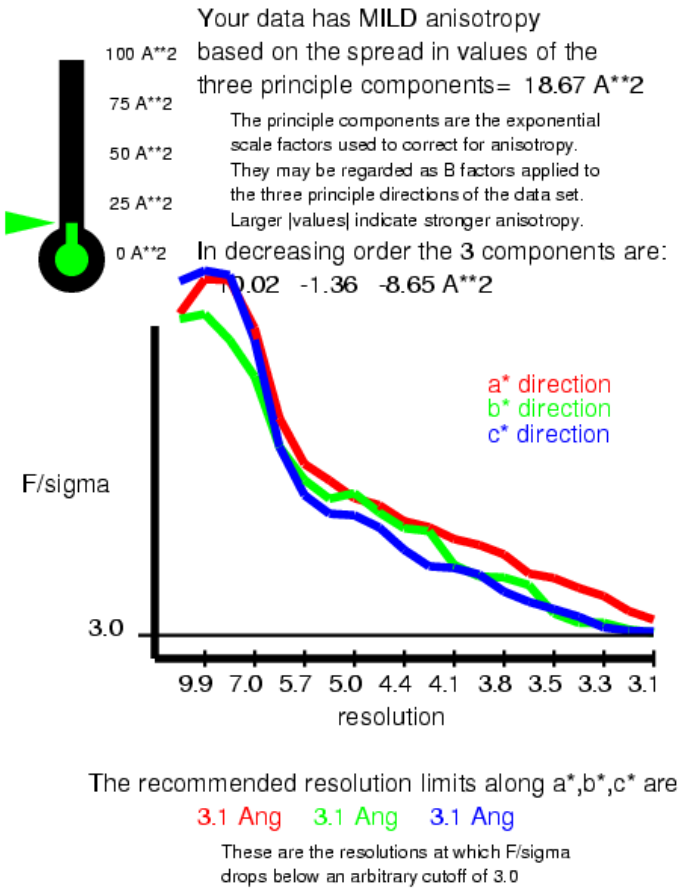


Fig. 24. Mild anisotropy in the diffraction data set of wild-type excision product complex

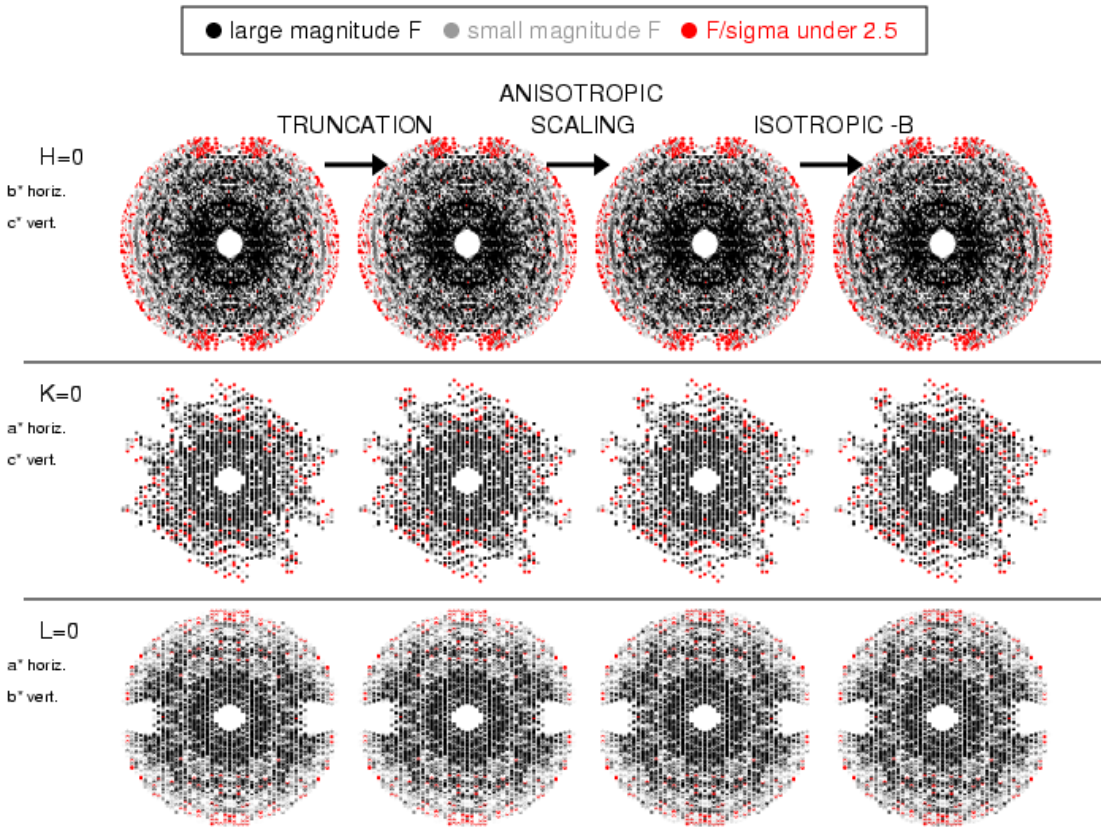


Fig. 25. Ellipsoidal truncation and anisotropic scaling of diffraction datasets of wild-type HIV-1 RT excision product complex

The dataset of the AZT-resistant excision product complex was strongly anisotropic, with diffraction limits of 3.2 Å along a*, 3.4 Å along the b* and 3.3 Å along the c* (Fig. 26). For this reason, data that fell outside the specified ellipsoid with dimensions 1/3.2, 1/3.4, 1/3.3 Å-1 along a*, b*, c*, respectively, were truncated. 111,124 reflections were in the initial dataset. 10,238 were discarded because they fell outside the specified ellipsoid. 100,886 reflections remain after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the data set. Lastly, an isotropic B of -7.86 Å 2 was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling. The pseudo precession images (Fig. 26) illustrate the individual steps.

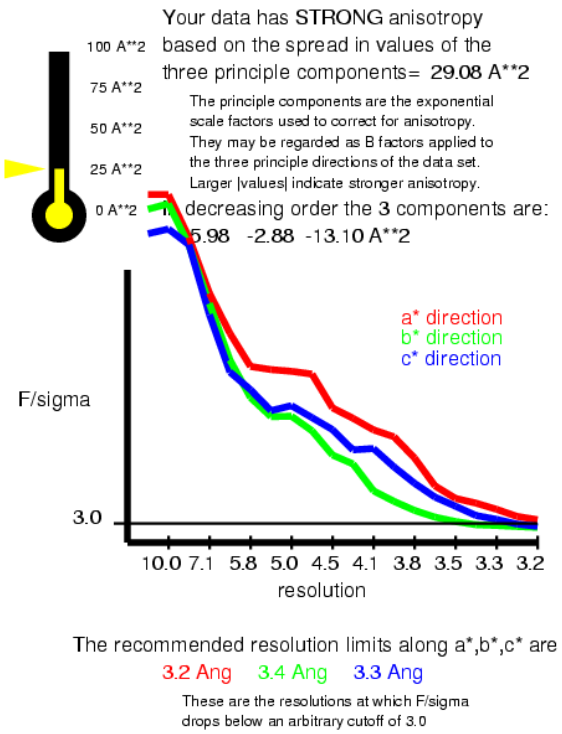


Fig. 26. Strong anisotropy in the diffraction data set of AZT-resistant HIV-1 RT excision product complex

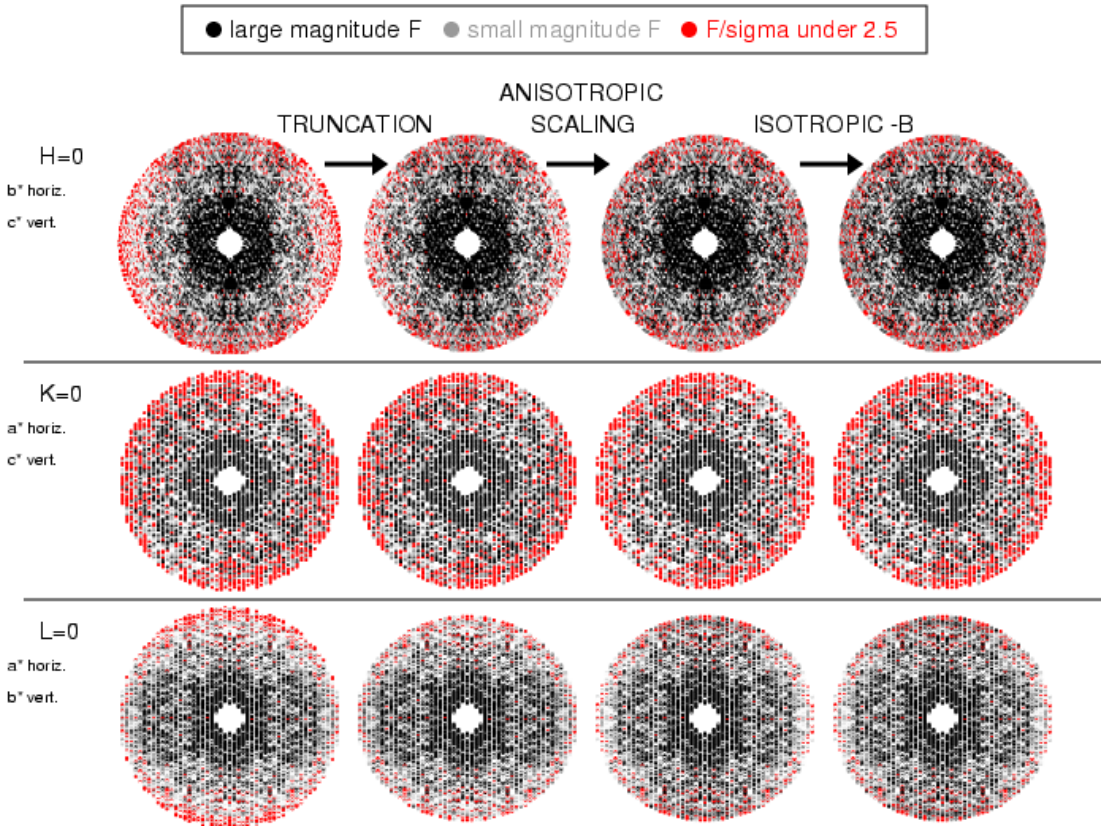


Fig. 27. Ellipsoidal truncation and anisotropic scaling of diffraction datasets of AZT-resistant HIV-1 RT excision product complex

9.2 Ellipsoidal truncation and anisotropic scaling of diffraction datasets of AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes

The AZT-resistant pre-translocation complex dataset has severe anisotropy with diffraction resolution limits of 3.7 Å along a*, 3.7 Å along the b* and 4.5 Å along the c*. (Fig. 28). There were 32,217 reflections in the initial data set. 4,931 reflections were discarded because they fell outside the specified ellipsoid with dimension 1/3.7, 1/3.7, 1/4.5 $Å^{-1}$ along a*, b*, c*, respectively. 27,286 reflections remain after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the data set. Lastly, an isotropic B of -16.24 $Å^2$ was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling. The pseudo precession images (Fig. 29) illustrate the individual steps.

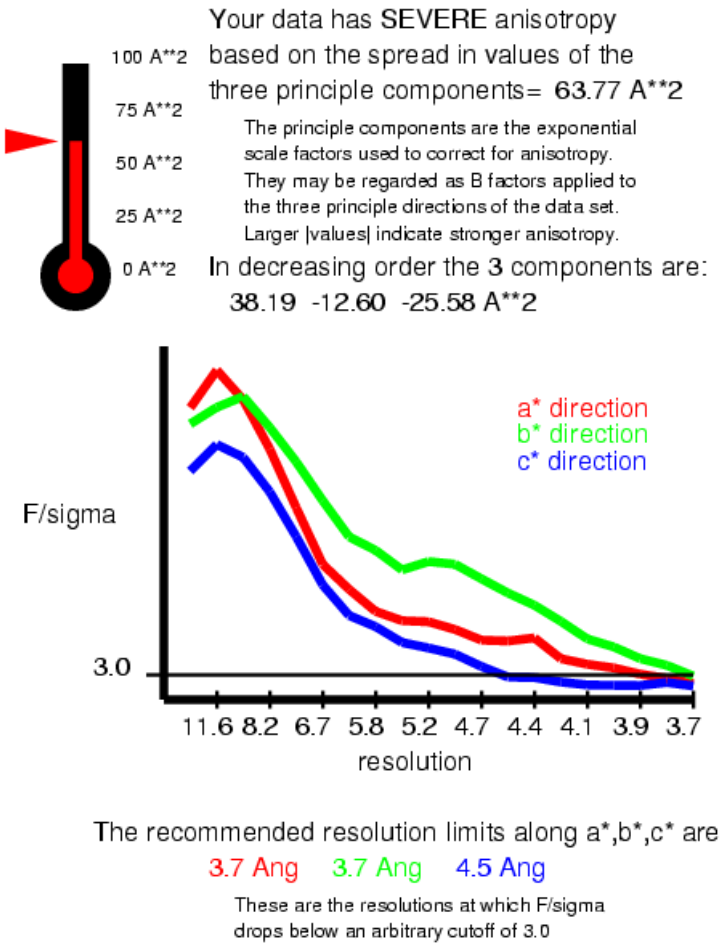


Fig. 28. Severe anisotropy in the diffraction data set of AZT-resistant HIV-1 RT excision product complex

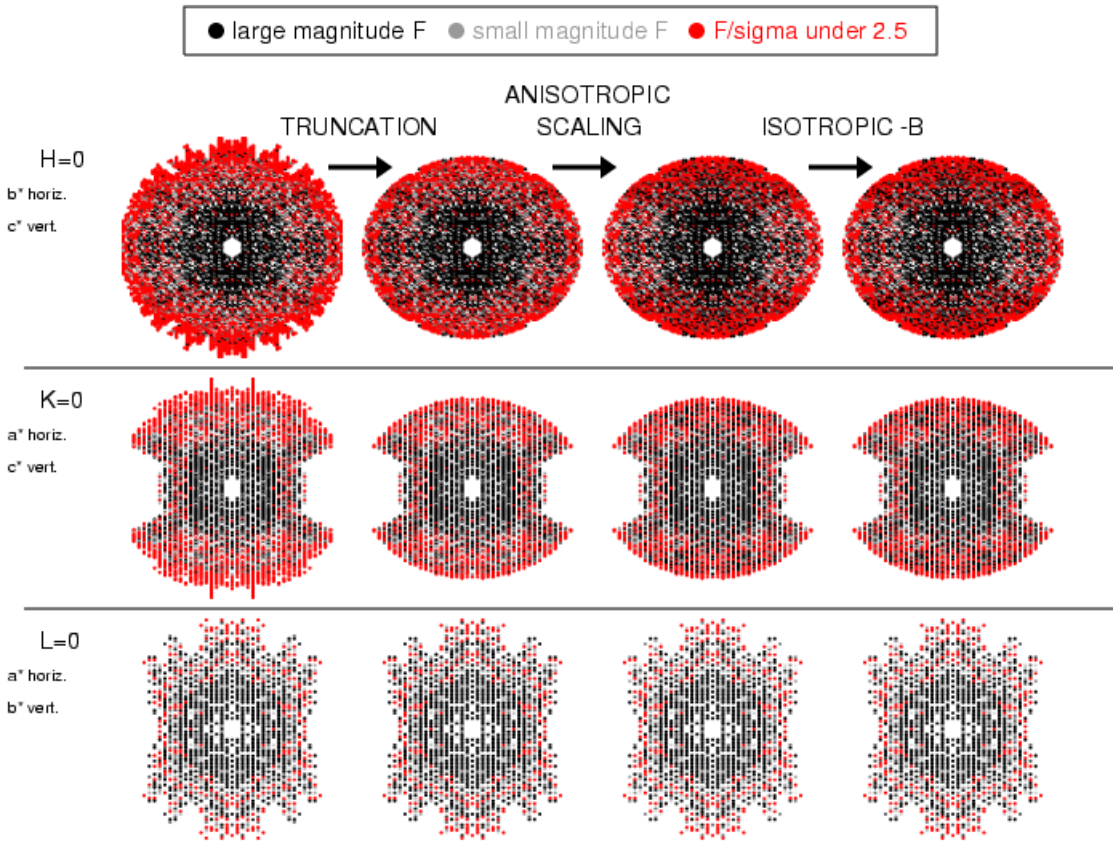


Fig. 29. Ellipsoidal truncation and anisotropic scaling of diffraction datasets of AZT-resistant HIV-1 RT pre-translocation complex

Fortunately, there is almost no anisotropy for the AZT-resistant HIV-1 RT post-translocation complex data set, so no data truncation and anisotropic scaling was applied to it (Fig. 30).

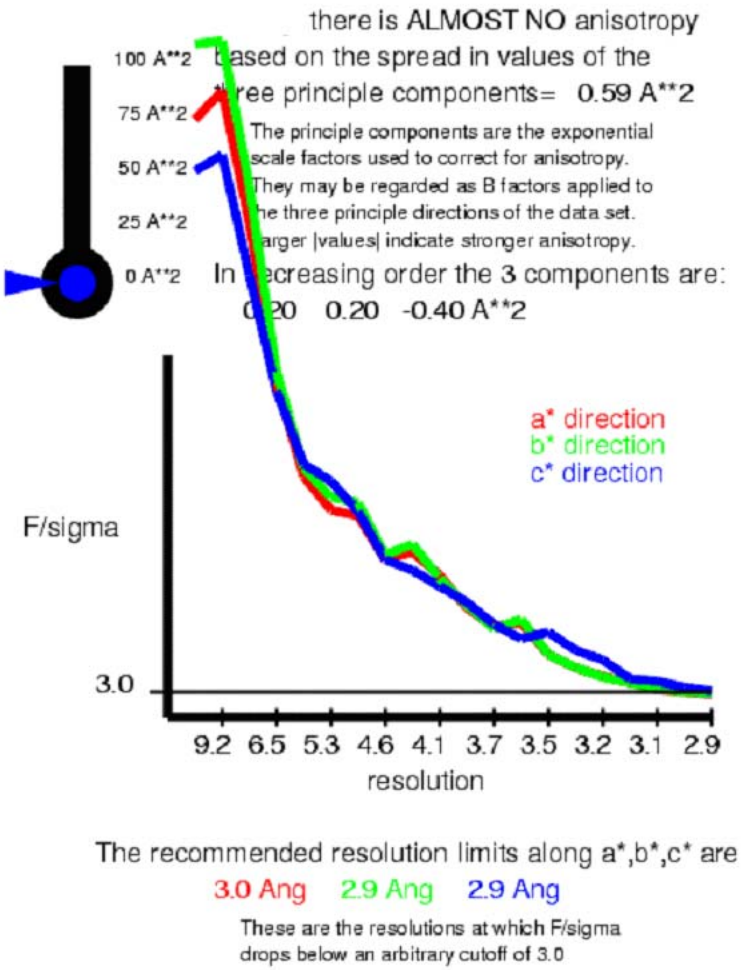


Fig. 30. No anisotropy in the diffraction data set of AZT-resistant HIV-1 RT post-translocation complex

9.3 Ellipsoidal truncation and anisotropic scaling of diffraction datasets of unliganded AZT-resistant HIV-1 RT

The data set for the unliganded AZT-resistant HIV-1 RT has only mild anisotropy with diffraction resolution limits of 2.6 Å along a*, 2.7 Å along the b*, and 2.7 Å along the c* directions (Fig. 31). 43,399 reflections were in the initial data set. 1,304 reflections were discarded because they fell outside the specified ellipsoid with dimension 1/2.6, 1/2.7, 1/2.7 $Å^{-1}$ along a*, b*, c*, respectively. 42,095 reflections remain after ellipsoidal truncation. Anisotropic scale factors were then applied to remove anisotropy from the data set. Lastly, an isotropic B of -5.85 $Å^2$ was applied to restore the magnitude of the high resolution reflections diminished by anisotropic scaling (Fig. 32).

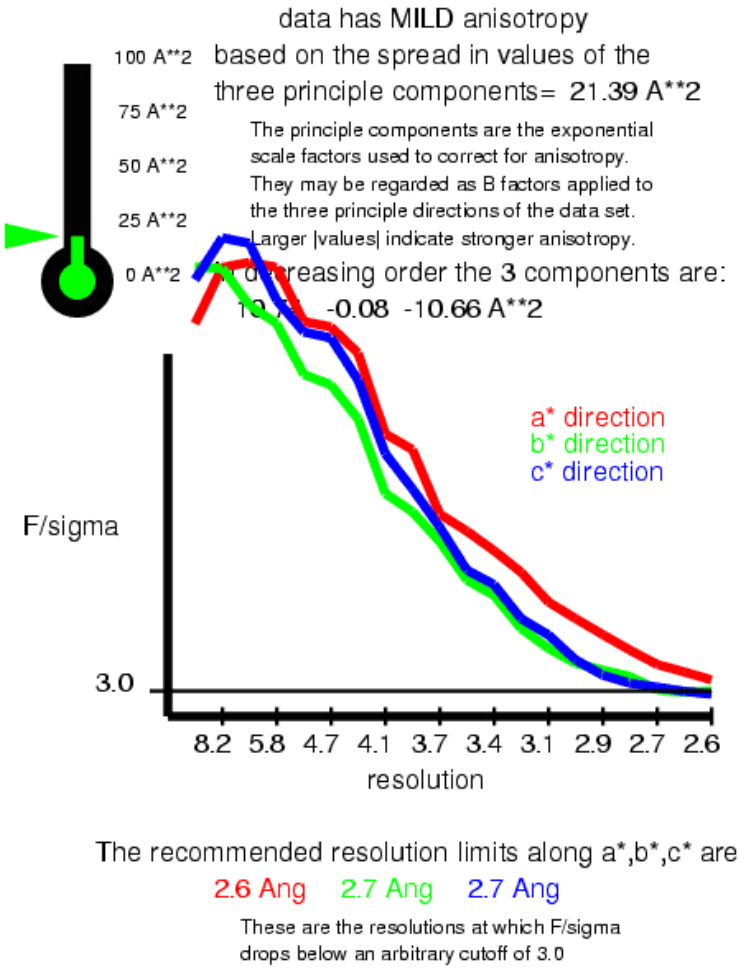


Fig. 31. Mild anisotropy in the diffraction data set of unliganded AZT-resistant HIV-1 RT

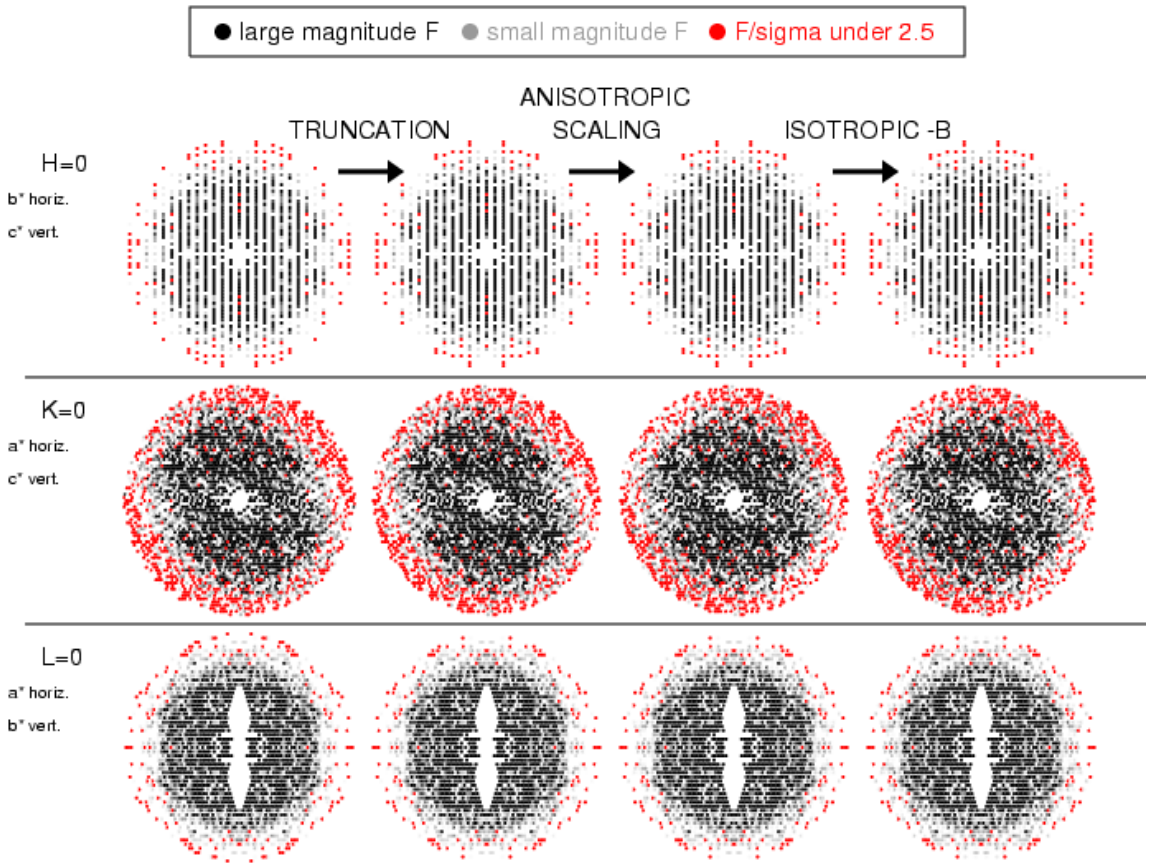


Fig. 32. Ellipsoidal truncation and anisotropic scaling of diffraction datasets of unliganded AZT-resistant HIV-1 RT

10. Molecular replacement phasing

10.1 Molecular replacement phasing of wild-type and AZT-resistant excision product complexes

Both structures were solved by the molecular replacement method with the same starting model: the well-ordered molecular copy of wild-type HIV-1 RT catalytic ternary complex (PDB id 1RTD). Rotation and translation search were performed using Phaser in the CCP4 suit (CCP4, 1994). Phaser is crystallographic program that employs maximum likelihood techniques to solve the phases of macromolecular crystal structures. According to the cell parameters and the molecular weight of the complex (~125 kDa), the asymmetric unit contains four complexes with a 3.4 $Å^3$/Dalton Matthews coefficient and 63% solvent content. Therefore, molecular replacement phasing should give four solutions to the asymmetric unit of each crystal. Solutions were searched one by one. The next solution was found by fixing the previous one. A flow diagram of automated molecular replacement using Phaser is shown in Fig. 33. The wild-type excision product complex dataset unambiguously gave all four solutions (Table 10), but the AZT-resistant excision product complex data only give three of them (Table 11). The fourth solution was generated using crystallographic symmetry operators. Phaser output a Z-score, number of clashes in the packing, and a refined LLG for each solution of rotation and translation function. The Z-score is the number of standard deviations above the mean value. RFZ is the rotation function Z-score, and TFZ is the translation function Z-score. For a rotation function, the Z-score of a correct solution may be under 4 and hide in the list

of Z-scores, so a translation function need to be performed to uncover the correct solution. For a translation function, the correct solution will generally have a Z-score over 5 and be well isolated from the rest of the solutions. LLG represents log likelihood gain. The likelihood is the probability that the data would have been measured, so it allows us to compare how well different models agree with the data. "The LLG is the difference between the likelihood of the model and the likelihood calculated from a Wilson distribution, so it measures how much better the data can be predicted with your model than with a random distribution of the same atoms" (Phaser Manual). The LLG should be a positive value for a reasonable solution.

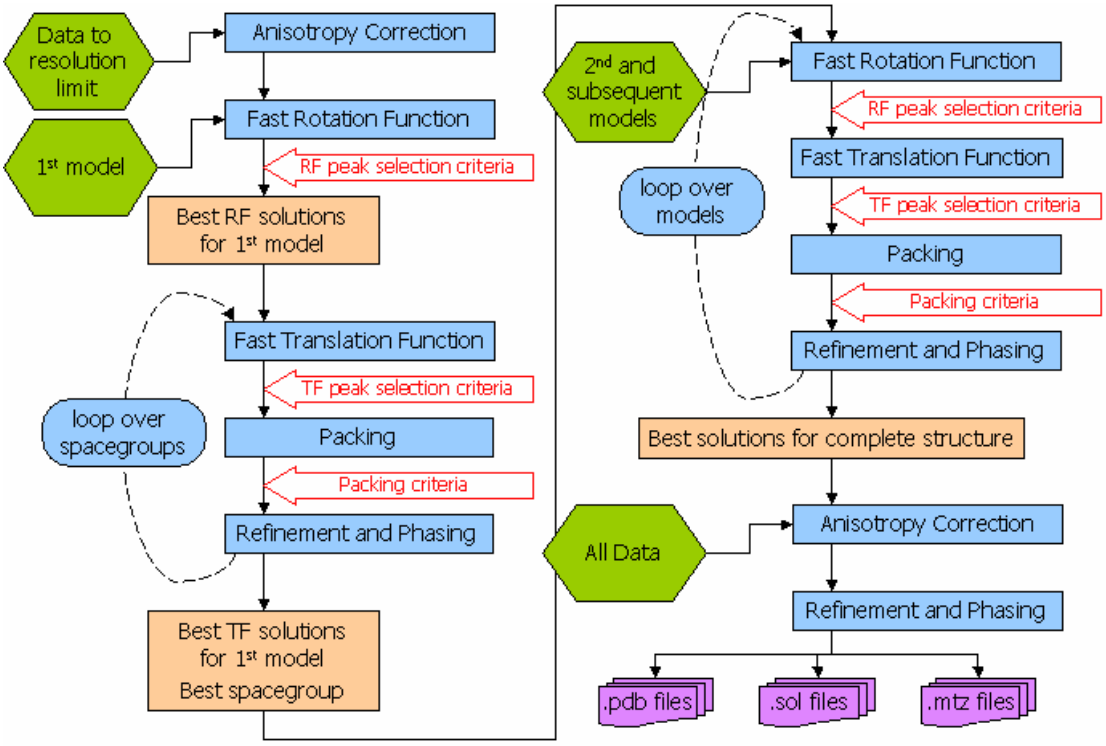


Fig. 33. A flow diagram of automated molecular replacement using Phaser □from the Phaser manual□

Table 10. Molecular replacement phasing of the wild-type HIV-1 RT excision product complex

```
SPACegroup HALL  P 2yb #P 1 21 1
SOLU SET   RFZ=34.6 TFZ=21.8 PAK=0 LLG=1149 RFZ=36.9 TFZ=66.3 PAK=0 LLG=4218
LLG=4582  RFZ=14.4 TFZ=50.3 PAK=1 LLG=6245 LLG=6718  RFZ=12.2 TFZ=41.4 PAK=2
LLG=7213 LLG=7689
SOLU 6DIM ENSE ensemble1 EULER  74.772   0.117  285.248 FRAC  0.50024
-0.00018  0.49963
SOLU 6DIM ENSE ensemble1 EULER 179.988  180.000   0.000 FRAC  1.00042
0.01060  0.04694
SOLU 6DIM ENSE ensemble1 EULER 241.663  48.509  253.131 FRAC -0.13327
0.12611  0.67458
SOLU 6DIM ENSE ensemble1 EULER  62.893  48.714  253.226 FRAC  0.63387
0.38481  0.12766
```

Table 11. Molecular replacement phasing of the AZT-resistant HIV-1 RT excision product complex

```
SPACegroup HALL  P 2yb #P 1 21 1

SOLU SET   RFZ=29.3 TFZ=30.1 PAK=0 LLG=1657 RFZ=31.5 TFZ=85.5 PAK=0 LLG=5919

LLG=6330  RFZ=7.6 TFZ=42.9 PAK=0 LLG=7217

SOLU 6DIM ENSE ensemble1 EULER  266.620   87.713  265.855 FRAC  0.78352

-0.00306  0.45640

SOLU 6DIM ENSE ensemble1 EULER   93.679   92.466   85.404 FRAC  0.29463

0.00299  0.00383

SOLU 6DIM ENSE ensemble1 EULER  134.177  135.667  279.054 FRAC -0.31765

0.14818  0.20919
```

10.2 Molecular replacement phasing of AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes

For the AZT-resistant HIV-1 RT pre-translocation complex, the search model was the crystallographic structure of HIV-1 RT-DNA/dTTP catalytic ternary complex (PDB id 1RTD). The rotational and translational search in Phaser gave an unambiguous solution (Table 12), as evidenced from the 24.8 of rotation Z-score, 49.0 of translation Z-score and 2573 of LLG (log likelihood gain). This molecular replacement solution has no packing clash in the unit cell (Table 12).

For the AZT-resistant HIV-1 RT post-translocation complex, the search model was the crystallographic structure of the wild-type HIV-1 RT post-translocation complex (PDB id 1N5Y). Rotational and translational search in Phaser gave an unambiguous solution (Table 13), as evidenced from the 41.2 of rotation Z-score, 72.5 of translation Z-score and 11774 of LLG (log likelihood gain). This molecular replacement solution has no packing clash in the unit cell (Table 13).

10.3 Molecular replacement phasing of unliganded AZT-resistant HIV-1 RT

The search model was the previously solved unliganded HIV-1 RT structure (PDB id 1DLO). Rotational and translational search gave an unambiguous solution (Table 14). The LLG is 6659. The rotational Z-score is 44.6, and the translational Z-score is 67.7.

Table 12. Molecular replacement phasing of the AZT-R HIV-1 RT pre-translocation complex

```
SPACegroup HALL  P 2ac 2ab #P 21 21 21

SOLU SET  RFZ=24.8 TFZ=49.0 PAK=0 LLG=2753 LLG=2753

SOLU 6DIM ENSE ensemble1 EULER  355.507   3.968   0.563 FRAC -0.62991

-0.48157  0.00791
```

Table 13. Molecular replacement phasing of the AZT-R HIV-1 RT post-translocation complex

```
SPACegroup HALL  P 32 2 (0 0 2) #P 32 1 2

SOLU SET  RFZ=41.2 TFZ=72.5 PAK=0 LLG=11529 LLG=11774

SOLU 6DIM ENSE ensemble1 EULER  239.912  180.000   0.000 FRAC -0.33291

0.33493  0.00034
```

Table 14. Molecular replacement phasing of the unliganded AZT-R HIV-1 RT

```
SPACegroup HALL  C 2y #C 1 2 1

SOLU SET  RFZ=44.6 TFZ=67.7 PAK=0 LLG=6659 LLG=6659

SOLU 6DIM ENSE ensemble1 EULER  26.552   0.163  333.501 FRAC -0.49559

0.00676 -0.48483
```

11. Structural refinement

11.1 Structural refinement of wild-type and AZT-resistant excision product complexes

The refinement was conducted using programs from CNS (Brünger et al., 1998). Rigid body refinement involved three stages. First, the whole asymmetric unit was regarded as a rigid body; then each molecule was regarded a rigid body; finally each complex was divided into multiple rigid bodies according to subdomains, domains, and DNA. Since the asymmetric unit had four complex copies, non-crystallographic symmetry (NCS) restraints were applied throughout refinement. The strict NCS could significantly increase the observations and improve the data to parameter ratio. The ensuing refinement involved several rounds of torsion angle annealing, energy minimization, density modification and averaging, and iterative model building. Density averaging can enhance signal and flatten random noise. Density averaging was usually conducted using the program DM in CCP4 by applying a combination of solvent-flattening, histogram matching, and density averaging. All model building was performed using the graphics software COOT (Emsley et al., 2004). Previous knowledge about base planarity, Watson-Crick base pairing, and sugar pucker conformations are also used to restrain the DNA. The maximum likelihood refinement target was usually MLF (amplitudes), but sometimes MLF and MLHL (amplitude and Hendrickson-Lattman coefficients) were alternatively used to improve the phase. NCS-restraint groups were setup as follows. For low-resolution structures, main chains are generally better defined than side chains, so

main-chain NCS restraints were separated from side-chain restraints. Furthermore, main chains and side chains were divided into multiple NCS-restraint groups according to domains, subdomains, and DNA. Each NCS-restraint group was assigned a weight according to the quality of their density maps. At the beginning, tight NCS-restraints were used to keep each NCS group almost the same. As refinement progressed, the NCS restraints were gradually loosened. A general diagram of refinement process is shown in Fig. 34.

