An Integrated Molecular Dynamics, Principal Component Analysis and Residue Interaction Network Approach Reveals the Impact of M184V Mutation on HIV Reverse Transcriptase Resistance to Lamivudine

Soumendranath Bhakat[a], Alberto J.M. Martin [b,c] and Mahmoud E.S. Soliman[a]*

[a] School of Health Sciences, University of KwaZulu-Natal, Westville, Durban 4001, South Africa

[b] Computational Biology Lab, Fundación Ciencia & Vida, Santiago, Chile

[c] Centro Interdisciplinario de Neurociencia de Valparaíso, Facultad de Ciencias, Universidad de Valparaíso, Chile

*Corresponding author: Mahmoud E.S. Soliman, email: soliman@ukzn.ac.za

Graphical Abstract
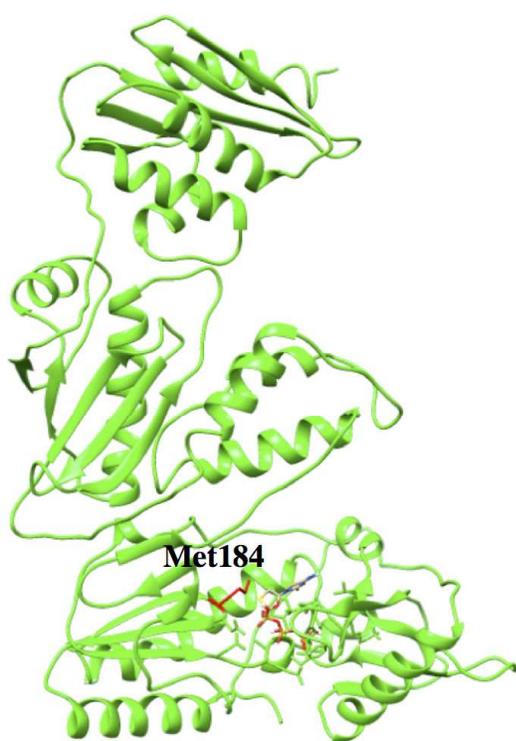
Wild type HIV-RT bound with 3TC

ΔG_{bind} **-46.5489±0.1251 kcal/mol**
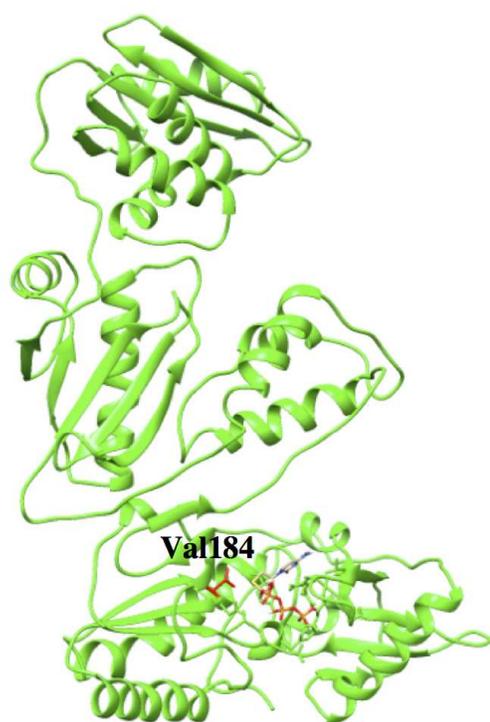
M184V variant HIV-RT bound with 3TC

ΔG_{bind} **-39.4238±0.1513 kcal/mol**

Abstract

The emergence of different drug resistant strains of HIV-1 reverse transciptase (HIV RT) remains a prime interest in relation to to viral pathogenesis as well as drug development. Amongst those mutations, M184V was found to cause a complete loss of ligand fitness. In this study, we report the first account of the molecular impact of M184V mutation on HIV RT resistance to 3TC (lamivudine) using an integrated computational approach. This involved molecular dynamics simulation, binding free energy analysis, Principle Component Analysis (PCA) and Residue Interaction Network (RIN). Results clearly confirmed that M184V mutation leads to steric conflict between 3TC and the beta branched side chain of valine, decreases the ligand (3TC) binding affinity by ~7 kcal/mol when compared to wild type, changes the overall conformational landscape of the protein and distorted the native enzyme residue-residue interaction network.

The comprehensive molecular insight gained from this study should be of great importance in understanding drug resistance against HIV RT as well as assist in the design of novel reverse transcriptase inhibitors with high ligand efficacy on resistant strains.

Keywords: M184V mutation; 3TC; molecular dynamics; binding free energy; principal component analysis; residue interaction network

Running title: Molecular Impact of M184V mutation on lamivudine resistance to RT

1. Introduction

Despite the substantial global effort to find a cure, HIV remains as one of the most challenging organisms to defeat. As of 2012 HIV/AIDS affected approximately 35.3 million people worldwide and resulted in about 1.6 million deaths (*1, 2*). Sub-Saharan Africa is the worst affected region with a whopping 22.9 million people affected with HIV/AIDS (*3*) South Africa is the most drastically affected country with around 6 million people living with HIV/AIDS(*4*).

Development of a target specific antiretrovirals is considered as one of the major breakthrough against HIV/AIDS(*5, 6*). These drugs act on different stages of viral life cycle, thereby halting HIV replication and resulted in decreased viral load. The current treatment against HIV/AIDS comprises of a drug cocktail containing different classes of target specific antiretrovirals (*3*). The multi-componenet cocktail, otherwise known as HARRT (Highly Active Antiretroviral Therapy) cocktail consists of different classes of inhibitors e.g. protease inhibitors (HIV-PRs), integrase inhibitors (HIV-INs), reverse transcriptase inhibitors (HIV-RTs) and virus-entry inhibitors(*7*). Reverse transcriptase inhibitors (HIV-RTs) used in HAART cocktail can further divided into two major categories: a) nucleoside analog reverse transcriptase inhibitors (NRTIs) and b) non-nucleoside reverse transcriptase inhibitors (NNRTIs)(*8, 9*). NRTIs played a crucial role in HAART therapy by acting as a competitive inhibitor of the active site, resulting in termination of synthesis of viral double stranded DNA. In order to act as a competitive inhibitor, NRTIs must be phosphorylated within the cell(*10*).

Rapid emergence of different mutant strains due to error proneness of HIV reverse transcriptase enzyme posses a great threat towards the treatment with NRTIs(*11, 12*). 3TC or lamivudine is one of the crucial NRTIs, which presents in different HAART cominations such as Combivir (GSK), Trizivir (GSK), Epzicom (in USA, GSK)/ Kivexa (in Europe, GSK)(*13*).

In *in-vivo* and *in-vitro* studies reported that single mutations at residue 184 of HIV-RT can cause high-level resistance to 3TC. It has been extensively studied that substitution of methionine with valine at point 184 is responsible for as much as 1000 fold increased resistance and 50 fold decreased activity towards 3TC-TP (3TC 5'-triphosphate)(*11, 14*). Due to the lack of 3TC-HIV RT crystal structure, wild or mutant, the exact molecular mechanism of M184V resistanc to lamivudine is not well understood. It has been believed that introduction of a beta-branched side chain (valine in M184V mutation) (**Figure 1**) creates a clash with the oxathiolane ring of 3TC, leading to dramatic loss in ligand binding affinity (*11, 15, 16*).
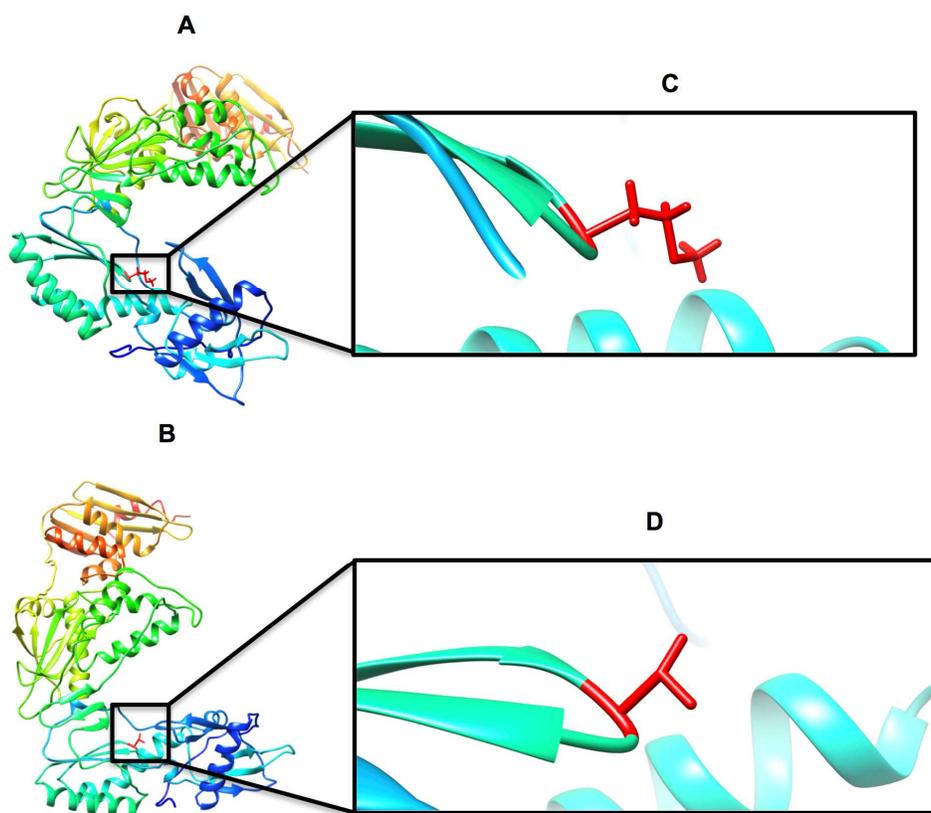
Figure 1. 3-D ribbon representation of HIV-RT wild (A) vs M184V point mutant (B). C and D showing atomic representation of methionine (C) and valine (D) at residue number 184.

In this work, we aim to provide a comprehensive understanding of the impact of the M184V mutation on the activity of lamivudine towards HIV RT. Findings from this study could be critical in future development of more potent and selective NRTIs.

Recent year's molecular dynamics simulation has become the close counterpart of experiment in understanding molecular insight of drug resistancy in a number of different biological systems (*17-23*).

Different post-dynamics analysis approaches has been widely used to gain an overall atomic level understanding from molecular dynamics simulations. Principal Component Analysis (PCA) or essential dynamics analysis is one of widely used post-dynamics analyses which helps understand dynamic behaviour of biological systems (*24*). PCA takes the trajectory of an MD simulation into account and reduces the dimensionality and complexity of the data to understand necessary concerted motions(*25, 26*). PCA has been extensively used to identify difference in motion in context of wild type and mutant/resistant strains(*21, 27*). It has also been studied that mutation(s) in a potein residue leads to the disruption in protein backbone.

