6. Results

6.1 HIV-1 pol gene sequence analysis

6.1.1 HIV-1 viral load and CD4 counts in the study population.

Study group 1 age and demographics

Among the 101 patients, there were 65 (64%) males and 36 (36%) females who had an age of range of 12-61 years. The study population consisted of individuals from South (n=89, 88%), and East (n=12, 12%) India

Study group 2 age and demographics

Among the 19 patients, there were 16 (84%) males and 3 (16%) females who had an age of range of 11-55 years. The study population consisted of individuals from South (n=17, 90%), East (n=1, 5%) and North (n=1, 5%) India.

Study group 1 viral load

The mean HIV-1 viral load in the treatment naïve group was log_{10} 5.78 copies/ml and with a median value of log_{10} 5.23 copies/ml. The viral load in this group ranged from log_{10} 3.04 – 7.0 copies/ml.

Study group 2 viral load

The mean HIV-1 viral load in this group was log_{10} 5.52 copies/ml with a median value of log_{10} 4.67 copies/ml. The viral load in this group ranged from log_{10} 3.51 – 6.37 copies/ml.
Study group 1 CD4 counts

The mean CD4 counts in this group was 301 cells/mm³ with a median value of 262 cells/mm³. The CD4 counts in this group ranged from 10 – 1037 cells/mm³.

Study group 2 CD4 counts

The mean CD4 counts in this group was 300 cells/mm³ with a median value of 282 cells/mm³. The CD4 counts in this group ranged from 26 – 1155 cells/mm³.

A scatter plot of the HIV-1 viral load and CD4 counts in group 1 and 2 are shown in Fig 8a and 8b

6.1.2 Phylogenetic analysis

HIV-1 pol gene sequence quality control

The phylogenetic trees (Fig. 9 A-F) revealed no cross-contamination as there were no strains showing 99% or more identity. Consensus sequence was created for the ten externally sequenced samples using BioEdit sequence alignment editor version 7.0.4.1 and aligned using ClustalW (http://www.ebi.ac.uk/Tools/clustalw/). The sequences were analyzed by Stanford HIV drug resistance database for genotypic drug resistance analysis (http://hivdb.stanford.edu). The mutation/polymorphism data for the ten strains from the external centre were 100% concordant with sequencing data from our centre.
Fig 8a: Viral Load Vs CD4 in ART naive individuals (Group1)

Fig 8b: Viral Load Vs CD4 in ART experienced individuals (Group 2)
6.1.3 HIV-1 pol gene mutations

6.1.3.1 HIV-1 pol gene mutations in ART naïve individuals

In the phylogenetic tree the study sequences were seen to segregate with consensus C at the major node. One strain was seen to closely match with the consensus A1 sequence. We obtained similar results using the REGA HIV-1 Subtyping Tool - v2.0 which identified hundred strains as clade C and the same strain as clade A1 based on Pr and RT sequence data.

Amplification and sequencing of the reverse transcriptase gene revealed hundred and one strains with amino acid substitutions at positions 35, 39, 245 and 272. Amino acid substitutions were also seen at positions 60 (n=95), 122 (n=96), 173 (n=97), 200 (n=87), 207 (n=97), 291 (n=95) and 293 (n=94). The most common amino acid substitutions seen in RT are shown in Table 8. Positions known to confer drug resistance to nucleoside reverse transcriptase inhibitors (NRTI) and non-nucleoside reverse transcriptase inhibitors (NNRTI) also had amino acid substitutions. We found substitutions at NRTI drug resistance position 41 and 77 (n=1 each) and for NNRTI, they were at positions 100 (n=1), 101 (n=6), 103 (n=2), 179 (n=7) and 181 (n=1). One strain had amino acid substitutions at positions 98 and 210 of the RT gene sequence. The substitutions at these positions which are indicated as cause of the development of drug resistance in the Stanford HIV drug resistance database are summarized in Table 1 and 2 in the Appendix II. This was obtained using the calibrated population resistance tool of the Stanford HIV drug resistance database. Table 1 is based on the surveillance drug resistance mutation list (SDRM) (2008) while Table 2 is based on the IAS mutation list (2007). The results displayed by Stanford HIV drug resistance database on submission of the sequences are shown in Appendix VII A and B.
### Table 8: Common amino acid substitutions seen in Reverse Transcriptase sequence of the ART naïve HIV-1 infected individuals (1-300 amino acid positions) (n=101)