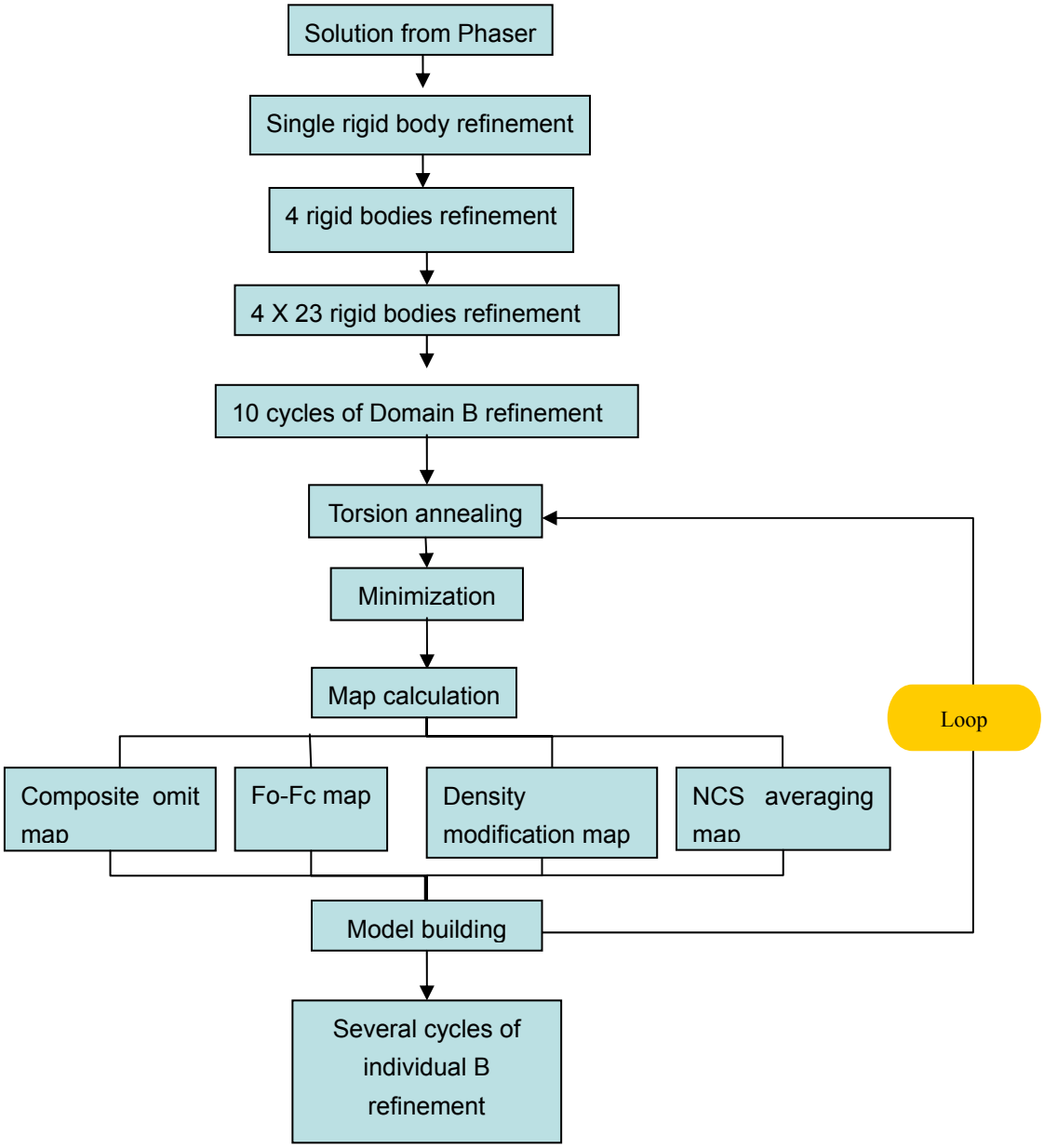


Fig. 34. A general diagram of structural refinement procedure

Water molecules were added into models based on Fo-Fc difference Fourier map using COOT at the final stage, followed by individual B-factor refinement in CNS. Each water molecule was validated according to following criteria: 1) The water peaks height were at least 3σ; 2) they had possible hydrogen bonding partners; 3) the putative hydrogen distances fell in the range of 2.3 Å – 3.5 Å; 4) their B factors were lower than 80 $Å^2$. Final refinement results are shown in Table 15. The wild-type excision product complex was determined to 3.15 Å resolution, and refined to R_{work} of 0.263 and R_{free} of 0.295. The AZT-resistant HIV-1 RT excision product structure was determined to 3.20 Å resolution, and refined to R_{work} of 0.278 and R_{free} of 0.308.

Table 15. Crystallographic structural refinement statistics		
	AZT-resistant excision product complex	Wild-type excision product complex
Resolution (Å)	25-3.2	25-3.15
No. reflections	111149	98589
R_{work}[a] / R_{free}[b]	0.276/0.307	0.266/0.294
No. atoms		
Protein	32140	31704
DNA/Ligand/ion	3839	3816
Water	91	74
R.m.s deviations		
Bond lengths (Å)	0.011	0.011
Bond angles (°)	1.413	1.423
Ramachandran plot quality		
Allowed (%)	99.7%	99.6%
Generously allowed (%)	0.2%	0.3%
Disallowed (%)	0.1%	0.1%

[a] $R_{work} = \Sigma \, ||F_o| - |F_c|| \, / \, \Sigma \, |F_o|$.

[b] For R_{free} calculations, 3% of data was excluded from refinement.

11.2 Structural refinement of AZT-resistant HIV-1 RT pre-translocation and post-translocation complexes

The crystal of the AZT-resistant HIV-1 RT pre-translocation complex was only diffracted to modest resolution (3.7 Å), so the data-to-parameter ratio of AZT-resistant HIV-1 RT pre-translocation complex is lower than those of the excision product complexes with resolution around 3.1 Å. The structural refinement of the AZT-resistant HIV-1 RT pre-translocation complex underwent a similar refinement process as the excision product complexes did, but much milder with only low temperature (600K) torsion annealing refinement implemented. Since the asymmetric unit contains two complexes, strict NCS restraints were implemented throughout the whole process. Additional DNA restraints were also added in order to increase the observations of DNA. The refinement first underwent three steps of multiple rigid bodies refinement. The ensuing refinement involved several rounds of energy minimization, density modification and averaging, and iterative model building. All the model building was performed using the graphics software COOT (Emsley et al., 2004). The MLF and MLHL (amplitude and Hendrickson-Lattman coefficients) were alternatively used to improve the phases. The NCS-restraint groups were setup in the same way as the excision product complexes. At the last step, group B-factors were refined. Individual B-factors were not refined and no water were added. This structure was refined to 3.7 Å with R_{factor} of 0.288 and R_{free} of 0.338, respectively. The final RMSD of bonds and angles are 0.0116 and 1.700, respectively (Table 16).

The AZT-resistant HIV-1 RT post-translocation complex diffracted to 2.9 Å resolution. The data-to-parameter ratio is good enough to conduct higher temperature torsion annealing refinement (1500 Kelvin), and individual B-refinement. Since the asymmetric unit contains only one complex, no NCS restraints were applied to the refinement process. The refinement followed the same procedure as the excision product complexes. Water peaks were searched and validated using COOT. This structure was determined to 2.9 Å, and refined to Rfactor of 0.257 and Rfree of 0.257. The RMSD bonds and angles are 0.012 and 1.666, respectively (Table 16).

11.3 Structural refinement of unliganded AZT-resistant HIV-1 RT

The asymmetric unit of unliganded AZT-resistant HIV-1 RT crystal contains only one RT complex. This crystal has even higher data-to-parameter ratio, so the simulated-annealing can tolerate a higher temperature. The structural refinement of unliganded AZT-resistant HIV-1 RT followed a procedure similar to what the excision product complexes followed, except that no NCS and DNA restraints were applied. At the final step, group B factors and individual B factors were refined. Water peaks were searched and validated using COOT. This structure was refined to 2.65 Å with R_{factor} of 0.251 and R_{free} of 0.294, respectively. The final RMSD of bonds and angles are 0.010 and 1.327, respectively (Table 16).

Table 16 Crystallographic structural refinement statistics			
	AZT-resistant pre-translocation complex	AZT-resistant post-translocation complex	Unliganded AZT-resistant HIV-1 RT
Resolution (Å)	30.0-3.70	25.0-2.90	20.0-2.65
No. reflections	31625	73103	41862
R_{work}[c] / R_{free}[d]	0.287/0.338	0.255/0.296	0.251/0.294
No. atoms			
Protein	15868	11420	8016
DNA/Ligand/ion	1804	918	N/A
Water	N/A	98	37
R.m.s deviations			
Bond lengths (Å)	0.012	0.012	0.010
Bond angles (°)	1.710	1.693	1.327
Ramachandran plot quality			
Allowed (%)	97.5%	99.2%	99.5%
Generously allowed (%)	1.7%	0.4%	0.4%
Disallowed (%)	0.8%	0.3%	0.1%

PART III. STRUCTURAL ANALYSES OF WILD-TYPE AND AZT-RESISTANT HIV-1 RT EXCISION PRODUCT COMPLEXES

1. Structural validation

The excision product complex is HIV-1 RT cross-linked to DNA in a complex with AZTppppA. Although the resolution of the crystal structures are around 3.2 Å, the overall quality of their electron density maps for RT and DNA is excellent. The composite annealed omit map (contoured at 1.0σ) of AZT-resistant excision product complex has clear electron density showing even the C-terminal residues 555-560 of the p66 subunit, which are missing in previously solved HIV-1 RT-DNA/dTTP ternary complex (PDB id 1RTD) and HIV-1 RT-DNA/tenofovir ternary complex (PDB id 1T05). For this reason, six more residues were included in the model at the C-terminus. Residues 218-230 of the p51 subunit are disordered in all previously solved HIV-1 RT structures. However, a portion of backbone in this region (residues 218-222) can be clearly observed in the electron density of the AZT-resistant excision product complex. Hence, five Ala were modeled at these positions of the p51 subunit of AZT-resistant RT. The wild-type RT of the wild-type excision product complex was truncated at residue 555. The residues 218-230 of p51 are disordered in the wild-type excision product complex. The geometric quality of protein was assessed using the structure validation tool PROCHECK (Laskowski et al., 1993). The nucleic was is validated by NDB NUCHECK (Gelbin et al., 1996; Clowney et al., 1996;

Schneider et al., 1997; Lu et al., 2003).

1.1 Structural validation of the wild-type HIV-1 RT excision product complex

The crystal structure of the wild-type excision product complex has excellent stereochemistry. The main Ramachandran plot shows that 99.6% of residues fall in the allowed regions (Fig. 35). Only four residues out of the total 4214 residues are in the disallowed regions. These residues are M184 of p66 subunits. M184 is located at a β-turn called $Y_{183}M_{184}D_{185}D_{186}$ motif. The $Y_{183}M_{184}D_{185}D_{186}$ motif is an essential element for polymerization or pyrophosphorolysis. The main-chain conformations of M184 from all previously solved RT structures actually have a very consistent distribution on the Ramachandran plot, in which φ and ψ are tightly around 60˚ and -120˚ (Fig. 36).

The torsion angles of all the base-paired nucleotides are distributed in the A-DNA region (yellow) or B-DNA region (blue) or A-B DNA region (green), as can be seen from Fig. 37. Nucleotide 703, 704 and 705 are located in the template overhang region, and are not base paired with nucleotides from the primer strand.

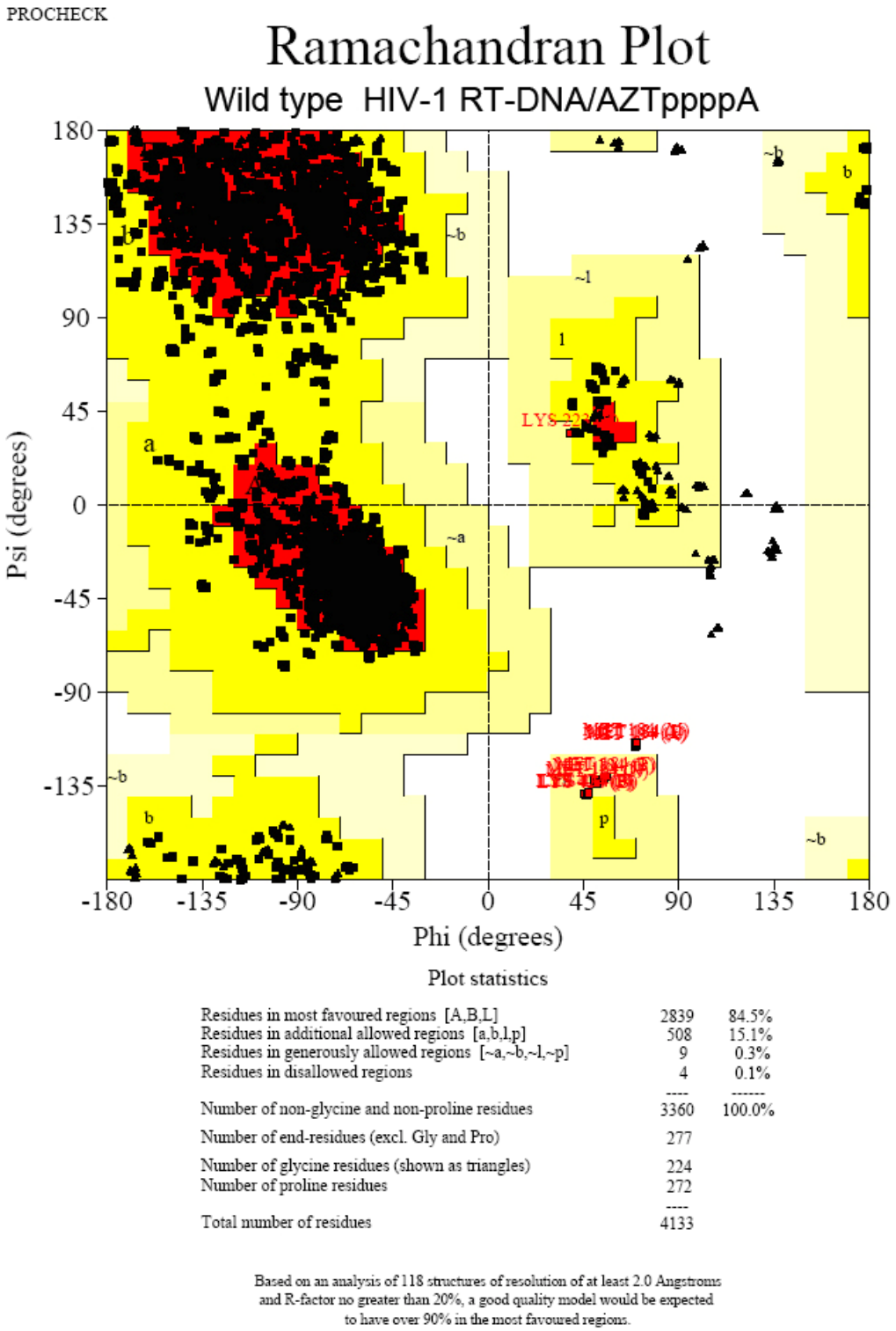


Fig. 35. Ramachandran plot of the wild-type excision product complex

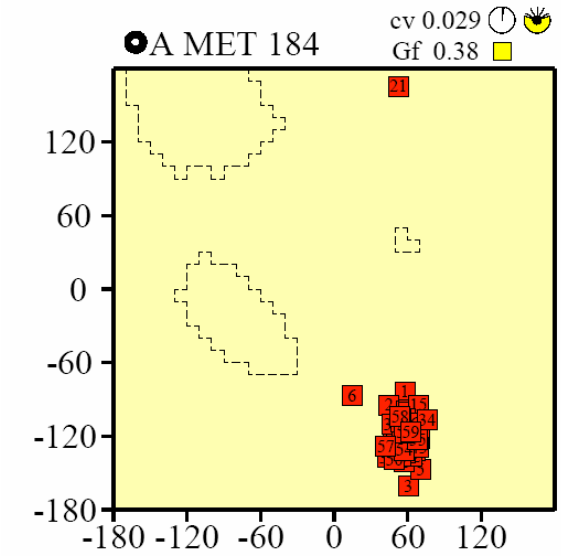


Fig. 36. Ramachandran plot of M184 from previously solved RT structures

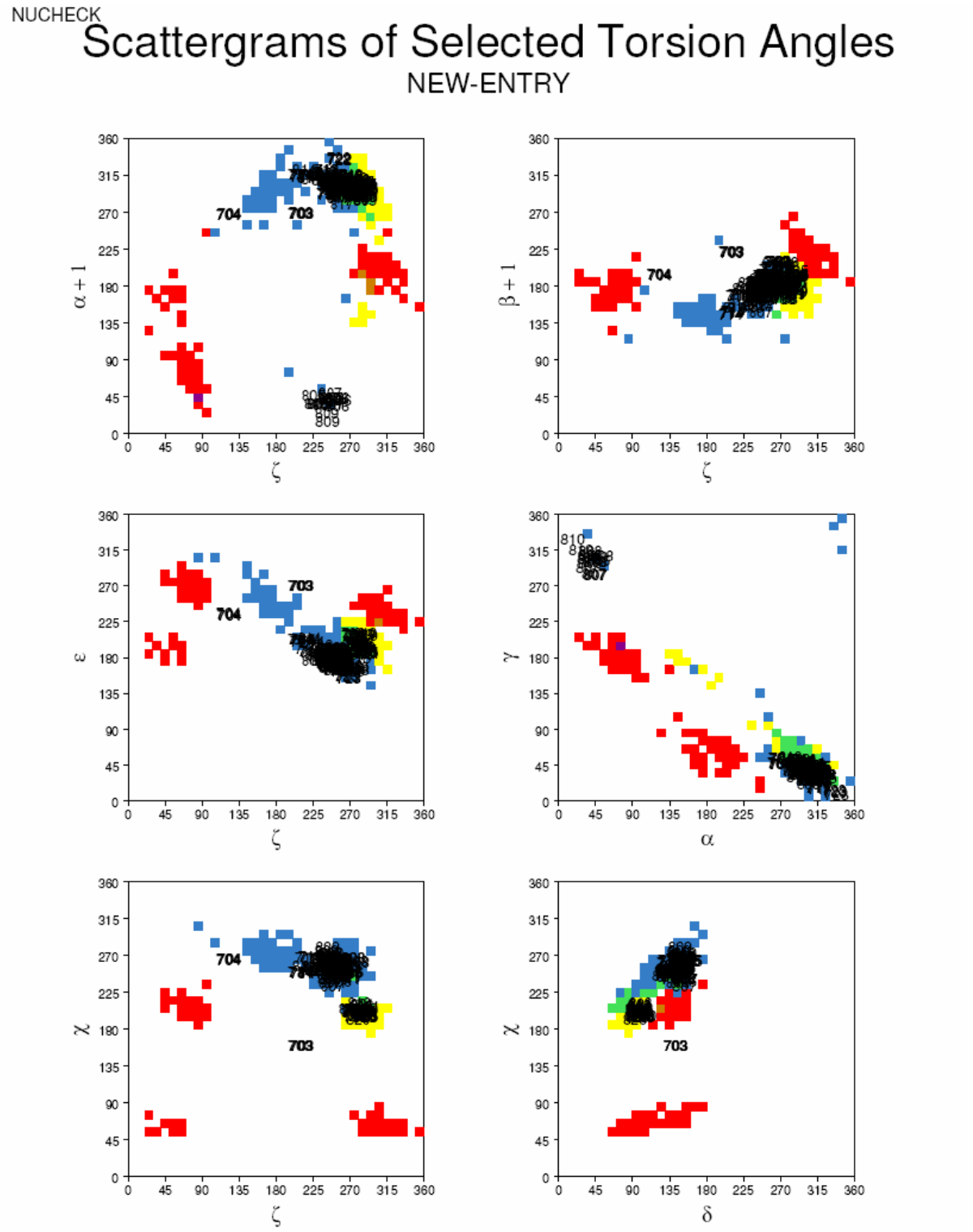
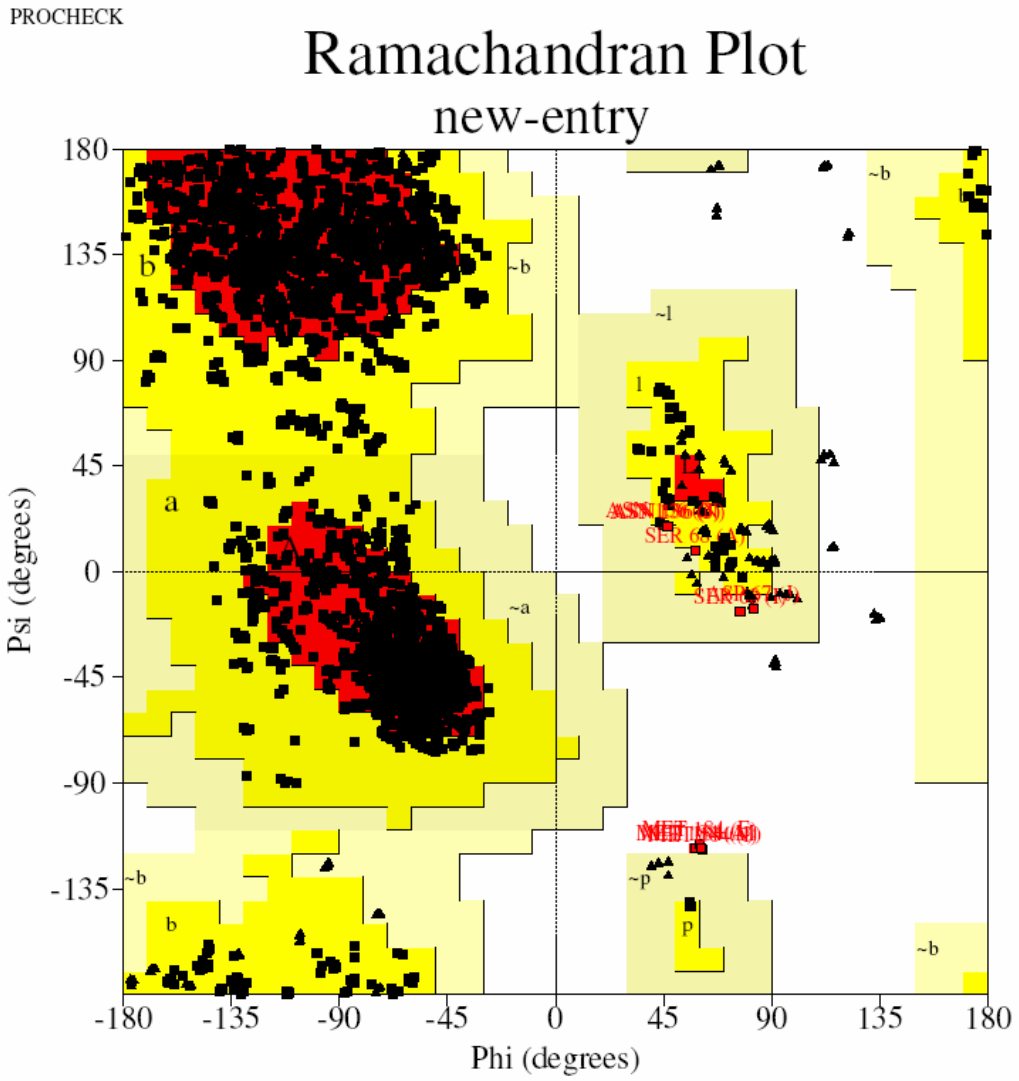


Fig. 37. Scattergrams of selected torsion angles of DNA of the wild-type HIV-1 RT excision product complex

1.2 Structural validation of the AZT-resistant HIV-1 RT excision product complex

The crystal structure of the AZT-excision product complex also has excellent stereochemistry. The main-chain Ramachandran plot shows that 99.6% of residues fall in the allowed regions and 0.1% of residues fall in the disallowed regions (Fig. 38). These residues in the disallowed region are Met184 of p66 subunits, consistent with the observation in the wild-type excision product complex. The main-chain torsion angles of both primary mutant R70 and Y215 have a tight distribution, and are distributed in the most favored regions (Fig. 39). The Chi-1 angle of both primary mutations has a relatively broad distribution, suggesting small heterogeneity. However, the Chi-2 angle of both primary mutations has a very tight distribution. The torsion angles of the base-paired nucleotides are distributed in the A-DNA region (yellow) or B-DNA region (blue) or A-B DNA region (green), as can be seen from Fig. 39.



Plot statistics

Residues in most favoured regions [A,B,L]	2832	83.2%
Residues in additional allowed regions [a,b,l,p]	560	16.5%
Residues in generously allowed regions [~a,~b,~l,~p]	6	0.2%
Residues in disallowed regions	4	0.1%
Number of non-glycine and non-proline residues	3402	100.0%
Number of end-residues (excl. Gly and Pro)	20	
Number of glycine residues (shown as triangles)	228	
Number of proline residues	272	
Total number of residues	3922	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig. 38. Ramachandran plot of the AZT-resistant excision product complex

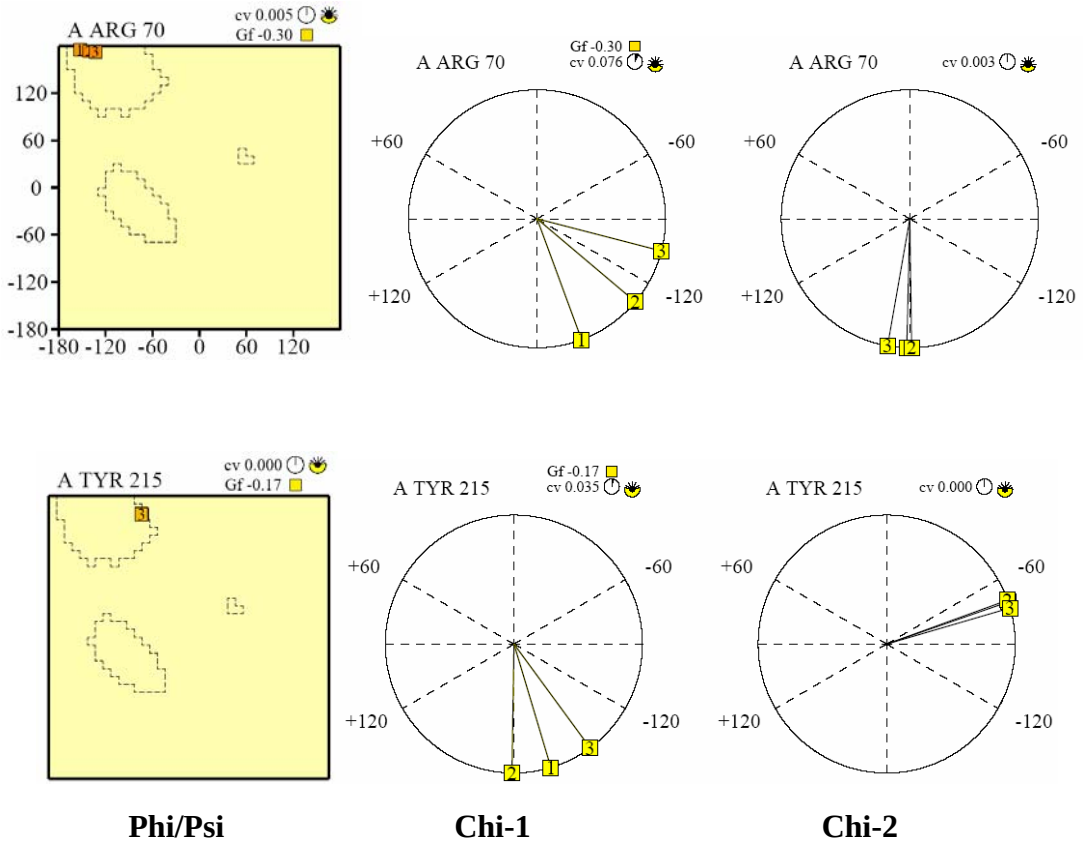


Fig. 39. Conformations of primary mutant R70 and Y215 in the crystallographic asymmetric unit of the AZT-resistant RT excision product complex. Only the first three complex copies in the crystallographic asymmetric unit are compared. The main-chain conformations of both primary mutations have a tight distribution. The Chi-1 angles of both side chains exhibit small heterogeneity, but the Chi-2 angles of both side chains have a tight distribution. The conformational wheel plot was generated by modifying the Program Procheck-NMR (Laskowski et al., 1996).

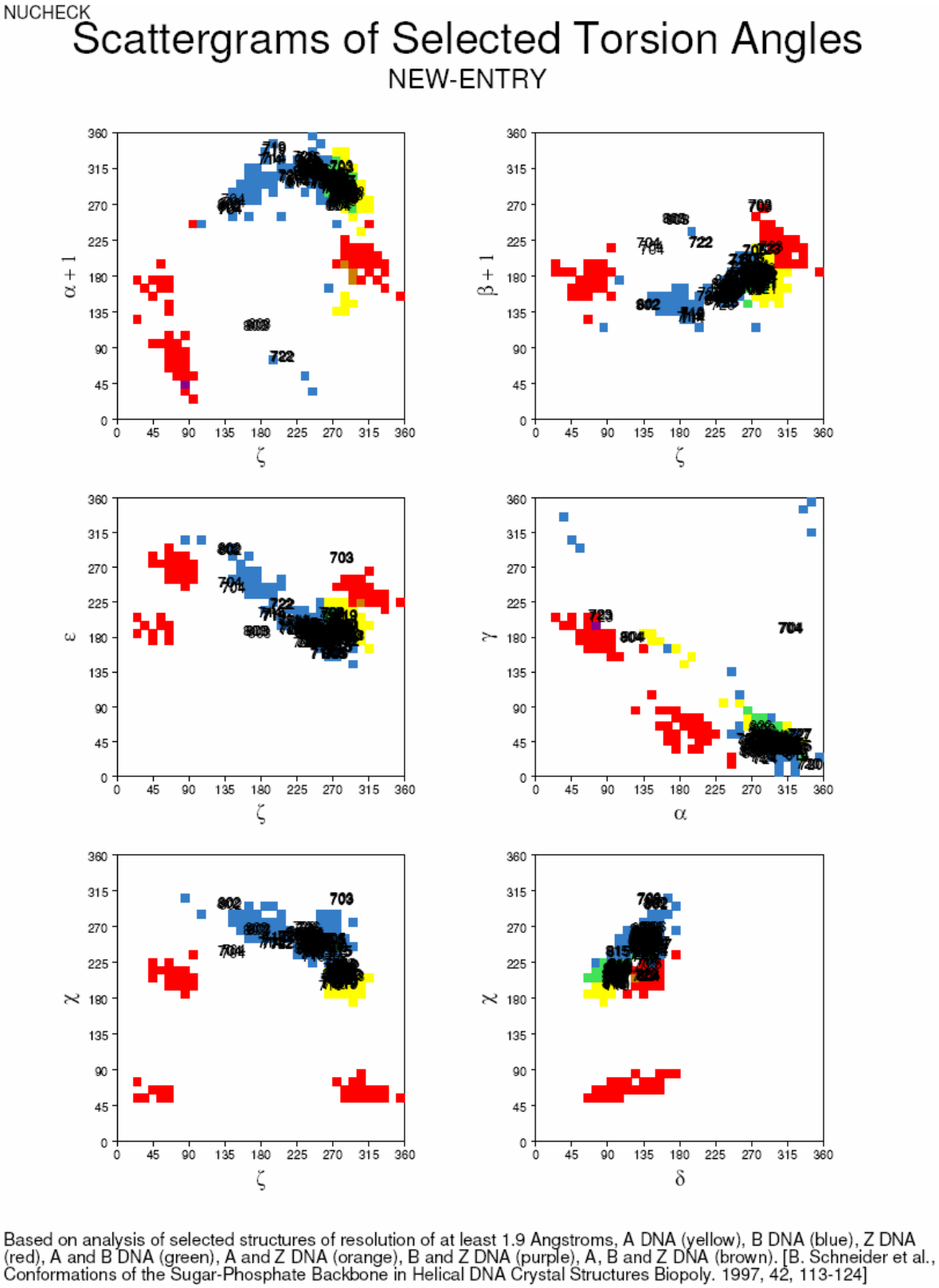


Fig. 40. Scattergrams of selected torsion angles of DNA of the AZT-resistant HIV-1 RT excision product complex

2. Space-group assignment (monoclinic *vs.* orthorhombic) of both wild-type and AZT-resistant excision product complexes

The space group assignments for both of the crystals turned out to be a barrier for the structural determination at the beginning. The space groups of both excision product complexes were considered likely to be orthorhombic, since the β angles of their crystal cells are very close to 90˚ (Table 17). Their crystal cells are similar to the one of HIV-1 RT-DNA/dTTP ternary complex in dimensions and angles, and the crystal of HIV-1 RT-DNA/dTTP ternary complex has been successfully assigned to the orthorhombic space group $P2_12_12_1$ (Huang et al., 1998). Furthermore, the crystals of both excision product complexes have the same look and shape as the HIV-1 RT-DNA/dTTP ternary complex crystals, which are plate clusters. Taken together, it is quite possible to think that the space groups should be orthorhombic. However, the scaling with orthorhombic space groups gives very high linear R-merge values for the data within the low resolution shells, indicating that crystals are possibly assigned wrong Laue symmetry (or point symmetry). Refinement with orthorhombic symmetry also led to high Rfree and Rwork values that couldn't be brought down after trying all the measures. To break through this barrier, we lower the crystal symmetry by assigning a monoclinic space group $P2_1$ that is a subgroup of $P2_12_12_1$. The monoclinic space group $P2_1$ gives reasonably low linear R-factors for low resolution data. The Rfree and Rwork also successfully dropped to reasonable low levels after extensive structural refinement. Although the symmetry reduction lowers the redundancy of the final reflection dataset, we have collected enough images to make

up for it. Their final data redundancy is more than 4, which means each reflection is measured more than 4 times on average.

TABLE 17 COMPARISON OF CELL PARAMETERS			
Dataset	AZT-R RT-DNA/AZTppppA	WT RT-DNA/AZTppppA	WT HIV-1 RT-DNA/dTTP
Space group	$P2_1$	$P2_1$	$P2_12_12_1$
Cell parameters			
a, b, c (Å)	78.7, 283.4, 155.2	78.5, 274.8, 152.4	78.5 150.0 279.3
α, β, γ (°)	90.0, 89.7, 90.0	90.0, 90.1, 90.0	90.0, 90.0, 90.0
NCS-related copies	4	4	2

According to their space group, cell parameters and the molecular weight of the complex (~125 kDa), the asymmetric unit contains four NCS-related copies. The Matthew's coefficient (V_M) is 3.39 $Å^3$/Dalton (Matthews, 1989). Assuming protein density is 1.34, the solvent content is about 63%.

3. Structural comparisons among the four NCS-related complex copies within the crystallographic asymmetric units of wild-type and AZT-resistant excision product complexes

The four NCS-related complex copies show almost the same overall conformation of enzyme and DNA. Superposition of the four copies of the wild-type excision product complex gives less than 0.4 Å of root-mean-square deviation (RMSD) for the overall Cα-atom displacements of the enzyme, and less than 0.2 Å of RMSD for the overall all-atom displacements of the DNA (Table 18). Superposition of the four copies of AZT-resistant excision complexes show less than 0.6 Å of RMSD for the overall Cα-atom displacements of the enzyme and less than 0.1 Å of RMSD for the overall all-atom displacements of the DNA (Table 19). The relative small overall RMSD may be caused by the NCS restraints implemented in the refinement process, although the structural modeling is based on less biased electron density maps including composite omit maps, difference maps, density modification maps, etc. The AZTppppA exhibits significantly higher RMSD than the RT and DNA across the four copies of the wild-type excision product complex. The reason is the AMP part of AZTppppA is ill-defined. This AMP part accounts for significant proportion of the

whole molecule. However, the four copies of AZTppppA in the crystallographic asymmetric unit of AZT-resistant excision product complex have consistent conformations with less than 0.3 Å of RMSD, which is in the variation range of DNA and RT.

Although the four NCS-related excision product complex copies of wild-type and AZT-resistant excision product complexes are almost identical in terms of overall conformation of enzyme and DNA, they differ in their overall structural order indicated by their electron density maps and average isotropic temperature factors. We have arranged the four NCS-related copies in an order in which the first copy has the best overall electron density and lowest average isotropic temperature factor, whereas the last copy has the lowest ordered electron density and highest average isotropic temperature factor. In the asymmetric unit of the wild-type excision product complex crystal (Table 20), the first complex copy has an average temperature factor of 49 $Å^2$ for the main-chain atoms, and the fourth complex copy has an average temperature factor of 102 $Å^2$ for its main-chain atoms. In the asymmetric unit of the AZT-resistant excision product complex crystal (Table 20), the four complex copies show an even wider range of order. The first copy has an average temperature factor of 57 $Å^2$ for the main-chain atoms, whereas the fourth copy is nearly disordered with an average temperature factor of 128 $Å^2$ for its main-chain atoms. The packing analysis showed the fourth copy makes fewer contacts with the others in both excision product complex crystals. This wide range of structural order might explain the

breakdown of the higher symmetric arrangement of RT complex in the crystal (from orthorhombic to monoclinic), and may also explain the diffraction anisotropy of the two excision product crystals, as discussed in the experimental section. Differences of NCS-related complexes in average temperature factors or residue conformations are normal phenomena, which were also reported in a HIV-1 RT structure (Huang et al., 1998) and other protein structures (Carfí et al., 1999; Kleywegt, 1996).