Residue Interaction Network (RIN) is also one of the modern tools that can be used to analyse the changes in protein interaction network structures upon mutation (*28*).

In this work, a validated in-silico model of 3TC-RT complexes (M184V mutant and wild type) were built up using ligand fitting approach (See Methods section). 50 ns MD simulations, binding free energy calculations, PCA and RIN analysis were employed to provide insight into the effect of M184V mutation on 3TC resistancy. To our knowledge, this report is the first account of such comprehensive computational analysis being used to reveal the precise impact of M184V mutation on lamivudine resistance to RT at molecular level.

The integrated computational approache applied in this study could serve as a powerful tool to understand effect of mutation on drug resistance and drug-protein interactions and can be applied in drug discovery and development workflow to develop potential inhibitors.

2. Computational Methods

2.1. Generation of 3D coordinates of 3TC-RT complexes

Due to the lack of the crystal structure of 3TC-RT complex, building initial the 3D coordinates for subsequent computational analysis was challenging. In many cases, initial ligand-protein complexes generated by molecular docking ends up with unrealistic ligand orientations in the binding cavity. To this end, and to generate more precise and reasonable initial ligand-bound coordinates, we opted for an in-house developed "ligand fitting" strategy based on the available crystal structure of RT-zidovudine (AZT) complex. "Ligand fitting" approach is fully explained in our previous report (*17*). The choice of AZT was based on that fact that AZT poses high level of structural similarity with 3TC (**Figure 2**). The crystal structure of HIV-1 RT in complex with AZT (PDB code: 1RTD) has been extracted from the PDB and used to generate initial 3D coordinates(*29*). This protein crystal structure consists of eight chains but only chain A bound with AZT was used to reduce the computational cost and time. The mutated residues associated with the crystal structure, has been modified to generate native wild type RT-AZT complex. Co-crystallized water molecules were deleted. The structure of 3TC was then superimposed against AZT (**Figure 3**), and later was then removed to generate initial 3TC conformation at the active site of HIV reverse transcriptase (RT). In-silico mutation was carried out to mutate methionine (MET) at position 184 to valine (Valine). Chimera package (*30*) was used to accompolish the above-mentioned modelling tasks.
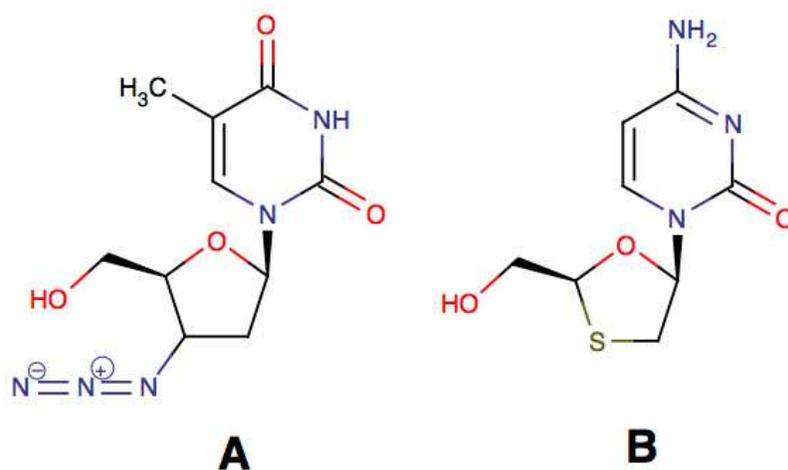
Figure 2. 2D chemical structures of AZT (zidovudine) and 3TC (lamivudine)
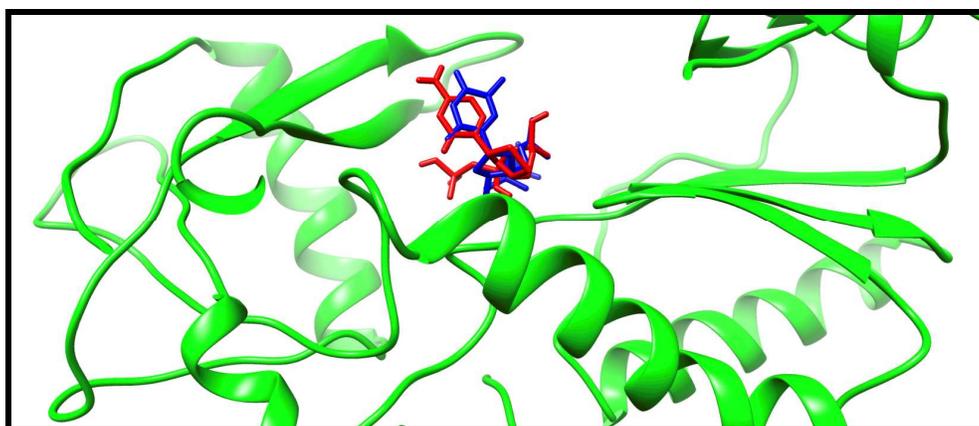
Figure 3. 3TC (Red) is superimposed against the Zidovudine/AZT (Blue) in the active site pocket of HIV-1 reverse transcriptase (PDB code: 1RTD)

2.2. Molecular Dynamics Simulation

The wild type and M184V variant of HIV RT in complex with 3TC were subjected to all-atom, unrestrained molecular dynamics simulations in explicit solvent using the GPU version PMEMD engine provided with Amber12(*31, 32*). The geometry of the ligands were optimized with Gaussian 09 at HF/6-31G* level(*13*). Finally antechamber module was used to generate atomic partial charges using restrained electrostatic potential (RESP) procedure(*33*) and force field parameters using GAFF(*34*). The ff99sb force field implemented with Amber12 was used to describe the protein systems (*35*). The Leap module integrated with Amber12 was used to add missing hydrogen atoms and counter ions for neutralisation. Prior to set up, both systems were immersed into an orthorhombic box with TIP3P (*36*) water molecules such that no protein atom was within 8 Å of any box edge. Long-range electrostatic interactions have been calculated using particle mesh Ewald (PME) method (*37*)with a direct space and vdW cut-off of 12 Å. Initial minimization with a restrained potential of 500 kcal/mol $Å^2$ applied to the solute, was performed for 1000 steps of steepest descent followed by an 1000 steps of conjugate gradients. A further 1000 steps of minimization was carried out using conjugate gradients algorithm. All minimizations were performed using CPU version of Amber12. Prior to minimization, a gradual heating from 0 to 300K was performed for both systems with a harmonic restraint of 5kcal/mol $Å^2$ applied to all the solute and a Langevin thermostat with a collision frequency of 1/ps using a canonical ensemble (NVT) MD simulation. Both systems were subsequently equilibrated at 300k in a NPT ensemble for 500 ps with no restrained and a Berendsen barostat was used to maintain the pressure at 1 bar. The SHAKE algorithm(*38*) was used to constrain the bonds of all hydrogen atoms, a time step of 2fs and SPFP precision model (*39*)was used for all MD runs. A production

MD run for continuous 50 ns was performed in a NPT ensemble with a target pressure of 1 bar and a pressure coupling constant of 2 ps.

The trajectories were analysed in every 1 ps and further analysis e.g. RMSD, RMSF, radius of gyration, solvent accessible surface area, hydrogen bond occupancy over time, inter atomic distances and PCA were carried out using PTRAJ and CPPTRAJ module implemented in Amber12. All visualisation and plots were carried out using Chimera molecular modelling tool (*30*) and Origin data analysis software(*40*) respectively.

2.3. Thermodynamic Calculations

The binding free energy profiles of 3TC bound wild type and M184V variants of HIV-RT were computed using the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) approach(*41-43*). The binding free energy was calculated taking in account 1000 snapshots from a 50ns production run. Binding free energy calculation is an end point energy calculation which offers a valuable insight into the association of protein-ligand complex. The following set of equations describes the calculation of the binding free energy:

$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand}$ (1)

$\Delta G_{bind} = E_{gas} + G_{sol} - T\Delta S$ (2)

$E_{gas} = E_{int} + E_{vdw} + E_{ele}$ (3)

$G_{sol} = G_{GB} + G_{SA}$ (4)

$G_{SA} = \gamma SASA$ (5)

Where E_{gas} signifies gas-phase energy; E_{int} signifies internal energy; and E_{ele} and E_{vdw} signifies the electrostatic and Van der Waals contributions, respectively. E_{gas} is evaluated directly from the FF99SB force field terms. The solvation free energy, denoted by G_{sol}, can be decomposed into polar and nonpolar contribution states. The polar solvation contribution, G_{GB}, is determined

by solving the GB equation, whereas, G_{SA}, the nonpolar solvation contribution is estimated from the solvent accessible surface area (SASA) determined using a water probe radius of 1.4 Å. T and S correspond to temperature and total solute entropy, respectively.

In order to determine the individual amino acid contribution towards total binding free energy profile between 3TC and wild-type/M184V variant of HIV-1 RT, a per-residue decomposition analysis of the interaction energy for each residue was carried out using MM/GBSA method.

2.4. Principle Component Analysis (PCA)

Before processing the MD trajectories for PCA, the 50ns MD trajectories of the wild and mutant complexes were stripped of solvent and ions using the PTRAJ module of AMBER 12. The stripped trajectories were then aligned against the fully minimized structure. PCA was performed for C-α atoms on 1000 snapshots each at 50ps intervals. Using in-house scripts, the first two principal components were calculated and the covariance matrices were generated. The first two principal components correspond to the first two Eigen vectors of the covariance matrix. The PCA scatter plots were then created using the Xmgrace program (http://plasma-gate.weizmann.ac.il/ Grace/). All structural diagrams were created using VMD(44). The porcupine plots were generated using the ProDy interface of normal mode wizard (NMW) of VMD(45).

2.5. Residue Interaction Networks (RIN) analysis

The average structure derived from the 50 ns trajectory of each system, wild and M184V mutant was used for constructing the RIN interactively in 2D graphs using RING [19]. The RINs used in this work were defined using PROBE [37] to identify interacting amino acids.(46) PROBE uses a small virtual probe (typically 0.25 Å) that is rolled around the van der Waals surface of each atom and interactions (contact dot) are detected if the probe touches another non-covalently

bonded atom. Once interaction amino acids have been defined, RING uses several tools to define non covalent interactions between them (e.g. interatomic contact, hydrogen bonds, salt bridges, pi-pi interaction etc.)