<table>
<thead>
<tr>
<th>S No.</th>
<th>Position</th>
<th>Amino Acid Change</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>V35</td>
<td>T(n=93), K(n=4), M(n=3), I(n=1)</td>
<td>101 (100%)</td>
</tr>
<tr>
<td>2</td>
<td>T39</td>
<td>D (n=68), E (n=16), N (n=14), I (n=1), DN (n=1), DE (n=1)</td>
<td>101 (100%)</td>
</tr>
<tr>
<td>3</td>
<td>S48</td>
<td>T (n=89)</td>
<td>89 (88%)</td>
</tr>
<tr>
<td>4</td>
<td>V60</td>
<td>I (n=95)</td>
<td>95 (94%)</td>
</tr>
<tr>
<td>5</td>
<td>D121</td>
<td>Y (n=66), H (n=12), C (n=2), DHNY (n=1)</td>
<td>81 (80%)</td>
</tr>
<tr>
<td>6</td>
<td>K122</td>
<td>E (n=95), A (n=1)</td>
<td>96 (95%)</td>
</tr>
<tr>
<td>7</td>
<td>K173</td>
<td>A (n=68), T (n=20), E (n=3), V (n=2), L (n=2), IKRT (n=1), APST (n=1)</td>
<td>97 (96%)</td>
</tr>
<tr>
<td>8</td>
<td>D177</td>
<td>E (n=83), Q (n=2), G (n=2)</td>
<td>87 (86%)</td>
</tr>
<tr>
<td>9</td>
<td>T200</td>
<td>A (n=89), E (n=2), V (n=2)</td>
<td>87 (86%)</td>
</tr>
<tr>
<td>10</td>
<td>Q207</td>
<td>E (n=75), G (n=10), K (n=5), A (n=3), N (n=3), R (n=1)</td>
<td>97 (96%)</td>
</tr>
<tr>
<td>11</td>
<td>V245</td>
<td>Q (n=87), H (n=4), E (n=4), R (n=1), K (n=2), EK (n=1), HQ (n=1), V→*EKQ (n=1)</td>
<td>101 (100%)</td>
</tr>
<tr>
<td>12</td>
<td>A272</td>
<td>P (n=101)</td>
<td>101 (100%)</td>
</tr>
<tr>
<td>13</td>
<td>E291</td>
<td>D (n=94), T (n=1)</td>
<td>95 (94%)</td>
</tr>
<tr>
<td>14</td>
<td>V292</td>
<td>I (n=92), ILV (n=2)</td>
<td>94 (93%)</td>
</tr>
<tr>
<td>15</td>
<td>I293</td>
<td>V (n=93), D (n=1)</td>
<td>94 (93%)</td>
</tr>
</tbody>
</table>
All the one hundred and one strains sequenced for the protease gene showed the amino acid substitution H69K. Ninety nine strains had an additional amino acid substitution at position 36. Other common sites of amino acid substitutions seen in the strains were at positions 15 (n=91), 19 (n=96), 41 (n=92), 63 (n=94) and 93 (n=96). The most common amino acid substitutions in the protease gene are listed in Table 9. Amino acid substitutions were also seen at positions 10 (n=3), 24 (n=3), 35 (n=12), 74 (n=4) and 82 (n=10). The substitutions in protease which are implicated in the emergence of drug resistance are summarized in table 1 and 3 in the Appendix II. Table 1 is based on the surveillance drug resistance mutation list (SDRM) (2008) while Table 3 is based on the IAS mutation list (2007).