Table 18 Structural superposition of the other NCS-related copies on the best-ordered copies within crystallographic asymmetric unit of wild-type excision product complex			
p66 subunit of wild-type RT	**Complex 2**	**Complex 3**	**Complex 4**
RMS XYZ DISPLACEMENT =	0.104	0.397	0.350
AVERAGE XYZ DISPLACEMENT =	0.094	0.342	0.302
MAXIMUM XYZ DISPLACEMENT =	0.255	1.158	0.943
DNA			
RMS XYZ DISPLACEMENT =	0.142	0.21	0.181
AVERAGE XYZ DISPLACEMENT =	0.113	0.176	0.151
MAXIMUM XYZ DISPLACEMENT =	0.677	0.811	0.738
AZTppppA			
RMS XYZ DISPLACEMENT =	1.548	0.781	0.471
AVERAGE XYZ DISPLACEMENT =	0.934	0.604	0.439
MAXIMUM XYZ DISPLACEMENT =	3.931	2.241	0.797

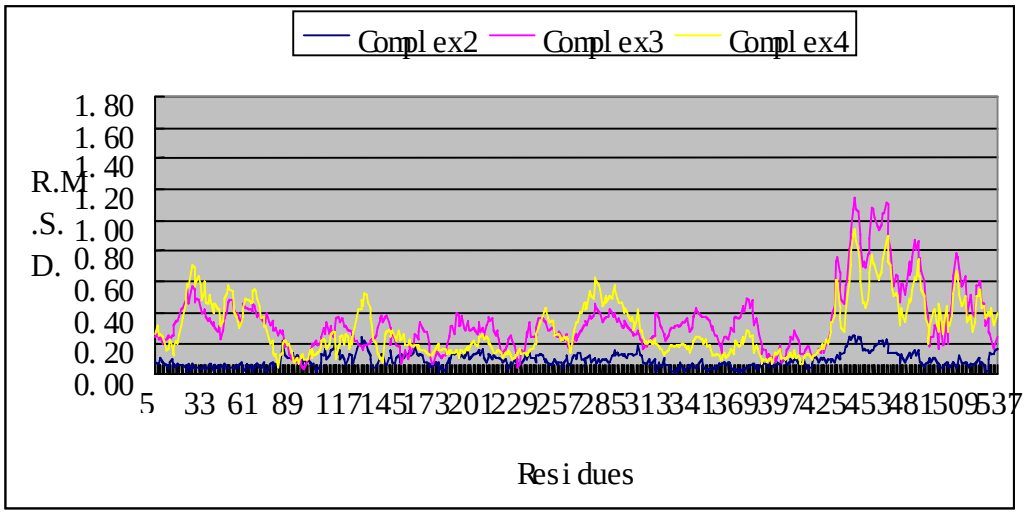


Fig. 41. The RMSD of the p66 subunit of the other three copies in reference to the p66 subunit of the best ordered complex copy

Table 19 Structural superposition of the other NCS-related copies on the best-ordered copies within crystallographic asymmetric unit of AZT-resistant excision product complex

P66 subunit of AZT-resistance RT	**Complex2**	**Complex3**	**Complex4**
RMS XYZ DISPLACEMENT =	0.140	0.602	0.463
AVERAGE XYZ DISPLACEMENT =	0.123	0.542	0.426
MAXIMUM XYZ DISPLACEMENT =	0.453	1.655	1.226
DNA			
RMS XYZ DISPLACEMENT =	0.079	0.117	0.095
AVERAGE XYZ DISPLACEMENT =	0.069	0.104	0.086
MAXIMUM XYZ DISPLACEMENT =	0.251	0.287	0.338
AZTppppA			
RMS XYZ DISPLACEMENT =	0.116	0.246	0.199
AVERAGE XYZ DISPLACEMENT =	0.107	0.226	0.180
MAXIMUM XYZ DISPLACEMENT =	0.239	0.526	0.414

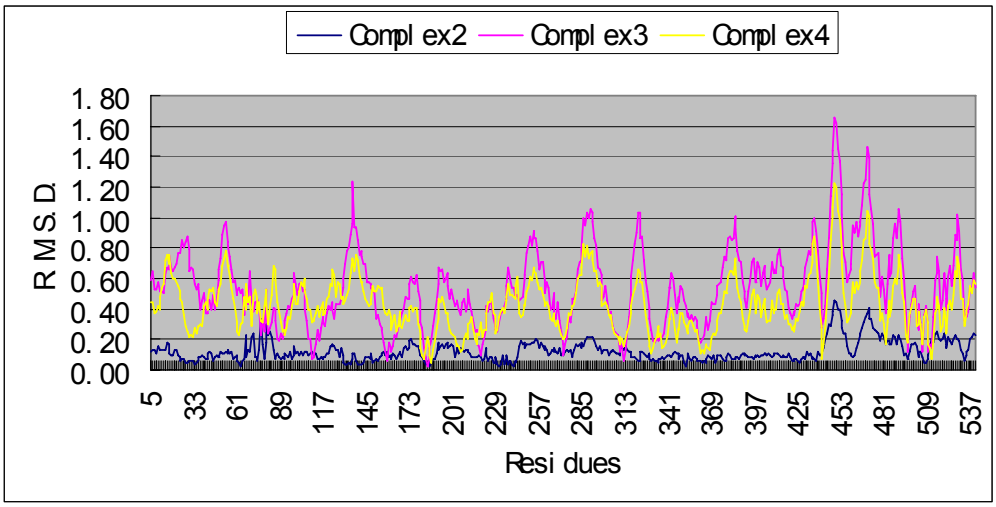


Fig. 42. The RMSD of the p66 subunit of the other three copies in reference to the p66 subunit of the best ordered complex copy

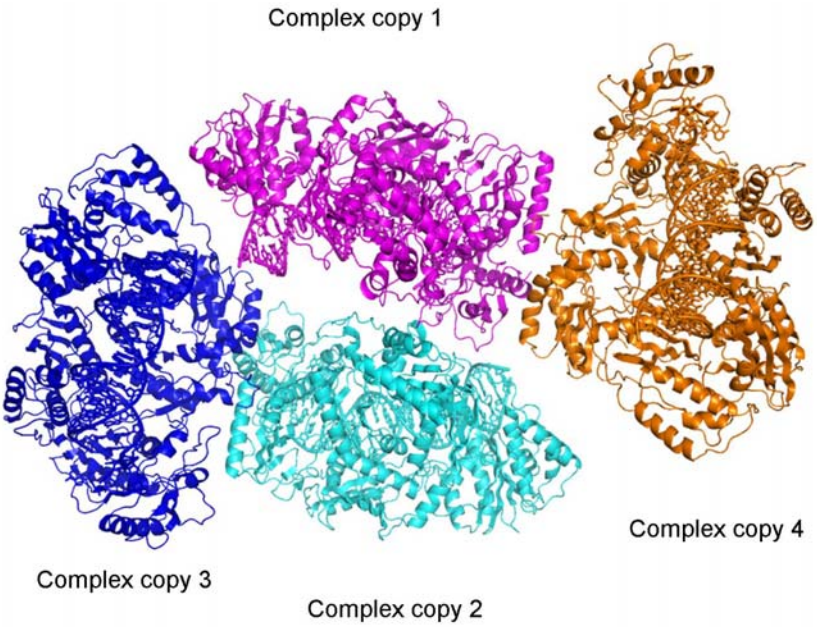
Table 20 Comparison of the average temperature factor (B factor) of each NCS-related complex copies within both crystallographic asymmetric units				
AZT-R excision product complex	Complex copy 1	Complex copy 2	Complex copy 3	Complex copy 4
Average B value for main chain	57.32	60.08	114.60	128.36
Average B value for side chain	59.31	61.95	114.74	127.95
Average B value for all atoms	58.33	61.03	114.67	128.15
Average RMS B for the main chain	1.19	1.11	0.50	0.45
Average RMS B for the side chain	1.75	1.68	0.71	0.66
Wild-type excision product complex				
Average B value for main chain	48.89	52.39	85.65	101.65
Average B value for side chain	50.11	53.58	85.97	101.97
Average B value for all atoms	49.51	53.00	85.81	101.82
Average RMS B for the main chain	1.18	1.16	0.86	0.66
Average RMS B for the side chain	1.59	1.46	0.86	0.70

4. Crystal packing analysis of wild-type and AZT-resistant excision product complexes

Both crystallographic asymmetric units contain four NCS-related copies of the excision product complex. The transformation relationships among these four NCS-related copies of the wild-type excision product complex are shown in Fig. 43. Complexes 1 and 2 are related by an 180˚ rotation. Complexes 1 and 4 are related by a 176˚ rotation. Complexes 2 and 3 are related by a 175˚ rotation. Complex 3 and 4 are related by an 180˚ rotation.

The transformation relationships among these four NCS-related copies of wild-type excision product complex are shown in Fig. 43. Complexes 1 and 2 are related by an 180˚ rotation. Complexes 1 and 4 are related by a 177˚ rotation. Complexes 2 and 3 are related by a 175˚ rotation. Complexes 3 and 4 are related by an 180˚ rotation.

A



B

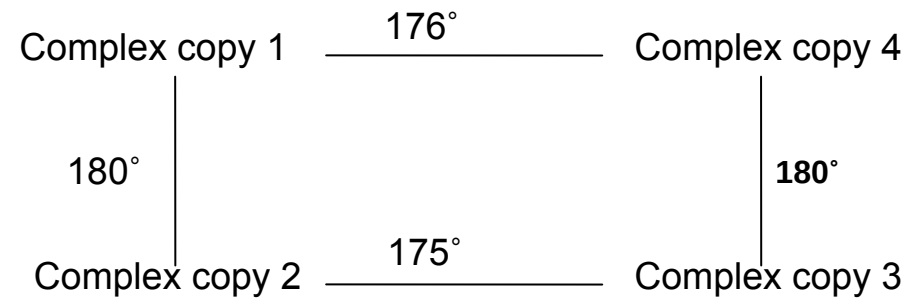
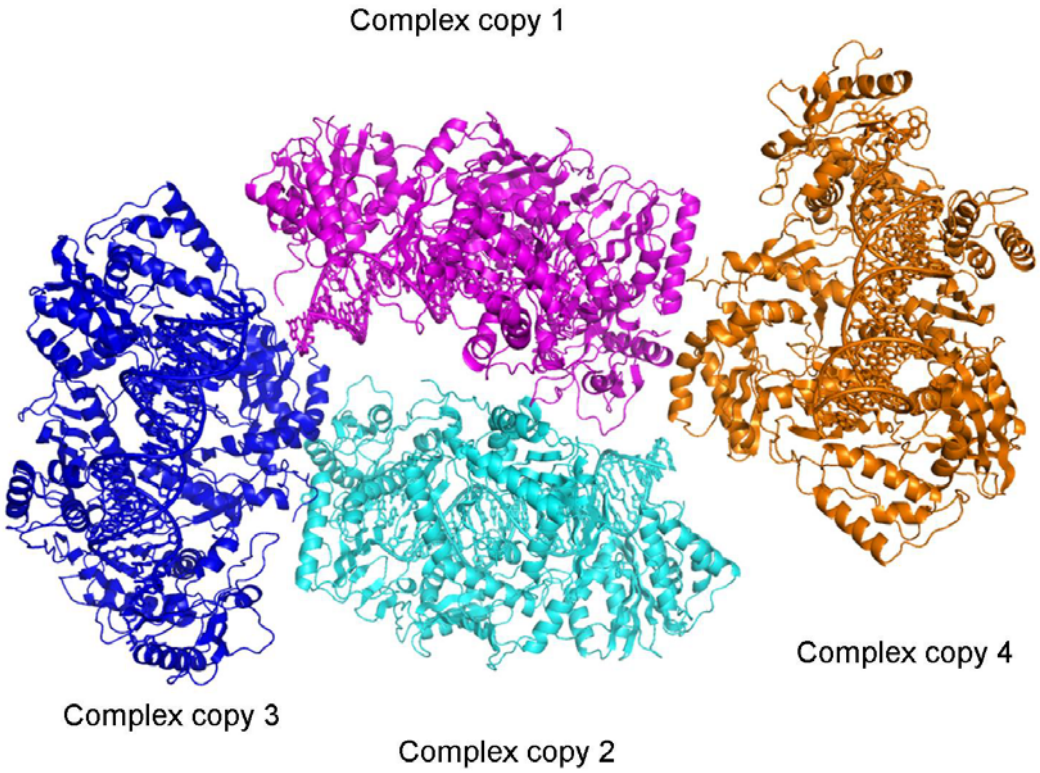


Fig. 43. Overall image of asymmetric unit of the wild-type excision product complex crystal

(A) overall image of all four complex copies in the asymmetric unit

(B) transformation relationships among the four complex copies

A



B

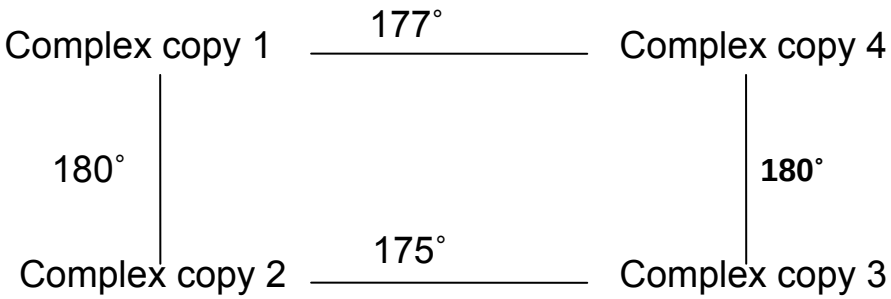


Fig. 44. Overall image of asymmetric unit of the AZT-resistant excision product complex crystal

(C) overall image of the four complex copies in the asymmetric unit

(D) transformation relationships among the four complex copies

Both crystals have high solvent content (63%). The four NCS-related complex copies are packed differently within the crystal. The complex 1 is a tight complex with the most crystal-packing interactions (Table 21), which is correlated to its low average temperature factor and best defined electron density. On the other hand, the complex 4 is a loose complex with the least crystal-packing interactions (Table 21), which explain why it has a high average temperature factor and the worst defined electron density. The fourth copies of both crystal structures are actually almost disordered.

Table 21 The number of crystallographic contacts	
	Number of crystallographic contacts
Crystallographic asymmetric unit of wild-type excision product complex	
Complex copy 1	216
Complex copy 2	165
Complex copy 3	170
Complex copy 4	81
Crystallographic asymmetric unit of AZT-resistance excision product complex	
Complex copy 1	210
Complex copy 2	147
Complex copy 3	140
Complex copy 4	78

The number of crystallographic contacts of one complex copy is defined by the number of atoms belonging to another complex copy or symmetry-related complex copy that are located at a distance < 4 Å.

5. Overview of the structures of the wild-type and AZT-resistant excision product complex

The representations of HIV-1 RT-DNA structures are traditionally oriented in a way showing the right-handed shape of polymerase domain, but this orientation can't give a clear view of polymerization/pyrophosphorolysis active site and AZT-resistance mutation sites. In order to achieve a better representation of polymerization/pyrophosphorolysis active and AZT-resistance mutation sites, the HIV-1 RT-DNA is shown in another orientation (Fig. 45).

The two excision product complexes are RT cross-linked to DNA in complex with the AZT excision product, AZTppppA. The AZTppppA molecules in both complexes are bound within a cleft between the fingers and palm subdomains, where the polymerase active site is located (Fig. 46 and Fig. 47). The AZTppppA can naturally be divided into ATP and AZTMP moieties according to the excision reaction, by which the structures will give insights about how RT uses ATP to excise the AZTMP. If we want to expand our thinking, the AZTppppA can also be divided into AZTTP part and AMP part, by which the structures will give insights about how AZTTP binds to RT/DNA as an incoming nucleotide. The four phosphate groups of AZTppppA are labeled according to the way by which we divide it (Fig. 46A). Since a primary aim is to explore the structural basis for the ATP-mediated pyrophosphorolysis, we divide AZTppppA into the ATP part and AZTMP part. Thus, we use α', β', and γ' to indicate the three phosphate groups of the ATP part, and α to indicate the phosphate group of

the AZTMP part. The nomenclature of the three phosphate groups of ATP are shown in the Fig. 46B.

6. Superposition of AZT-resistant and wild-type excision product complexes reveals a significant repositioning of the ATP moiety of AZTppppA

Structural comparison of the two excision product complexes showed that the wild-type and the AZT-resistant RT shared almost the same overall conformation with only 0.6 Å RMSD, when Cα atoms of their best ordered copy were compared. No significant main-chain displacement or rearrangement was observed within the polymerization/pyrophosphorolysis catalytic site or around the sites of resistance mutations (Fig. 49). This suggests that the AZT-resistance mutations have no significant effect on either overall or local backbone conformations of RT. This similarity indicates pyrophosphorolytic excision elements conform to a stringent architecture, as also can be confirmed that no AZT-resistance mutations occur directly in the polymerase/pyrophosphorolysis catalytic site.

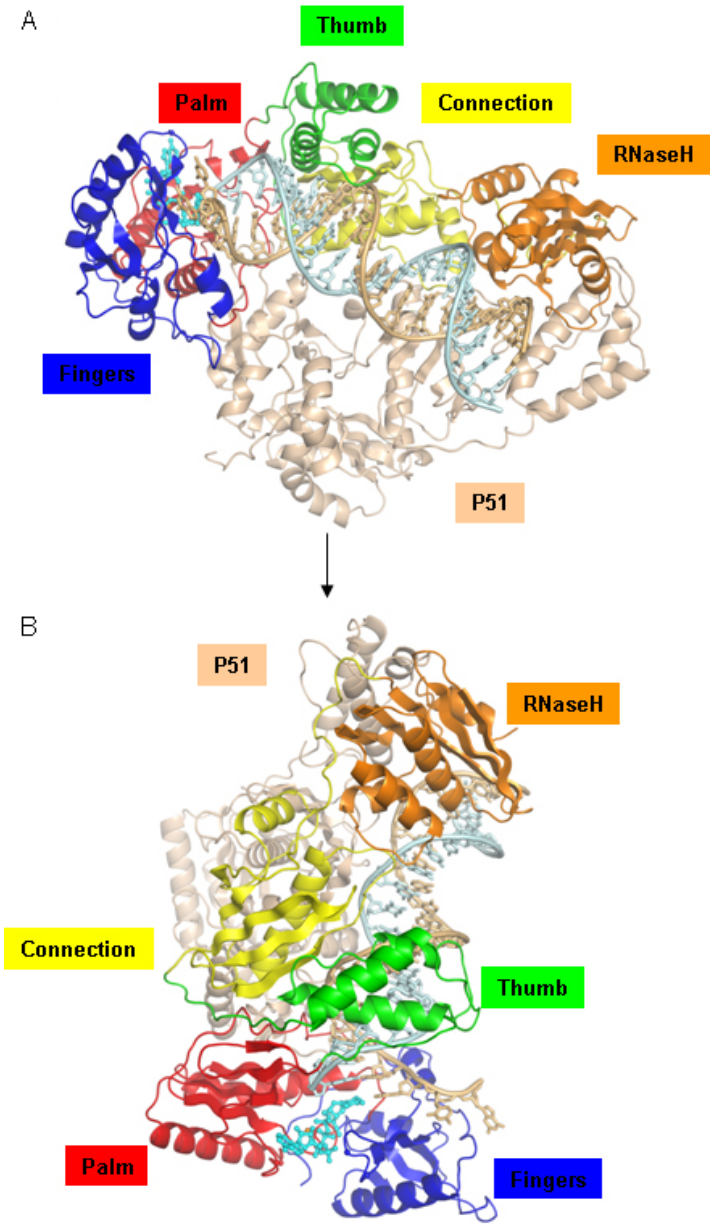
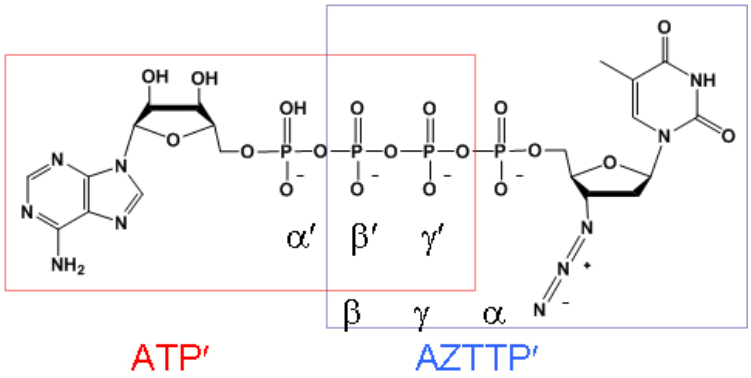


Fig. 45. Viewing orientation of HIV-1 RT-DNA/nucleotide

(A) traditional view of HIV-1 RT-DNA

(B) the viewing orientation used in this dissertation

A



B

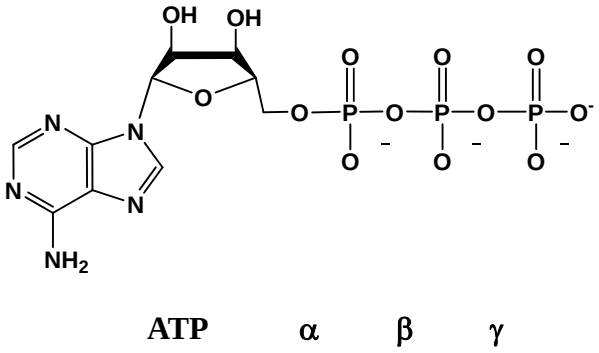


Fig. 46. Nomenclature of the phosphate groups in AZTppppA and ATP

(A) The four phosphate groups are named based on the two components ATP and AZTTP

(B) The nomenclature of the three phosphate groups of ATP

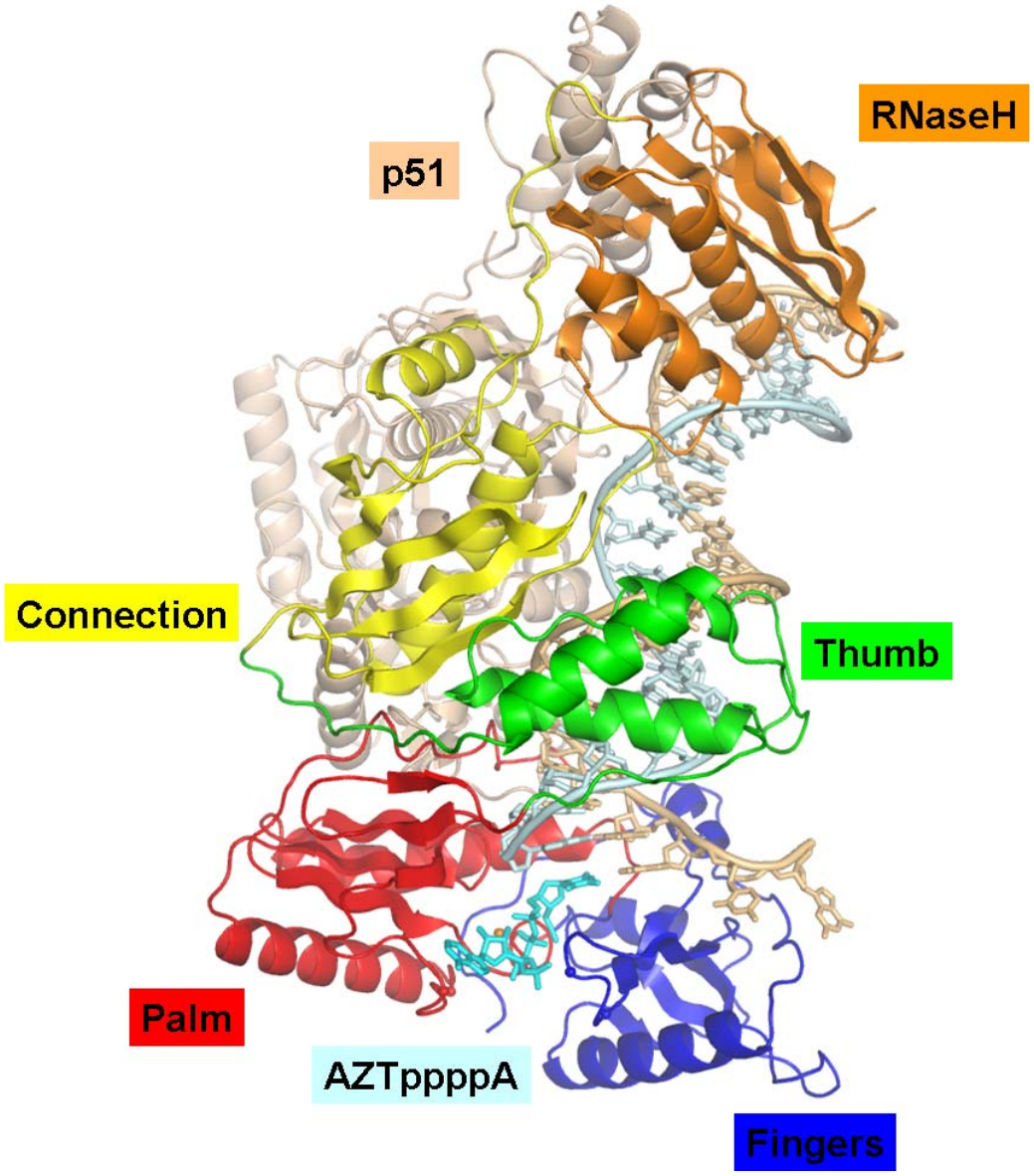


Fig. 47. Overall structure of wild-type HIV-1 RT-DNA/AZTppppA

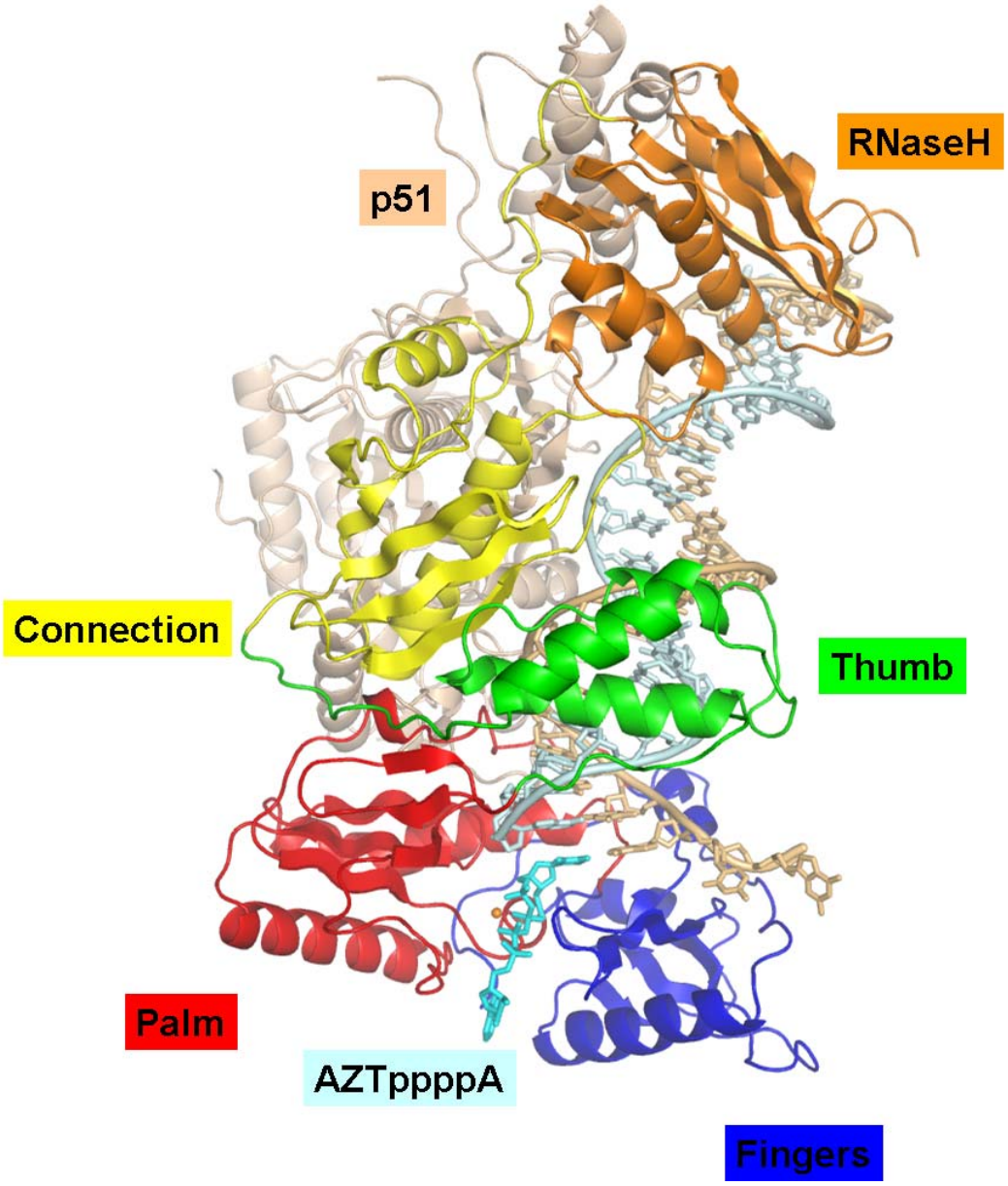
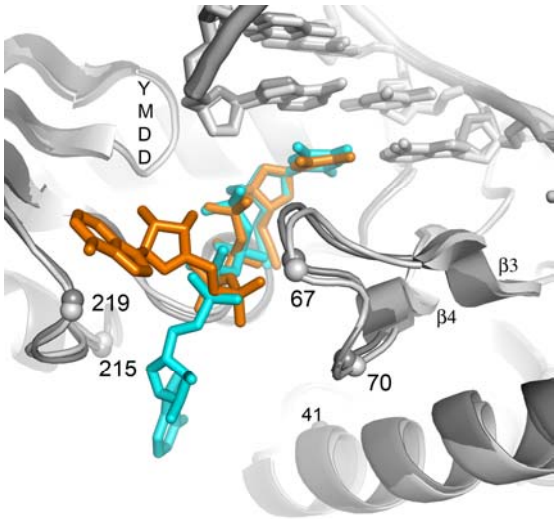


Fig. 48. Overall structure of AZT-resistant HIV-1 RT-DNA/AZTppppA

A



B

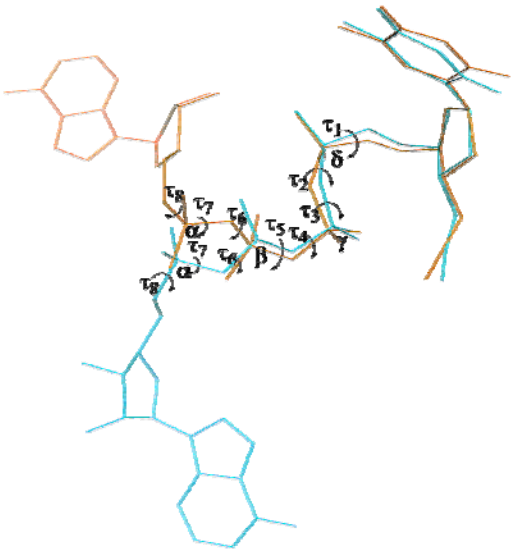


Fig. 49. Structural comparison of wild-type and AZT-resistant HIV-1 RT excision product complexes

(A) Superposition of the wild-type (orange) and the AZT-resistant (cyan) excision product complex reveals the repositioning of the ATP moiety of AZTppppA.

(B) Close-up view of superposed AZTppppA in the wild-type (orange) and the AZT-resistant (cyan) excision product complex.

Table 22. The torsion angles of phosphate groups of AZTppppA in wild-type and AZT-resistant excision product complexes.

Wild-type excision product complex	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8
complex copy 1	-65	108	26	-129	**-130**	-64	-80	-69
complex copy 2	-65	107	25	-128	**-129**	-67	-96	-71
complex copy 3	-64	106	14	-122	**-132**	-64	-77	-74
complex copy 4	-65	108	18	-123	**-130**	-67	-67	-67
Average	-65	107	21	-126	**-130**	-66	-80	-70
AZT-resistant excision product complex								
complex copy 1	-86	80	56	-112	**139**	178	-63	153
complex copy 2	-89	90	54	-113	**145**	176	-60	151
complex copy 3	-84	90	56	-113	**139**	178	-63	151
complex copy 4	-86	91	55	-110	**145**	175	-62	153
Average	-86	88	55	-112	**142**	177	-62	152

The torsion angles of phosphate groups are counted from the consistent part (AZTMP part) to the divergent part (ATP part). The torsion angles significantly differ at τ_5 representing the torsion angle of β'-phosphate group around the β',γ'-phosphodiester bond. The τ_5 differs about 90˚ in wild-type RT *versus* AZT-resistant RT.

Although the superposition of the two excision product complexes shows that both active sites align very well including the YMDD motif, AZTMP component, and β3-β4 region (Fig. 50), but it also reveals significant repositioning of the ATP moiety of AZTppppA in AZT-resistant RT *versus* wild-type RT (Fig. 51). The β',γ'-pyrophosphate of the ATP moiety remains coordinated with Mg^{2+} that is required for the catalysis, but the torsion angle of the β'-phosphate group around the β',γ'-phosphodiester bond (τ_5) differs about 90˚ in the AZT-resistant RT *versus* wild-type RT complex (Table 22). This difference in torsion angle τ_5 is the primary reason that the adenosine components are more than 10 Å apart in AZT-resistant RT versus wild-type RT. The repositioning of the ATP moiety in the same overall RT architecture lends us to propose that ATP binds differently to AZT-resistant RT *versus* wild-type RT.

7. Overall structural comparison of excision product complex with the wild-type HIV-1 RT polymerization catalytic complex (PDB id 1RTD) and the wild-type pre-translocation complex (PDB id 1N6Q)

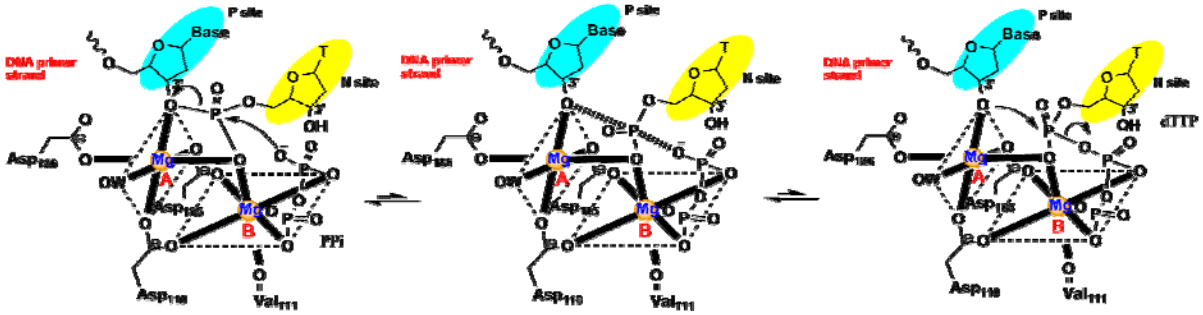
The enzymatic active site is composed of catalytic site and substrate binding site. The polymerization catalytic site of HIV-1 RT includes $Y_{183}M_{184}D_{185}D_{186}$ motif, β3-β4 region (residues 58 to 77) of the fingers subdomain, and p66 residues D110 and V111. The $Y_{183}M_{184}D_{185}D_{186}$ motif contains catalytic residues (D185 and D186) used for DNA polymerization and pyrophosphorolysis. The configuration of the $Y_{183}M_{184}D_{185}D_{186}$ motif is affected by assembling and disassembling the coordination

of Mg^{2+} (A) with D_{185} and D_{186}, since D185 and D186 are coordinate with Mg^{2+} (A). The β3-β4 region contains residues involved in binding the incoming dNTP.

7.1 Structural comparison of excision product complexes with the wild-type HIV-1 RT-DNA/dTTP catalytic complex

The wild-type HIV-1 RT-DNA/dTTP complex has a dNTP substrate trapped at the polymerization catalytic state. Since pyrophosphorolysis is the reverse of the polymerization reaction, the wild-type HIV-1 RT polymerization catalytic complex (HIV-1 RT-DNA/dTTP) can be also thought as an excision catalytic complex formed through pyrophosphorolysis with PPi. The Mg^{2+} (A) and Mg^{2+} (B) in the polymerization catalytic complex coordinate with all the polymerization/pyrophosphorolysis elements such as D185, 3'-OH, phosphate groups of dNTP, and so on. This two-magnesium ion coordination can also be seen from other polymerization catalytic structures from HIV or Dpo4 (Das et al., 2007; Vaisman et al., 2005). The polymerization by HIV-1 RT very possibly involves periodic assembling and disassembling of Mg^{2+} coordination, coincident with periodic conformational change of the $Y_{183}M_{184}D_{185}D_{186}$ motif. Polymerization is characterized by the addition of a dNMP, including the formation of a new phosphodiester bond at the primer 3'-end. On the other hand, the pyrophosphorolysis is characteristic of the break of the phosphodiester bond at the primer 3'-end. No matter what the reaction is (forward or backward), bond formation and break are performed on the same scaffold constructed by the coordination of Mg^{2+} with other

reaction elements. In other word, the the conformation of RT very possibly remains constant at pre-pyrophosphorolysis, pyrophosphorolysis, and post-pyrophosphorolysis states (Fig. 50).



Pre-pyrophosphorolysis (Post-polymerization) Transition state Post-pyrophosphorolysis (Pre-polymerization)

Fig. 50. Catalytic process of pyrophosphorolysis by PPi. Catalysis occurs on a scaffold constructed by the two Mg^{2+}. The structure of the enzyme may remain very similar in pre-pyrophosphorolysis, transition and post-pyrophosphorolysis states.