2.5.1. Interactive visual analysis of residue networks

The residue interaction networks (RIN) generated from the averaged MD structures were used to visualize the network using RINAnalyzer(*47*) plugin integrated with Cytoscape(*48*). Simple visual inspection of the different RINs allows to easily extract and define changes in the networks that reflect physico-chemical alterations in the structures represented in the RINs.

3. Results and Discussion

3.1. Post-dynamic analysis: Wild type versus M184V mutant

3.1.1. Root of mean square fluctuation (RMSF)

Figure 4 shows the residue-based root mean square fluctuations (RMSF) of the wild-type and M184V mutant over the 50 ns simulation run. As evident from Figure 4, it is clear that M184V has altered the overall protein flexibility when compared to the wild type. Most noticeably, the amino acids region 170-220 (which contains Met184/Val184). One possible explanation is that the Val184 residue in the mutant interacts less with the nearby residues than methionine.
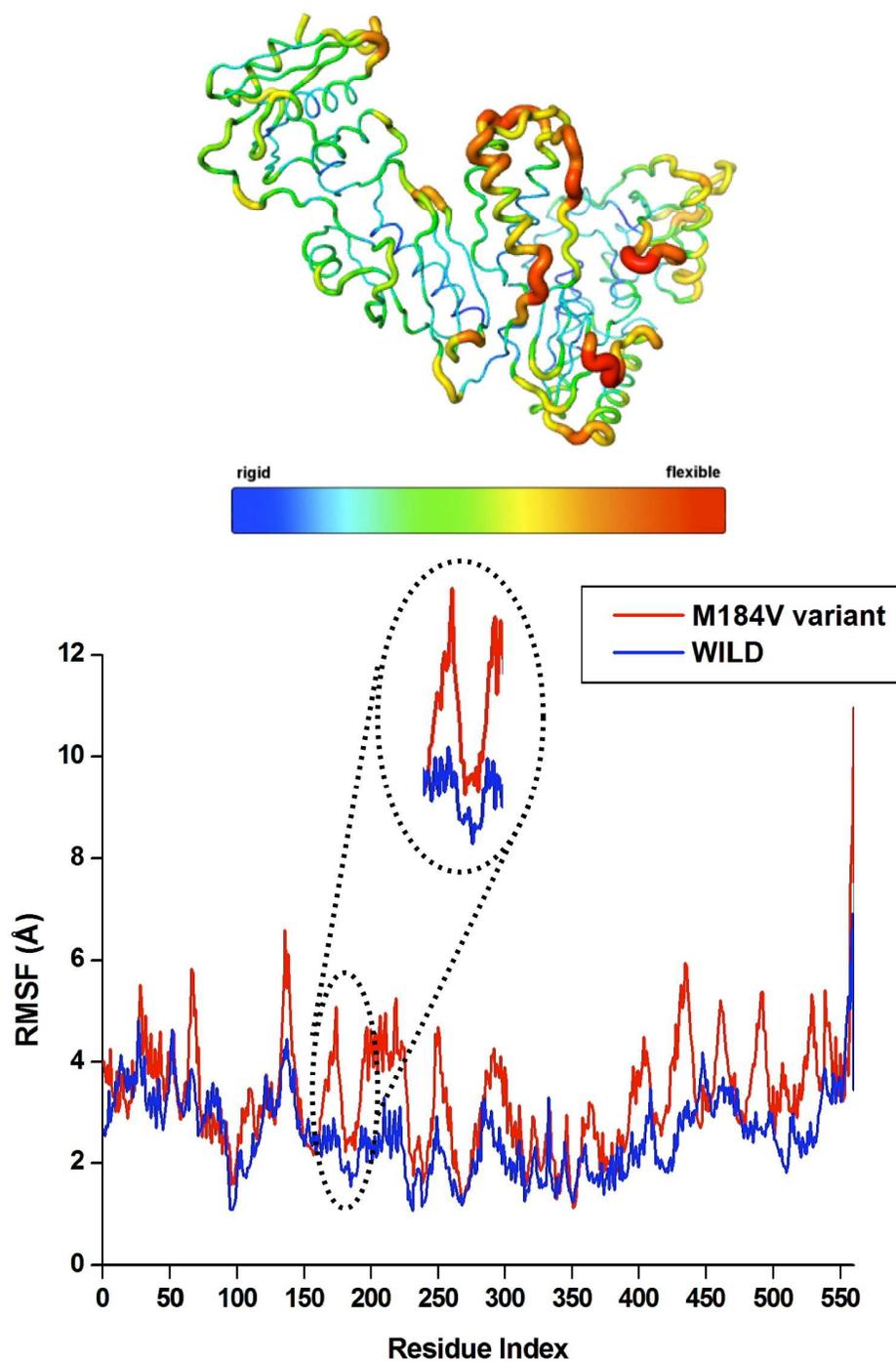
Figure 4. Per-residue amino acid fluctuation for 3TC bound with wild and M184V variant of

HIV-RT. The zoomed plot highlighting more fluctuation at region 170-200.

Interestingly, as indicated above, the M184V mutation was also found to impact the overall dynamics of some distal amino acid regions (**Figure 4**). For instances, residues ranging from 2-50 exhibit larger fluctuations while residues 300-400 were found to be less fluctuating relative to the wild type.

3.1.2. Radius of Gyration

Radius of gyration of native wild type and M184V variant complexed with 3TC were carried out at 300K. In the conformational analysis, radius of gyration defined as the moment of inertia of group of atoms from its centre of mass. **Figure 5** highlights a comparative radius of gyration of Cα atoms between wild and M184V variant. It clearly reflects the highly unstable nature of mutant complex, as compared to wild type. This is in accordance with our assumption that mutation decreases the interaction among neighbouring amino acids, which leads to unstable moment of inertia.
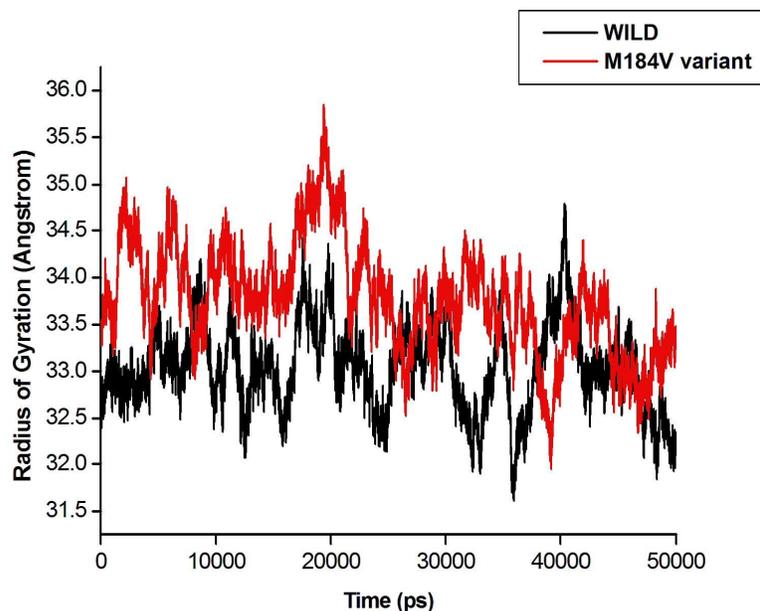
Figure 5. Radius of gyration of C-alpha atoms of native wild type and M184V variant HIV-RT complexed with 3TC

3.1.3. Solvent Accessible Surface Area (SASA)

Measurement of SASA provides insight into compactness of hydrophobic core, which plays a key factor in biomolecular stability(*49*). The alteration of solvent accessible surface area (SASA) between wild type and M184V variant with simulation time is depicted in **Figure 6**. M184V variant mutant exhibited a comparatively higher SASA value than native wild type. A large breathing of SASA in case of mutant verified the unstable nature of M184V HIV-RT-3TC complex. This fact further justifies findings from RMSF and radius of gyration.
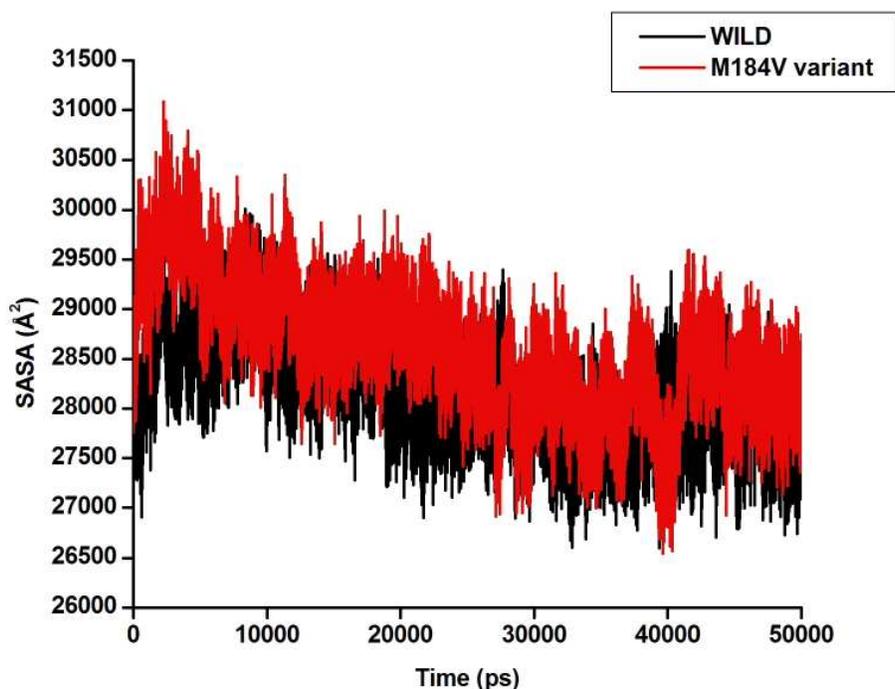
Figure 6. SASA ($Å^2$) over simulation time (ps) for both M184V variant and native wild type HIV-RT complexed with 3TC.