6.1.2 HIV-1 pol gene mutations in ART failure individuals

In the phylogenetic tree (Fig.9 B) the study sequences were individually distinct and seen to segregate with consensus C at the major node indicating robust sequencing information. Similar results were obtained using the REGA HIV-1 Subtyping Tool - v2.0.

Amplification and sequencing of the reverse transcriptase gene revealed nineteen strains with M184V substitution. The other amino acid substitutions observed were M41L (n=5), D67N (n=8), K70R (n=7), K103N (n=9), Y181C (n=4), Y188L (n=4), G190A (n=5), T215F (n=4) and T215Y (n=5). Among the observed substitutions K103N, Y181C, Y188L and G190A have implicated for NNRTI resistance while M41L, D67N, K70R, T215F and T215Y have been implicated for NRTI resistance. Among the 21 strains, 20 showed resistance to both NRTIs and NNRTIs while only 1 strain showed resistance only to the NNRTIs. The list of
amino acid substitutions seen in RT known to affect susceptibility to NRTIs and NNRTIs are shown in Table 10. The information on the frequency of mutations observed in the RT of the strains generated based on the drug resistance mutation list of the IAS-USA major (2007) is shown in Table 4 of the Appendix II.

The RT gene showed major mutations at 21 sites with 25 mutational patterns while the Pr gene showed 4 mutational patterns at 4 sites. Among the 21 strain, 5 strains showed resistance to PIs. These 5 strains were all from patients who had been exposed to PIs during their course of infection. The treatment histories of all the individuals in group 2 have been summarized in table. These 5 strains showed 3 class resistance i.e., they were resistant to NRTIs, NNRTIs and PIs. The most common substitution responsible for resistance to PIs was M46I (n=4) in these strains. The other substitutions observed were D30N and L90M (n=2, each). The substitutions in protease which are implicated in the emergence of drug resistance are summarized in Table 10. The information on the frequency of mutations observed in the Pr of the study strains generated based on the drug resistance mutation list of the IAS-USA major (2007) is shown in Table 5 of the Appendix II.

6.2 Comparison of HIV-1 drug resistance algorithms

There were discrepancies in the subtyping (clade identification) results obtained from the Stanford HIV drug resistance database (SHDB) and Geno2Pheno (G2P) databases. The subtyping feature was not available in DR_SEQAN (DS). While the SHDB assigned 121 strains as clade C and 1 as clade A, G2P subtyped them as clade C (n=86), CRF 08_BC (n=29), clade D (n=4) clade H (n=2) and CRF 03_AB (n=1). All the 21 strains in the ART
Table 9: Common amino acid substitutions seen in Protease sequence (1-99 amino acid positions) of the ART naïve HIV-1 infected individuals (n=101)

<table>
<thead>
<tr>
<th>S No.</th>
<th>Position</th>
<th>Amino Acid Change</th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T12</td>
<td>S (n=63), P (n=8), A (n=3), TAPS (n=4), TIKR (n=1)</td>
<td>79 (78%)</td>
</tr>
<tr>
<td>2</td>
<td>I15</td>
<td>I→V</td>
<td>91 (90%)</td>
</tr>
<tr>
<td>3</td>
<td>L19</td>
<td>I (n=64), T (n=27), V (n=2), LIV (n=3)</td>
<td>96 (95%)</td>
</tr>
<tr>
<td>4</td>
<td>R41</td>
<td>R→K</td>
<td>92 (91%)</td>
</tr>
<tr>
<td>5</td>
<td>M36</td>
<td>I (n=79), V (n=16), L(n=3) ILV (n=1)</td>
<td>99 (98%)</td>
</tr>
<tr>
<td>6</td>
<td>L63</td>
<td>P (n = 60), T (n =14), S (n =9), A (n = 3), H (n = 3), APST (n = 2), Q,N,V (n = 1 each)</td>
<td>94 (93%)</td>
</tr>
<tr>
<td>7</td>
<td>K69</td>
<td>H→K</td>
<td>101 (100%)</td>
</tr>
<tr>
<td>8</td>
<td>I93</td>
<td>I→L</td>
<td>96 (95%)</td>
</tr>
</tbody>
</table>