The excision product complexes can be regarded as an intermediate in which Mg^{2+} (A) coordination is disassembled after pyrophosphorolysis. Cα superposition of the structures of HIV-1 RT in the excision product complexes and the polymerization catalytic complex showed less than 0.8 Å of RMSD. The $Y_{183}M_{184}D_{185}D_{186}$ motifs and fingers subdomains of the two complexes aligned very well with no significant displacement (Fig. 51). The backbone conformations of the other catalytic elements of polymerization/pyrophosphorolysis are also almost identical. The configuration of the fingers subdomain of this polymerization catalytic complex has been referred to as the "closed" configuration (Huang et al., 1998) in relative to an "open" configuration of the fingers subdomain in unliganded RT. The functional part of the fingers subdomain is the β3-β4 region that accommodates residues interacting with the incoming dNTP. The β3-β4 region of the excision product complexes also aligned well with the β3-β4 region of the polymerization catalytic complex with less than 0.4 Å RMSD. The difference can be attributed to the flexibility of the β3-β4 loop. All the data together suggest that the the two excision product complexes very possibly share the same overall conformation as the enzymes in pre-pyrophosphorolysis, transition, and post-pyrophosphorolysis states.

7.2. Structural comparison of excision product complexes with wild-type pre-translocation complex

The AZT excision model proposes that the wild-type HIV-1 RT pre-translocation complex (complex N) is the ATP-binding intermediate for the pyrophosphorolytic

excision. It will be also interesting to determine if there are conformational differences of RT in the pre-translocation complex (before excision) and the excision product complexes (after excision).

Overall superposition of the structures of HIV-1 RT after excision (excision product complexes) and before excision (pre-translocation complex) showed that they are similar to each other with less than 0.8 Å RMSD for Cα atoms (Fig. 51), if the fingers subdomains of p66 were excluded from the superposition. However, the RMSD for Cα atoms was more than 2.0 Å, if the fingers subdomain of the p66 subunit were included in the superposition. This suggests a significant conformational difference in fingers subdomain. The fingers subdomain is in an open configuration in the wild-type pre-translocation complex; however the fingers subdomain has a closed configuration in the excision product complex. The open configuration of the subdomain in the wild-type pre-translocation complex might be caused by the crystal packing; the details are discussed in the next section. Structural comparison also shows that the Cα of D185 is displaced ~0.6 Å on average relative to its position in wild-type post-translocation complex, which is comparable to the displacement (0.5 Å) of Cα of D185 of polymerization catalytic complex relative to the wild-type post-translocation complex (Fig. 52).

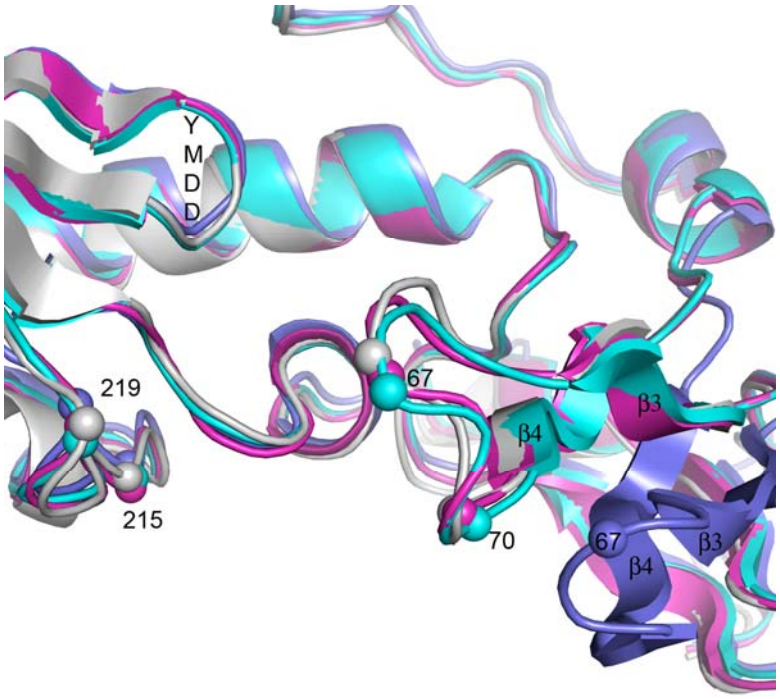


Fig. 51. Cα superposition of RTs from AZT-resistant excision product complex (magenta), wild-type excision product complex (cyan), wild-type polymerization catalytic complex (grey), and wild-type pre-translocation complex (blue)

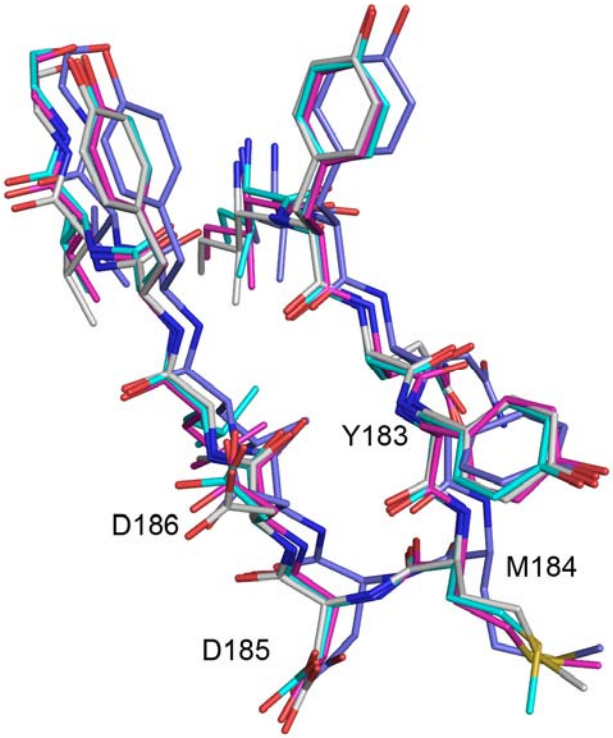


Fig. 52. Close-up view of superposed YMDD motif from AZT-resistant excision product complex (magenta), wild-type excision product complex (cyan), wild-type polymerization catalytic complex (grey), and wild-type pre-translocation complex (blue)

8. The four AZTppppA binding sites in the asymmetric unit of excision product complexes

As discussed before, the AZTppppA in both complexes are bound within the same cleft between the fingers and palm subdomain. The AZTMP moiety of AZTppppA is bound at the polymerization catalytic site. Since AZTppppA can be a polymerization substrate (Dharmasena et al., 2007; Meyer et al., 2007), in order to ensure formation of an excision product complex, the primer 3'-end was terminated with dideoxyadenosine triphosphate. The dideoxyadenosine-terminated primer 3'-end is located at the priming site (P site) of HIV-1 RT, and base-paired with template; both wild-type and AZT-resistance excision product complexes have clear electron density for the dideoxyadenosine.

8.1 The four AZTppppA binding sites in the asymmetric unit of the wild-type excision product complex

The four AZTppppA binding sites from the four copies of wild-type excision product complex are shown in Fig. 53. Except for the adenosine component, the AZTppppA molecules are well defined across all the four copies. The adenosine component is ill-defined in the context of wild-type RT. Although the first and third copies have weak continuous electron density extending to the adenine group, the position of adenine group can be roughly defined, since the round electron density blobs corresponding to the adenine group can allow multiple rotational fits. This explains why the structural superposition shows that the ATP moiety has increased RMSD from the ribose group and reach the highest level of more than 2.0 Å of RMSD at the

adenine group (Fig. 54) . In the first and third complex copies, water peaks that were found to bridge the ribose of adenosine component with neighboring residues may help to explain their relatively better electron density, but the water peaks could be artificial at this resolution level, besides their positions are not consistent with each other.

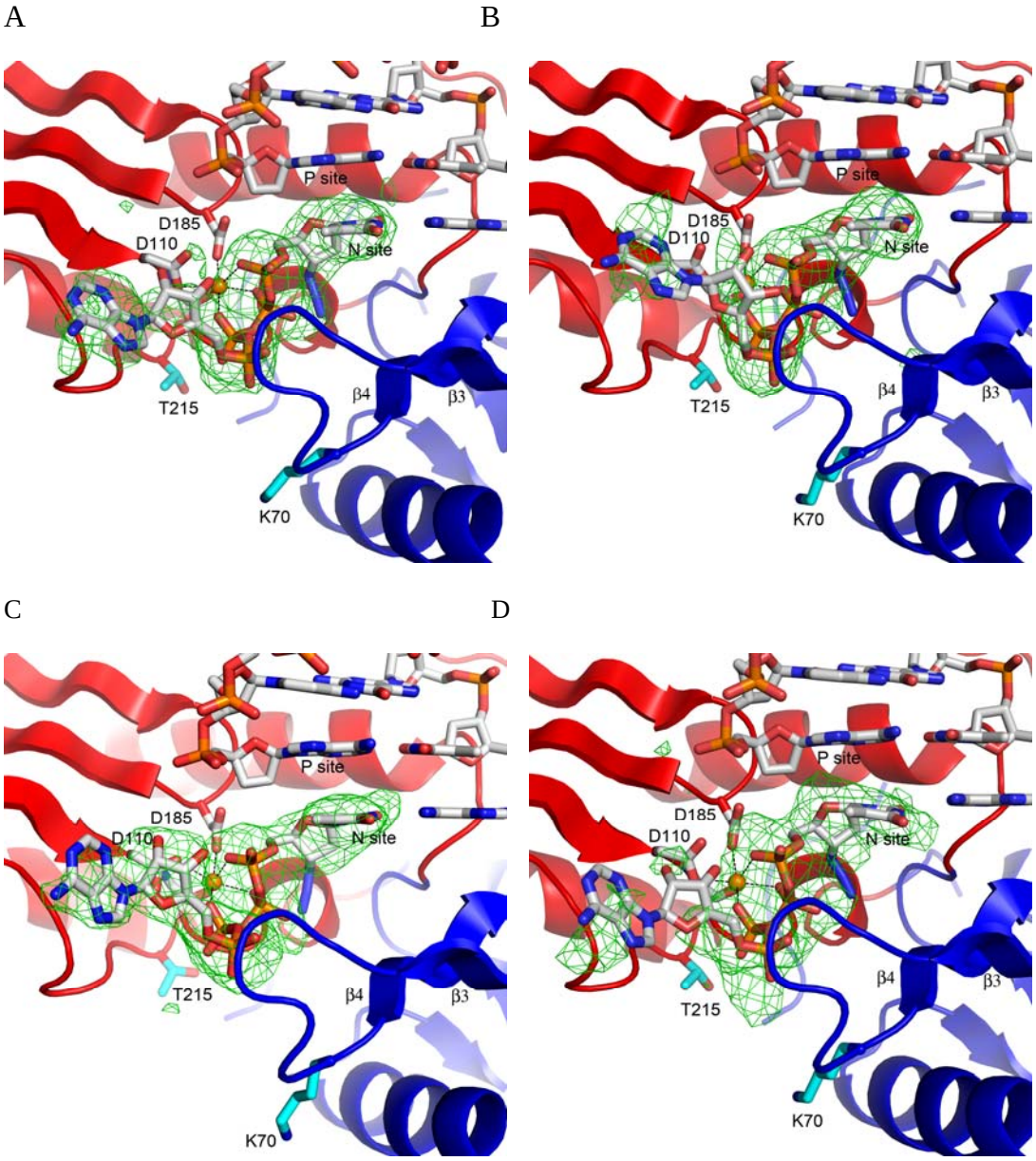
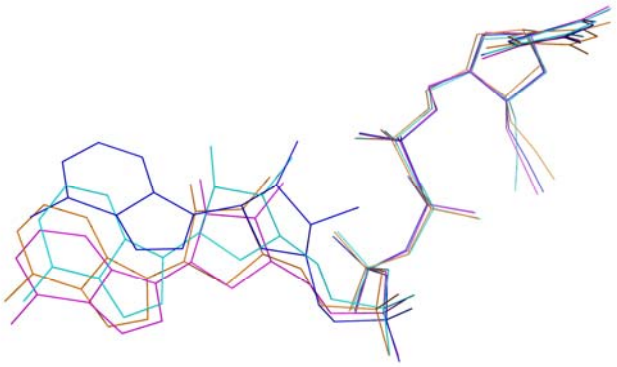
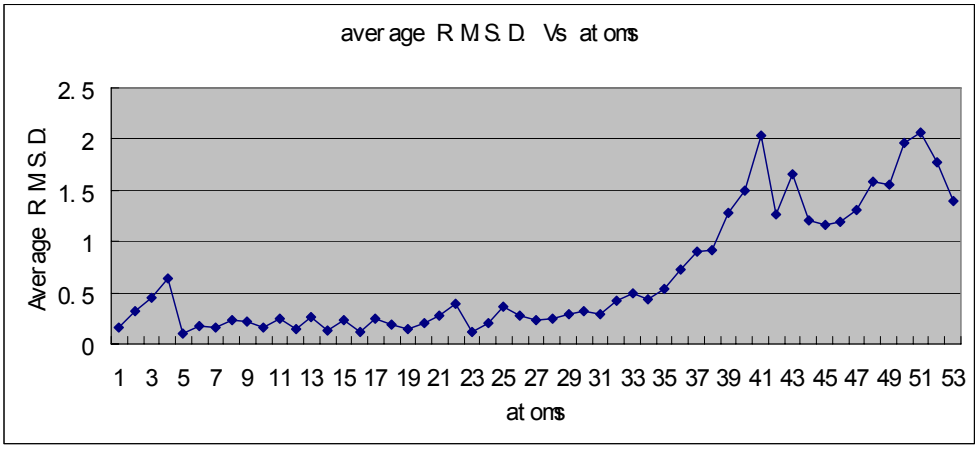


Fig. 53. Simulated-annealing difference omit map of AZTppppA from the four NCS-related copies of wild-type excision product complex (contoured at 3.0σ)

A



B



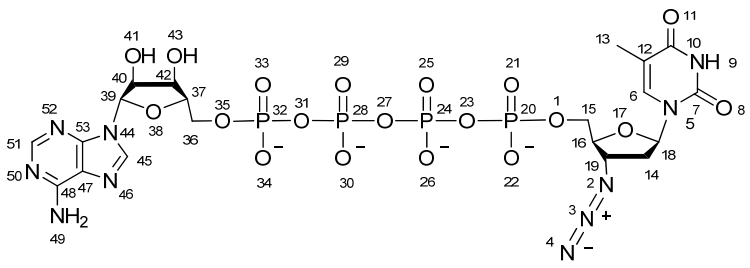


Fig. 54. Structural superposition of the four NCS-related AZTppppA molecules within the asymmetric unit of the wild-type excision product complex

(A) Structural superposition of the four NCS-related AZTppppA molecules

(B) Average RMSD at each atom

The four NCS-related AZTppppA are colored magenta, cyan, blue, and orange in order. Superposition is based on AZTMP moiety together with β',γ'-phosphate group of ATP moiety.

8.2 The four AZTppppA binding sites in the asymmetric unit of the AZT-resistant excision product complex

The four AZTppppA binding sites from the four copies of the AZT-resistant excision product complex are shown in Fig. 55. Unlike the AZTppppA in the context of wild-type RT, the AZTppppA has consistently well-defined electron densities in all the four copies of the AZT-resistant RT excision product complex. The position of the whole AZTppppA molecule is well defined, as evidenced from the structural superposition that shows that the RMSD is less than 0.1 Å (Fig. 56). The adenosine group is as well defined as the AZTMP, since the RMSD for the adenosine group is within the variation range of RMSD for the AZTMP moiety. The only difference in polymerase domain between the AZT-resistance RT and wild-type RT is the presence or absence of the five AZT resistance mutations, so the ordering of the adenosine component in AZT-resistance RT should be attributed to the AZT-resistance mutations.

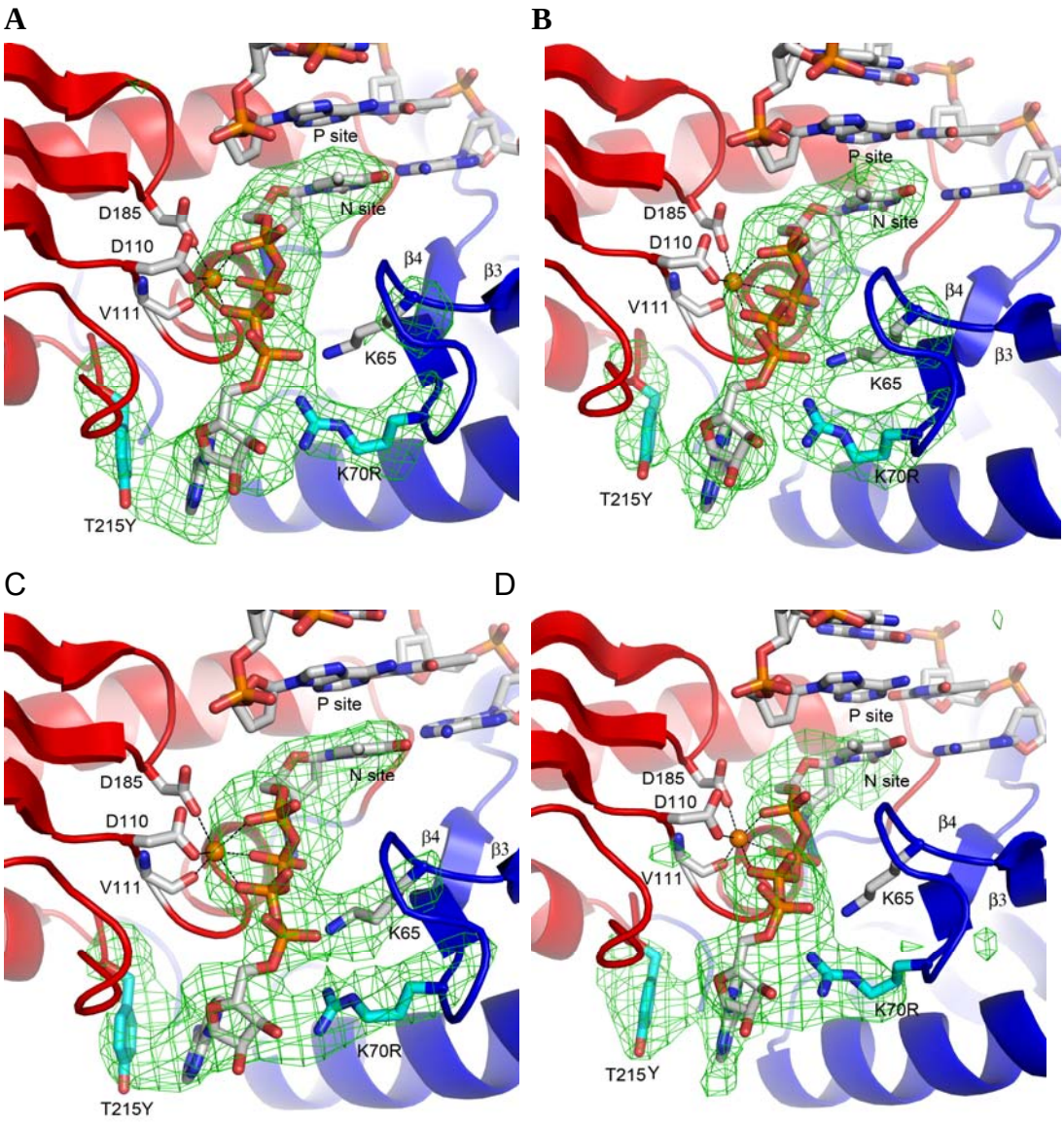
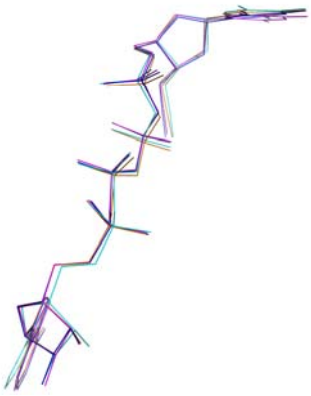
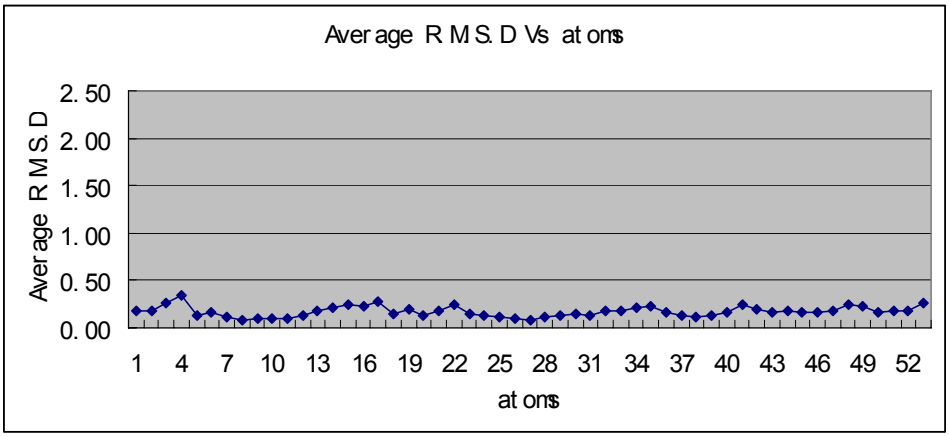


Fig. 55. Simulated-annealing difference omit map of AZTppppA from the four NCS-related copies of the AZT-resistant excision product complex

A



B



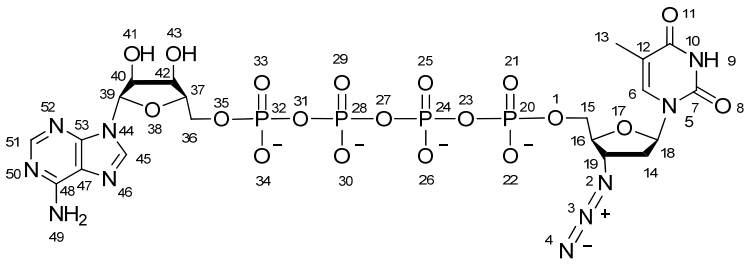


Fig. 56. Structural superposition of the four NCS-related AZTppppA molecules in the AZT-resistant RT excision product complex

(A) Structural superposition of the NCS-related AZTppppA molecules

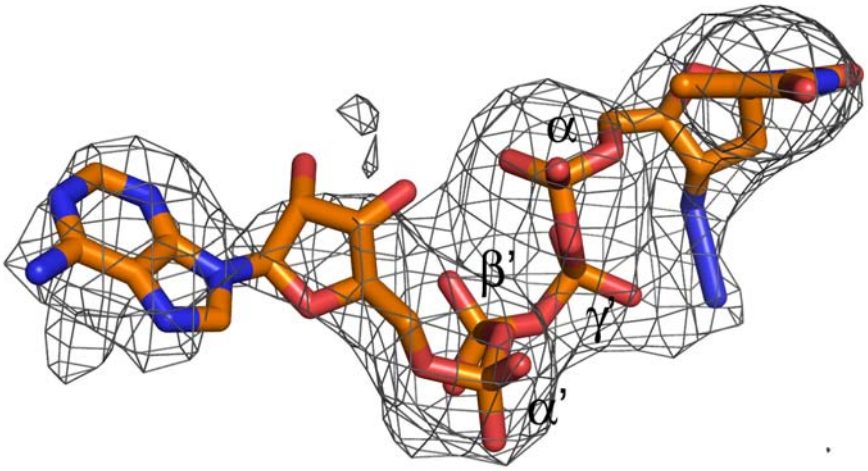
(B) Average RMSD at each atom

The four NCS-related AZTppppA are colored in magenta, cyan, blue, and orange in order.

9. Interactions involved in binding of AZTppppA in the wild-type and AZT-resistant excision product complexes

Since the best ordered complex copy has the best-defined electron density, it should provide the most reliable information on the interactions of AZTppppA. Thus, unless otherwise noted all the interactions discussed in binding of AZTppppA to RT are derived from the best ordered complex copy; the annealed difference omit maps of AZTppppA in wild-type and AZT-resistant RT are shown in Fig. 57. All the interactions involved in binding AZTppppA to wild-type and AZT-resistant RT are summarized in Fig. 58.

A



B

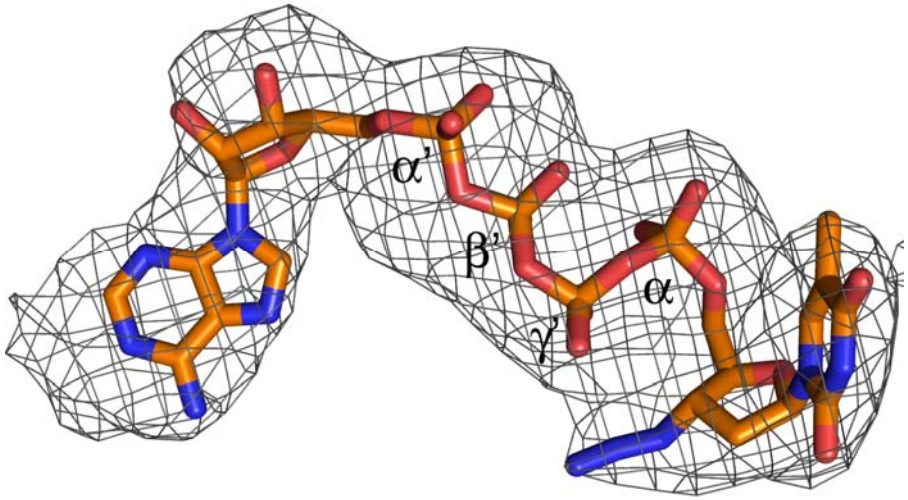
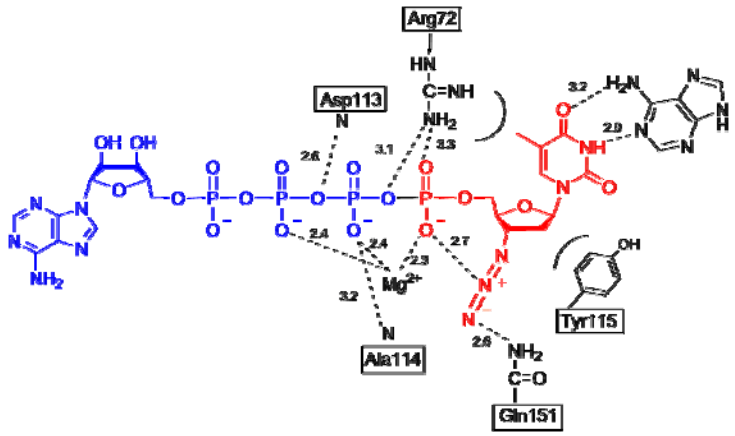


Fig. 57. Annealed difference omit map of AZTppppA in best-ordered wild-type (A) and AZT-resistant (B) excision product complexes (contoured at 3.0σ).

A



B

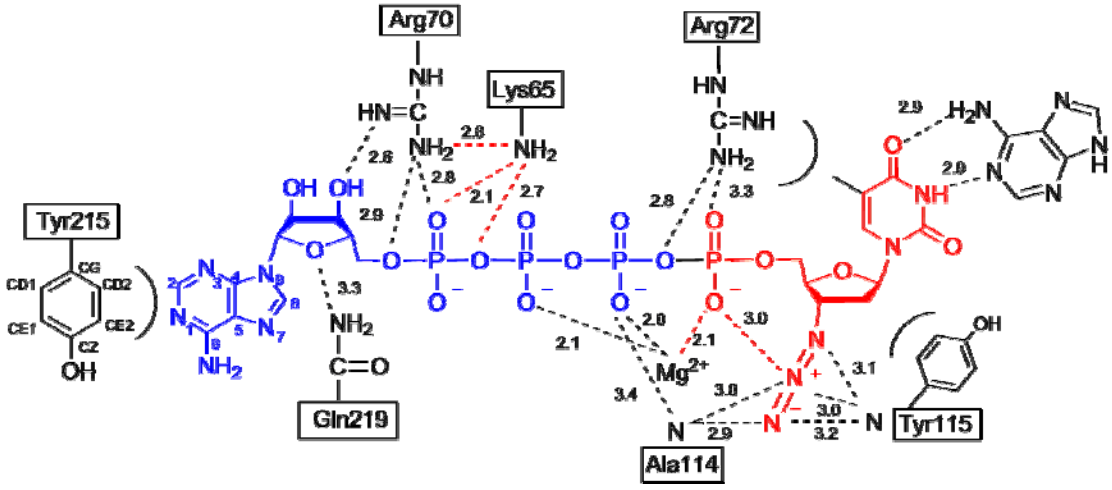


Fig. 58. AZTppppA interactions in wild-type and AZT-resistant excision product complexes (distance cutoff is 3.5 Å)

(A) Interactions of AZTppppA in the wild-type excision product complex.

(B) Interactions of AZTppppA in the AZT-resistant excision product complex.

Distance cutoff is 3.5 Å. Distances colored in red dash line are derived from the third NCS-related complex copy, and suggest that the residue K65 can compete R70 for the α-phosphate group of ATP moiety of AZTppppA.

9.1 interactions involved in binding of the AZTMP moiety of AZTppppA in the wild-type and the AZT-resistant HIV-1 excision product complex

The AZTMP moiety occupies the N site and is base-paired with the template. In comparison with the AZTMP part before excision, the ribose of the AZTMP moiety after excision undergoes ~30˚ of rotation resulting from the cleavage of the phosphodiester bond (Fig. 59). The conformation of the 3'-azido groups significantly differs in the excision product complexes (after excision) *versus* the wild-type HIV-1 RT pre-translocation complex (before excision). We believe this difference is caused by the flexibility of the 3'-azido group. In comparison with the dTTP at the N site, the binding of the AZTMP moiety and the following two phosphate groups (equivalent to AZTTP) to RT is similar to the binding of dTTP to the RT (Huang et al., 1998) (Fig. 59), and the ribose ring of AZTMP moiety aligns well with the ribose ring of dTTP.

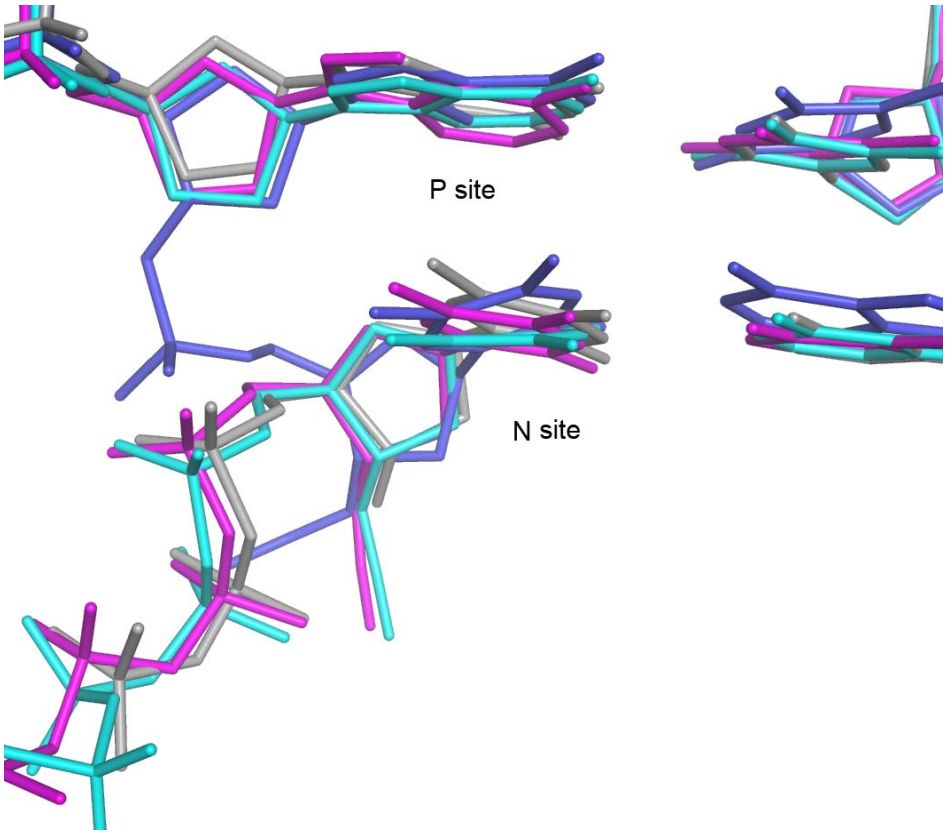


Fig. 59. Structural superposition of nucleotides at the N site from different RT-DNA complexes

Wild-type excision product complex (after excison; cyan); AZT-resistant excision product complex (after excision; magenta); polymerization catalytic complex (grey); and wild-type pre-translocation complex (before excision; blue).

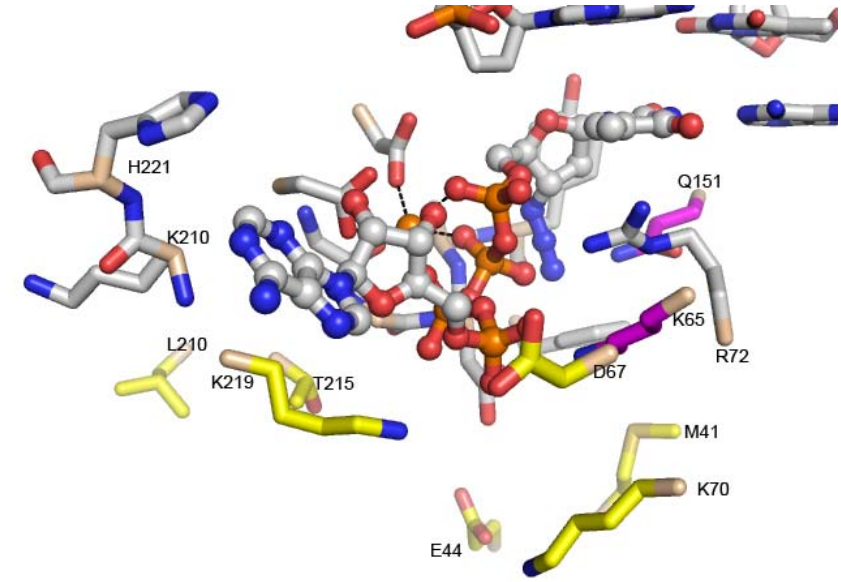
Interactions involved in binding of the AZTMP moiety to RT are common for both excision product complexes except for the 3’-azido group (Fig. 58). The guanidinium group of R72 stacks with the base of the AZTMP moiety, and forms salt bridges with the phosphate group of the AZTMP moiety. The aromatic side chain of Y115 stacks with the dideoxyribose ring of the AZTMP moiety. Although the 3’-azido group interacts with the α-phosphate group of the AZTMP moiety in both wild-type and AZT-resistant excision product complexes, it is observed that the 3’-azido group interacts with the side chain of Q151 in the wild-type excision product complex instead of the main-chain nitrogens of Y115 and A114 in the AZT-resistant excision product complex (Fig. 58). These interactions related to the 3’-azido group can possibly enhance the overall binding of AZTTP to HIV-1 RT during polymerization. A previous study has also shown that the 3’-azido group can interact with Ala114 main atoms and the AZTMP phosphate group (Sarafianos et al., 2002).

9.2 interactions involved in binding of the ATP moiety to the wild-type HIV-1 RT

Interactions involved in binding of the ATP moiety to wild-type RT are shown in Fig. 58. Besides the coordination of β', γ'-pyrophosphate part of the ATP moiety with the magnesium ion, the main-chain amide nitrogens of Asp113 and Ala114 lie on the bottom of the AZTppppA-binding cleft, about 2.56 and 3.19 Å, respectively, away from the bridging and non-bridging oxygens of the γ'-phosphate group of ATP moiety (Fig. 60B). Residues 219, 220, 221, and 228 are near the adenosine component of the ATP moiety of AZTppppA. There may possibly be some interactions between these residues and adenine group of AZTppppA, but the specific interactions can not be confidently defined, since adenine group is ill-defined in the four NCS-related complex copies of wild-type excision product complex. These residues may primarily provide a hydrophobic environment for the adenine group (Fig. 60A). The side chain of K65 is disordered in the wild-type HIV-1 excision product complex. No interactions are observed between the ATP moiety and any of side chains of the resistance mutation sites in the wild-type excision product complex. The two primary sites of resistance mutations (K70 and T215) in the wild-type excision product complex are even more than 7 Å away from the ATP moiety. The lack of interactions for the AMP part of the ATP moiety explains why its position is ill defined. These data suggest that the AMP part of ATP moiety may have no significant contributions to the binding of ATP to wild-type RT. The binding of ATP to the wild-type enzyme appears to rely mostly on the coordination of its β,γ-pyrophosphate moiety with a Mg^{2+}

ion. This observation can explain *in vitro* pyrophosphorolysis excision results showing that wild-type RT lacks significant selectivity among nucleotide triphosphates, nucleotide diphosphates, and pyrophosphate (PPi) (Meyer et al., 1998). The observation that PPi has somewhat higher excision efficiency than ATP or other nucleotides (Meyer et al., 1998; Boyer et al., 2001) can be explained as follows. The AMP part of ATP moiety makes no contribution to binding of ATP to wild-type RT, whereas PPi alone can more easily diffuse and coordinate with the magnesium ion to initiate excision.