3.1.4. Hydrogen bond formation

Formation of hydrogen bonds between residues considered as a major force on controlling stable conformation of protein structure. Hydrogen bond formation during simulation time was monitored for both wild and M184V variant HIV-RT complexed with 3TC (**Figure 7**). The

mutant exhibited comparatively lower hydrogen bond formation during simulation time when compared to wild type.
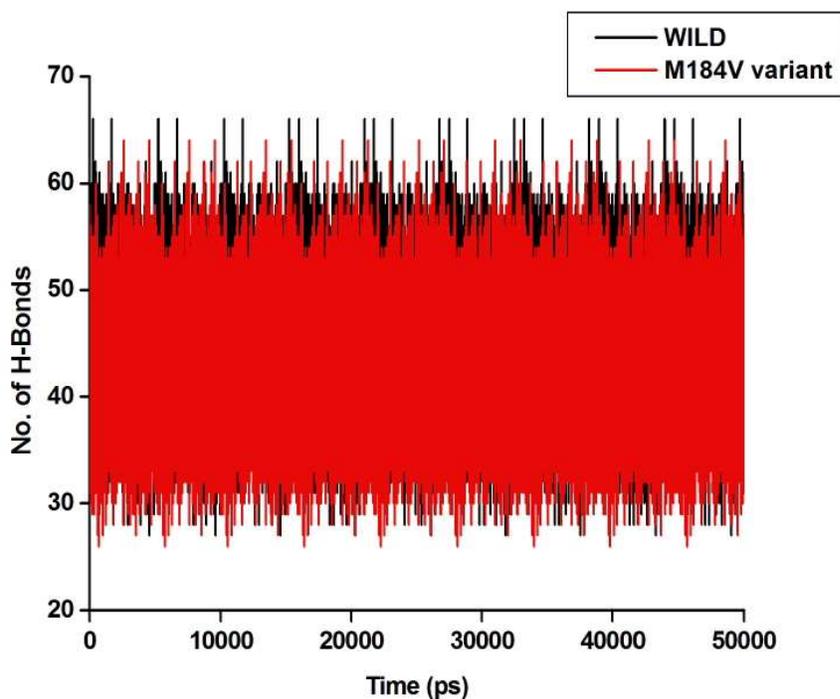
Figure 7. No. of H-bond formation over the simulation time between native wild and M184V variant HIV-RT complexed with 3TC.

3.1.5. Per-residue Secondary Structure Elements (SSE)

Further information on structural plasticity comparing native wild type and M184V variant of HIV-RT was obtained from careful understanding of per-residue occupancy of different secondary structural elements (SSE) in both cases. **Figure 8** showed that at an average occupancy of secondary structure components such as α-helix, 3_{10}-helix, parallel beta sheet, pi-helix and turn were less abundant in case of M184V mutant when compared to the native wild type, though a very slight difference was observed in case of antiparallel-beta sheet component. The per-residue occupancy of secondary structural elements further highlighted the difference in

SSE between wild and M184V variant, which further verify the loss of stability of protein structure due to mutation.
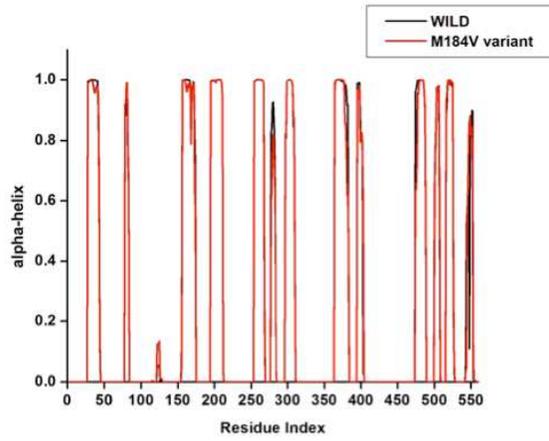
A
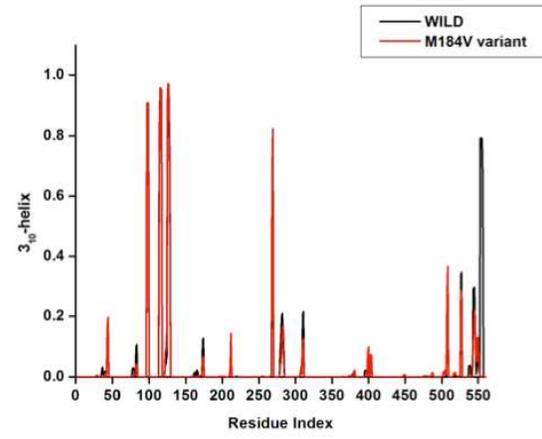
B
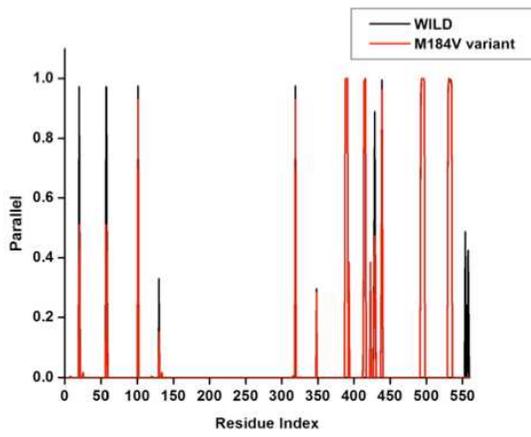
C
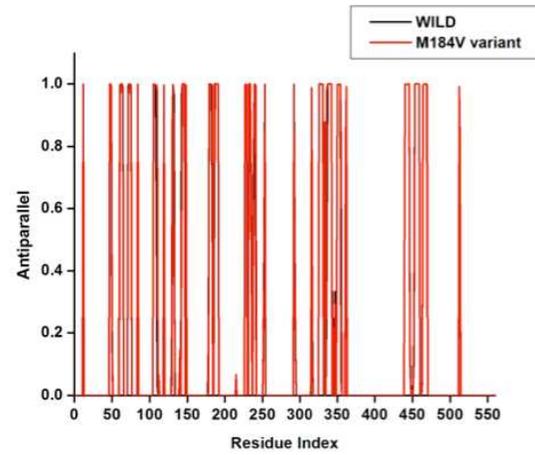
D
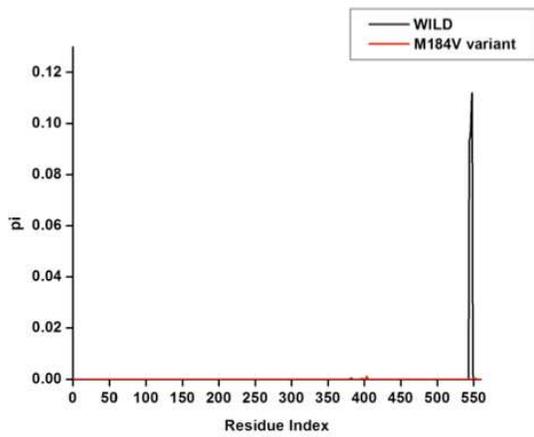
E
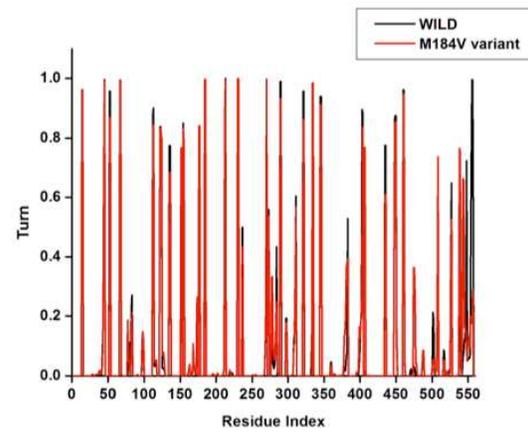
F

Figure 8. Comparative per-residue secondary structure occupancy between native wild and

M184V variant HIV-RT complex. A,B,C,D,E,F denotes alpha helix, 310 helix, parallel beta

sheet, antiparallel beta sheet, pi-helix and turn respectively.

3.1.2. Steric conflict between beta-branched Val184 and oxathiolane ring of lamivudine

Steric conflict between the oxathiolane ring of 3TC and the side chain of beta branched amino

acid Val at position 184 has been previously assumed to be responsible for the perturbed

inhibitor binding. However, this assumption was based on static models (*15*). To verify this

assumption, we monitored the distance between the (O) atom of oxathiolane ring and the Cβ of

the amino acid residues at position 184 (Met in case of wild type and Val in case of mutant)

during the period of MD simulation. **Figure 9** suggests that substitution of methionine to valine

at position 184 resulting in an increased steric conflict with the ligand that continues to exist

during the duration of the MD simulation, this fact is evident from the increased separation

between the (O) atom of oxathiolane ring and the Cβ of the amino acid residue of Val184. These

findings are in a great accordance with the assumption of Sarafianos and Huang *et al* (*13*).
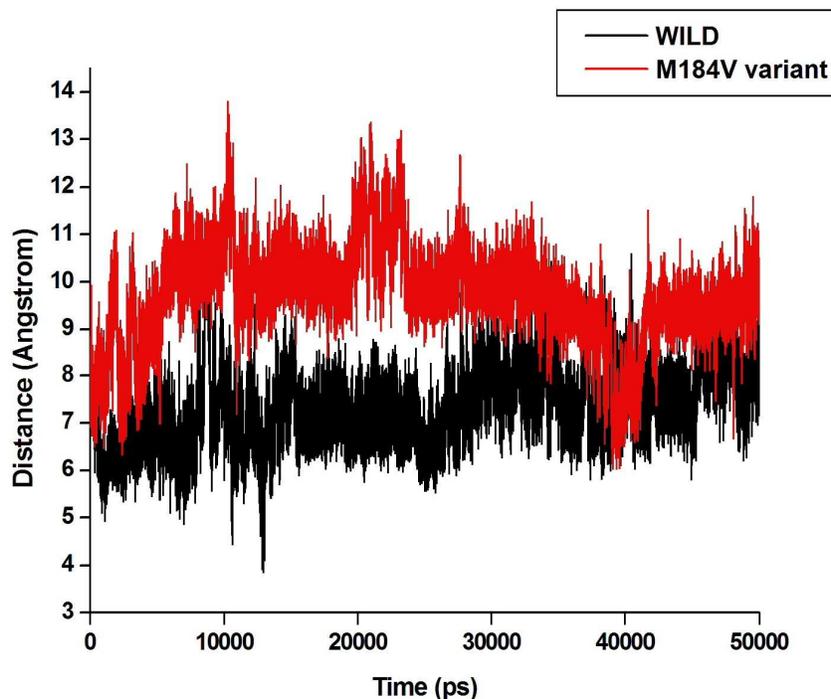
Figure 9. The distance between the (O) atom of oxathiolane ring present in 3TC and the Cβ of the amino acid residue at position 184 (Met in wild type and Val in mutant)

3.1.3. MM/GBSA binding free energy calculations

All the components of molecular mechanics and solvation energy representing average quantities calculated over the 50 ns simulation using the MM/GBSA technique are listed in **Table 1**.

Table 1 MM/GBSA based binding free energy profile of 3TC bound with wild and M184V variant of HIV-RT.