Table 10: List of the mutations observed in the Reverse Transcriptase (1-300 aa positions) and Protease (1-99 aa positions) sequences of ART experienced HIV-1 infected individuals (n=19) as ascertained by the Stanford HIV Drug resistance database for individual sequences (GenBank ID) along with the ART regimen

<table>
<thead>
<tr>
<th>Sr No.</th>
<th>GenBank ID</th>
<th>Pr Mutations</th>
<th>RT Mutations</th>
<th>ART regimen in the last 6 months</th>
<th>WHO Clinical Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EU037773</td>
<td></td>
<td>NRTI: M184V</td>
<td>d4T + 3TC+ NVP</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td>EU037776</td>
<td></td>
<td>NNRTI: Y188L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>EU037772</td>
<td>D30N,M46I, N88D</td>
<td>NRTI: M184V</td>
<td>ABC + ddi + NFV</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NNRTI: Y188L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Substitution(s)</td>
<td>Drug Combination</td>
<td>Previous Exposure</td>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>------------------------------------</td>
<td>-------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>EU781846</td>
<td>NRTI: M184V, NNRTI: K103N</td>
<td>AZT + dDI + NFV</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>EU781845</td>
<td>NRTI: M184V, NNRTI: K103N</td>
<td>d4T + 3TC</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>EU781844</td>
<td>NRTI: M184V, T215Y, NNRTI: K103N</td>
<td>d4T + 3TC + NVP</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>EU781843</td>
<td>NRTI: M41L, D67N, M184V, T215Y, NNRTI: K103N</td>
<td>AZT + 3TC + NVP</td>
<td>Stage 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Previous regimen 2003: d4T + 3TC + NVP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(oral Candidiasis, Chronic Diarrhea, Weight Loss)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>EU781842</td>
<td>NRTI: M41L, D67N, M184V, L210W, T215Y, NNRTI: K103N</td>
<td>NFV + AZT + 3TC + NVP</td>
<td>Stage 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Previously exposure to ATV - 10 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Oral Ulcers, Candidiasis, Aphthous Ulcers)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>EU781841</td>
<td>NRTI: M184V, NNRTI: K103N</td>
<td>d4T + 3TC</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>EU781840</td>
<td>NRTI: M184V, NNRTI: V106A, P225H</td>
<td>d4T + 3TC + NVP</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>EU781839</td>
<td>NRTI: D67N,</td>
<td>d4T + 3TC + EFV</td>
<td>Stage 1</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Protease Residues</td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>Treatment</td>
<td>Stage</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
<td>-------</td>
<td>--------</td>
<td>-----------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>EU781838</td>
<td>K70R, M184V, K219Q</td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>NVP+AZT+3TC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D67N, K70R, M184V, K219Q</td>
<td>G190A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU781837</td>
<td>M184V</td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>d4T + 3TC+ NVP</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M184V</td>
<td>Y181I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU781835</td>
<td>M46I, V82A</td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>IDV+ddl+d4T</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M41L, D67N, L210W,T215Y</td>
<td></td>
<td>Y188L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU781834</td>
<td>M41L, V108I, M184V,T215Y</td>
<td></td>
<td>Y181C</td>
<td>d4T + 3TC+ NVP</td>
<td>Stage 1</td>
</tr>
<tr>
<td>EU037779</td>
<td></td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>d4T + 3TC+ NVP</td>
<td>Stage 1</td>
</tr>
<tr>
<td></td>
<td>K103N, V108I</td>
<td></td>
<td>Y188L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU037774</td>
<td></td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>d4T + 3TC+ NVP</td>
<td>Data Missing</td>
</tr>
<tr>
<td></td>
<td>K70R, M184V, T215F</td>
<td></td>
<td>K103N, V108I</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU037778</td>
<td></td>
<td>NRTI:</td>
<td>NNRTI:</td>
<td>AZT + 3TC+ EFV</td>
<td>Stage 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2006 : EFV+AZT</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Accession</td>
<td>NRTI</td>
<td>NNRTI</td>
<td>Treatment</td>
<td>Stage</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>----------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>18</td>
<td>EU037777</td>
<td>M46I, L90M</td>
<td></td>
<td>AZT+3TC+NFV (Previous regimens:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2001: AZT+3TC</td>
<td>Stage 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2003: NFV+AZT +3TC+NVP</td>
<td>(Disseminated TB, Distal oesophagus ulcers)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2004: IDV+AZT +3TC</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>EU781836</td>
<td>L74LV, M184V</td>
<td>A98AG, Y181C</td>
<td>d4T+3TC+NVP</td>
<td>Data Missing</td>
</tr>
</tbody>
</table>