A



B

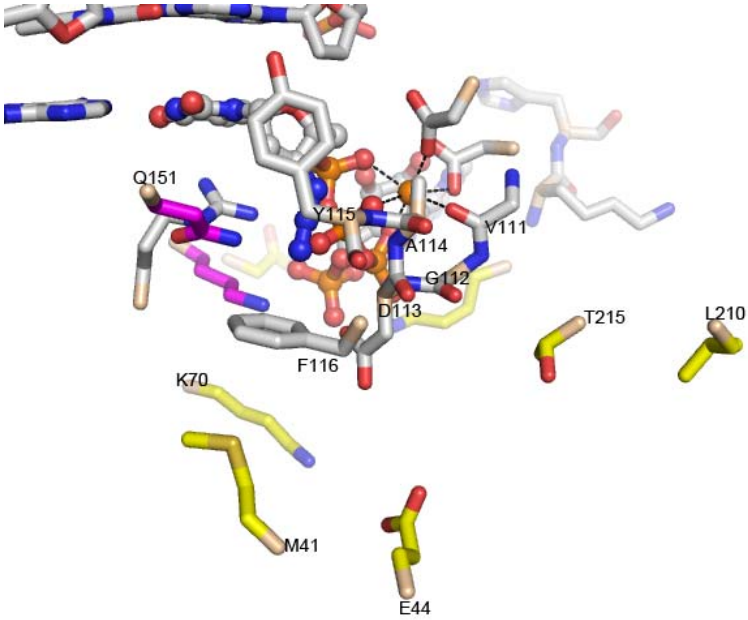


Fig. 60. The AZTppppA binding site in the wild-type RT excision product complex

(A) Front view

(B) Back view

Side-chain carbons of AZT-resistance mutation sites are colored in yellow; Side-chain

carbons of K65 and Q151 are colored in magenta; others are colored in gray.

9.3 Interactions involved in binding of the ATP moiety to the AZT-resistant HIV-1 RT

The β',γ'-pyrophosphate part of ATP moiety has similiar coordination and geometry in AZT-resistant RT as it is in the wild-type RT. However the AMP part of the ATP moiety is positioned in a space surrounded by the AZT-resistance mutations including primary, secondary, and other associated AZT-resistance mutations. The mutant R70 side chain interacts extensively with the ATP moiety (Fig. 61). One guanidinium η-nitrogen of R70 forms a hydrogen bond with the 3'-OH of the ATP moiety, and the other guanidinium η-nitrogen of R70 forms electrostatic interactions with a non-bridging oxygen of the α-phosphate and the ribose O5 of ATP moiety. The aromatic ring of Y215 stacks with the adenine ring of the ATP moiety and creates π-π interactions, giving solid proof to the previous conjectured role of T215Y mutation (Boyer et al., 2001; Yahi et al., 2005). The π-π stacking primarily occurs with the six-membered ring of the adenine group (Table 23), which explains why purine nucleotides have higher excision rate than pyrimidine nucleotides (Meyer et al., 1999); pyrimidines do not contain a second ring in the analogous location. Recent biochemical studies have shown that mutation K70E can substantially impair the ability of AZT-resistance RT to excise the terminal AZTMP from the primer 3'-end (Sluis-Cremer et al., 2006). Since the side-chain charge changes from positive to negative, we expect the mutation K70E will disfavor the binding of ATP to RT by repelling the negatively charged α-phosphate of ATP.

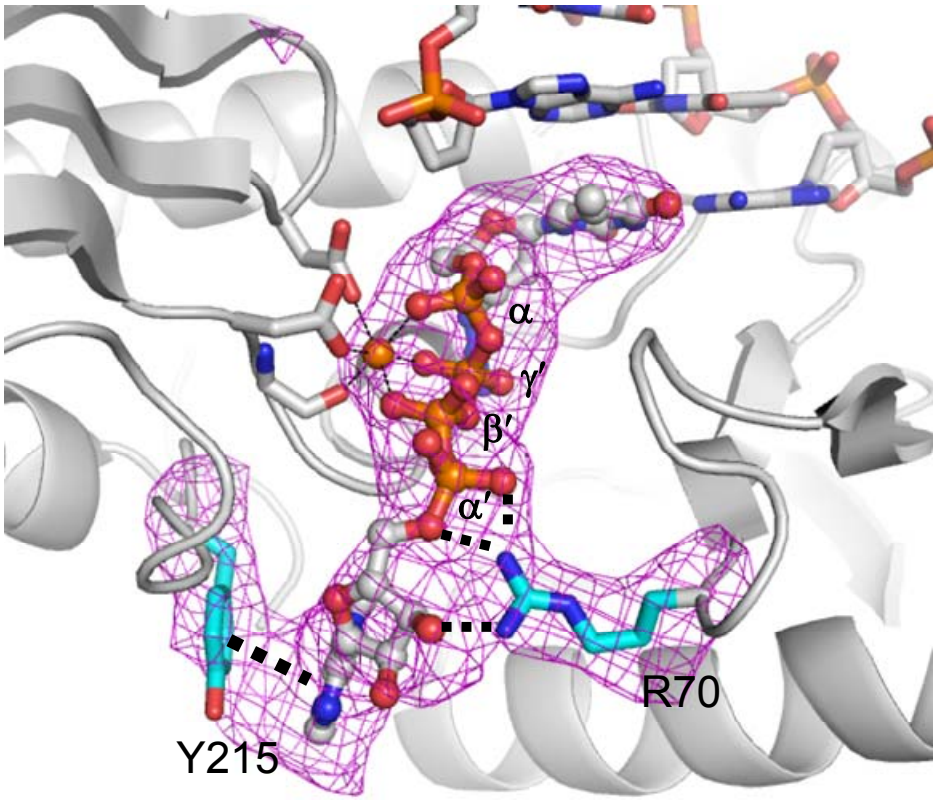
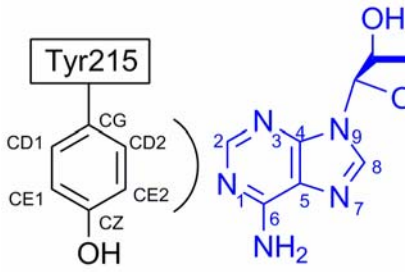


Fig. 61. Simulated-annealing difference omit map of AZTpppppA, R70, and Y215 (contoured at 4.0σ)

Table 23. The contacts between Tyr215 and AZTppppA

Tyr215	AZTppppA	distance (Å)	Tyr215	AZTppppA	distance (Å)
CG [C]:	N1[N]:	4.48	CE2[C]:	N1[N]:	4.01
	C5[C]:	4.45		N3[N]:	4.15
	C6[C]:	4.14		C2[C]:	3.90
	N6[N]:	4.22		C4[C]:	4.48
CD1[C]:	N1[N]:	3.92		C6[C]:	4.35
	C6[C]:	3.76	CZ [C]:	N1[N]:	3.39
	N6[N]:	3.63		N3[N]:	4.21
CE1[C]:	N1[N]:	3.34		C2[C]:	3.51
	C2[C]:	3.97		C6[C]:	3.99
	C6[C]:	3.69	OH [O]:	N1[N]:	3.59
	N6[N]:	3.82		N3[N]:	4.27
CD2[C]:	C5[C]:	4.48		C2[C]:	3.41
	C6[C]:	4.41			



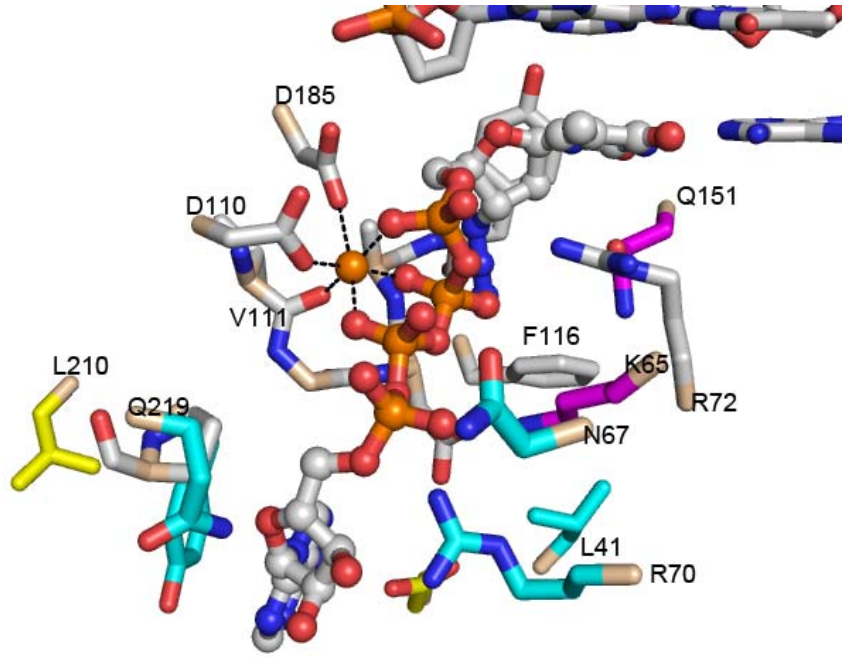
The π-π interaction mainly occurs between the six-membered adenine ring of the ATP moiety of AZTppppA (distance cutoff at 4.5 Å). The coordinates of T215Y and AZTppppA are from the third AZT-resistant excision product complex copy that presents the best π-π interaction in the crystal asymmetric unit.

Some secondary mutations can assist in binding the ATP moiety based on the structural data. The side chain of mutation K219Q is about 3.2 Å away from the O4 of the ribose ring, and may act as a platform for stabilizing the ATP ribose ring position and conformation. Although the mutation D67N is about 8 Å away from the α-phosphate group of the ATP moiety, its side chain faces directly toward this α-phosphate group and lies on the same side of the AZTppppA-binding cleft as the side chains of K65, K66, and R70 (Fig. 62A). This mutation may provide a better environment for accommodating negatively charged ATP, since the D67N mutation neutralizes the negative charge of D67. Mutation M41L is proximal to F116 and E44, with a distance of about 3.9 Å and 3.4 Å at the closest, respectively. Furthermore, the side chain of F116 is adjacent to the side chain of Q151 and the 3'-azido group of the AZTMP moiety, with distances of about 3.1 Å and 3.0 Å at the closest, respectively (Fig. 62B). The mutated residue L41 may have indirect impact on the interactions between Q151 and the 3'-azido group; the 3'-azido group interacts differently with Q151 in the context of wild-type enzyme and AZT-resistant enzyme (Fig. 58). L210W is a secondary AZT-resistance mutation that does not exist in the AZT-resistant RT used in our experiment. The L210W mutation usually occurs in combination with T215Y, and can significantly enhance AZT resistance (Hooker et al., 1996). The structure of the AZT-resistant excision product complex reveals that the L210 is adjacent to Y215 (Fig. 62B). Modeling shows that the aromatic side chain of W210 may stack with the aromatic side chain of Y215 at a distance of about 3.3 Å and

enhance the π-π stacking between Y215 and ATP, consistent with previous modeling studies of the L210W mutation (Hooker et al., 1996; Yahi et al., 2005).

E44A/D is a mutation strongly associated with primary mutation T215Y instead of K70R (Romano et al., 2002; Marcelin et al., 2004). E44A/D contributes to the resistance of NRTIs by enhancing the ATP-mediated excision in the presence of primary mutation T215Y (Girouard et al., 2003). In the structure of the AZT-resistant HIV-1 RT excision product complex, residue E44 is proximal to the adenine of the ATP moiety, about 3.6 Å from adenine N6 of the ATP moiety (Fig. 58B; Fig 62B). This mutation could possibly improve the π-π stacking by disrupting its interaction with the adenine group of ATP, since Y215 and E44 are on either sides of the adenine group. The past and current viewpoints about the roles played by AZT-resistance mutations are summarized in Table 24.

A



B

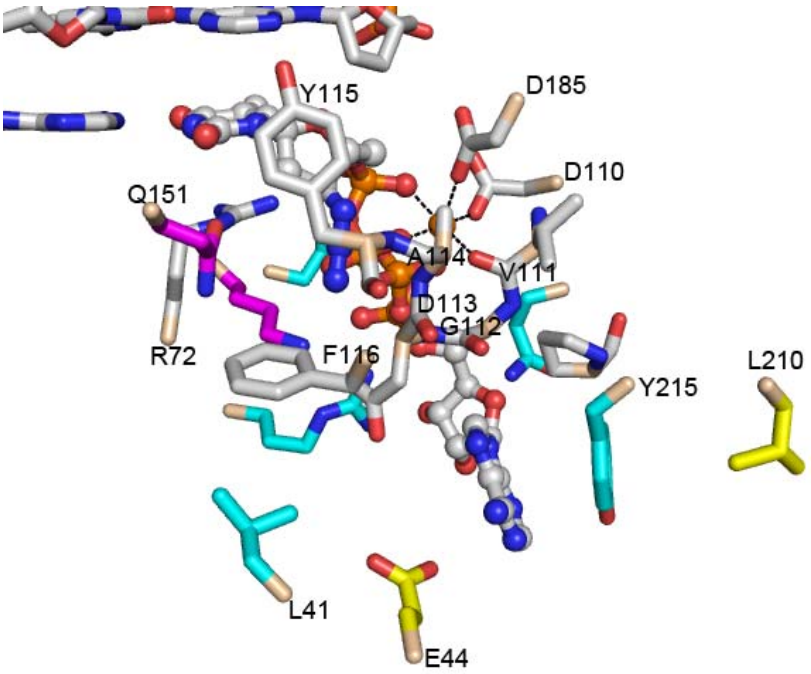


Fig. 62. The AZTppppA binding site in AZT-resistant RT

(A) Front view

(B) Back view

Side-chain carbons of AZT-resistance mutants and associated mutation sites are colored in cyan and yellow, respectively; side-chain carbons of R65 and Q151 are colored in magenta; side chain carbons of other residues are colored in gray.

The electron density maps show that the first three complex copies fortuitously give different snapshots of the K65 and R70 side-chain interactions with AZTppppA (Fig. 55). The fourth complex has a very high B factor, so it was not included from our structural analysis. These snapshots together reveal the dynamic interactions of K65 and R70, and may help to decipher the antagonistic relationship between the K65R mutation and AZT-resistance mutations. The first copy has the strongest electron density showing interactions of R70 with the α'-phosphate group of ATP moiety, but no electron density showing the side chain of K65. The second copy has weaker electron density showing the interactions of R70 with the α'-phosphate group of ATP moiety, but continuous electron density covering the whole side chain of K65. The third copy has no electron density showing the interactions of R70 with the α'-phosphate group of ATP moiety, but has the strongest electron density showing the side-chain of K65 and its interaction with the α'-phosphate group of ATP moiety. The R70 side chain is displaced more than 3.6 Å away from the non-bridging oxygen of α'-phosphate group of ATP moiety. These snapshots show two things. One is the residue K65 can interact α-phosphate group of ATP. The other one is that the K65 side chain can compete with the mutant R70 for the non-bridging oxygen of the α-phosphate group of the ATP moiety (Fig. 55). The basis for this competition possibly is that the non-bridging oxygen of the α'-phosphate group of the ATP moiety is unable to accommodate two positively charged nitrogen-containing groups simultaneously, perhaps owing to electrostatic repulsion. The K65R mutation was originally isolated from patients treated with ddC or ddI, and subsequently shown to confer resistance to TDF and foscarnet (Gu et al., 1994; Gallant et al., 2003; Meyer et al., 2003). Biochemical studies have shown that the K65R is antagonistic to ATP-mediated pyrophosphorolysis and can re-sensitize AZT to AZT-resistance mutations (Miller et al., 2005). The introduction of a more positively-charged side chain will make the interactions of residue 65 with the α-phosphate group stronger. It may cause the following effects: disrupting the interaction of R70 with the α-phosphate group of the ATP by displacing the R70 side chain, tilting the adenine group to attenuate the π-π stacking between ATP and Y215, and/or even tilting the

β,γ-pyphosphate group of ATP. As a result, K65R mutation will reduce the enhanced excision caused by AZT-resistance mutations.

Table 24 The summary of roles played by AZT-resistance mutations

Resistance mutations	Past viewpoint	Current viewpoint
T215Y/F	π-π stacks with aromatic adenine group of ATP	Mainly π-π stacks with the six membered-ring of adenine group of ATP
K70R	Mutant R70 may make a new hydrogen bond with γ-phosphate group of dNTP or D113	Interacts with the α-phosphate group and 3'-OH group of ATP
M41L	Causes indirect rearrangement of dNTP-binding pocket through residue 116	L41 is proximal to E44 and F116, and may disrupt the interactions between Q151 and the 3'-azido group.
L210W	π-π stacks with T215Y, and enhances the π-π stack between T215Y and adenine group of ATP	Same
D67N	Disrupts interaction with residue 219	N67 facilitates ATP access and binding by neutralizing its negative charge
K219Q	Affects interactions of residue 219 with residue 67 across the cleft, as well as formation of the 3' pocket	Interacts with O4 of ribose ring of ATP, and may stabilize the ribose ring
E44A/D	Affects interactions of residue 219 with residue 67 across the cleft, as well as formation of the 4' pocket	About 3.6 Å from adenine N6 of the ATP moiety, and E44 may interfere with the π-π stacking between Y215 and ATP.

10. Comparison of the ATP binding sites in wild-type and AZT-resistant RT

As discussed before, the AMP part of ATP moiety assumes different conformations in wild-type HIV-1 RT versus AZT-resistant HIV-1 RT as reflected in the difference in the torsional angle of the β'-phosphate group of the ATP moiety. This conformational difference in the AMP part should be mainly caused by primary mutations, since the major difference in the two binding site is the absence and presence of the primary resistance mutations. The electrostatic potential surface of both wild-type and AZT-resistant RT in the excision product complexes clearly shows two sites that can potentially accommodate ATP (Fig. 63). In the wild-type excision product complex, Site I has a broad hydrophobic surface toward the adenine plane that can help to trap the ATP, but Site II is filled with only solvent (Fig. 60A). In the AZT-resistance excision product complex, Site I is similar to the one in the wild-type enzyme, but Site II is filled with the side chains of K70R and T215Y (Fig. 60B). This suggests that the AZT-resistance mutations do not alter the ATP-binding site in wild-type RT, but rather create a new site that has more interactions with ATP. The primary resistance mutations interact extensively with the AMP part of ATP in site II, and dramatically enhance the binding of ATP to AZT-resistant RT. This analysis in consistent with recent biochemical studies that have shown the AZTppppA has more than 100-fold greater binding affinity to AZT-resistant HIV-1 RT than to wild-type HIV-1 RT (Dharmasena et al., 2007; Meyer et al., 2007). We believe that the ATP has potential to bind both sites. It prefers Site I in wild-type RT, and Site II in AZT-resistant RT. The mutant R70 side chain might contribute more to ATP-mediated

pyrophosphorolysis than that of Y215, since R70 has more extensive interactions with ATP and is closer to the reacting pyrophosphate moiety. Once the positions of the α-phosphate and ribose of ATP are locked, the β,γ-pyrophosphate moiety can be positioned for excision, even if the adenine component is disordered or has multiple conformations. This analysis is consistent with primer unblocking assays based on single nucleotide excision. These assays have shown that the RT variant containing all of the resistance mutations has the highest excision rate, followed by variants containing only K70R and D67N, or only T215Y and K219Q mutations (Meyer et al., 1999).

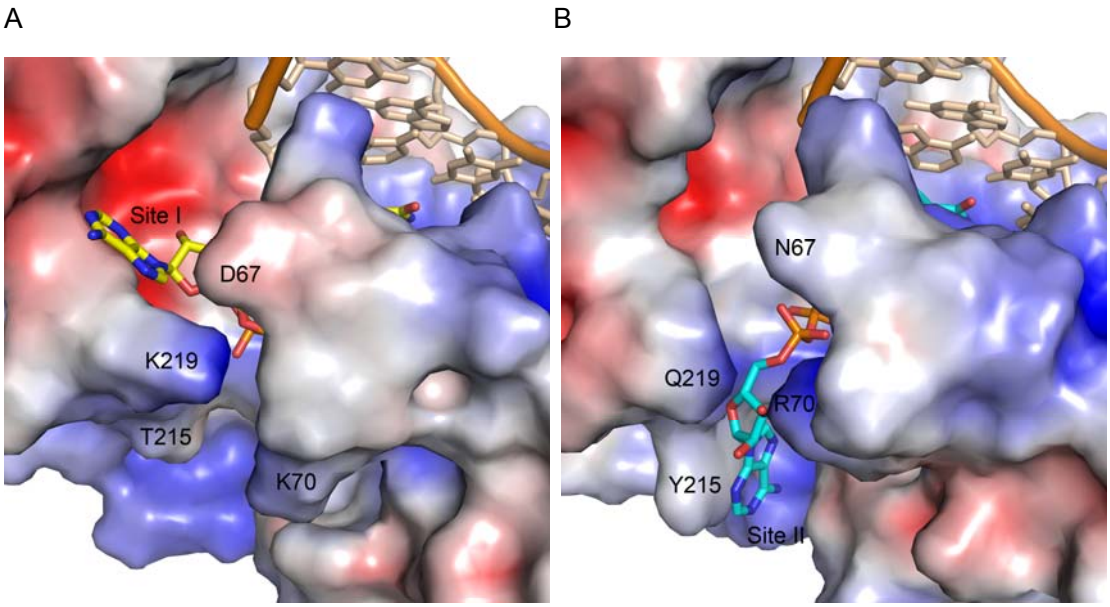


Fig. 63. Electrostatic potential surfaces of AZTppppA binding sites in wild-type and AZT-resistant RT

(A) The electrostatic potential surface of the wild-type excision product complex

(B) The electrostatic potential surface of the AZT-resistant excision product complex. The carbon atoms of AZTppppA from the wild-type excision product complex are colored in yellow; the carbon atoms of AZTppppA from the AZT-resistant excision product complex are colored in cyan.

11. Two Mg^{2+} ion catalysis

Similar to the previously solved HIV-1 RT-DNA/tenofovir ternary complex (Tuske et al., 2004), only one Mg^{2+} ion was observed in both excision product complexes. The Mg^{2+} has the same coordination and geometry in the two excision product complexes and HIV-1 RT-DNA/tenofovir ternary complex. The Mg^{2+} is structurally similar to the Mg^{2+} B of the polymerization catalytic ternary complex (Huang et al., 1998) and the Mn^{2+}(B) of a RT/ATP complex (Das et al., 2007), but catalyzes the pyrophosphorolytic excision instead of binding dNTP. This Mg^{2+} is coordinated by two carboxylate groups from D110 and D185, and two non-bridging oxygens from the β- and γ-phosphate groups of the ATP moiety (Fig. 62A). A non-bridging oxygen from the phosphate group of the AZTMP moiety and a backbone carbonyl from V111 then cap the bottom and top to complete the octahedral coordination arrangement (Fig. 62A).

Although only one magnesium ion was observed in these two excision product complexes, we believe that pyrophosphorolysis employs a two-metal-ion mechanism, since polymerization is catalyzed by two metal ions. If the two-metal-ion forward reaction has the lowest activation energy, then the two-metal-ion back reaction should also have the lowest activation energy according to the principle of microscopic reversibility. The absence of one magnesium ion in the excision product complexes is possibly due to the experimental design in which the primer 3'-end has been terminated by a dideoxynucleotide (to prevent incorporation) that lacks the 3'-OH to

coordinate the second metal ion. During ATP-mediated pyrophosphorolysis, the two Mg^{2+} ions assemble the β,γ-pyrophosphate moiety of ATP together with the scissile phosphate group of the AZTMP of the primer 3'-end, so that the γ-phosphate group of ATP can cleave the phosphodiester bond of the terminal AZTMP (Fig. 64).

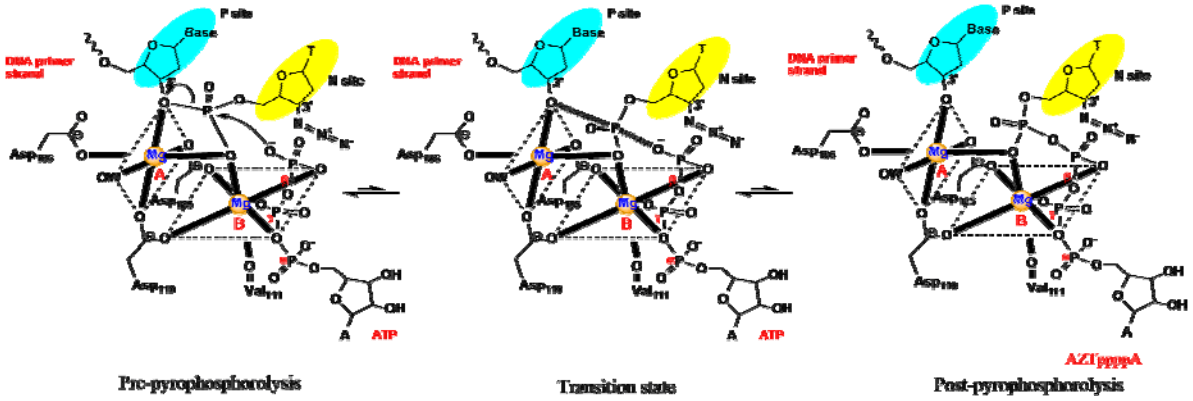


Fig. 64. The model of two metal ion catalysis of ATP-mediated pyrophosphorolysis

12. The pathway of ATP-mediated pyrophosphorolysis

The pyrophosphate moiety of ATP during the ATP-mediated excision experiences four stages: access into the binding site, pre-excision, post-excision (staying coordinated with Mg^{2+} ions as part of AZTppppA), and leaving as part of AZTppppA. The two excision product complexes provide us with images of the pyrophosphate moiety in post-excision state, and can be used to deduce the structure of the pre-excision state of pyrophosphate moiety, since it stays coordinated with the Mg^{2+} ions in both states. The crystal structure of a T7 RNA polymerase pre-insertion complex with PPi shows that the PPi is bound to Arg627 and is coordinated with a Mg^{2+}ion. A water molecule replaces the α-phosphate of incoming NTP to complete the octahedral geometry of Mg^{2+}ion coordination, but the position of β,γ phosphates and Mg^{2+}ion are away from where they are in polymerization state. The Arg627 is structurally analogous to Lys65 of HIV-1 RT. This snapshot represents a state after the formation of PPi, but it also raises a possibility that the Mg^{2+}ion first assembles the pyrophosphate moiety of ATP together with a water molecule instead of the scissile phosphate of primer 3'-end. The scissile phosphate of the primer 3'-end then replaces the water molecule and elicits the excision reaction. A crystal structure of a foot-and-mouth disease virus (FMDV) polymerase elongation complex also provides us an image of PPi binding. In this crystal structure, the PPi binds to Arg168 and Ala243 through charged interaction, but no metal ion is observed in the binding site. This snapshot shows a state in the late stage of PPi leaving, but possibly also represents an early state of PPi accessing the binding site. The Arg168 is structurally similar to

Arg627 of T7 RNA polymerase and Lys65 of HIV-1 RT.

A complete pathway of ATP-mediated pyrophosphorolysis excision has already been proposed (Sarafianos et al., 2002; 2004). Our results add more details to this ATP-mediated pyrophosphorolysis excision pathway. The steric conflict between the azido group of AZT and the α-phosphate group of the incoming dNTP destabilize the dead end complex, and hence the primer 3'-end has more chances to access to the N site through thermal motion. The Brownian motion drives fingers opening and closing; the fraction of pre-translocation intermediate with fingers in a closed conformation is the ATP-binding intermediate. The primary AZT-resistance mutations facilitate the ATP binding and access of pyrophosphate moiety into the binding site. The pyrophosphate moiety is firstly stabilized in the binding site through charged interactions with R70 and K65. The Mg^{2+}ion then assembles the reaction components through octahedral coordination. It is possible that a water molecule is assembled first, and is then replaced by the scissile phosphate of the primer 3'-end. No matter what path the scissile phosphate takes, the γ-phosphate is finally able to get close to the scissile phosphate and cleave the phosphodiester bond that links the AZTMP with the upstream nucleotide. After excision, the pyrophosphate moiety became part of excision product AZTppppA, and temporarily stays coordinated with Mg^{2+}ions. The leaving of AZTppppA is also possibly through interaction with Lys65 and the Brownian movement of the fingers.

13. Other factors that affect the pyrophosphorolytic excision

During ATP-mediated pyrophosphorolysis, the most critical step for completing octahedral coordination is trapping of the primer 3'-end, thus any factors affecting the position of primer 3'-end will affect the ATP-mediated pyrophosphorolysis. Another critical step is the positioning of the β, γ-pyrophosphate moiety of ATP, which can be affected by the interactions of the adenosine and α-phosphate moiety of ATP. Any factors that affect positioning of reaction elements will have an effect on the pyrophosphorolytic rate.

The position of the primer 3'-end determines the position of the scissile phosphate. The position of the primer 3'-end can be affected by the "primer grip" and "template grip". The primer grip is the β12-β13 hairpin (residues 227-235). Hydrophobic residues Phe227, Trp229, and Leu234 of the primer grip also lie at the non-nucleoside drug-binding site. However, none of the most common mutations that confer resistance to NNRTIs are found in the primer grip, suggesting that the enzymatic activity can barely tolerate conformational alterations of primer grip. The ``template grip" includes regions that interact with the DNA sugar-phosphate backbone of the first four nucleotides of the template strand. These include β4 and αB (residues 73-83) of the fingers subdomain and β5a (residues 86-90), β5a-β5b connecting loop (residues 91-93), and β8-αE (residues 148-154) connecting loop of the palm subdomain. Foscarnet is a pyrophosphate analog. Some foscarnet-resistance mutations introduced by site-directed mutagenesis are located in the "template grip"

regions, for example mutations at codons 72, 89, 90, 151, 154. The residue 89 has direct contact with phosphate backbone of template. The mutations W88S, W88G, Q161L, and H208Y were originally identified in clinical isolates of HIV-1 from foscarnet-treated patients (Mellors et al., 1995). The structures of excision product complexes show that Q161 may indirectly interacts with Y115 through M184. Although H208 is not in the primer grip, its side chain is sandwiched by the aromatic side chains of Trp212 and Phe171. H208Y mutation possibly creates an opportunity for π-π stacking among these residues, and causes local rearrangement, which will displace the primary mutation T215Y. The aromatic side chain of Y115 lies flat against the deoxyribose of AZTMP component, so its position will be very critical to the primer 3'-end. The nature of primer 3'-end base and the six base pairs upstream of it can also strongly affect the pyrophosphorolysis excision, and the efficiency of pyrophosphorolysis can vary several hundred-fold in different sequence contexts (Boyer et al., 2002; Meyer et al., 2004; Sluis-Cremer, et al., 2005). The L74 side chain supports the template base that is base-paired with the primer 3'-end base. The L74V mutation shortens the side chain by about 1.5 Å, and thus may shift the position of the template base and, in turn, the primer 3'-end. L74V is a resistance mutation to ABC, ddI and 3TC, but biochemical data show that it can severely impair the pyrophosphorolysis excision ability of AZT-resistance mutations (Miranda et al., 2005; Frankel et al., 2005).

PART IV. STRUCTRAL ANALYSES OF AZT-RESISTANT HIV-1 RT PRE-TRANSLOCATION COMPLEX, AZT-RESISTANT HIV-1 RT POST-TRANSLOCATION COMPLEX AND UNLIGANDED AZT-RESISTANT HIV-1 RT

The AZT-resistant HIV-1 RT contains five resistance mutations including M41L, D67N, K70R, T215Y and K219Q. The pre-translocation and post-translocation complexes are so called for the reason that their AZTMP terminated primer 3'-end is in pre-translocation and post-translocation state, respectively. The pre-translocation and post-translocation complexes are also referred to as complex N and complex P (Sarafianos et al., 2002), respectively, since the pre-translocated primer 3'-end is at the N site and post-translocated primer 3'-end is at the P site. These two complexes were trapped by cross-linking the AZT-resistant HIV-1 RT (a cysteine was introduced into the thumb subdomain of p66 at position 258) to a modified dsDNA (a thiol-bearing tether was attached to a guanosine of primer). The primer was designed to have a modified guanosine at the fifth upstream nucleotide from the 3'-end. Since the specificity of this cross-linking reaction required the modified guanosine at the sixth register upstream from the P site, the post-translocation complex were trapped by incorporation of an AZTMP followed by translocation (Fig. 65A), and the pre-translocation complex were trapped by incorporation of a dAMP followed by translocation and incorporation of an AZTMP (Fig. 65B).

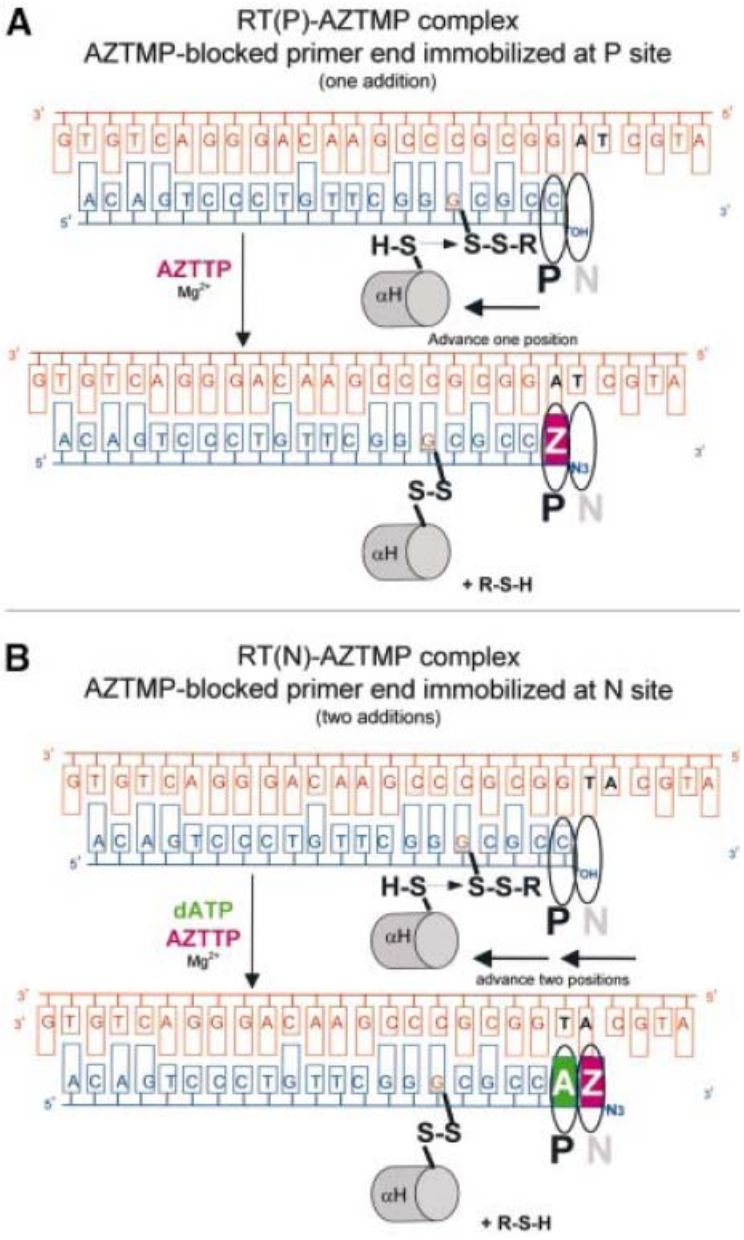


Fig. 65. Scheme for trapping the pre-translocation complex and post-translocation complex (Sarafianos et al., 2002)

The AZT-resistant pre-translocation complex was determined to 3.7 Å resolution, and refined to R_{work} of 0.288 and R_{free} of 0.338, respectively. The AZT-resistant post-translocation complex was determined to 2.90 Å, and refined to R_{work} of 0.257 and R_{free} of 0.257, respectively. The unliganded AZT-resistant RT was determined to 2.65 Å, and refined to R_{work} of 0.251 and R_{free} of 0.294, respectively.

1. Structural validation

1.1 AZT-resistant HIV-1 RT pre-translocation complex

Although the structure is only determined to 3.7 Å, the crystal structure of AZT-resistant HIV-1 RT has relatively good geometry quality. The Ramachandran plot shows that 97.5% of residues fall in the allowed regions and 0.8% of residues fall in the disallowed regions (Fig. 66).