Complexes	ΔG_{bind}	ΔE_{ele}	ΔE_{vdW}	ΔG_{gas}	ΔG_{sol}
Wild	-46.5489±0.1251	-25.0332±0.1442	-48.3440±0.1012	-73.3772±0.1851	26.8283±0.1104
M184V variant	-39.4238±0.1513	-8.0217±0.1586	-43.5869±0.1322	-51.6086±0.2062	12.1847±0.1432

The calculated binding free energy (ΔG_{bind}) between 3TC and wild-type RT is -46.5489 kcal/mol compared to -39.4238 kcal/mol in the case of M184V varint HIV-RT. Such a large reduction in binding affinity (~7 kcal/mol) as a result of mutation could significantly impair ligand binding and thus the effectiveness of 3TC against the M184V mutant decreased significantly – this is in great accordance with experimental evidence that M184V mutation leads to a ~50-fold diminished sensitivity towards 3TC 5'-triphosphate (3TC-TP) and as much as ~1000 fold increased resistancy towards 3TC with an 50% increase in inhibitory constant (IC_{50}) of 3TC in tissue culture evaluations (*65*). The calculated van der Waals contributions (ΔE_{vdW}) to the total binding free energy in the 3TC bound wild-type RT complex (-48.3440 kcal/mol) is higher than that for the 3TC bound M184V mutant RT complex (-43.5869 kcal/mol). On another hand, the calculated electrostatic contributions (ΔE_{ele}) to the binding free energy for 3TC bound M184V mutant RT complex (-8.0217 kcal/mol) are lower compared to that for the 3TC bound wild type RT complex (-25.0332 kcal/mol). In the 3TC bound wild type RT complex, the calculated solvation contributions ($\Delta G_{sol.}$ $\Delta G_{sol} = \Delta G_{SA} + \Delta G_{GB}$) to the binding free energy (26.8283 kcal/mol) are higher than that for the M184V variant (12.1847 kcal/mol). The free energy components shown in Table 1 suggest that the majority of the favourable contributions observed for 3TC binding arise from ΔE_{vdW} and $\Delta E_{ele.}$ The remarkable difference in the Δg_{gas} (~22 kcal/mol) as a result of mutation also confirmed that M184V mutation could lead to significant loss of efficacy of 3TC which further supports previously reported experimental findings.

3.1.4. Per-residue interaction energy decomposition analysis

We further decomposed the total binding free energy into contributions from each HIV-RT amino acid residues from both wild and M184V variant. In **Table 2**, the comparison of protein-ligand interaction spectra between wild-type and M184V mutant is highlighted. It can be observed from the energy decomposition analysis that in case of 3TC bound wild-type HIV-RT system, the major contributions were -3.782(vdW) and -4.962 (elec), -2.491(vdW) and -0.305 (elec); 0.446(vdW) and 0.069(elec) kcal/mol from amino acid residues Arg72, Tyr115, and Met184, respectively (Table 2). On the other hand, there were some minor contributions towards the interaction energy from residues Asp110 (-0.159 kcal/mol), Ala114 (-0.861 kcal/mol) and Phe116 (-0.857 kcal/mol).

Table 2 Highlighting prominent residues for the 3TC-RT interaction on the basis of decomposed van-der Waals and electrostatic energy (kcal/mol) contribution

Residues	van-der Waals energy (kcal/mol)	Electrostatic energy (kcal/mol)
Arg72	-3.782±0.497	-4.962±1.398
	-1.883±1.315*	-1.847±3.2638
Met184	-0.446±0.376	0.069±0.143
Val184	-0.903±0.376*	-0.077±0.512*
Tyr115	-2.491±0.835	-0.305±0.306
	-2.034±0.652*	-0.197±0.527*

*M184V mutant variant

As depicted in **Table 2**, the decomposed van der Waals and electrostatic energies for residues 72, 115 and 184 in the 3TC bound M184V mutant RT complex implies that the change in the van

der Waals contribution is what is mostly responsible for the decrease in the overall interaction energy. Furthermore, when Met mutated to Val at position 184, the prominent van der Waals contributions towards the total binding free energy from increased by two fold with slight increase in electrostatic contribution but it ultimately lowers down the total van der waals contributions of other prominent residues which ultimately affects the total binding free energy. This finding stands consistent with a decreased van der Waals contributions from Arg72 (-3.782 to -1.883 kcal/mol) and Tyr115 (-2.491 to -2.034 kcal/mol) and an overall loss of van der Waals contribution towards binding free energy (**Table 1, 2**).

Interestingly, as evident from **Table 2**, however mutation of Met to Val at position 184 has improved the binding at the site of mutation - this might be due to better hydrophobic interactions with the Val side chain but the overall drug binding affinity is reduced due to the negative impact of the mutation on the binding affinity of the nearby active nearby residues as mutation at position 184 affects the highly conserved area of RT enzyme.
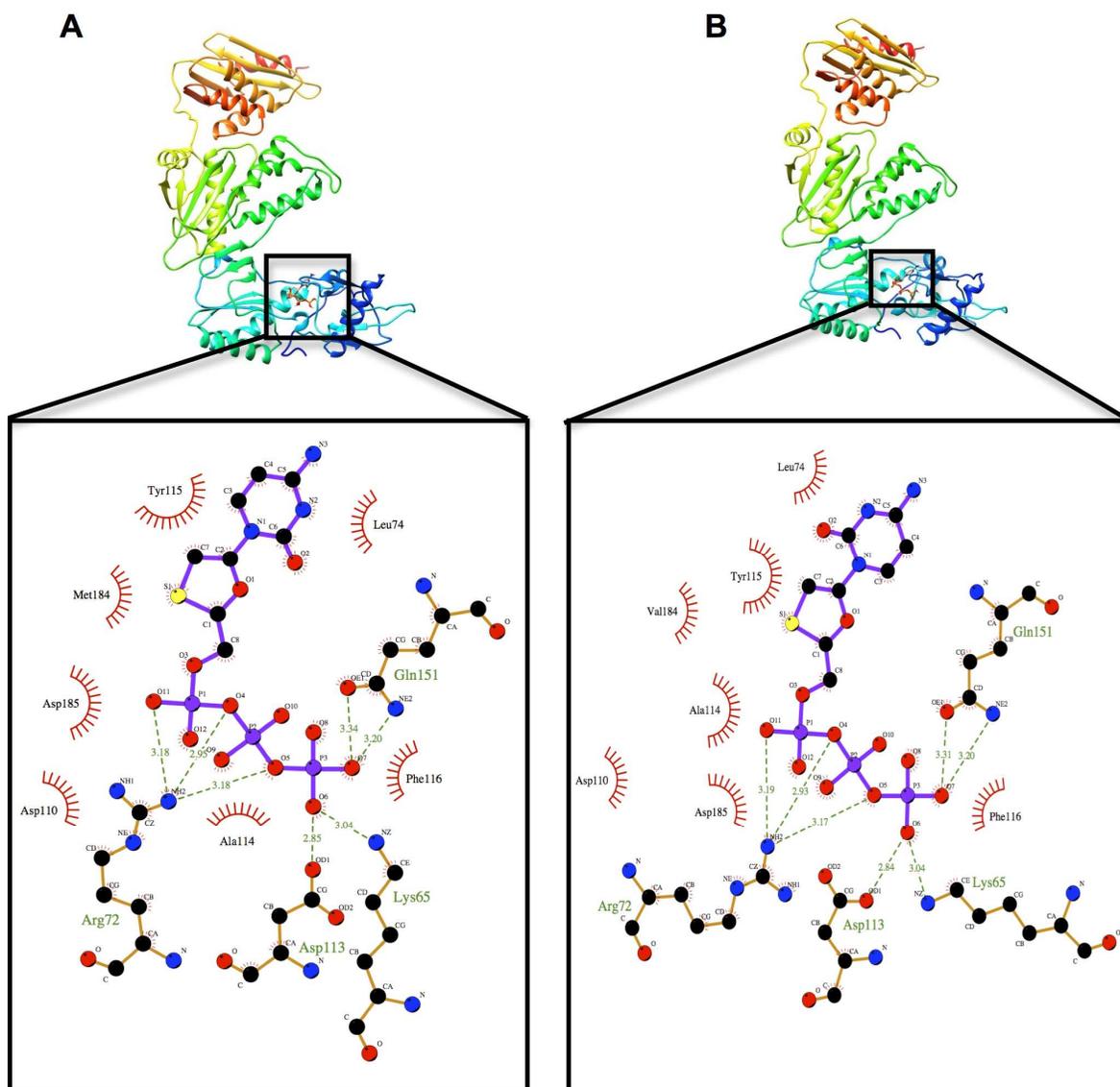
Figure 10. The representative sructure of wild type (A) and M184V variant (B) of HIV-RT complexed with 3TC

3.2. Principle component analysis (PCA)

Figure 11 shows a PCA scatter plot generated for the 3TC bound wild type and M184V mutant RT complexes showing a significant difference between both systems as evident from the characteristic structures plotted along the direction of two principal components. From the scatter

plot, it is clear that eigenvectors computed from the MD trajectory for both systems are quite

varied which clearly indicates difference in protein motion between wild and M184V mutant.
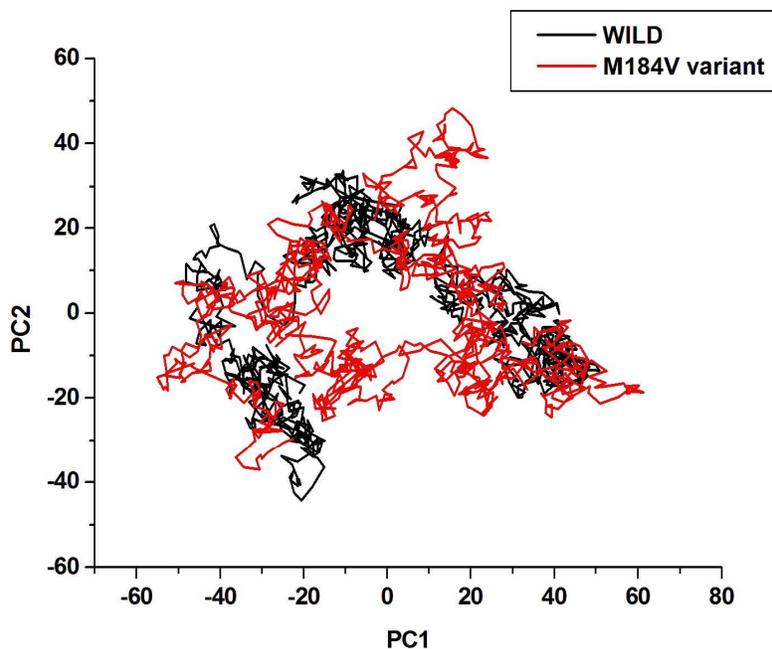
Figure 11. PCA scatter plot of 1000 snapshots alomg first two principal componenets, PC1 and
PC2 respectively showing difference in motion between wild and M184V variant over
simulation time.