**NRTI**: Nucleoside/Nucleotide Reverse Transcriptase Inhibitors

**NNRTI**: Non-Nucleoside Reverse Transcriptase Inhibitors
experienced group strains were classified as subtype C by SHDB. G2P classified 19 as clade C and 2 as CRF 08_BC. The subtyping results from SHDB were completely concordant with the subtype identification using the REGA HIV-1 subtyping tool - v2.0.

The mean observed concordance between SHDB and G2P was 86% while between SHDB and DS was 43%. The highest concordance between SHDB and G2P was seen for NNRTIs for which it was 89.6% and the least was 81.8% which was observed for NRTIs. The PIs had the lowest level of concordance between SHDB and DS (20.9%) whereas the highest was seen for NRTIs (58.4%). When the level of concordance were seen based on exposure to ART, the G2P was found to have a better level of concordance (76.8%) to SHDB as compared to SHDB Vs DS (33.7%). However, when only the genotypic interpretation and not interpretation based on fold increase (phenotypic IC50 values) was taken for DS the level of concordance for the ART experienced group rose to 70%. Similarly when only the genotypic interpretation was taken into account for the treatment naive group the level of concordance between SHDB Vs DS was 95.6%. Among the different classes of drugs, when comparing SHDB and G2P, the best concordance was seen for 3TC (98.4%) among NRTIs, NVP (91%) among NNRTI and LPV (95.9%) among PIs. The lowest concordance was seen for d4T (57.4%) among NRTIs, DLV (87.7%) among NNRTI and SQV (76.2%) among PIs. While comparing the concordance between SHDB and DS among the different classes of drugs, the best concordance was seen for 3TC (94.1%) among NRTIs, NVP (67.2%) among NNRTI and ATV (82.4%) among PIs. The lowest concordance was seen for AZT and ddI (4.2%) among NRTIs, DLV (47.7%) among NNRTI and LPV (0.8%) among PIs.
A predominant number of sequences that were submitted to G2P and DS showed a variation at positions 3 (V3I) and 37 (S37N) in the protease and position 214 (L214F) in the RT. Strains submitted to SHDB primarily showed variations at positions 122 (K122E), 272 (A272P) and 277 (K277R).

The DS algorithm was unable to generate a susceptibility profile for three strains from the treatment failure group. The error message generated for these strains was ‘6.- Overflow’. The mean percentage concordance between the different algorithms for each group of drugs and for the ART exposed, unexposed group are summarized in Table 11. The results displayed by G2P on submission of the sequences are shown in Appendix VII C and D.