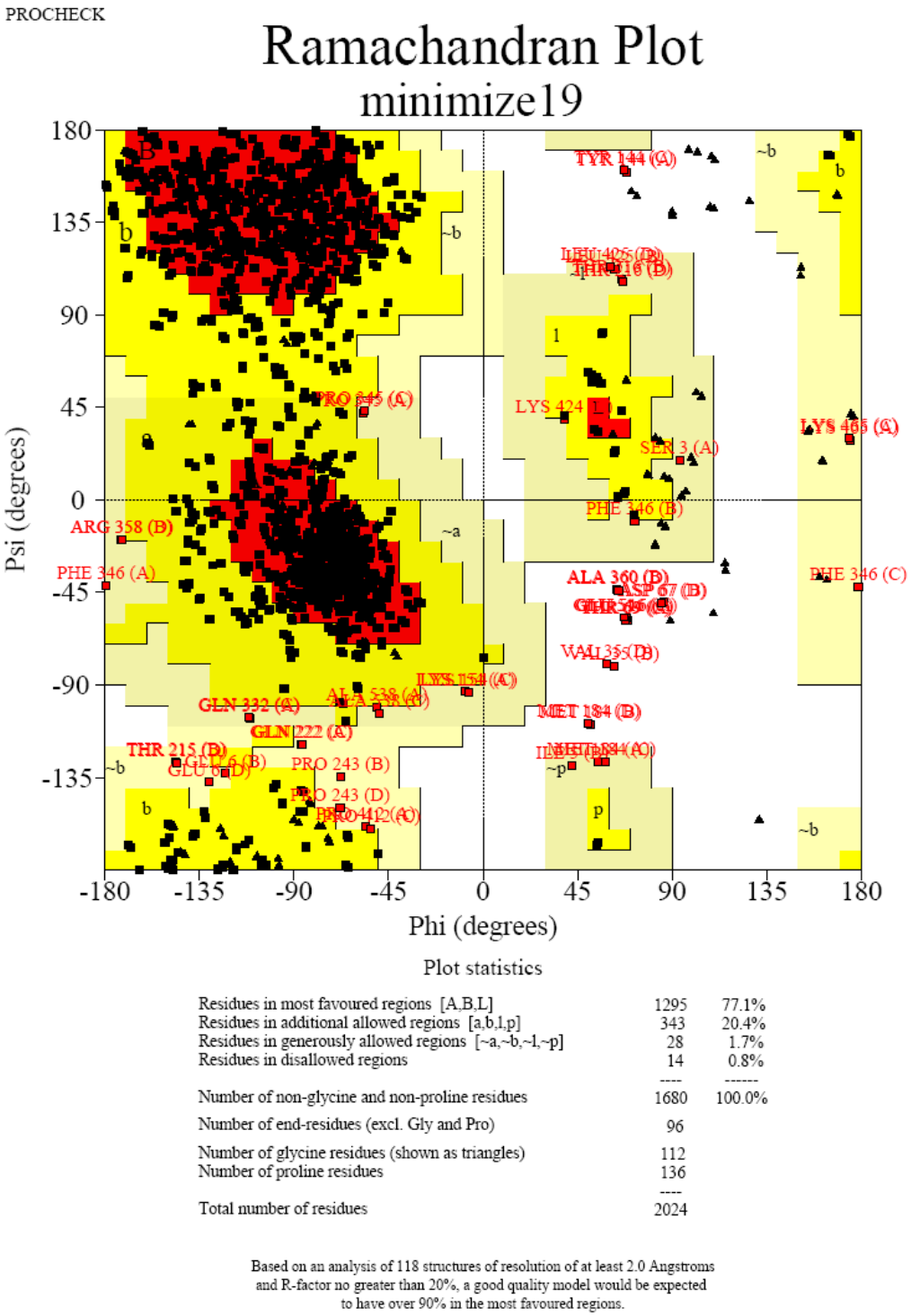
1.2 AZT-resistant HIV-1 RT post-translocation complex

The crystal structure of AZT-excision product has good stereochemistry. The Ramachandran plot shows that 99.2% of residues fall in the allowed regions and 0.3% of residues fall in the disallowed regions (Fig. 67).

1.3 Unliganded AZT-resistant HIV-1 RT

The crystal structure of unliganded AZT-resistant HIV-1 RT also has good stereochemistry. The Ramachandran plot shows that 99.5% of residues fall in the allowed regions and 0.1% of residues fall in the disallowed regions (Fig. 68). The

residues in the disallowed region is Met184 of p66.



Plot statistics

Residues in most favoured regions [A,B,L]	1295	77.1%
Residues in additional allowed regions [a,b,l,p]	343	20.4%
Residues in generously allowed regions [~a,~b,~l,~p]	28	1.7%
Residues in disallowed regions	14	0.8%
Number of non-glycine and non-proline residues	1680	100.0%
Number of end-residues (excl. Gly and Pro)	96	
Number of glycine residues (shown as triangles)	112	
Number of proline residues	136	
Total number of residues	2024	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig. 66. Ramachandran plot of AZT-resistant HIV-1 RT pre-translocation complex

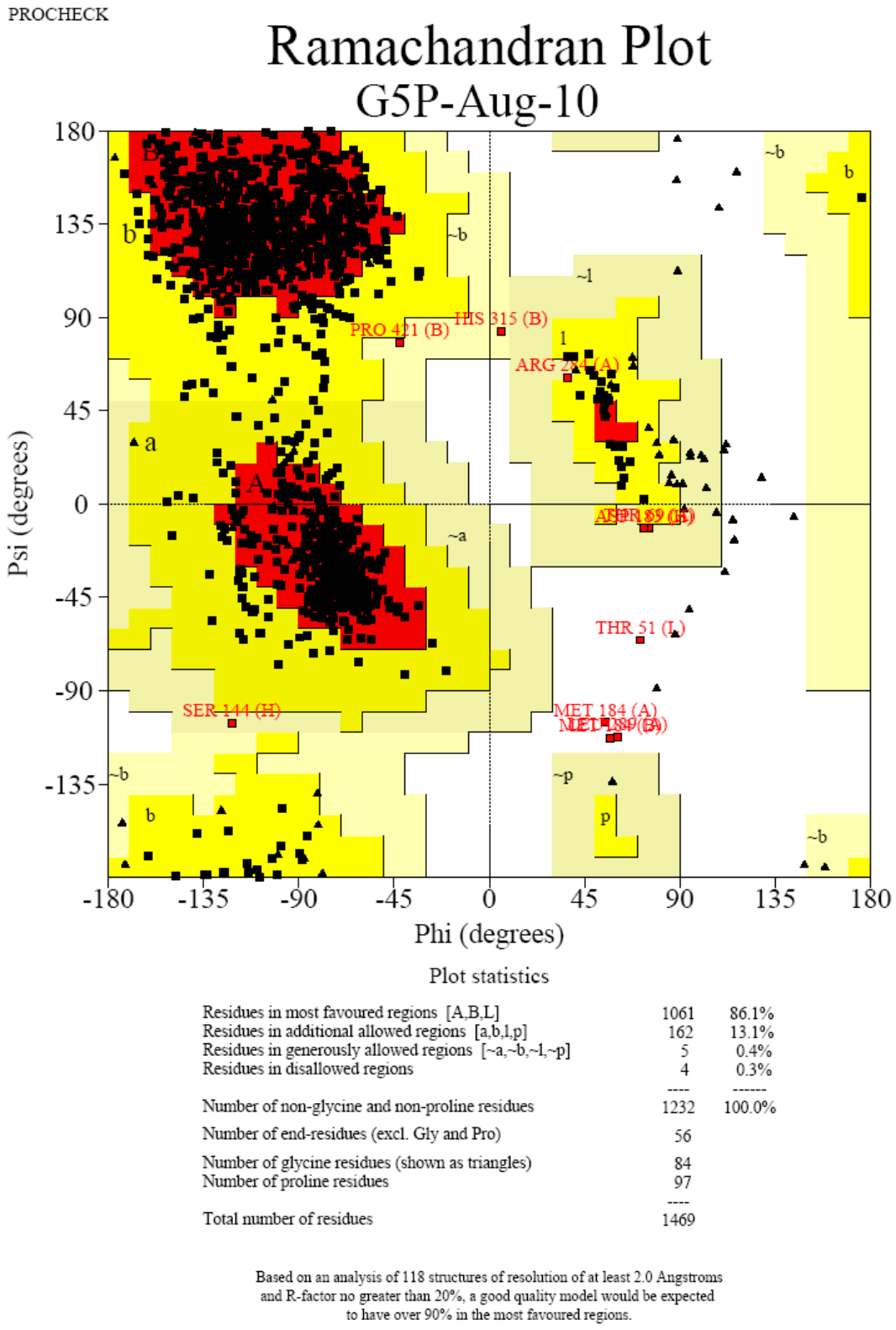
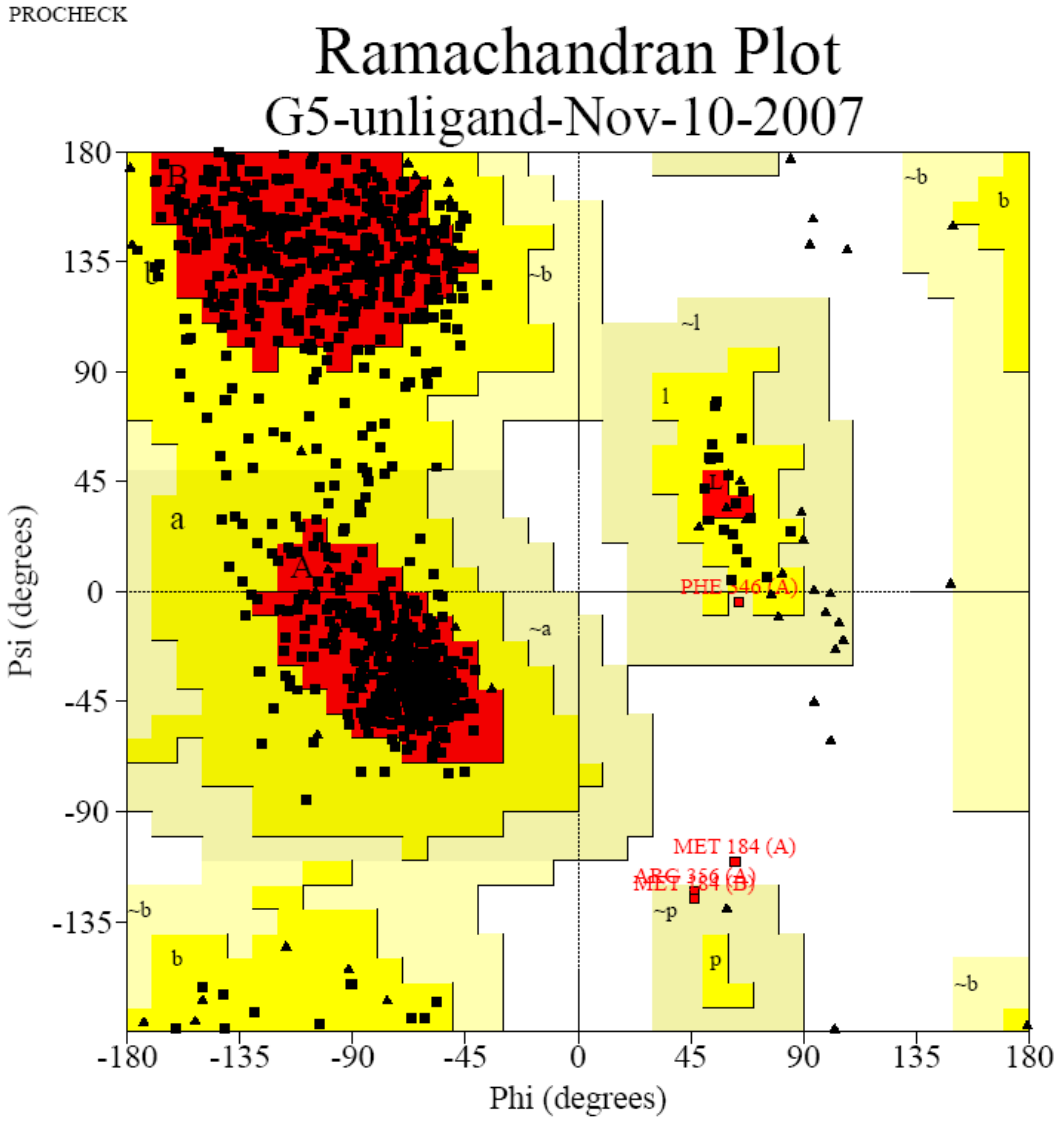


Fig. 67. Ramachandran plot of AZT-resistant HIV-1 RT post-translocation complex



Plot statistics

Residues in most favoured regions [A,B,L]	728	85.8%
Residues in additional allowed regions [a,b,l,p]	116	13.7%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.4%
Residues in disallowed regions	1	0.1%
Number of non-glycine and non-proline residues	848	100.0%
Number of end-residues (excl. Gly and Pro)	39	
Number of glycine residues (shown as triangles)	56	
Number of proline residues	72	
Total number of residues	1015	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Fig. 68. Ramachandran plot of unliganded AZT-resistant HIV-1 RT

1.4 Torsional angles of primary resistant mutant residues R70 and Y215

The torsional angles of primary resistant mutant residues R70 and Y215 of the post-translocation complex and the unliganded RT are shown in Fig. 69. The main-chain torsional angles of R70 and Y215 are distributed in the most favored region. The side-chain torsional angles of R70 and Y215 also have good distributions, since neither of them is in a cis-conformation. The post-translocation complex and unliganded RT are intermediates before ATP binding; however the excision product complex is an intermediate after ATP binding. The distributions of side-chain torsional angles of R70 and Y215 before ATP binding and after ATP binding have been compared (Fig. 69). It has been found that the Chi-1 angle of R70 and the Chi-2 angle of Y215 are distributed differently before ATP binding and after ATP binding, suggesting conformational differences linked to the binding of ATP to RT.

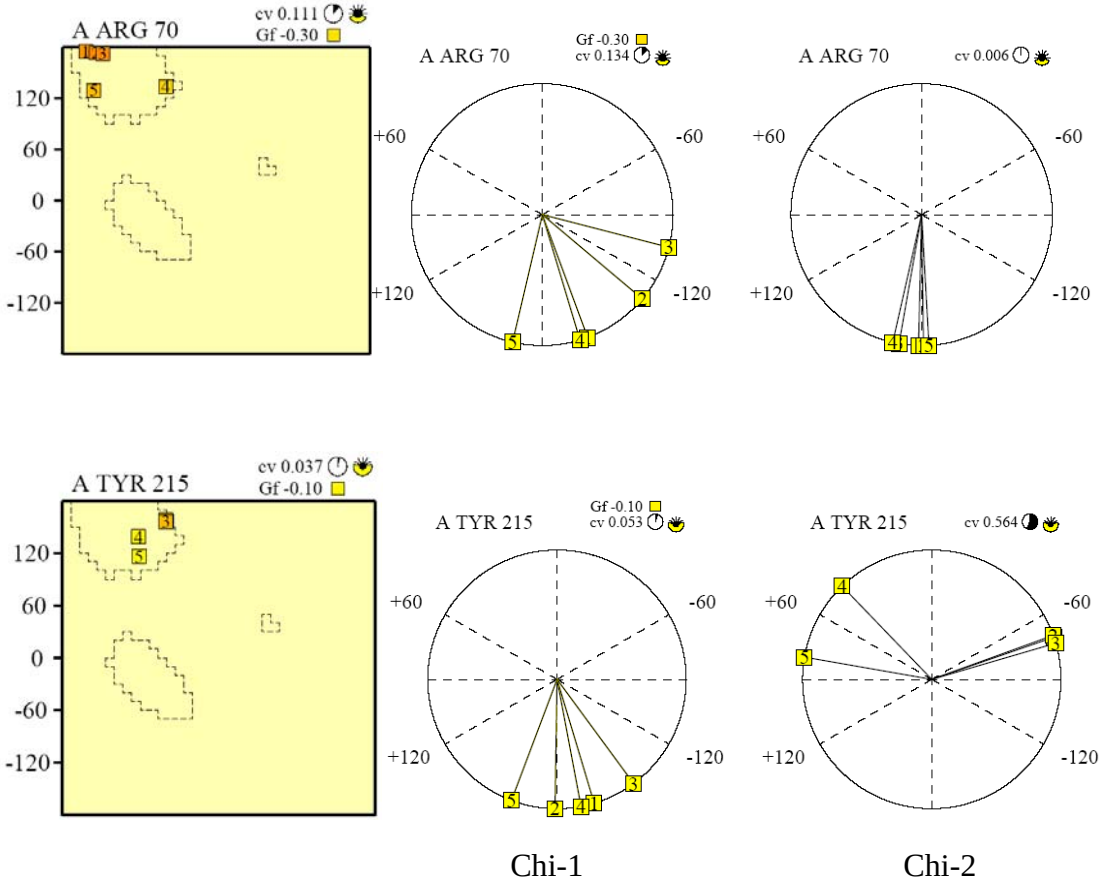


Fig. 69. Torsion angles of primary mutant residues R70 and Y215

1, 2, 3 indicate the first three complex copies of the AZT-resistant HIV-1 RT excision product complex; 4 indicates the AZT-resistant HIV-1 RT post-translocation complex; 5 indicates the unliganed AZT-resistant HIV-1 RT.

2. Structure of the AZT-resistant HIV-1 RT pre-translocation complex

2.1 The AZT-resistant HIV-1 RT pre-translocation complex was crystallized without a monoclonal antibody fragment Fab28 and in space group $P2_12_12_1$

The AZT-resistant HIV-1 RT pre-translocation complex has the AZTMP-terminated primer at the N site, and is the acceptor for ATP to initiate excision according to the excision model (Boyer et al., 2001; Sarafianos et al., 2002; Fig. 70). The AZT-resistant HIV-1 RT pre-translocation complex was crystallized differently from the reported wild-type HIV-1 RT pre-translocation complex. The AZT-resistant HIV-1 RT pre-translocation complex was crystallized without the Fab monoclonal antibody fragment and in space group of $P2_12_12_1$; however, the reported wild-type HIV-1 RT pre-translocation complex was crystallized with a Fab28 and in space group of $P3_212$ (Sarafianos et al., 2002). On the other hand, the AZT-resistant HIV-1 RT pre-translocation complex was crystallized similarly to the reported wild-type HIV-1 RT ternary complex. Both of them are crystallized in the space group of $P2_12_12_1$, and share similar cell parameters. Both crystallographic asymmetric units have two NCS-related complexes, and the two NCS-related complexes differ significantly in their overall order.

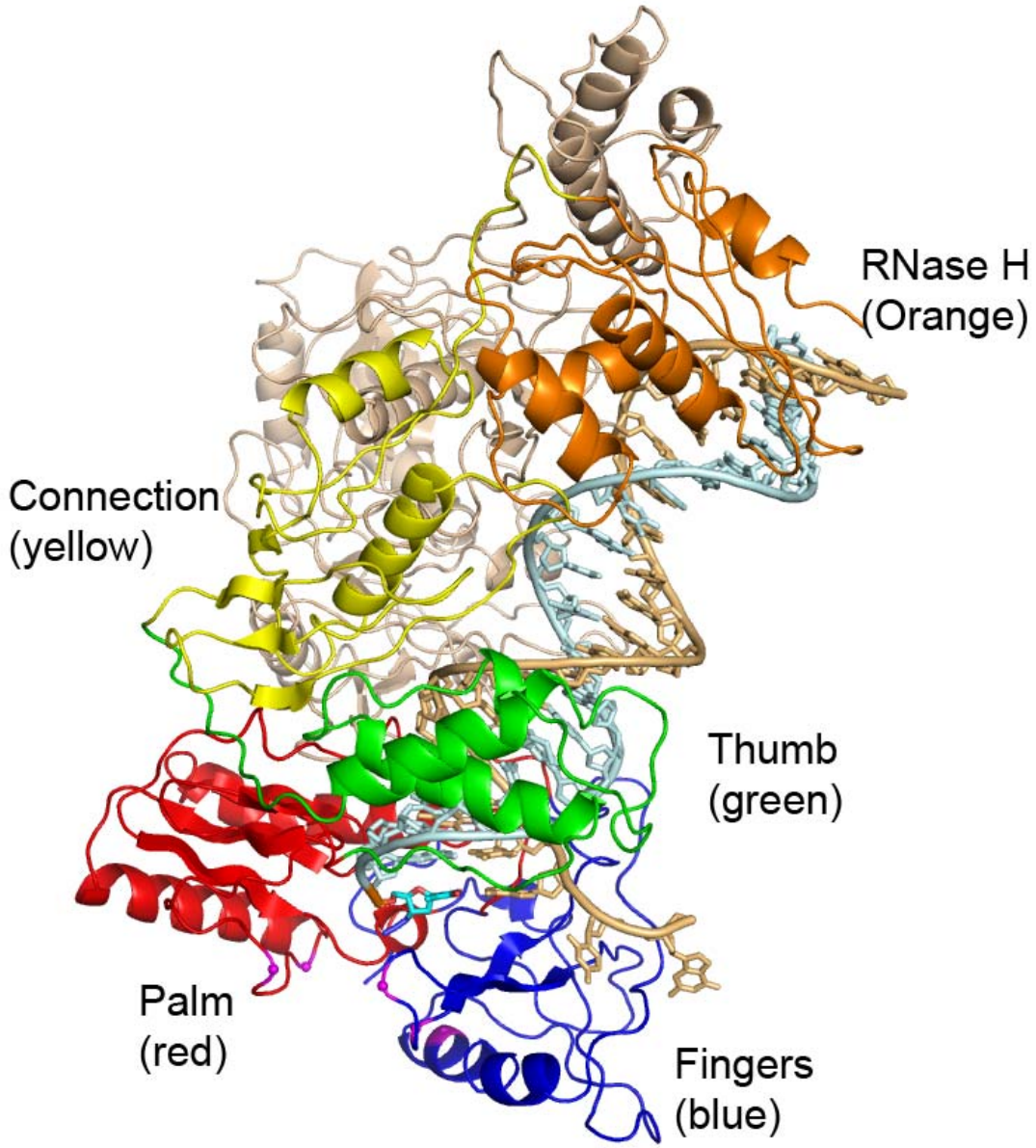


Fig. 70. Overall structure of the AZT-resistant pre-translocation complex. The AZTMP at the primer 3'-end is colored in cyan-blue-red. The Cα of AZT-resistance mutations are showed in sphere and colored in magenta.

Although the resolution is at 3.7 Å, the electron density map for the AZT-resistant pre-translocation complex is of good quality, allowing confident identification of DNA, protein main chain, and some side chains. The electron density is clear for the newly incorporated AZTMP, except for the 3'-azido group (Fig. 71). This terminal AZTMP of primer 3'-end is bound at the N site and base-paired with the template, in a fashion similar to what is observed in the wild-type pre-translocation complex. A non-bridging oxygen of the terminal AZTMP is in close contact with the carboxylate group of Asp185 with a distance of about 2.5 Å. This close contact is also observed in the reported wild-type pre-translocation complex. The base of the terminal AZTMP stacks with the guanidinium group of R72, and the ribose ring of the terminal AZTMP stacks with the aromatic side chain of Y115. However, these interactions of R72 and Y115 are not observed in the reported wild-type pre-translocation complex. The side chains of the resistance mutant residues are unfortunately ill-defined, although there are clear electron densities for their main-chain atoms.

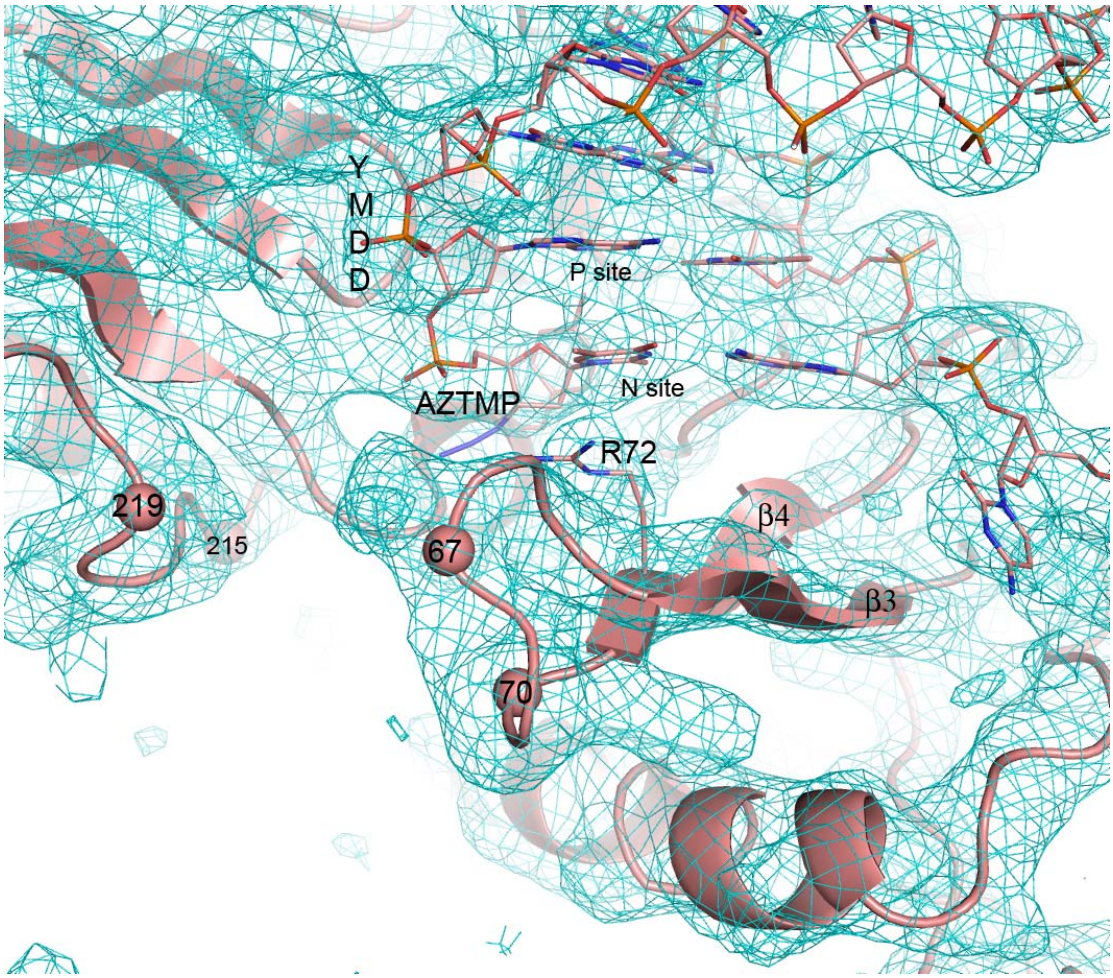


Fig. 71. 2Fo-Fc Fourier map (contoured at 1.4σ) of the polymerase active site of the AZT-resistant pre-translocation complex.

2.2 The AZT-resistant HIV-1 RT pre-translocation complex has the fingers subdomain in a closed configuration

Overall superposition of the AZT-resistant and the wild-type pre-translocation complexes gives 2.1 Å of root-mean-square deviation (RMSD) for the Cα atoms. However, the RMSD of these two structures drops to 0.9 Å if the fingers subdomain of p66 is excluded from the overall superposition. This observation indicates a motion of fingers subdomain in the AZT-resistant *versus* the wild-type pre-translocation complex. The configuration of the fingers subdomain in the reported structure of HIV-1 RT polymerization ternary complex (HIV-1 RT-DNA/dTTP) has been referred to as a "closed" configuration, since the fingers subdomain closes down on the incoming nucleotide. On the other hand, the configuration of the fingers subdomain in the reported structure of wild-type pre-translocation complex has been referred to as an "open" configuration, since it moves outward from the N site (Sarafianos et al., 2002). Structural superposition of this AZT-resistant pre-translocation complex on the wild-type pre-translocation complex (open fingers) and the polymerization ternary complex (closed fingers) shows that the fingers subdomain of this AZT-resistant pre-translocation complex aligns well with the fingers subdomain of the polymerization ternary complex, which suggests that the fingers subdomain of this AZT-resistant pre-translocation complex assumes a closed configuration instead of an open configuration (Superposition is based on 107-112 and 155-215) (Fig.72).

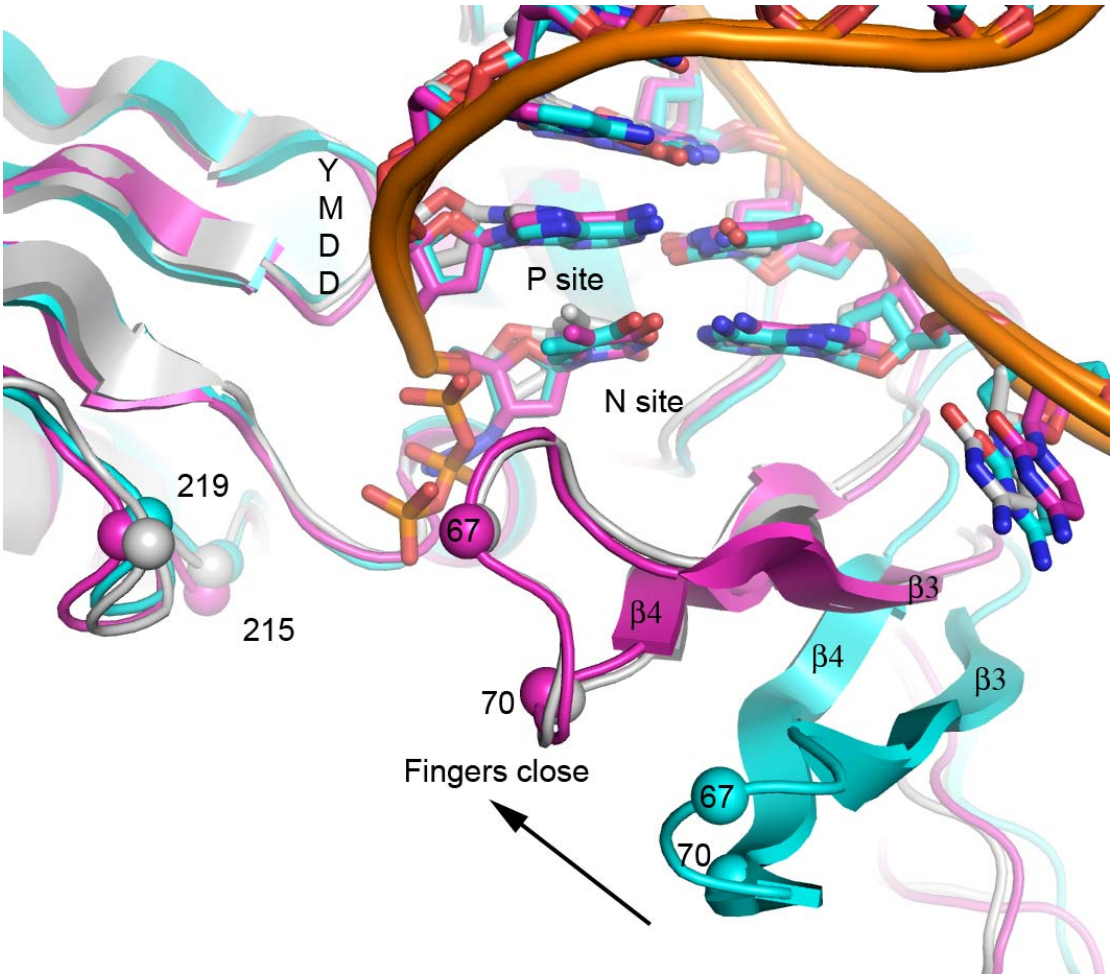


Fig. 72. Structural superposition of the AZT-resistant pre-translocation complex (magenta), the wild-type pre-translocation complex (cyan) and the polymerization catalytic ternary complex (grey). The fingers subdomain assumes an open configuration in the reported structure of the wild-type pre-translocation complex, however a closed configuration in the AZT-resistant pre-translocation complex.

The closed configuration of the fingers subdomain is functionally significant for the AZT excision and DNA polymerization. The structure of the AZT-resistant HIV-1 RT excision product complex showed that the β3-β4 region (residue 57-75) of the fingers subdomain interacts with both the terminal AZTMP and ATP. The structure of the polymerization ternary complex has showed that it extensively interacts with the incoming dNTP. These interactions are essential for achieving the enzymatic activity in excision or polymerization. Docking of an ATP in the assumed binding site of the wild-type pre-translocation complex shows that the β3-β4 region is too distant to interact with ATP or with AZTMP. It is only in the closed configuration that the fingers subdomain can interact with ATP. Since the fingers subdomain can assume either a closed or an open configuration, we expect that the pre-translocation complex with the fingers in a closed configuration is the preferred acceptor for ATP to initiate excision.

Why does the fingers subdomain assume a closed configuration in AZT-resistant pre-translocation complex, however an open configuration in the wild-type pre-translocation complex? Is one of them artificial, resulting from the cross-linking or crystal packing? The AZT-resistance mutations seem to have nothing to do with this change in the configuration, since the mutation sites are not involved in any interactions that linking the fingers subdomain with the other parts of HIV-1 RT in both the wild-type and the AZT-resistant pre-translocation complexes. A power stroke model has been proposed to explain the translocation mechanism. The power

stroke model is derived from crystal images of T7 RNA polymerase complexes (Yin et al., 2004). In this model, the chemical energy from nucleotide triphosphate hydrolysis is transformed into conformational changes of the fingers subdomain that drive polymerase translocation following or concomitant with the release of PPi. It means that the fingers subdomain should be in an open configuration and primer 3'-end should be translocated to the P site after the release of PPi. The primer can't backtrack to the previous state without energy input. According to this model, the configuration of the fingers subdomain in both the AZT-resistant and wild-type pre-translocation complex are artifacts, resulted from the cross-linking of DNA to RT that prevents the DNA translocation. However, they could be real, if the conformational change of the fingers subdomain is not involved in the translocation. Another translocation mechanism, Brownian ratchet mechanism, has also been proposed to describe the translocation mechanism. The Brownian-ratchet model (Astumian 1997 and Schafer et al. 1991) proposes that the translocation between the post-translocation and the pre-translocation states is a Brownian motion driven by thermal energy, and the translocation is biased in one direction by incoming dNTP or NTP that acts like a pawl in a ratchet to prevent backtracking (Reeder and Hawley 1996; Komissarova and Kashlev 1997; Nudler et al., 1997; Landick 1997; Landick 1998; Guajardo and Sousa 1997; Guajardo et al. 1998). More and more data have recently been provided to support the Brownian-ratchet model. The site-specific foot printing analyses with foscarnet, a pyrophosphate analogue, has revealed that the foscarnet can selectively stabilize HIV-1 RT in a pre-translocation state. The

foscarnet acts like a pawl of a rachet to prevent primer translocation to the P site. This study suggests the translocation of HIV-1 RT matches the Brownian-ratchet model (Marchand et al., 2007). Recent studies of halted T7 RNA polymerase elongation complexes has also suggest the T7 RNA polymerase translocates actually via a passive Brownian ratchet mechanism. Studies of single molecules of *Escherichia coli* RNA polymerase also suggest that the RNA polymerase acts as a Brownian ratchet that is driven forward by the binding of incoming nucleoside triphosphates (Abbondanzieri et al., 2005). This Brownian ratchet model is also consistent with the ATP-mediated excision model (Boyer et al., 2001; Sarafianos et al., 2002) that assumes the primer 3'-end can shuttle between N site and P site freely. Taken together, these data suggest that the fingers configuration has nothing to do with the DNA translocation, and are not driven by the hydrolysis of incoming dNTP.

The open configuration of the fingers subdomain in the reported structure of wild-type pre-translocation complex may be attributed to the crystal packing. The wild-type pre-translocation complex was co-crystallized with an antibody fragment Fab and in space group $P3_212$. Packing analysis shows residue 49-53 and residue 139-141 have extensive contacts with a symmetry-generated Fab heavy chain (Table 25). Some significant interactions can be clearly identified from the electron density map. The main chain oxygens of residue 140 and 141 interact with Q16 side chain of Fab H chain. The side chain of E53 of the fingers subdomain electronstatically interacts with the side chain of Arg19 of Fab H chain. These interactions will be lost,

if the wild-type pre-translocation complex (open fingers configuration) is replaced with the AZT-resistant RT pre-translocation complex (closed fingers configuration) in the crystal lattice of wild-type RT pre-translocation complex. Furthermore, the total number of contacts also decreases from 43 to 29. Thus, we think the open configuration of the fingers subdomain in the reported structure of wild-type pre-translocation complex may have be stabilized by the crystal packing.