In most cases a few low frequency normal modes are enough to capture major protein motions

along specific directions, which are represented by eigenvectors. In this case we generate

porcupine plots with three low frequency modes (Mode=1, 2 & 3) to visualize the comparative

motion difference between 3TC bound wild type and mutant complexes (**Figure 12, 14**). The

eigenvectors show a clear difference in direction of motion, which is consistent with the PCA

scatter plot and residue based mobility plots across three different modes (**Figure 13, 15**). **Table**

3 highlights list of residues which contributed prominently towards the majority of motions

across different normal modes in each case. It was observed that residues at M184V variant of

HIV-RT-3TC complex are bit more mobile than native wild type across three normal modes.

Methionine (magneta) at position 184 in HIV-RT 3TC (red) complex

Mode 1   Mode 2   Mode 3
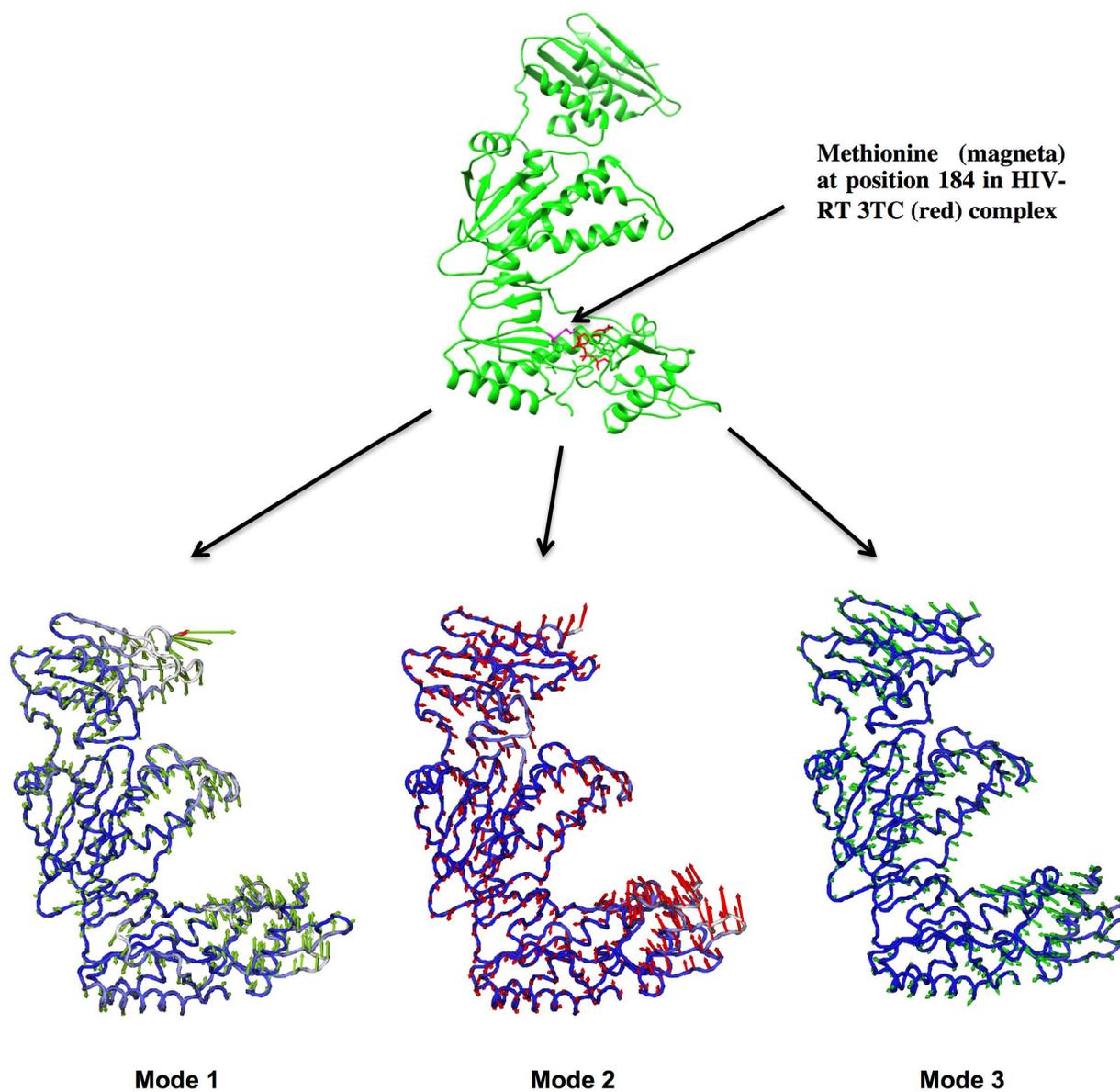
Figure 12. Porcupine plots showing prominent motions related with wild type HIV-RT bound with 3TC. Yellow, red, green represents eigenvectors showing direction of prominent motions across mode 1, 2 and 3 respectively.
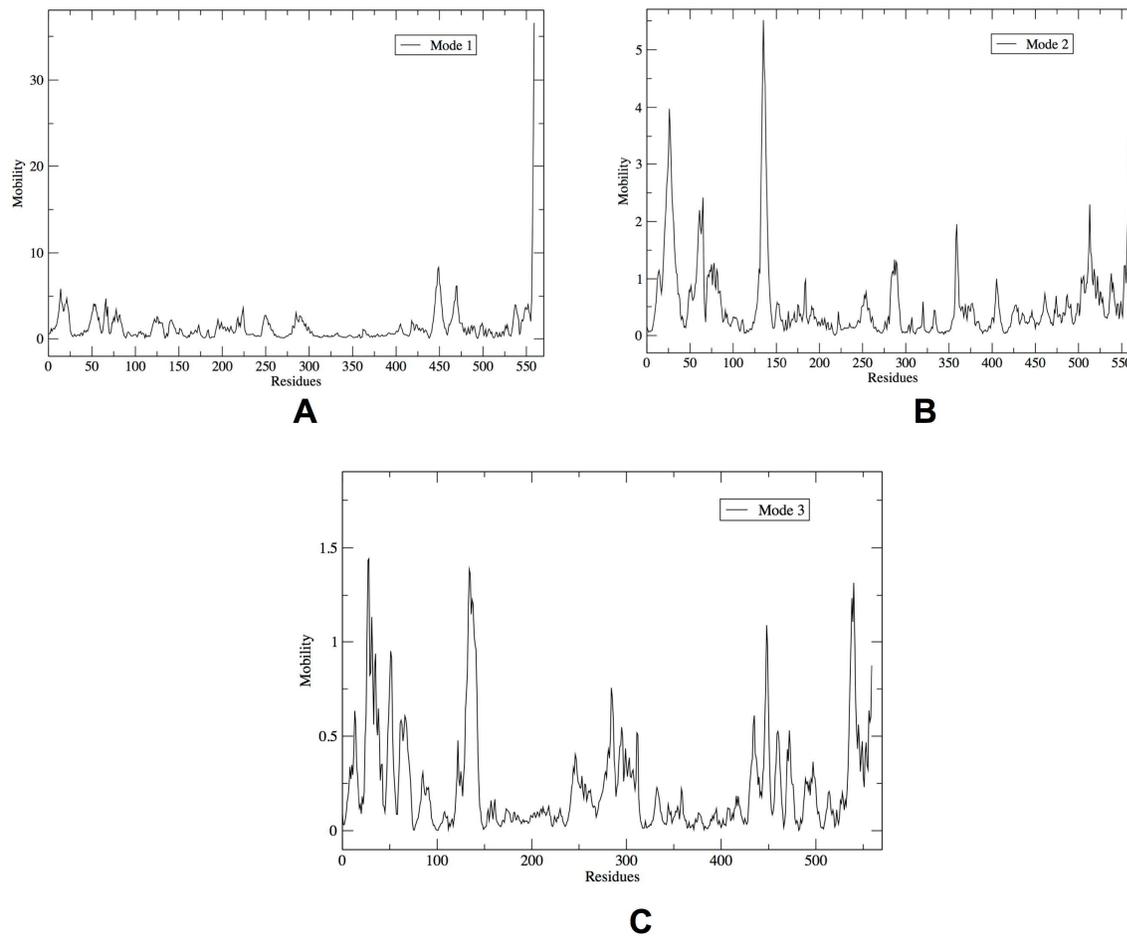
Figure 13. Residue based mobility plot of HIV-RT bound with 3TC showing mobility of different residues across different modes

Valine (magneta) at position 184 in HIV-RT 3TC (red) complex

Mode 1   **Mode 2**   **Mode 3**
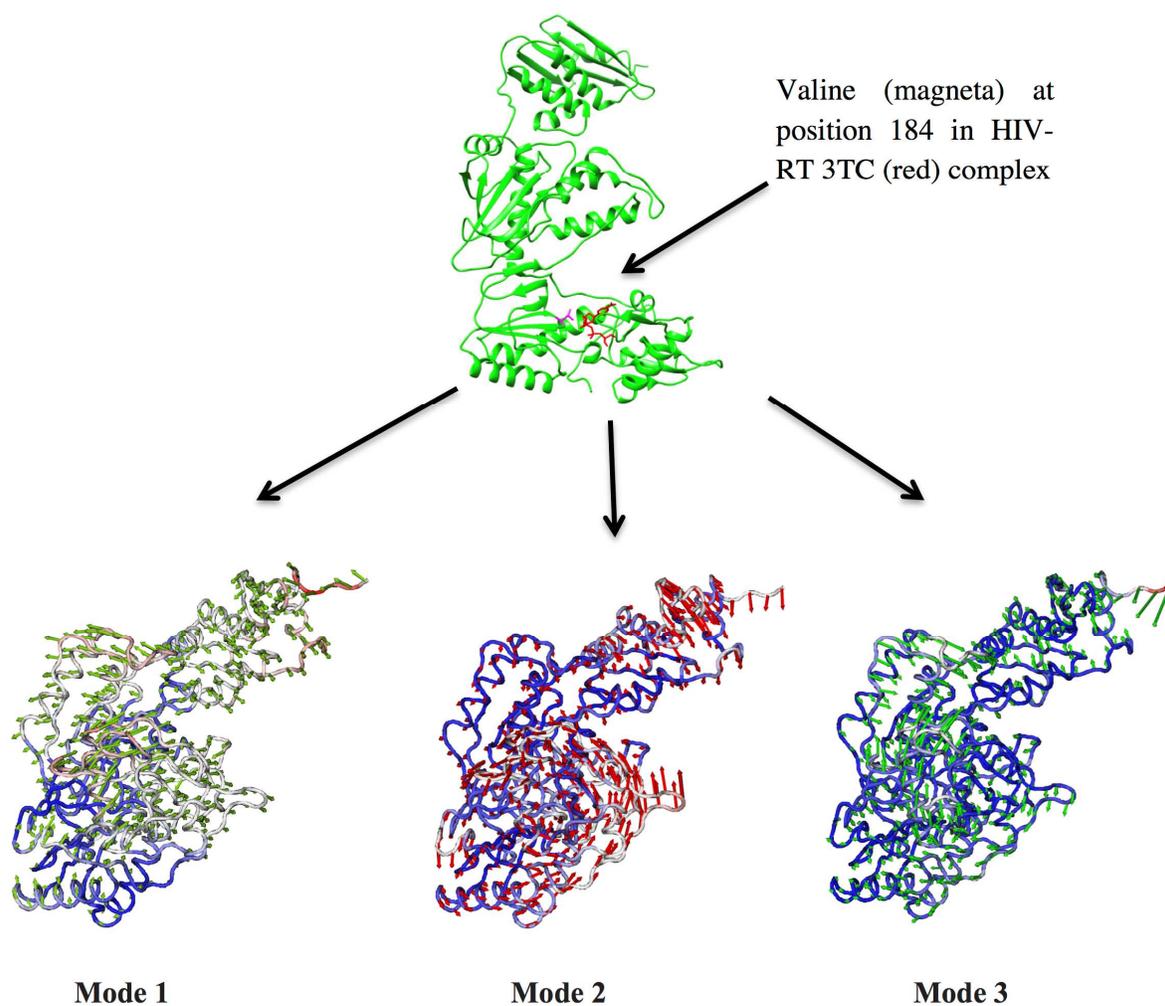
Figure 14. Porcupine plots showing prominent motions related with mutant (M184V) type HIV-RT bound with 3TC. Yellow, red, green represents eigenvectors showing direction of prominent motions across mode 1, 2 and 3 respectively.