6.3 Construction of HIV-1 Reverse Transcriptase 3D models

Eight mutations found in the Clade-C strains that confer resistance to both the classes of reverse transcriptase inhibitors viz., NRTIs and NNRTIs were investigated for their plausible role in affecting the structural integrity of the RT. Mutations that confer resistance to these two groups of drugs are distinct. Hydrogen bond interactions and van der Waals contacts involving the site of mutations were investigated using the PyMOL (DeLano Scientific LLC, San Carlos, California). In the PyMOL software to determine hydrogen bonding interaction a distance cutoff of 4 Å was used along with a donor-acceptor bond angle: > 90°. Distance cutoff: Sum of van der Waals radii ± 0.5 Å. In the analysis that was undertaken this was the criterion for recognizing the role of van der Waals forces in the 3D models.
Table 11: Comparison of the mean percentage concordance (MPC) between the different algorithms for each group of drugs among the ART experienced and naive groups

<table>
<thead>
<tr>
<th>MPC for drug categories</th>
<th>1 Vs 2 (n=122)</th>
<th>1 Vs 3 (n=119)</th>
<th>ART experienced</th>
<th>ART naïve (n=101)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 Vs 2 (n=21)</td>
<td>1 Vs 3 (n=18)</td>
</tr>
<tr>
<td>NRTI</td>
<td>81.83%</td>
<td>58.4%</td>
<td>73%</td>
<td>19.45%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNRTI</td>
<td>89.61%</td>
<td>56.79%</td>
<td>71.43%</td>
<td>72.22%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>88.25%</td>
<td>20.86%</td>
<td>83.33%</td>
<td>28.69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall MPC</td>
<td>85.96%</td>
<td>43.07%</td>
<td>76.82%</td>
<td>33.70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MPC: Mean Percentage Concordance. The data was analysed using Epi Info v 6.03 to calculate the percentage concordance (observed agreement) for each drug across the algorithms. The mean of the percentage concordance for the susceptibility pattern of 122,119, 21 and 18 strain specific observations are shown in the table above.

1 Vs 2: SHDB Vs G2P
1 Vs 3: SHDB Vs DS
1 Vs 3*: SHDB Vs DS (using interpretation based on DR_SEQAN algorithm only and not based on fold increase of IC₅₀)

# DS was not able to generate an output for three strains
The changes in the local interaction patterns were then interpreted with the help of 3D models for these mutations. The study models were built by mutating the template structures downloaded from the PDB. The mutations that were incorporated into the template were those observed in treatment naive and treatment failure patients from study groups 1 and 2. The different atoms in the figures have been depicted using the CPK model, where nitrogen is blue, oxygen is red, carbon is grey/black, sulphur is yellow and hydrogen is white. The amino acid structures, the 3 and 1 letter codes amd the polarity are shown in Appendix V.

\textbf{i) K101H}

The location of Lys at position 101 in the RT crystal structure of the wild type native molecule without its ligand (drug) \{PDB ID: 1RTJ\} is shown in Fig. 10A. In the ligand unbound RT structure, the side chain of Tyr at position 181 points into the hydrophobic pocket where the NNRTIs bind (Fig. 10B-E). The hydrophobic pocket of the RT to which the NNRTIs bind is located between the $\beta$6-$\beta$10-$\beta$9 and $\beta$12-$\beta$13-$\beta$14 of the p66 subunit. The pocket is about 10 Å from the aspartic acid triad at positions 110, 185 and 186 as reported earlier (232). This native conformation is stabilized by the hydrogen bonds formed by the side chain hydroxyl group of Tyr at position 181 with the side chain N-H group of Asn at position 136 of the p55 subunit of the RT (PDB Ids: 1DLO, 1RTD) and with side chain carboxyl of Glu at position 138 of the p55 subunit [1HMV] (Fig. 10B-D). The side chain of Lys at position 101 interacts with Pro at position 321 through van der Waals forces. In some structures (1DLO, 1RTD), it also interacts with the side chain carboxyl of Glu 28 through a weak salt bridge. Salt bridge denotes a relatively weak ionic bond between positively and negatively charged side chains of amino acids in a protein, which contribute to the stability of protein structures. (Fig. 10B, D). In the template/primer bound RT structure (1RTD), Lys 101 side chain makes van der Waals contacts with the side chain of Tyr 319 (Fig.10D). In
the structure of RT (1RTJ), which largely corresponds to open conformation, Lys 101 forms a salt bridge with Glu 28. It interacts with Glu 138 through water mediated hydrogen bonds (Fig. 10E).