Table 25 Crystal contacts of the fingers subdomain with the H chain of Fab

Fingers subdomain residues	H chain residues of Fab	Distance (Å)	Symmetry
49(LYS). / CA [C]:	85(MET). / CE [C]:	3.94	X,X-Y,-Z
49(LYS). / CB [C]:	85(MET). / CE [C]:	3.58	X,X-Y,-Z
	85(MET). / CB [C]:	3.86	X,X-Y,-Z
49(LYS). / CG [C]:	86(THR). / OG1[O]:	3.94	X,X-Y,-Z
49(LYS). / CD [C]:	86(THR). / OG1[O]:	2.85	X,X-Y,-Z
49(LYS). / CE [C]:	86(THR). / OG1[O]:	2.84	X,X-Y,-Z
49(LYS). / NZ [N]:	86(THR). / CB [C]:	2.95	X,X-Y,-Z
	86(THR). / OG1[O]:	1.89	X,X-Y,-Z
	86(THR). / CG2[C]:	3.01	X,X-Y,-Z
49(LYS). / C [C]:	85(MET). / CE [C]:	3.06	X,X-Y,-Z
49(LYS). / O [O]:	85(MET). / SD [S]:	3.69	X,X-Y,-Z
	85(MET). / CE [C]:	2.82	X,X-Y,-Z
	85(MET). / CG [C]:	3.83	X,X-Y,-Z
50(ILE). / N [N]:	85(MET). / CE [C]:	3.34	X,X-Y,-Z
50(ILE). / CA [C]:	85(MET). / CE [C]:	3.42	X,X-Y,-Z
50(ILE). / C [C]:	85(MET). / CE [C]:	3.03	X,X-Y,-Z
50(ILE). / O [O]:	85(MET). / CE [C]:	2.94	X,X-Y,-Z
	17(PRO). / CG [C]:	3.93	X,X-Y,-Z
51(GLY). / N [N]:	85(MET). / CE [C]:	3.63	X,X-Y,-Z
52(PRO). / CD [C]:	17(PRO). / CG [C]:	3.87	X,X-Y,-Z
52(PRO). / CG [C]:	19(ARG). / CB [C]:	3.91	X,X-Y,-Z
53(GLU). / CG [C]:	19(ARG). / NH1[N]:	3.62	X,X-Y,-Z
	83(ASN). / ND2[N]:	3.88	X,X-Y,-Z
53(GLU). / CD [C]:	19(ARG). / CZ [C]:	3.49	X,X-Y,-Z
	19(ARG). / NH1[N]:	2.51	X,X-Y,-Z
	83(ASN). / ND2[N]:	3.47	X,X-Y,-Z
53(GLU). / OE1[O]:	19(ARG). / NH2[N]:	3.05	X,X-Y,-Z

	19(ARG). / CZ [C]:	2.42	X,X-Y,-Z
	19(ARG). / NH1[N]:	1.66	X,X-Y,-Z
	19(ARG). / CD [C]:	3.63	X,X-Y,-Z
	19(ARG). / NE [N]:	3.28	X,X-Y,-Z
	83(ASN). / ND2[N]:	3.95	X,X-Y,-Z
53(GLU). / OE2[O]:	19(ARG). / NH1[N]:	3.06	X,X-Y,-Z
	83(ASN). / CG [C]:	4.00	X,X-Y,-Z
	83(ASN). / ND2[N]:	3.33	X,X-Y,-Z
139(THR). / CB [C]:	13(GLN). / OE1[O]:	3.77	X,X-Y,-Z
139(THR). / OG1[O]:	13(GLN). / CB [C]:	3.92	X,X-Y,-Z
	13(GLN). / CG [C]:	3.11	X,X-Y,-Z
	13(GLN). / CD [C]:	2.90	X,X-Y,-Z
	13(GLN). / OE1[O]:	2.64	X,X-Y,-Z
	13(GLN). / NE2[N]:	3.80	X,X-Y,-Z
139(THR). / CG2[C]:	16(GLN). / CG [C]:	3.66	X,X-Y,-Z
	16(GLN). / CD [C]:	3.76	X,X-Y,-Z
	16(GLN). / NE2[N]:	3.92	X,X-Y,-Z
	13(GLN). / CG [C]:	3.89	X,X-Y,-Z
140(PRO). / CB [C]:	16(GLN). / OE1[O]:	3.89	X,X-Y,-Z
140(PRO). / CG [C]:	16(GLN). / CD [C]:	3.95	X,X-Y,-Z
	16(GLN). / OE1[O]:	3.33	X,X-Y,-Z
140(PRO). / O [O]:	16(GLN). / CD [C]:	3.77	X,X-Y,-Z
	16(GLN). / NE2[N]:	3.16	X,X-Y,-Z
141(GLY). / C [C]:	16(GLN). / NE2[N]:	3.81	X,X-Y,-Z
141(GLY). / O [O]:	16(GLN). / CD [C]:	3.71	X,X-Y,-Z
	16(GLN). / OE1[O]:	3.74	X,X-Y,-Z
	16(GLN). / NE2[N]:	2.86	X,X-Y,-Z

Distance cutoff = 4 Å

The closed configuration of the fingers subdomain in the pre-translocation complex is possibly a dominant configuration, according to the crystal packing analyses of the AZT-resistant pre-translocation complex. This AZT-resistant pre-translocation complex is crystallized without the antibody fragment Fab and in space group $P2_12_12_1$. The closed fingers subdomain makes no contributions to the packing interactions in the crystal lattice of the AZT-resistant pre-translocation complex (distance cutoff 4.5 Å). If the AZT-resistant HIV-1 RT (closed configuration) is replaced with the HIV-1 RT from the structure of the wild-type pre-translocation complex (open configuration) in the crystal lattice of AZT-resistant pre-translocation complex, there are no steric conflicts caused by the crystal packing. It seems that the fingers subdomain can be either a closed or an open configuration in the crystal lattice of the AZT-resistant pre-translocation complex. The fact that the AZT-resistant pre-translocation complex selects the closed configuration may suggest that the closed configuration is a dominant configuration. It is also implies that polymerization or ATP-mediated pyrophosphorolysis may occur without a major opening of the fingers subdomain.

3. Structure of the AZT-resistant HIV-1 RT post-translocation complex

The AZT-resistant HIV-1 RT post-translocation complex is AZT-resistant HIV-1 RT in complex with a post-translocation AZTMP-terminated DNA. This AZT-resistant HIV-1 RT post-translocation complex has a lot in common with the wild-type HIV-1 RT post-translocation complex (Sarafianos et al., 2002). Both of them are crystallized with a monoclonal antibody fragment Fab28 and in space group $P3_212$.

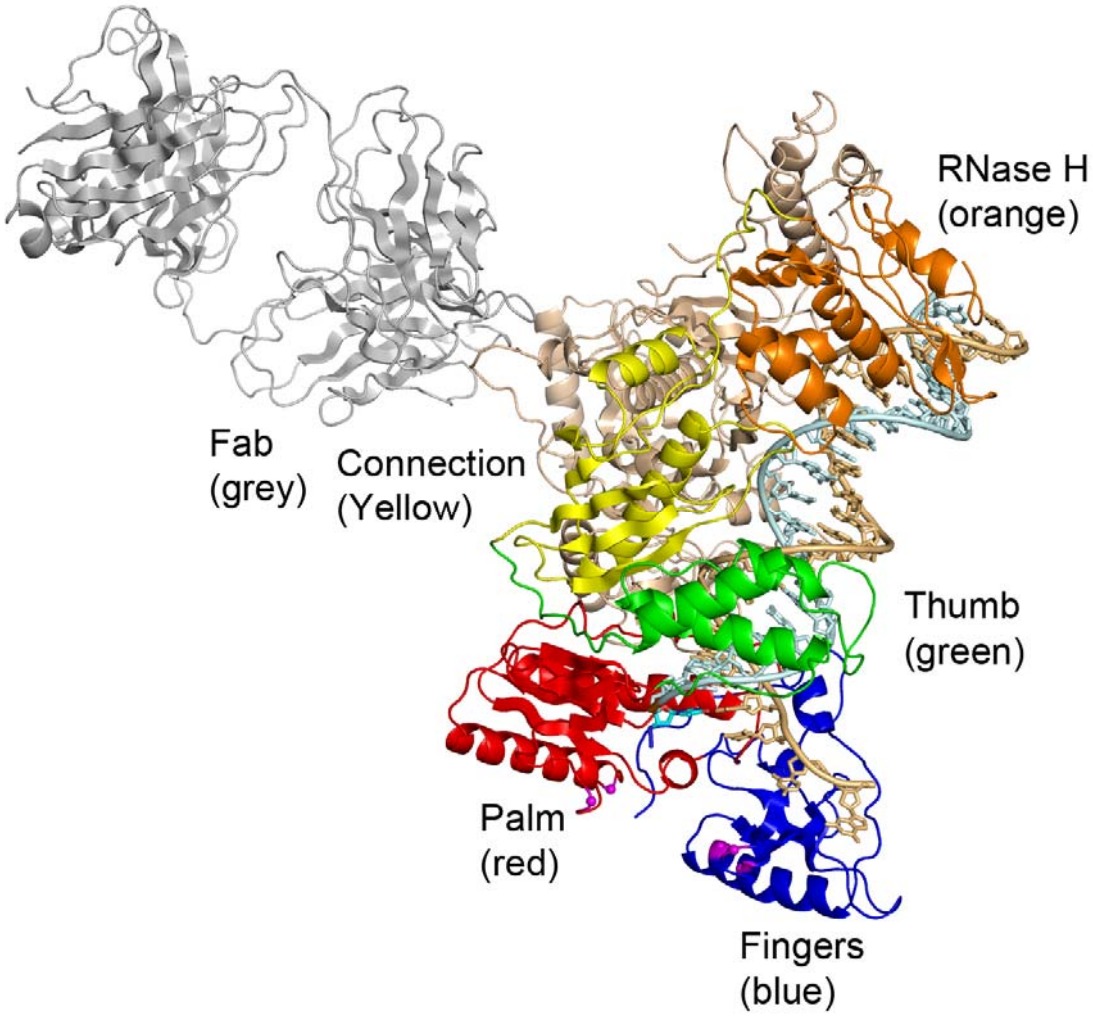


Fig. 73. Overall structure of the AZT-resistant HIV-1 RT post-translocation complex

The AZT-resistant and the wild-type post-translocation complexes are structurally similar to each other with 0.40 Å of RMSD for the Cα atoms. Close-up viewing shows that there are no significant main-chain rearrangements around the site of the AZT-resistance mutations and catalytic site (Fig. 74). The YMDD motifs of the two post-translocation complexes align very well with no significant displacement (Fig.74). Both the AZT-resistant and the wild-type post-translocation complexes have fingers subdomain in an open configuration (Fig. 74). All of these similarities suggest that AZT-resistance mutations have no effect on either the local backbone conformation or overall conformation.

The 2Fo-Fc Fourier map (sigma weighted) shows clear electron density for the AZTMP, except for the 3'-azido group (Fig. 75). This terminal AZTMP is bound at the P site and base paired with the template, in a fashion similar to what is observed in the wild-type post-translocation complex. The azido group has a weak electron density. The conformation that is most consistent with the annealed difference omit map of AZTMP is different from the one in the wild-type post-translocation complex. In this conformation, the azido group interacts with OD1 and OD2 of D185 through its second nitrogen (positively charged) instead of its first nitrogen, with distances of 2.9 Å and 3.0 Å, respectively. If an incoming dNTP is modeled into the N site, the azido group of this terminal AZTMP will point toward the incoming dNTP, and cause a steric conflict (1.9 Å) between its third nitrogen (negatively charged) and one of the non-bridging oxygens of the α-phosphate group of the incoming dNTP. This

modeling study is consistent with previous modeling studies based on the wild-type post-translocation complex (Boyer et al., 2001; Sarafianos et al., 2002). The incoming dNTP was modeled by assuming that the YMDD motif is not displaced when a dNTP is bound at the N site (alignment of post-translocation complex and polymerization ternary complex based on residues 107-112 and 155-215 of p66).

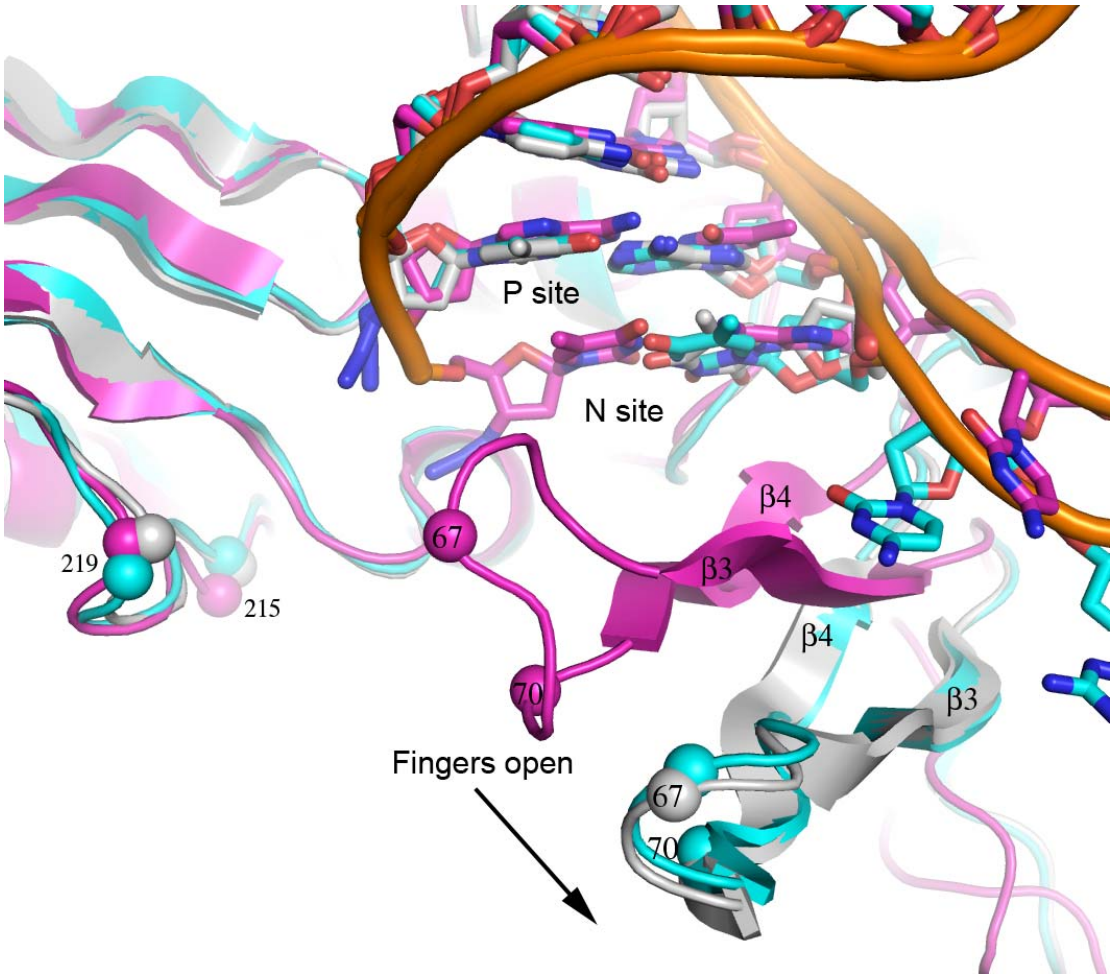


Fig. 74. Structural superposition of the AZT-resistant post-translocation complex (cyan), the wild-type post-translocation complex (grey) and the AZT-resistant pre-translocation complex (magenta). The fingers subdomain assumes an open configuration in the AZT-resistant post-translocation complex instead of a closed configuration in the AZT-resistant pre-translocation complex (magenta).

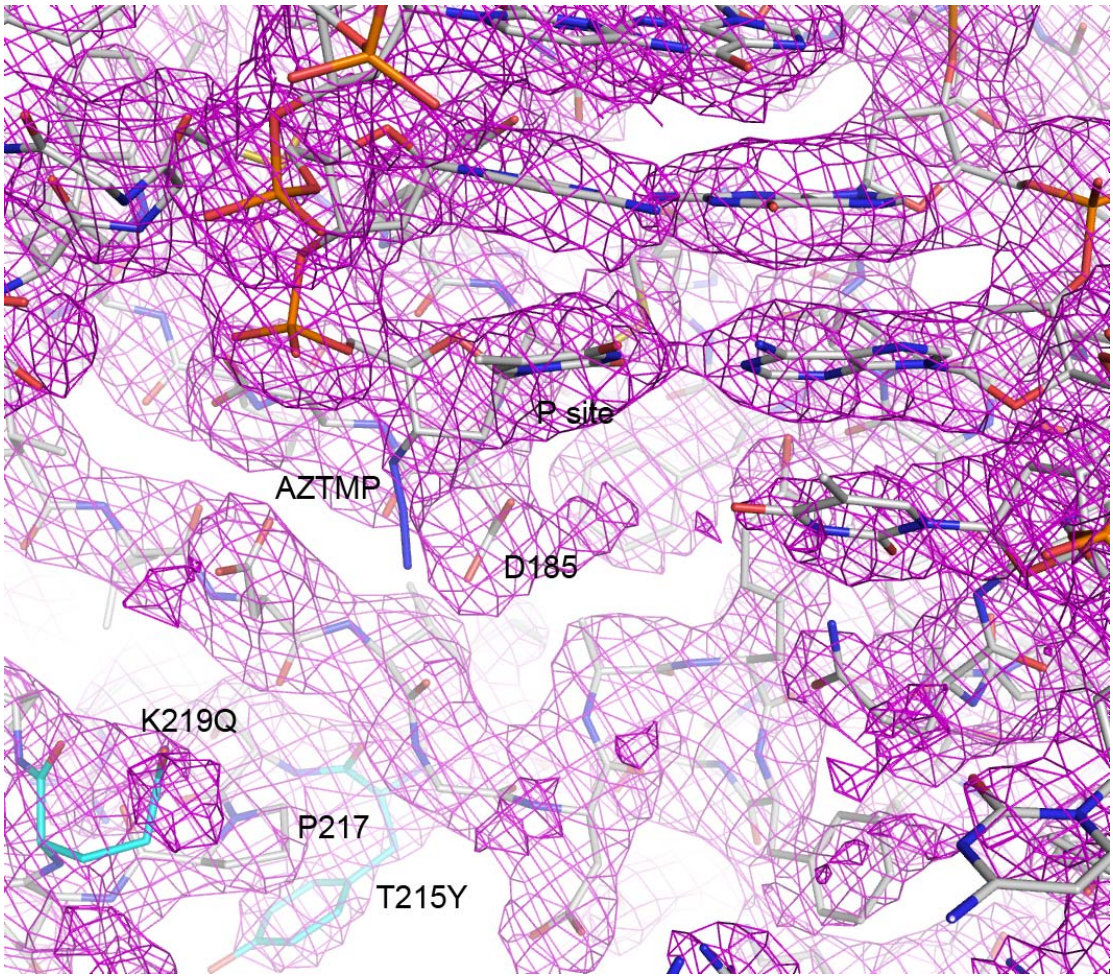


Fig.75. 2Fo-Fc electron density map (contoured at 1.4σ) of the polymerase active site of the AZT-resistant post-translocation complex.

The simulated-annealing omit maps show clear electron density for the side chains of resistant mutants in p66 (Fig. 76). The side chains of mutant residues R70, Y215 and L41 are well defined, but the side chains of mutant residues N67 and Q219 are relatively ill-defined. The R70 side chain stacks with the R72 side chain through their guanidinium groups. The aromatic side chain of Y215 stacks with the five-membered ring of residue Pro217. These conformations of R70 and Y215 are not consistent with what are observed in the AZT-resistant excision product complex. In the AZT-resistant excision product complex, the aromatic side chain of Y215 stacks with the adenine group of the AZTppppA, and the R70 side chain interacts with the α-phosphate group and 3'-OH group of the ATP moiety of AZTppppA. Since the post-translocation and excision product complexes represent the intermediates before ATP binding and after ATP binding, respectively, these conformational differences in primary mutant residues may be caused by the binding of ATP to RT.

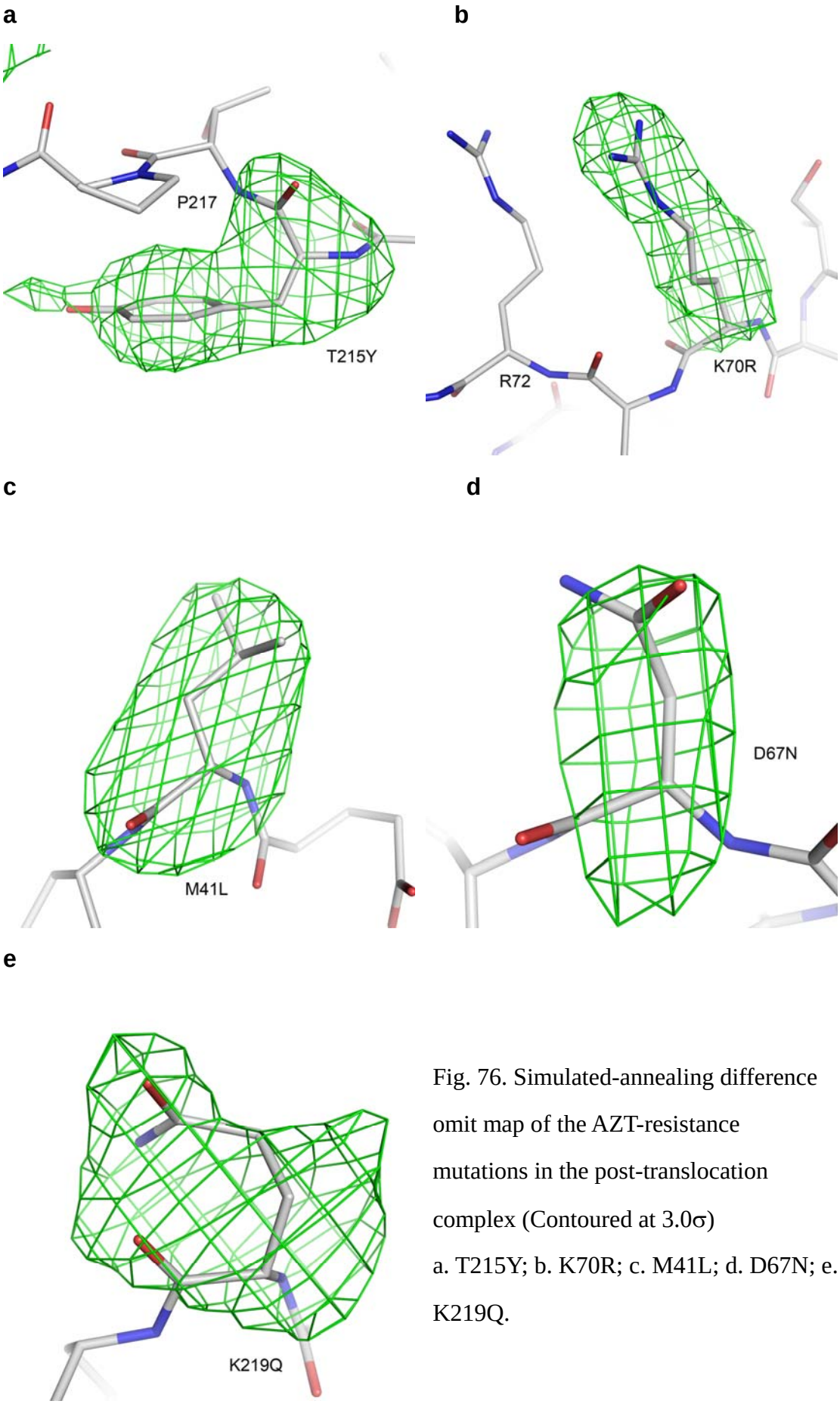


Fig. 76. Simulated-annealing difference omit map of the AZT-resistance mutations in the post-translocation complex (Contoured at 3.0σ)
a. T215Y; b. K70R; c. M41L; d. D67N; e. K219Q.

4. Structure of the unliganded AZT-resistant HIV-1 RT

The simulated-annealing omit maps show electron densities for the resistance mutations in p66 (Fig. 77). The primary mutant residue Y215 is clearly defined. Its aromatic side chain stacks with P217, in a fashion similar to what is observed in the AZT-resistant post-translocation complex. The side chain of primary mutant R70 has weaker electron density, and is less ordered than it is in the AZT-resistant post-translocation complex. The side chain of R70 is exposed to solvent region, instead of stacking with R72 observed in the AZT-resistant post-translocation complex. The side chain of R72 is actually disordered in this unliganded AZT-resistant HIV-1 RT. The mutant L41, N67 and Q219 have relatively good electron densities.

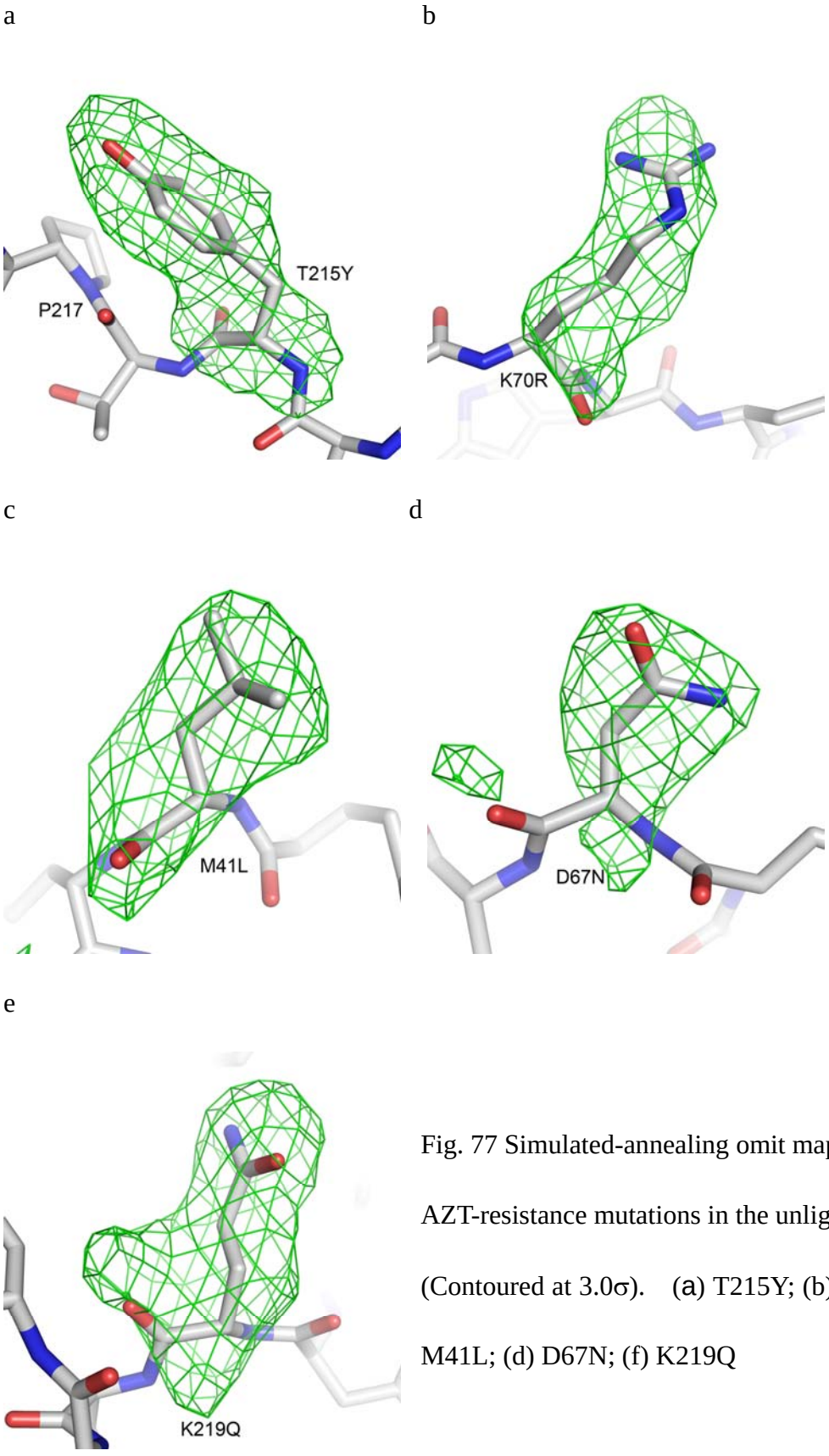


Fig. 77 Simulated-annealing omit maps of the AZT-resistance mutations in the unliganded RT (Contoured at 3.0σ). (a) T215Y; (b) K70R; (c) M41L; (d) D67N; (f) K219Q

The structure of unliganded AZT-resistant HIV-1 RT is superimposed on the reported structure of unliganded wild-type HIV-1 RT (PDB id: 1DLO) (Fig. 75). There is no significant main-chain displacement in the region around the mutation M41L. It is logical, since residue 41 is located at a structurally conservative helix (α-Helix A). Although the regions around the mutations D67N and K70R are displaced in the AZT-resistant HIV-1 RT *versus* wild-type HIV-1 RT, these regions are part of β3-β4 connecting loop, so it is difficult to tell whether this rearrangement is caused by the mutations themselves, or the flexibility of the loop. Another conformational change is observed around the regions containing K219Q (residue 216 to 226), which is proximal to the primary mutation T215Y and β11. It is notable that the main chain oxygen of G112 is actually flipped toward the active site in the structure of the AZT-resistant RT (Fig. 79). Similar observation of this flipped conformation of G112 has been reported in the structure of an AZT-resistant HIV-1 RT (mutations including D67N, K70R, T215Y and K219Q) in complex with a non-nucleoside inhibitor (PDB id 1RT3; Ren et al., 1998). It has been proposed that this conformational change is induced by the mutations K219Q and T215Y. However, the exact mechanisms underlying this rearrangement are undefined at that time.

In the wild-type unliganded RT structure (PDB id: 1DLO), a four strand β–sheet (β11–β6–β10–β9) is clearly seen in the active site (Fig. 79). However, in the structure of AZT-resistant unliganded HIV-1 RT, the β-hydrogen bonding between

β11−β6 is disrupted. The β11 strand is proximal to the mutations K219Q and T215Y. Since the only difference in this region between the wild-type and AZT-resistant enzyme is presence and absence of these two mutations, we expect the disruption of β-hydrogen bonding between β11−β6 is caused by mutations T215Y and K219Q. This induced main-chain rearrangement can somehow extend to the active site, which may be the reason of what appears to be the most important conformational change at residues 112 and 113. Both of these two residues are proximal to the polymerase active site and, more importantly, at the bottom of the dNTP-binding site. Their amido groups face toward the phosphate groups of incoming dNTP in the structure of wild-type HIV-1 RT ternary complex. However, the flipped main-chain of G112 will cause the carbonyl oxygen to face toward to the phosphate groups of incoming dNTP, resulting in unfavorable charge repulsion. A β-hydrogen bond is formed between residue G112 and D185, after G112 is flipped toward D185 in this unliganded AZT-resistant variant structure. This interaction may also have an impact on polymerization activity. Previous biochemical data have already showed that the mutations D67N, K70R, T215F, and K219Q give a three-fold increase in Ki for the AZTTP. Although a similar change in this strucuture was reported in the strucuture of an AZT-resistant RT/NNRTI complex (Ren et al., 1998), it is not clear if the conformational change was due to the NNRTI or due to the AZT-resistance mutations. Our structure unambiguously shows that the AZT-resistance mutations by themselves can cause a change in that area

The conformation of G112 is also significant to excision-based NRTI resistance. Preliminary experiments from Dr. Stephen H. Hughes laboratory that were designed based on our structural data confirm the role of G112 in ATP-based excision, as mutation at residue112 appears to lose the ability to excise AZT-terminated primers (Boyer and Hughes, unpublished results, personal communication).

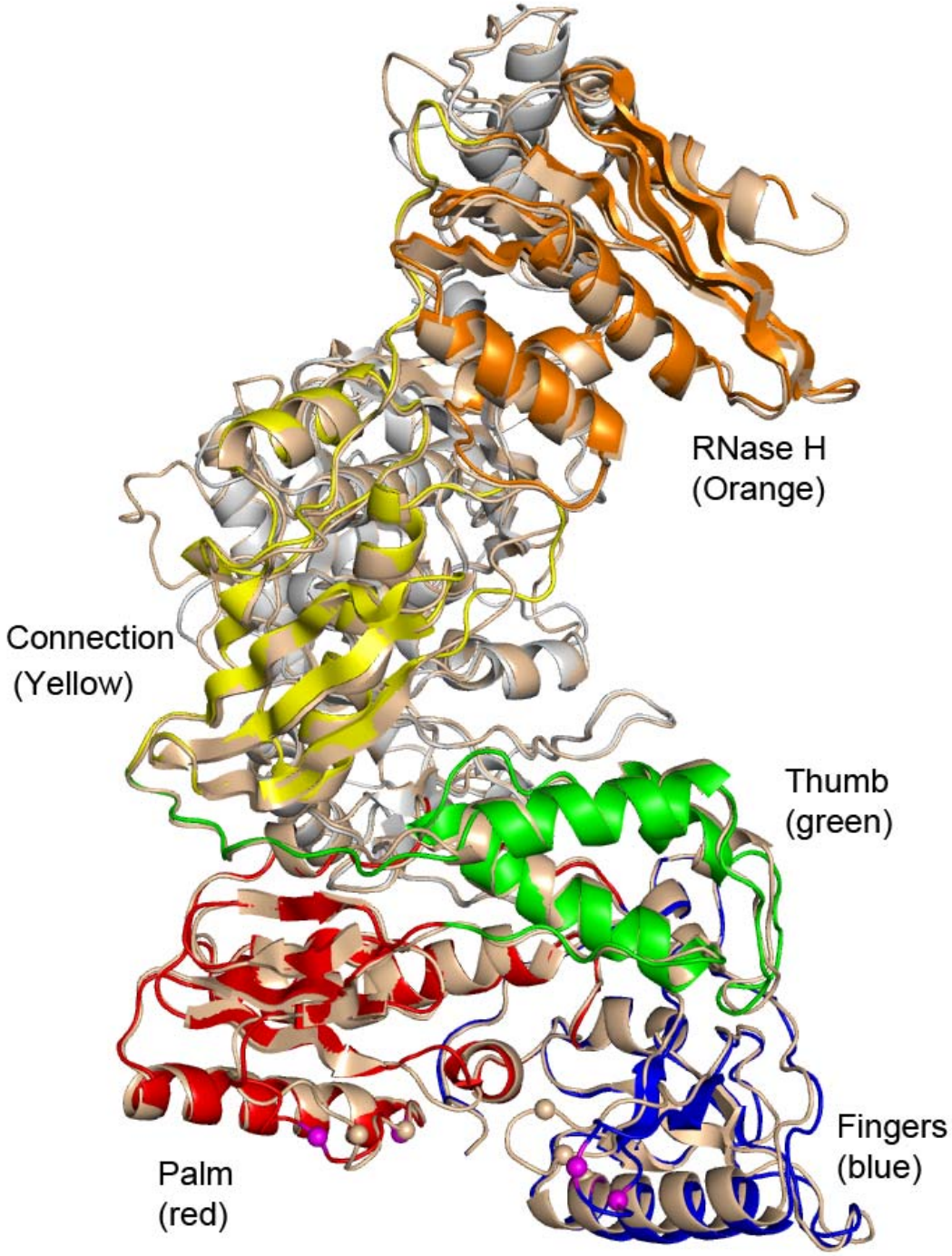


Fig. 78. Superposition of the AZT-resistant HIV-1 RT and the wild-type HIV-1 RT.

The palm, thumb, and connection subdomain are aligned very well.

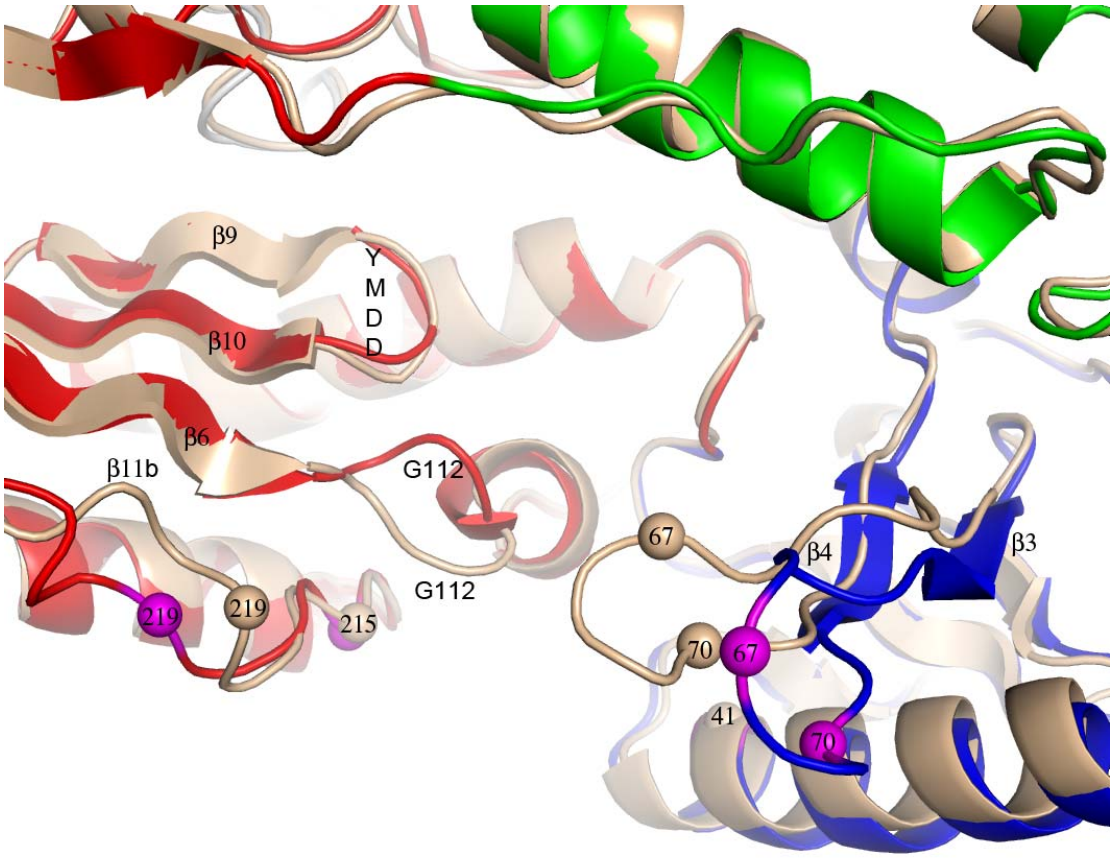


Fig. 79. Close-up view of the superposed active sites of the AZT-resistant HIV-1 RT and the wild-type HIV-1 RT. Conformational changes may be induced by the T215Y and K219Q mutations. The p66 subunit of the wild-type HIV-1 RT (PDB id: 1DLO) is colored wheat. The p66 subunit of the AZT-resistant HIV-1 RT is colored according to its subdomains or domains. A four strand β-sheet (β11-β6-β10-β9, yellow) is clearly seen in the active site of wild-type HIV-1 RT (1DLO). However, the β-hydrogen bonding between β11b-β6 is disrupted in the AZT-resistant HIV-1 RT, resulting from the local conformational changes that may be induced by T215Y and K219Q. These conformational changes may somehow extend to the active site, resulting in the flip of residue G112 toward D185.