A

B

C
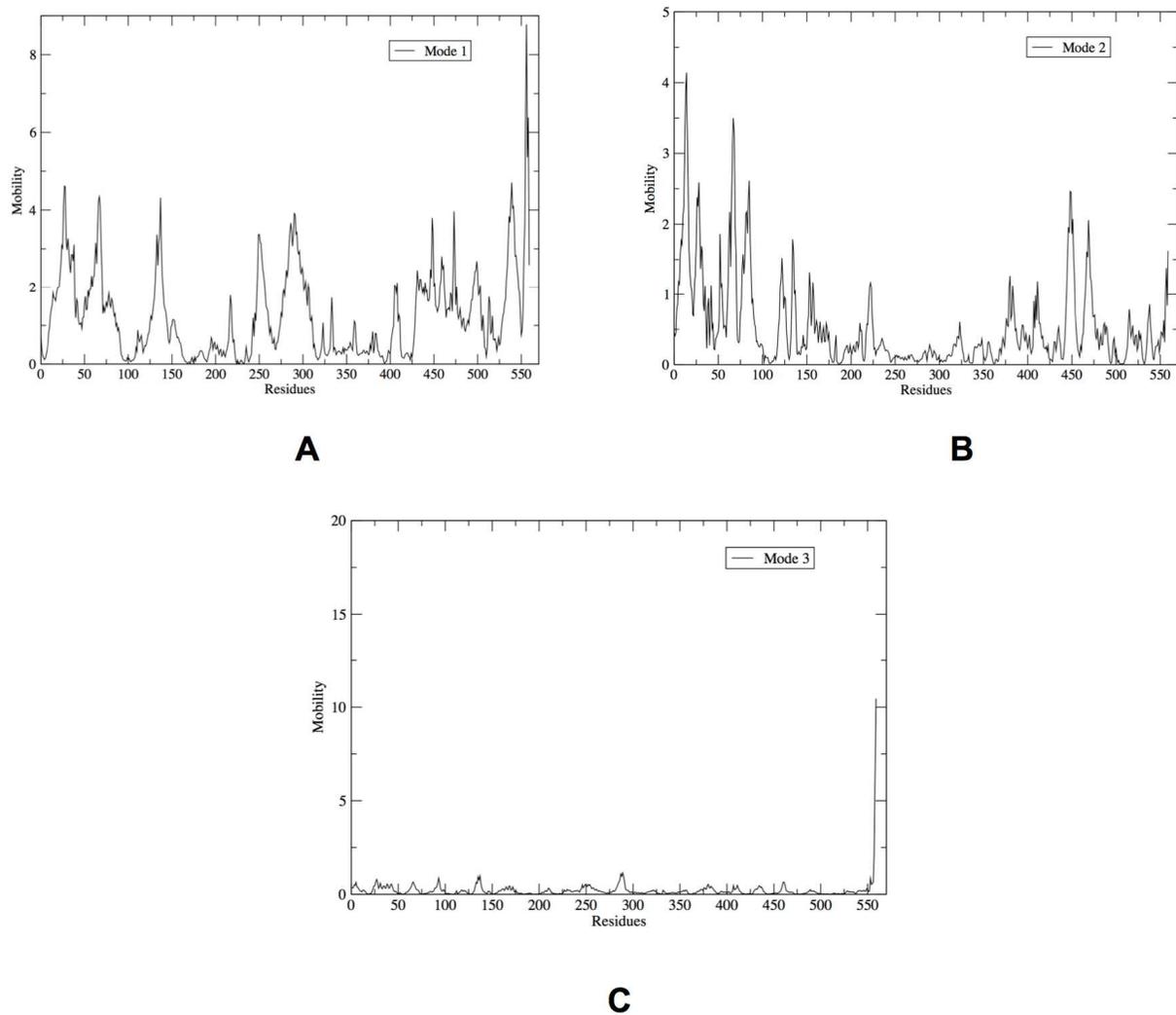
Figure 15. Residue based mobility plot of mutant (M184V) HIV-RT bound with 3TC showing

mobility of different residues during MD

Table 3 Prominent residues contributing towards motions of wild and M184V variant of HIV-

RT complexed with 3TC across different modes

Complexes	Normal Modes	Residue Numbers with Highest Mobility
WILD	1	560,449,470,14
	2	26,135, 359, 513, 560
	3	28, 134, 284, 448, 540, 560
M184V variant	1	28, 67, 137, 250, 291, 448, 473, 539, 555, 556
	2	14, 28, 67, 85, 122, 134, 222, 380, 411, 448, 449, 469, 560
	3	292, 560

3.3. Residue interaction network (RIN)

The residue interaction network analysis of a protein backbone is a new strategy to identify key residue interactions and can be used to explore the differences between different proteins including wild type and mutants. In this work we investigated the relationship between key residues of the wild type and M184V mutant by generating residue interaction networks using the representative structures from the 50ns of MD. As evident from the RIN plot (**Figure 16**), the M184V mutation has distorted the overall residue interaction network when compared to wild type. For instance, as shown in Figure **16**, there is a hydrogen bond interaction between Tyr115 and Met184 whereas in the case of the mutant the hydrogen bond interaction with Tyr115 changes to just a close atom interaction. It is quite interesting to observe that in the case of the wild type, Pro157 builds a close atom interaction with Met184 whereas in the case of the mutant

there is a close atom interaction instead between Gln161 and Val184. Again, the M184V

mutation affects the interaction network and consequently the drug binding landscape.
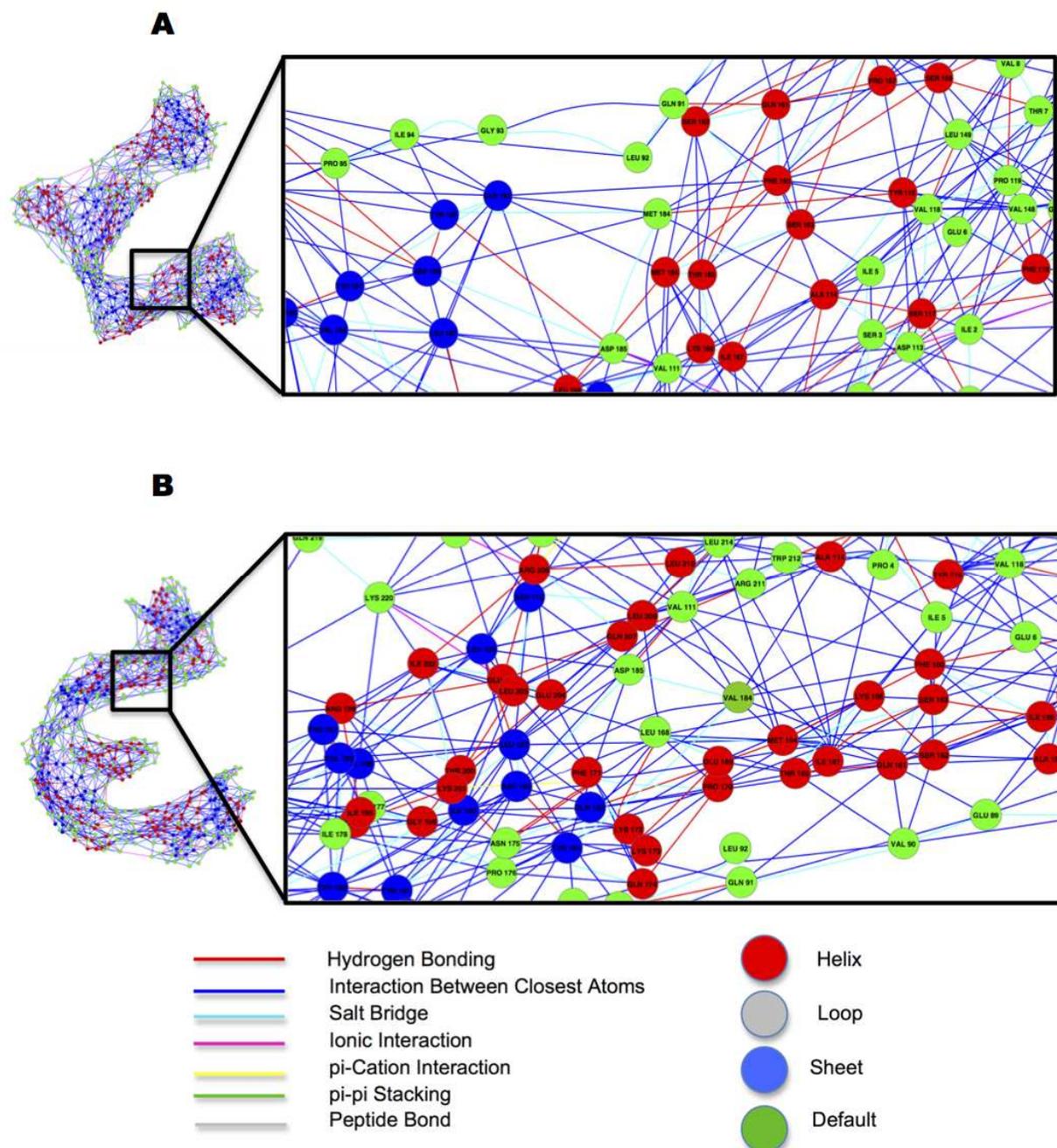
Figure 16. Residue Interaction Network (RIN) comparison between wild and M184V variant of HIV-RT complexed with 3TC, highlighting changes in network interaction at point 184

4. Conclusions

The precise molecular understanding of the serious impact of a single M184V on the RT resistance to 3TC is lacking in literature. In this report, we embarked on a wide range of computational approaches in order to provide a multidimensional insight into M184V resistance to 3TC. Molecular dynamics simulations, binding free energy calculations, principle component analysis (PCA and residue interaction network (RIN) led us to several findings that clearly explained the serious impact of M184V mutation on drug resistance. These findings verified that mutation decreased drug binding affinity by ~**7** kcal/mol, distorted the ligand optimum orientation in the RT active site, affected the overall enzyme conformational landscape and distorted the atomic interaction network with ligand. These findings stand consistent with the reported drastic effect of M184V mutation in 3TC binding with HIV-1 RT and provide a molecular level understanding of the impact of mutation. The findings of this report can provide potential key insights for further design of novel NRTIs to combat resistant strains.