In the case of efavirenz and nevirapine ligand bound RT structures (1IKW, 1VRT), Lys 101 forms salt bridges with Glu 138. The side chain carboxyl of Glu 138 also interacts with Gln136 through a NH---O hydrogen bond (Fig. 10F, H). Efavirenz directly interacts with the main chain carbonyl and amino groups of Lys 101 through hydrogen bonds. Nevirapine forms water mediated hydrogen bonds with the main chain amino group and carboxyl groups of Lys 101 (Fig. 10F, H). The conformational changes that occur in the RT molecule at the NNRTI binding pocket upon ligand binding are shown in Fig. 10G. The salt bridge that Lys 101 forms with Glu 138 is lost (Fig. 10I) as a result of the mutation Lys 101 → His (K101H) due to the shorter side chain of Histidine. This might alter the geometry of the binding pocket. The loss of this salt bridge might also affect the hydrogen bond interaction between Glu 138 and Gln 136 and their side chains are likely to become more flexible. This increases the possibilities of interaction between Glu 138/Gln 136 and Tyr 181, thereby affecting the inhibitor entry into the pocket. The direct or water mediated interactions with efavirenz and nevirapine also get affected because of changes in the side chain conformation.

**ii) M230L**

The Met to Lys mutation at position 230 (M230L) in the efavirenz bound RT structure is shown in Fig. 11A. In the ligand unbound RT structures in the closed conformation (1DLO, 1HMV), Met 230 interacts with the side chain of Pro 95 through van der Waals forces (Fig. 11B). The side chain carboxyl of Asp 186 and the side chain amino group of Trp 266 forms hydrogen bonds with the main chain amino and carbonyl group of Met 230, respectively.
The sulphur atom of Met 230 interacts with the π clouds of Tyr 183 and Trp 229 (<5Å distance). π electrons are delocalized electrons over the aromatic rings which constitute the π cloud (a simple aromatic ring such as benzene has six π electrons). In ligand unbound RT structure close to open conformation (1RTJ), Met 230 interacts with Trp 266 side chain through sulphur-π cloud interaction (Fig. 11C).

In RT bound to efavirenz and nevirapine (1IKW, 1VRT), Met 230 sulphur atom is located close to the side chain of Tyr 183 (Fig. 11D, E). The side chain aromatic rings of Tyr 183 and Trp 229 interact through π-π stacking. The NH group of Trp 229 can also interact with the π cloud of Tyr 183 (Fig.11E).

When the NNRTI enters the hydrophobic pocket, the side chains of Tyr 181 and Tyr 188 move upwards, giving space for the incoming inhibitor. This movement is accompanied by changes in side chain conformations of Trp 229, Tyr 183 and Phe 227. The side chain conformations of Tyr 183 and Trp 229 are stabilized by their interactions among themselves as well as with Met 230. The conformational changes of the residues that occur in the NNRTI binding pocket are shown in Fig.11D. The mutation M230L results in the loss of the above mentioned interactions, as Leu is smaller in size compared to Met and can thus affect the entry of the inhibitor (Fig.11F). The smaller size of Leu increases the distance of both Tyr 183 and Trp 229 to Leu 230. Hence, the interactions that stabilizes the RT molecule to allow entry of the NNRTI into the hydrophobic pocket does not occur. The interaction with Trp 229 which is in a loop joining the palm and thumb sub-domains of RT might have a role in stabilizing the structure at different conformation i.e., open (substrate unbound) and closed (substrate bound). Thus, this mutation could also alter the flexibility of RT.
Fig 10. (A) Location of Lys 101 in the unliganded RT structure, 1RTJ. p66 subunit is coloured in yellow while the p55 subunit is in pink (Same subunit colouring code is followed for the other figures). The part of the NNRTI entry site in the p66 subunit and the side chain of Lys 101 are coloured green. The part of drug entry site in the p55 subunit is coloured in orange.