5. The induced-fit of the side chains of primary resistance mutations and ATP

As mentioned before, the AZT-resistant unliganded RT, post-translocation complex and pre-translocation complex are intermediates before ATP binding; however the AZT-resistant excision product complex represents an intermediate after ATP binding. The previous distribution analyses of torsional angles of R70 and Y215 before ATP binding and after ATP binding have revealed that the two primary resistant residues may undergo conformational changes induced by ATP binding. In the unliganded AZT-resistant RT, the Y215 side chain stacks with P217, and the R70 side chain is exposed to the solvent. In the AZT-resistant post-translocation complex, the Y215 side chain still stacks with the P217, and the R70 stacks with R72 by their guanidino groups. Although we don't know the conformations of the side chains of the two primary mutants in the AZT-resistant pre-translocation complex, we believe that Y215 side chain still stacks with P217, considering its hydrophobic property. The side chains of Y215 and R70 can not be defined in the solved AZT-resistant excision product complex, which may be caused by low resolution or low data incompleteness. It is notable that the Y215 side chain of the AZT-resistant excision product complex stacks with the adenine group of the ATP moiety of the excision product, AZTppppA, instead of the P217; and the R70 side chain interacts with the α-phosphate group and the 3'-OH group of the ATP moiety instead of stacking with the R72 side chain or being exposed to the solvent. These data suggest that the binding of ATP to RT is a process concomitant with the induced-fit between the primary resistance mutations and the ATP. Although the main chains of primary resistance mutations are

relatively constant, the side chains of the primary resistance mutations can adjust their conformations to interact with ATP. These interactions efficiently enhance the binding affinity. It is only in these sites that side chains of R70 and Y215 can extensively interact with ATP without significant conformational changes of main chains. That is the reason why the AZT selects these two sites as primary mutation sites and selects Tyr and Arg as their respective mutations residues.

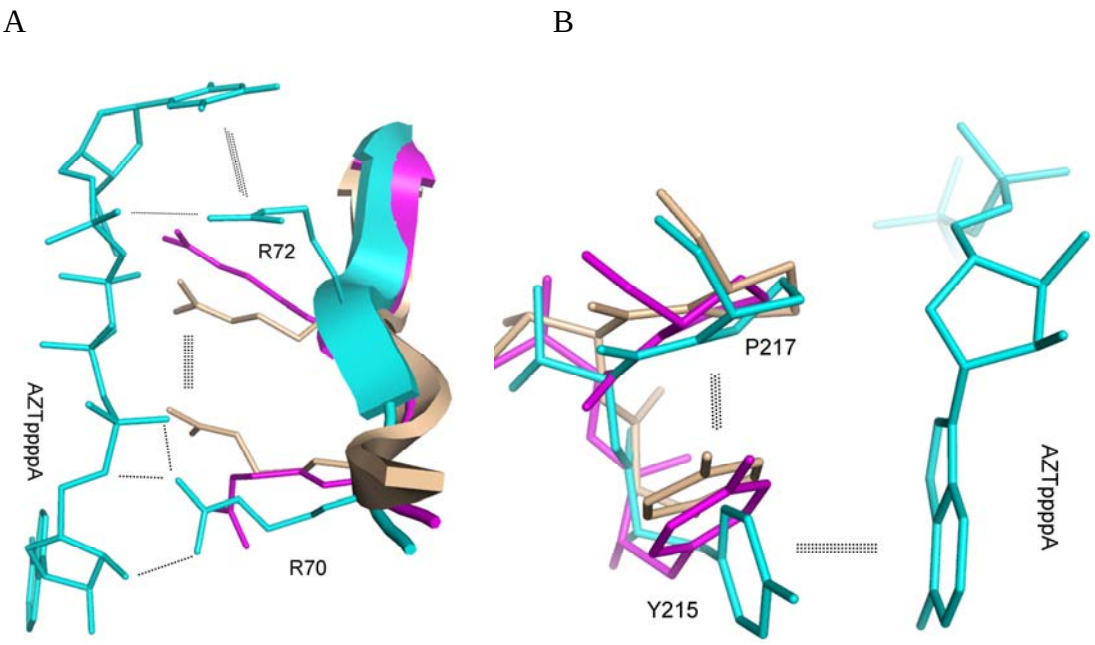


Fig. 80. Conformational changes of the primary resistance mutations induced by the binding of ATP to HIV-1 RT

(A) Conformation of R70 side chain in different intermediates (Superposition is based on residue 55-80 of the p66 fingers subdomain)

(B) Conformation of Y215 side chain in different intermediates (Superposition is based on residue 107 to 115 and 151 to 215 of the p66 palm subdomain**)**

AZT-resistant HIV-1 RT, AZT-resistant HIV-1 RT post-translocation complex and AZT-resistant HIV-1 RT excision product complexes are colored magenta, wheat, and cyan, respectively.

PART V. CONCLUSIONS

I have described five crystallographic structures of HIV-1 RT intermediates involved in the ATP-mediated pyrophosphorolytic excision of AZT in this dissertation, including wild-type HIV-1 RT excision product complex, AZT-resistant HIV-1 RT excision product complex, AZT-resistant HIV-1 RT pre-translocation complex, AZT-resistant HIV-1 RT post-translocation complex, and unliganded AZT-resistant HIV-1 RT. These excision intermediates cover almost every step of the AZT excision mechanism, and give a relatively complete picture of ATP-mediated excision by HIV-1 RT. These structures have not only given direct support to some ideas proposed before (e.g. T215Y), but also provided some new insights into the structural mechanism of HIV-1 RT resistance to AZT. I hope our studies in the AZT resistance mechanism can help to develop drugs that can counterattack the drug resistance or re-sensitize the current nucleotide drugs to their respective resistant HIV-1 RT variants.

How does ATP bind to HIV-1 RT? These structural studies have shown that the ATP moiety of the excision product, AZTppppA, assumes different conformations in wild-type and AZT-resistant HIV-1 RT. Based on this structural data, I propose that ATP binds differently to AZT-resistant RT *versus* wild-type RT. The difference in the conformation of ATP occurs largely in the AMP moiety instead of the β,γ-pyrophosphophate moiety of ATP; the conformation of β,γ-pyrophosphophate moiety of ATP is constrained by the Mg^{2+} ion coordination in both wild-type and AZT-resistant RT. The difference in the conformation of the AMP moiety is mainly

attributed to rotation of the torsional angle of the β-phosphate group of ATP around the β,γ-phosphodiester bond. The torsional angles of the β-phosphate group differ by 90˚ degrees in wild-type *versus* AZT-resistant RT, resulting in more than 10 Å separation between the adenine groups in these two complexes. In the wild-type HIV-1 RT, the AMP moiety of ATP binds to a site away from the AZT-resistance mutation sites. There are no significant interactions between the AMP moiety and AZT-resistance mutation sties in wild-type RT excision product complex are observed. In the AZT-resistant HIV-1 RT, the AMP moiety binds to a site surrounded by AZT-resistance mutations. The primary mutations extensively interact with the AMP moiety. The secondary mutations also directly or indirectly interact with the AMP moiety. Since the only difference between the two enzymes is the presence and absence of AZT-resistance mutations, this conformational difference in ATP should be attributable to the AZT-resistance mutations.

How do the two primary resistance mutations (T215Y and K70R) enhance the ATP-mediated excision? In the background of wild-type enzyme, the AMP moiety of ATP makes no significant contributions to the binding of ATP. The binding of ATP to the wild-type enzyme mostly relies on the coordination of its β,γ-pyrophosphate moiety with an Mg^{2+} ion. That is why wild-type RT has no significant selectivity for nucleotide triphosphates, nucleotide diphosphates, and pyrophosphate (PPi). The observation that PPi has somewhat higher excision efficiency than ATP or other nucleotides can be explained by the lack of adenosine component that makes PPi easier

to diffuse to the Mg^{2+} ion coordination site. In the background of AZT-resistant enzyme, the AMP moiety makes significant contribution to the binding of ATP. The AMP moiety directly contacts with the two primary AZT-resistance mutations, and is surrounded by secondary resistance mutations. The aromatic ring of the primary mutation T215Y stacks with the adenine ring of the AMP moiety of ATP and creates π-π stacking. The π-π stacking primarily occurs to the six-membered ring of the adenine group, which may help explain why purine nucleotides have higher excision rate than pyrimidine nucleotides: pyrimidines do not contain a second ring in the analogous location. The other primary mutation, K70R, has extensive interactions with the AMP moiety. One of the η-nitrogens forms electrostatic interactions with bridging and non-bridging oxygen of the α-phosphate group of ATP, and the other η-nitrogen forms hydrogen bonds with the ribose 3'-OH group of the AMP moiety. These interactions enhance the binding affinity of ATP. That is why the primary AZT-resistance mutations can enhance the ATP-mediated pyrophosphorolytic excision.

What roles do the secondary and other associated AZT-resistance mutations play? It is noteworthy that the secondary mutations are around the ATP-binding site or primary mutations in the AZT-resistant RT. The secondary resistance mutations can assist in the binding of ATP to the resistant RT in either a direct or an indirect way. The side chain of mutant Q219 may interact with the O4 of the ribose ring, and act as a platform for stabilizing the ATP ribose ring position and conformation. The mutant

N67 side chain may provide a better environment for accommodating negatively charged ATP, since D67N mutation neutralizes the negative charge of D67. The mutant L41 may have an indirect impact on the interactions between Q151 and the 3'-azido group, as is observed that 3'-azido group interacts differently with Q151 in the context of wild-type enzyme and AZT-resistant enzyme. The aromatic side chain of L210W may stack with the aromatic side chain of mutant Y215; hence enhance the π-π stacking between Y215 and ATP. E44A/D can possibly improve the π-π stacking by disrupting its interaction with the adenine group of ATP, since Y215 and E44 are on either sides of the adenine group.

How are K70E and K65R antagonistic to the resistance effect of AZT-resistance mutations? Mutation K70E changes the side-chain charge from positive to negative, so it will disfavor the binding of ATP to RT by repelling the negatively charged α-phosphate group of ATP. As a result, this mutation will interfere in binding of ATP to RT, and/or positioning of β,γ-pyrophosphate moiety of ATP. The K65 side chain can bind to the α-phosphate group of ATP, and compete with the primary mutant R70 side chain for this α-phosphate group. The more chemically interactive side chain conferred by mutation K65R will possibly enhance its interaction with the α-phosphate group of ATP and, consequently, disrupt the interaction of R70 side chain with this α-phosphate group. This enhanced interaction may also alter the position of α-phosphate group. The effect could possibly extend to the adenine end of ATP to

attenuate the π-π stacking, and/or to the β,γ-pyrophosphate moiety end of ATP to deflect the precise positioning of β,γ-pyrophosphate moiety necessary for excision.

What is the acceptor of ATP in ATP-mediated excision? It has already been established that the pre-translocation complex is the acceptor of ATP for excision. However, the previously solved structure of wild-type pre-translocation complex has the fingers subdomain in an open configuration. The fingers subdomain accommodates residues critical for the binding of ATP to RT. When the fingers subdomain is in an open configuration, the fingers subdomain can not interact with the ATP. It is only when the fingers subdomain is in a closed configuration that the ATP can be stabilized in its binding site. The structure of the AZT-resistant pre-translocation complex has the fingers in a closed configuration. Crystal packing analysis of this AZT-resistant pre-translocation complex shows that the fingers subdomain can assume either a closed or an open configuration in the crystal lattice, but it selects close configuration. Crystal packing analysis of the previously solved structure of the wild-type pre-translocation complex shows that the open configuration of fingers domain may be stabilized by a Fab antibody fragment through crystal packing interactions. I think the open configuration may be the dominant configuration for the fingers subdomain in the absence of nucleic acid template-primer. I also think that the configuration of the subdomain may actually stay in a largely closed configuration during polymerization or ATP-mediated pyrophosphorolytic excision, but more evidence is needed to clarify the possibilities.

Are there any conformational changes related to AZT-resistance mutations during this excision process? I found that the AZT-resistance mutations have no effect on the local backbone conformation and catalytic site of HIV-1 RT when comparing states following polymerization or excision, but the side chains of primary resistant mutant residues (Y215 and R70) do experience significant conformational changes during excision, which is induced by the binding of ATP to RT. These conformational changes are caused by the extensive interactions between ATP and the side chains of the two primary mutations.

References

Abbondanzieri, E,A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature. 438, 460-465.

Arion, D., Kaushik, N., McCormick, S., Borkow, G., and Parniak, M.A (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.

Arion, D., Sluis-Cremer, N., and Parniak, M.A (2002). Mechanism by which phosphonoformic acid resistance mutations restore 3'-azido-3'-deoxythymidine (AZT) sensitivity to AZT-resistant HIV-1 reverse transcriptase. J. Biol. Chem. 275, 9251-9255.

Arnold, E., Das, K., Ding, J., Yadav, P.N., Hsiou, Y., Boyer, P.L., and Hughes, S.H. (1996). Targeting HIV reverse transcriptase for anti-AIDS drug design: structural and biological considerations for chemotherapeutic strategies. Drug Des. Discov. 13, 29-47.

Balzarini, J., Naesens, L., Aquaro, S., Knispel, T., Perno, C., De Clercq, E., and Meier, C. (1999). Intracellular metabolism of CycloSaligenyl 3'-azido-2', 3'-dideoxythymidine monophosphate, a prodrug of 3'-azido-2', 3'-dideoxythymidine (zidovudine). Mol. Pharmacol. 56, 1354-1361.

Basavapathruni, A., Bailey, C.M., and Anderson, K.S (2004). Defining a molecular mechanism of synergy between nucleoside and nonnucleoside AIDS drugs. J. Biol. Chem. 279, 6221-6224.

Berkhout, B. (1999). HIV-1 evolution under pressure of protease inhibitors: climbing the stairs of viral fitness. J. Biomed. Sci. 6, 298-305.

Berman HM, Gelbin A, Westbrook J. (1996). Nucleic acid crystallography: a view from the nucleic acid database. Prog. Biophys. Mol. Biol. 66, 255-288.

Boucher, C.A., Lange, J.M., Miedema, F.F., Weverling, G.J., Koot, M., Mulder, J.W., Goudsmit, J., Kellam, P., and Larder, B.A. (1992). Tersmette M. HIV-1 biological phenotype and the development of zidovudine resistance in relation to disease progression in asymptomatic individuals during treatment. AIDS. 6, 1259-1264.

Boucher, C.A., O'Sullivan, E., Mulder, J.W., Ramautarsing, C., Kellam, P., Darby, G., Lange, J.M., Goudsmit, J., and Larder, B.A. (1992). Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. J. Infect. Dis. 165,105-110.

Boyer, P.L., Imamichi, T., Sarafianos, S.G., Arnold E., and Hughes, S.H. (2004). Effects of the Delta67 complex of mutations in human immunodeficiency virus type 1 reverse transcriptase on nucleoside analog excision. J. Virol. 78, 9987-9997.

Boyer, P.L., Imamichi, T., Sarafianos, S.G., Arnold, E., and Hughes, S.H. (2004). Effects of the Delta67 complex of mutations in human immunodeficiency virus type 1 reverse transcriptase on nucleoside analog excision. J. Virol. 78, 9987-9997.

Boyer, P.L., Sarafianos, S.G., Arnold, E., and Hughes, S.H. (2001). Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. J. Virol. 75, 4832-4842.

Boyer, P.L., Sarafianos, S.G., Arnold, E., Hughes, S.H. (2002). Nucleoside analog resistance caused by insertions in the fingers of human immunodeficiency virus type 1 reverse transcriptase involves ATP-mediated excision. J. Virol. 76, 9143-9151.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) Crystallography & NMR system (CNS): a new software system for macromolecular structure determination. Acta Crystallogr. D 54, 905–921.

CCP4 (COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4) (1994). The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. D50, 760-763.

Cleland, A., Watson, H.G., Robertson, P., Ludlam, C.A., Brown, and A.J. (1996). Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. J Acquir Immune Defic Syndr Hum Retrovirol. 12, 6-18.

Collins, M., Sondel, N., Cesar, D., and Hellerstein, M. (2004). Effect of nucleoside reverse transcriptase inhibitors on mitochondrial DNA synthesis in rats and humans. J. Acquir. Immune. Defic. Syndr. 37, 1132-1139.

Crumpacker, C.S. (1992). Mechanism of action of foscarnet against viral polymerases. Am. J. Med. 92, 3S-7S.

Delaugerre, C., Roudiere, L., Peytavin, G., Rouzioux, C., Viard, J.P., and Chaix, M.L. (2005). Selection of a rare resistance profile in an HIV-1-infected patient exhibiting a failure to an antiretroviral regimen including tenofovir DF. J. Clin. Virol. 32, 241-244.

Dharmasena, S., Pongracz, Z., Arnold, E., Sarafianos, S.G., and Parniak, M.A. (2007). 3'-Azido-3'-deoxythymidine-(5')-tetraphospho-(5')-adenosine, the product of ATP-mediated excision of chain-terminating AZTMP, is a potent chain-terminating substrate for HIV-1 reverse transcriptase. Biochemistry 46, 828-836

Ding, J., Das, K., Hsiou, Y., Sarafianos, S.G., Clark, A.D., Jacobo-Molina, A., Tantillo, C., Hughes, S.H., and Arnold, E. (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 A resolution. J. Mol. Biol. 284,1095-1111.

Feng, Z., Westbrook, J., and Berman, H.M. (1998) NUCheck. NDB-407. Rutgers University, New Brunswick, NJ.

Gallant, J.E., and Deresinski, S. (2003). Tenofovir disoproxil fumarate. Clin. Infect. Dis. 37, 944-950.

Girouard, M., Diallo, K., Marchand, B., McCormick, S., and Götte, M. (2003). Mutations E44D and V118I in the reverse transcriptase of HIV-1 play distinct mechanistic roles in dual resistance to AZT and 3TC. J. Biol. Chem. 278, 34403-34410.

Goff, S.P. (1990). Retroviral reverse transcriptase: synthesis, structure, and function. J. Acquir. Immune. Defic. Syndr. 3, 817-831.

Gu, Z., Fletcher, R.S., Arts, E.J., Wainberg, M.A., and Parniak, M.A. (1994) The K65R mutant reverse transcriptase of HIV-1 cross-resistant to 2', 3'-dideoxycytidine, 2',3'-dideoxy-3'-thiacytidine, and 2',3'-dideoxyinosine shows reduced sensitivity to specific dideoxynucleoside triphosphate inhibitors in vitro. J. Biol. Chem. 269, 28118-28122.

Guajardo, R., and Sousa, R. (1997). A model for the mechanism of polymerase translocation. J. Mol. Biol. 265, 8–19.

Guajardo, R., Lopez, P., Dreyfus, M. and Sousa, R., (1998). NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation. J. Mol. Biol. 281, 777–792.

Han, Q., Gaffney, B., and Jones, R.A. (2006). One-Flask Synthesis of Dinucleoside Tetra- and Pentaphosphates. Organic Letters 8, 2075-2077.

Hanna, G.J., Johnson, V.A., Kuritzkes, D.R., Richman, D.D., Brown, A.J., Savara, A.V., Hazelwood, J.D., and D'Aquila, R.T. (2000). Patterns of resistance mutations selected by treatment of human immunodeficiency virus type 1 infection with zidovudine, didanosine, and nevirapine. J. Infect. Dis. 181, 904-911.

Hooker, D.J., Tachedjian, G., Solomon, A.E., Gurusinghe, A.D., Land, S., Birch, C., Anderson, J.L., Roy, B.M., Arnold, E., and Deacon, N.J. (1996). An in vivo mutation from leucine to tryptophan at position 210 in human immunodeficiency virus type 1

reverse transcriptase contributes to high-level resistance to 3'-azido-3'-deoxythymidine. J. Virol. 70, 8010-8018.

Hou, X. (2006). Synthesis of DNA and RNA Fragments with a Guanine Thioalkyl Tether, and Synthesis of [2-13C,NH2-15N,6-18O]-Deoxyguanosine. PhD dissertation, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey.

Hsieh, J.C., Zinnen, S., and Modrich, P. Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase.

Hsiou, Y., Ding, J., Das, K., Clark, A.D., Hughes, S.H., and Arnold, E. (1996). Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure. 15, 853-860.

Huang, H., Chopra, R., Verdine, G.L., and Harrison, S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282, 1669-1675.

Huang, H., Harrison, S.C., and Verdine, G.L. (2000). Trapping of a catalytic HIV reverse transcriptase-template:primer complex through a disulfide bond. Chem. Biol. 7, 355-364.

Imamichi, T., Berg, S.C., Imamichi, H., Lopez, J.C., Metcalf, J.A., Falloon, J., and Lane, H.C. (2000). Relative replication fitness of a high-level 3’-azido-3’-deoxythymidine-resistant variant of human immunodeficiency virus type 1 possessing an amino acid deletion at codon 67 and a novel substitution (Thr-->Gly) at codon 69. J. Virol. 74,10958-10964.

Imamichi, T., Sinha, T., Imamichi, H., Zhang, Y.M., Metcalf, J.A., Falloon, J., Lane, H.C. (2000). High-level resistance to 3’-azido-3’-deoxythimidine due to a deletion in the reverse transcriptase gene of human immunodeficiency virus type 1. J. Virol. 74, 1023-1028.

Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., Hizi, A., Hughes, S.H., and Arnold, E. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc. Natl. Acad. Sci. 90, 6320-6324.

Jeeninga, R.E., Keulen, W., Boucher, C., Sanders, R.W., and Berkhout, B. (2001). Evolution of AZT resistance in HIV-1: the 41-70 intermediate that is not observed in vivo has a replication defect.Virology 283, 294-305.

Jonckheere, H., Anne, J., and De Clercq, E. The HIV-1 reverse transcription (RT)

process as target for RT inhibitors. (2000). Med Res Rev. 20,129-54.

K. Cowtan (1994). Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 31, 34-38

Kati, W.M., Johnson, K.A, Jerva, L.F., and Anderson, K.S. (1992). Mechanism and fidelity of HIV reverse transcriptase. J. Biol. Chem. 267, 25988-25997.

Kellam, P., Boucher, C.A., and Larder, B.A. (1992). Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. Proc. Natl. Acad. Sci. 89, 1934-1938.

Keulen, W., Boucher, C., and Berkhout, B. (1996). Nucleotide substitution patterns can predict the requirements for drug-resistance of HIV-1 proteins. Antiviral Res. 31, 45-57

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., and Steitz, T.A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science. 256, 1783-1790.

Komissarova, N. and Kashlev, M. (1997). RNA polymerase switches between inactivated and activated state by translocating back and forth along the DNA and the RNA. J. Biol. Chem. 272, 15329–15338.

Landick, R. (1997). RNA polymerase slides home: pause and termination site recognition. Cell 88, 741–744.

Larder, B.A. (1994). Interactions between drug resistance mutations in human immunodeficiency virus type 1 reverse transcriptase. J. Gen. Virol. 75, 951-957.

Larder, B.A., and Kemp, S.D. (1989). Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). Science. 246, 1155-1158.

Larder, B.A., Bloor, S., Kemp, S.D., Hertogs, K., Desmet, R.L., Miller, V., Sturmer, M., Staszewski, S., Ren, J., Stammers, D.K., Stuart, D.I., and Pauwels, R. (1999). A family of insertion mutations between codons 67 and 70 of human immunodeficiency virus type 1 reverse transcriptase confer multinucleoside analog resistance. Antimicrob Agents Chemother. 43, 1961-1967.

Laskowski, R. A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR 8, 477-486.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J.

Appl. Cryst. 26, 283-291.

Le Grice, S.F.J. Human immunodeficiency virus reverse transcriptase. (1993) In: Skalka AM. Goff SP. Ed. Reverse transcriptase. pp. 163-191. Plainview, New York: Cold Spring Harbor Laboratory Press.

Lin, P.F., Samanta, H., Rose, R.E., Patick, A.K., Trimble, J., Bechtold, C.M., Revie, D.R., Khan, N.C., Federici, M.E., Li, H., et al. (1994). Genotypic and phenotypic analysis of human immunodeficiency virus type 1 isolates from patients on prolonged stavudine therapy. J. Infect. Dis. 170,1157-1164.

Lu X.J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res. 31, 5108-5121.

Lu, H., Macosko, J., Habel-Rodriguez, D., Keller, R.W., Brozik, J.A., and Keller, D.J. (2004). Closing of the fingers domain generates motor forces in the HIV reverse transcriptase. J. Biol. Chem. 279, 54529-54532.

Madrid M, Jacobo-Molina A, Ding J, and Arnold E. (1999). Major subdomain rearrangement in HIV-1 reverse transcriptase simulated by molecular dynamics. Proteins. 35, 332-337.

Marchand, B., Tchesnokov, E.P., and Götte, M. (2007). The pyrophosphate analogue foscarnet traps the pre-translocational state of HIV-1 reverse transcriptase in a Brownian ratchet model of polymerase translocation. J. Biol. Chem. 282, 3337-3346.

Mas, A., Parera, M., Briones, C., Soriano, V., Martinez, M.A., Domingo, E., and Menendez-Arias, L. (2000). Role of a dipeptide insertion between codons 69 and 70 of HIV-1 reverse transcriptase in the mechanism of AZT resistance. EMBO. J. 19, 5752-5761.

Meyer, P.R., Lennerstrand, J., Matsuura, S.E., Larder, B.A., and Scott, W.A. (2003). Effects of dipeptide insertions between codons 69 and 70 of human immunodeficiency virus type 1 reverse transcriptase on primer unblocking, deoxynucleoside triphosphate inhibition, and DNA chain elongation. J. Virol. 77, 3871-3877.

Meyer, P.R., Matsuura, S.E., Mian, A.M., So, A.G., and Scott, W.A. (1999). A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. Mol. Cell 4, 35-43

Meyer, P.R., Matsuura, S.E., So, A.G., and Scott, W.A. (1998). Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. Proc. Natl. Acad. Sci. 95, 13471-13476.

Meyer, P.R., Matsuura, S.E., Zonarich, D., Chopra, R.R., Pendarvis, E., Bazmi, H.Z., Mellors, J.W., and Scott, W.A. (2003). Relationship between 3'-azido-3'-deoxythymidine resistance and primer unblocking activity in foscarnet-resistant mutants of human immunodeficiency virus type 1 reverse transcriptase. J. Virol. 77, 6127-6137.

Miller, V., Ait-Khaled, M., Stone, C., Griffin, P., Mesogiti, D., Cutrell, A., Harrigan, R., Staszewski, S., Katlama, C., Pearce, G., and Tisdale, M. (2000). HIV-1 reverse transcriptase (RT) genotype and susceptibility to RT inhibitors during abacavir monotherapy and combination therapy. AIDS 14, 163-171.

Mitsuya, H., Weinhold, K.J., Furman, P.A., St Clair, M.H., Lehrman, S.N., Gallo, R.C., Bologesi, D., Barry, D.W., and Broder, S. (1985). 3'-azido-3;-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. Med. Sci. 82, 7096-7100.

Nudler, E., Mustaev, A., Lukhtanov, E. and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell 89, 33–41.

Oberg, B. (1989). Antiviral effects of phosphonoformate (PFA, foscarnet sodium). Pharmacol. Ther. 40, 213-285.

Odriozola, L., Cruchaga, C., Andreola, M., Dolle, V., Nguyen, C.H., Tarrago-Litvak, L., Perez-Mediavilla, A., and Martinez-Irujo, J.J. (2003). Non-nucleoside inhibitors of HIV-1 reverse transcriptase inhibit phosphorolysis and resensitize the 3’-azido-3’-deoxythymidine (AZT)-resistant polymerase to AZT-5’-triphosphate. J. Biol. Chem. 278, 42710-42716.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol. 276, 307-326.

Parker, W., White, E., Shaddix, S., Ross, L., Buckheit, R., Germany, J., Secrist, J., Vince, R., Shannon, W. (1991). Mechanism of inhibition of human immunodeficiency virus type 1 reverse transcriptase and human DNA polymerases alpha, beta, and gamma by the 5'-triphosphates of carbovir, 3'-azido-3'-deoxythymidine, 2',3'-dideoxyguanosine and 3'-deoxythymidine. A novel RNA template for the evaluation of antiretroviral drugs. J. Biol. Chem. 266, 1754-1762.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A.T., and Berman, H.M. (1996). New parameters for the refinement of nucleic acid-containing structures. Acta. Crystallogr. D. 52, 57-64.

Rang, H.P., Dale, M.M., and Ritter, J.M. (1995). Pharmacology, 3rd edition, Pearson

Professional Ltd.

Reeder, T.C. and Hawley, D.K. (1996). Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. Cell 87, 767–777.

Ren, J., Esnouf, R.M., Hopkins, A.L., Jones, E.Y., Kirby, I., Keeling, J., Ross, C.K., Larder, B.A., Stuart, D.I., and Stammers, D.K. (1998). 3'-Azido-3'-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. Proc. Natl. Acad. 95, 9518-9523.

Richman, D.D. (2001) HIV chemotheraphy, Nature 410, 995-1001.

Richman, D.D., Guatelli, J.C., Grimes, J., Tsiatis, A., and Gingeras, T. (1991). Detection of mutations associated with zidovudine resistance in human immunodeficiency virus by use of the polymerase chain reaction. J. Infect. Dis. 164, 1075-1081.

Rittinger, K., Divita, G., and Goody, R.S. (1995). Human immunodeficiency virus reverse transcriptase substrate-induced conformational changes and the mechanism of inhibition by nonnucleoside inhibitors. Proc. Natl. Acad. Sci. U S A. 92, 8046-8049.

Roberts, J.D., Bebenek, K., and Kunkel, T.A. (1998). The accuracy of reverse transcriptase from HIV-1. Science 242, 1171-1173.

Rodgers, D.W., Gamblin, S.J., Harris, B.A., Ray, S., Culp, J.S., Hellmig, B., Woolf, D.J., Debouck, C., and Harrison, S.C. (1995). The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. U S A. 92, 1222-1226.

Ross, L., Gerondelis, P., Liao, Q., Wine, B., Lim, M., Shaefer, M., Rodriguez, A., Limoli, K., Huang, W., Parkin, N.T., Gallant, J. and Lanier. R. (2005). Selection of the HIV-1 reverse transcriptase mutation K70E in antiretroviral-naive subjects treated with tenofovir/abacavir/lamivudine therapy. Antivir. Ther. 10, S102.

Sarafianos, S.G., Clark, A.D., Das, K., Tuske, S., Birktoft, J.J., Ilankumaran, P., Ramesha, A.R., Sayer, J.M., Jerina, D.M., Boyer, PL., Hughes, S.H., and Arnold, E. (2002). Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. EMBO J. 21, 6614-6624.

Sarafianos, S.G., Clark, A.D., Tuske, S., Squire, C.J., Das, K., Sheng, D., Ilankumaran, P., Ramesha, A.R., Kroth, H., Sayer, J.M., Jerina, D.M., Boyer, P.L., Hughes, S.H., and Arnold, E. (2003). Trapping HIV-1 reverse transcriptase before and after translocation on DNA. J. Biol. Chem. 278, 16280-16288.

Sarafianos, S.G., Das, K., Ding, J., Boyer, P.L., Hughes, S.H., and Arnold, E. (1999). Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at

the polymerase active site. Chem Biol. 6, R137-146.

Sarafianos, S.G., Das, K., Hughes, S.H., and Arnold, E. (2004). Taking aim at a moving target: designing drugs to inhibit drug-resistant HIV-1 reverse transcriptases. Curr. Opin. Struct. Biol. 14, 716-730.

Sarafianos, S.G., Hughes, S.H., and Arnold, E. (2004). Designing anti-AIDS drugs targeting the major mechanism of HIV-1 RT resistance to nucleoside analog drugs. Int. J. Biochem. Cell Biol. 36,1706-1715.

Sluis-Cremer, N., Arion, D., Kaushik, N., Lim, H., and Parniak, M.A. (2000). Mutational analysis of Lys65 of HIV-1 reverse transcriptase. Biochem. J. 348, 77-82.

Sluis-Cremer, N., Arion, D., Parikh, U., Koontz, D., Schinazi, R.F., Mellors, J.W., and Parniak, M.A. (2005). The 3'-azido group is not the primary determinant of 3'-azido-3'-deoxythymidine (AZT) responsible for the excision phenotype of AZT-resistant HIV-1. J. Biol. Chem. 280, 29047-29052.

Sluis-Cremer, N., Sheen, C.W., Zelina, S., Torres, P.S., Parikh, U.M., and Mellors, J.W. (2006). Molecular mechanism by which the K70E mutation in human immunodeficiency virus type 1 reverse transcriptase confers resistance to nucleoside reverse transcriptase inhibitors. Antimicrob Agents Chemother 51, 48-53.

Spence, R.A., Kati, W.M., Anderson, K.S., and Johnson, K.A. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. Science. 267, 988-993.

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R.G., Boyer, P.L., Hughes, S.H., Pauwels, R., Andries, K., Janssen, P.A., and Arnold, E. (1994). 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. 243, 369-387.

Tao, Y., Farsetta, D.L., Nibert, M.L., and Harrison, S.C. (2002). RNA Synthesis in a Cage—Structural Studies of Reovirus Polymerase λ3. Cell, 111, 733-745.

Tuske, S., Sarafianos, S.G., Clark, A.D., Ding, J., Naeger, L.K., White, K.L., Miller, M.D., Gibbs, C.S., Boyer, P.L., Clark, P., Wang, G., Gaffney, B.L., Jones, R.A., Jerina, D.M., Hughes, S.H., and Arnold, E. (2004). Structures of HIV-1 RT-DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. Nat. Struct. Mol. Biol. 11, 469-474.

Vagin, A., and Teplyakov, A. (1997). MOLREP：an automated program for molecular replacement. J. Appl. Cryst. 30, 1022-1025.

Wang, G. (2004). Synthesis of a 13C, 15N Multilabeled Hammerhead RNA Motif for 15N NMR Metal Binding Studies; and Synthesis of DNA Fragments with a Guanine Thioalkyl Tether. PhD dissertation, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey.

Whitcomb, J.M., and Hughes, S.H. (1992). Retroviral reverse transcription and integration: progress and problems. Annu. Rev. Cell. Biol. 8, 275-306.

White, K.L., Chen, J.M., Margot, N.A., Wrin, T., Petropoulos, C.J., Naeger, L.K., Swaminathan, S., and Miller, M.D. (2004). Molecular mechanisms of tenofovir resistance conferred by human immunodeficiency virus type 1 reverse transcriptase containing a diserine insertion after residue 69 and multiple thymidine analog-associated mutations. Antimicrob. Agents. Chemother. 48, 992-1003.

Yahi, N., Fantini, J., Henry, M., Tourrès, C., and Tamalet, C. (1999). Mutation patterns of the reverse transcriptase and protease genes in human immunodeficiency virus type 1-infected patients undergoing combination therapy: survey of 787 sequences. J. Clin. Microbiol. 37, 4099-4106.

Yarchoan, R., Mitsuya, H., Myers, C., and Broder, S. (1989). Clinical pharmacology of 3'-azido-2',3'-dideoxythymidine (zidovudine) and related dideoxynucleosides. N. Engl. J. Med. 321, 726-738.

Yin, Y.W., and Steitz, T.A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. Cell 6, 393-404.

VITA

XIONGYING TU

EDUCATION

- Ph.D., Chemistry, Expected 05/2008, Rutgers University, Piscataway, NJ
- M.S., Molecular Biology & Biochemistry, 07/2002, University of Science and Technology of China, Hefei, China
- B.S., Bioengineering, 07/1997, Nanchang University, Nanchang, China

PUBLICATIONS

- Tu X, et al., The structural mechanism of ATP-mediated pyrophosphorolysis in HIV-1 reverse transcriptase. To be submitted.
- Lou X, Liu Q, Tu X, Wang J, et al., The Atomic Resolution Crystal Structure of Atratoxin Determined by Single Wavelength Anomalous Diffraction Phasing J. Biol. Chem. (2004), Vol. 279, 39094-39104.
- Wang J, Guo M, Tu X, et al., Purification, partial characterization, crystallization and preliminary X-ray diffraction of two cysteine-rich secretory proteins from Naja naja atra and Trimeresurus stejnegeri venoms Acta Cryst. (2004). D60, 1108-1111.
- Lou X, Tu X, et al., Purification, N-terminal sequencing, crystallization and preliminary structural determination of atratoxin-b, a short-chain alpha-neurotoxin from Naja naja atra venom Acta Cryst. (2003). D59, 1038-1042.
- Tu X, Huang Q, et al., Purification, N-terminal sequencing, crystallization and preliminary X- ray diffraction analysis of atratoxin, a new short-chain alpha-neurotoxin from the venom of Naja naja atra, Acta Cryst. (2002). D58, 839-842