5. Acknowledgements

6. Conflicts of Interest

Authors declare no conflicts of interest.

7. Supplementary Informations

RMSD (Figure S1), potential energy (kcal/mol) (Figure S2) corresponding PDB files in text format were provided as a supplementary material to this manuscript. The trend in binding free energy over time (Figure S3, Table 1) as well as the pictorial representation of methionine and valine at position 184 in comparison with 3TC was also provided (Figure S4 and Figure S5) as a supplementary manuscript.

References

1.   http://www.unaids.org/en/resources/campaigns/globalreport2013/factsheet/.
2.   http://aids.gov/federal-resources/around-the-world/global-aids-overview/
3.   http://www.avert.org/hiv-and-aids-africa.htm.
4.   Morah, E. U. (2007) Are People Aware of Their HIV-positive Status Responsible for Driving the Epidemic in Sub-Saharan Africa? The Case of Malawi, *Development Policy Review 25*, 215-242.
5.   Broder, S. (2010) The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic, *Antiviral Research 85*, 1-18.
6.   http://www.who.int/hiv/topics/treatment/en/
7.   Young, B. (2005) Review: mixing new cocktails: drug interactions in antiretroviral regimens, *AIDS patient care and STDs 19*, 286-297.
8.   Cihlar, T., and Ray, A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine, *Antiviral Research 85*, 39-58.
9.   de Bethune, M.-P. (2010) Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989-2009), *Antiviral Research 85*, 75-90.
10.   Meteer, J. D., Schinazi, R. F., Mellors, J. W., and Sluis-Cremer, N. (2014) Molecular mechanism of HIV-1 resistance to 3 '-azido-2 ', 3 '-dideoxyguanosine, *Antiviral Research 101*, 62-67.
11.   Diallo, K., Gotte, M., and Wainberg, M. A. (2003) Molecular impact of the M184V mutation in human immunodeficiency virus type 1 reverse transcriptase, *Antimicrobial Agents and Chemotherapy 47*, 3377-3383.
12.   Gallant, J. E. (2006) The M184V mutation: what it does, how to prevent it, and what to do with it when it's there, *The AIDS reader 16*, 556-559.
13.   Alvir, J. M. J., Lieberman, J. A., Safferman, A. Z., Schwimmer, J. L., and Schaaf, J. A. (1993) CLOZAPINE-INDUCED AGRANULOCYTOSIS - INCIDENCE AND RISK-FACTORS IN THE UNITED-STATES, *New England Journal of Medicine 329*, 162-167.
14.   Choo, H., Chong, Y. H., and Chu, C. K. (2003) The role of 2 ',3 '-unsaturation on the antiviral activity of anti-HIV nucleosides against 3TC-resistant mutant (M184V), *Bioorganic & Medicinal Chemistry Letters 13*, 1993-1996.
15.   Sarafianos, S. G., Das, K., Clark, A. D., Ding, J. P., Boyer, P. L., Hughes, S. H., and Arnold, E. (1999) Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids, *Proceedings of the National Academy of Sciences of the United States of America 96*, 10027-10032.
16.   Boyer, P. L., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2002) The M184V mutation reduces the selective excision of zidovudine 5 '-monophosphate (AZTMP) by the reverse transcriptase of human immunodeficiency virus type 1, *Journal of Virology 76*, 3248-3256.
17.   Ahmed, S. M., Kruger, H. G., Govender, T., Maguire, G. E. M., Sayed, Y., Ibrahim, M. A. A., Naicker, P., and Soliman, M. E. S. (2013) Comparison of the Molecular Dynamics and Calculated Binding Free Energies for Nine FDA-Approved HIV-1 PR Drugs Against Subtype B and C-SA HIV PR, *Chemical Biology & Drug Design 81*, 208-218.

18. Elmore, D. E., and Dougherty, D. A. (2001) Molecular dynamics simulations of wild-type and mutant forms of the Mycobacterium tuberculosis MscL channel, *Biophysical Journal 81*, 1345-1359.

19. Barreca, M. L., Lee, K. W., Chimirri, A., and Briggs, J. M. (2003) Molecular dynamics studies of the wild-type and double mutant HIV-1 integrase complexed with the 5CITEP inhibitor: Mechanism for inhibition and drug resistance, *Biophysical Journal 84*, 1450-1463.

20. El-Bastawissy, E., Knaggs, M. H., and Gilbert, I. H. (2001) Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature, *Journal of Molecular Graphics & Modelling 20*, 145-154.

21. Wan, H., Hu, J.-p., Tian, X.-h., and Chang, S. (2013) Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d, *Physical Chemistry Chemical Physics 15*, 1241-1251.

22. Chen, X., Zhu, S., Wang, S., Yang, D., and Zhang, J. (2014) Molecular dynamics study on the stability of wild-type and the R220K mutant of human prion protein, *Molecular Simulation 40*, 504-513.

23. Gu, R.-X., Liu, L. A., Wang, Y.-H., Xu, Q., and Wei, D.-Q. (2013) Structural Comparison of the Wild-Type and Drug-Resistant Mutants of the Influenza A M2 Proton Channel by Molecular Dynamics Simulations, *Journal of Physical Chemistry B 117*, 6042-6051.

24. David, C. C., and Jacobs, D. J. (2014) Principal Component Analysis: A Method for Determining the Essential Dynamics of Proteins, In *Protein Dynamics: Methods and Protocols* (Livesay, D. R., Ed.), pp 193-226.

25. Amadei, A., Linssen, A. B. M., deGroot, B. L., vanAalten, D. M. F., and Berendsen, H. J. C. (1996) An efficient method for sampling the essential subspace of proteins, *Journal of Biomolecular Structure & Dynamics 13*, 615-625.

26. Daidone, I., and Amadei, A. (2012) Essential dynamics: foundation and applications, *Wiley Interdisciplinary Reviews-Computational Molecular Science 2*, 762-770.

27. Laine, E., de Beauchene, I. C., Perahia, D., Auclair, C., and Tchertanov, L. (2011) Mutation D816V Alters the Internal Structure and Dynamics of c-KIT Receptor Cytoplasmic Region: Implications for Dimerization and Activation Mechanisms, *Plos Computational Biology 7*.

28. Martin, A. J. M., Vidotto, M., Boscariol, F., Di Domenico, T., Walsh, I., and Tosatto, S. C. E. (2011) RING: networking interacting residues, evolutionary information and energetics in protein structures, *Bioinformatics 27*, 2003-2005.

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

30. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *Journal of Computational Chemistry 25*, 1605-1612.

31. Goetz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, *Journal of Chemical Theory and Computation 8*, 1542-1555.

32.   Palmer, S. E., McLean, R. M., Ellis, P. M., and Harrison-Woolrych, M. (2008) Life-threatening clozapine-induced gastrointestinal hypomotility: An analysis of 102 cases, *Journal of Clinical Psychiatry 69*, 759-768.

33.   Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) APPLICATION OF THE MULTIMOLECULE AND MULTICONFORMATIONAL RESP METHODOLOGY TO BIOPOLYMERS - CHARGE DERIVATION FOR DNA, RNA, AND PROTEINS, *Journal of Computational Chemistry 16*, 1357-1377.

34.   Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, *Journal of Computational Chemistry 25*, 1157-1174.

35.   Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins-Structure Function and Bioinformatics 78*, 1950-1958.

36.   Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER, *Journal of Chemical Physics 79*, 926-935.

37.   Harvey, M. J., and De Fabritiis, G. (2009) An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware, *Journal of Chemical Theory and Computation 5*, 2371-2377.

38.   Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.

39.   Le Grand, S., Goetz, A. W., and Walker, R. C. (2013) SPFP: Speed without compromise-A mixed precision model for GPU accelerated molecular dynamics simulations, *Computer Physics Communications 184*, 374-380.

40.   http://www.originlab.com/

41.   Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S. H., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Accounts of Chemical Research 33*, 889-897.

42.   Massova, I., and Kollman, P. A. (2000) Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding, *Perspectives in Drug Discovery and Design 18*, 113-135.

43.   Tsui, V., and Case, D. A. (2001) Theory and applications of the generalized Born solvation model in macromolecular Simulations, *Biopolymers 56*, 275-291.

44.   Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *Journal of Molecular Graphics & Modelling 14*, 33-38.

45.   Bakan, A., Meireles, L. M., and Bahar, I. (2011) ProDy: Protein Dynamics Inferred from Theory and Experiments, *Bioinformatics 27*, 1575-1577.

46.   Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms, *Journal of Molecular Biology 285*, 1711-1733.

47.   Doncheva, N. T., Klein, K., Domingues, F. S., and Albrecht, M. (2011) Analyzing and visualizing residue networks of protein structures, *Trends in Biochemical Sciences 36*, 179-182.

48.   Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks, *Genome Research 13*, 2498-2504.

49.   George Priya Doss, C., Rajith, B., Chakraboty, C., Balaji, V., Magesh, R., Gowthami, B., Menon, S., Swati, M., Trivedi, M., Paul, J., Vasan, R., and Das, M. (2014) In silico profiling and structural insights of missense mutations in RET protein kinase domain by molecular dynamics and docking approach, *Molecular bioSystems 10*, 421-436.