(B-F) Local hydrogen bond interactions involving K101 in unliganded RT structure 1DLO, 1HMD, 1RTD, 1RTJ and in Nevirapine bound RT 1VRT

(G) Structural superposition of 1IKW (yellow) and 1RTJ (green) showing changes in residue conformations.

(H) Interaction pattern involving K101 in Efavirenz (shown in orange) bound RT (1IKW)

(I) Changes due to K101H mutation. Site of mutation (here and in the rest of the figures) is shown in pink. All the figures are rendered in PyMol.

Fig.11 (A) Location of M230 (green) in Efavirenz (in orange) bound RT (1IKW) Interactions involving M230 in unliganded RT structures

(B) 1DLO and (C) 1RTJ.

(D) RT structure almost in open conformation (1RTJ, green) superposed on Efavirenz (in orange) bound RT (1IKW, yellow)

(E) Interactions involving M230 in Nevirapine bound RT (1VRT)

(F) Model for M230L mutation.

Fig.12 (A) Interactions involving Tyr 318 in Efavirenz (orange) bound RT (1IKW) and

(B) Changes due to Y318F mutation.

(C) Local interaction network involving Lys238 in Efavirenz bound RT structure (1IKW)

(D) Changes due to K238T mutation

(E) Comparison of side chain conformations of residues in the region involving Lys238, of Efavirenz bound RT, 1IKW (yellow) and of the unliganded form, 1RTJ (green)

(F) Interactions involving K238 in Nevirapine bound RT (1VRT).
**Fig 10. (A)** Location of Lys 101 in the unliganded RT structure, 1RTJ. p66 subunit is coloured in yellow while the p55 subunit is in pink (Same subunit colouring code is followed for the other figures). The part of the NNRTI entry site in the p66 subunit and the side chain of Lys 101 are coloured green. The part of drug entry site in the p55 subunit is coloured in orange.

**(B-F)** Local hydrogen bond interactions involving K101 in unliganded RT structure 1DLO, 1HMV, 1RTD, 1RTJ and in Nevirapine bound RT 1VRT

**(G)** Structural superposition of 1IKW (yellow) and 1RTJ (green) showing changes in residue conformations.

**(H)** Interaction pattern involving K101 in Efavirenz (shown in orange) bound RT (1IKW)

**(I)** Changes due to K101H mutation. Site of mutation (here and in the rest of the figures) is shown in pink. All the figures are rendered in PyMol.

**Fig.11 (A)** Location of M230 (green) in Efavirenz (in orange) bound RT (1IKW) Interactions involving M230 in unliganded RT structures

**(B)** 1DLO and (C) 1RTJ.

**(D)** RT structure almost in open conformation (1RTJ, green) superposed on Efavirenz (in orange) bound RT (1IKW, yellow)

**(E)** Interactions involving M230 in Nevirapine bound RT (1VRT)

**(F)** Model for M230L mutation.

**Fig.12 (A)** Interactions involving Tyr 318 in Efavirenz (orange) bound RT (1IKW) and

**(B)** Changes due to Y318F mutation.

**(C)** Local interaction network involving Lys238 in Efavirenz bound RT structure (1IKW)

**(D)** Changes due to K238T mutation

**(E)** Comparison of side chain conformations of residues in the region involving Lys238, of Efavirenz bound RT, 1IKW (yellow) and of the unliganded form, 1RTJ (green)

**(F)** Interactions involving K238 in Nevirapine bound RT (1VRT).
Fig. 13 (A) Location of V75 with respect to dNTP (orange) binding site. Some residues in the vicinity of V75 involved in dNTP binding are also shown.

(B) Interaction between Val 75, Phe77 and neighbors

(C) Model for V75M mutation

(D) Residues in the vicinity of V75 involved in template RNA binding.
FIG. 14(A) Superposition of the structures with PDB IDs 2AQU (red), 1HHP (white) and 1TW7 (green) which have closed, semi-open and open conformations respectively.

(B) The sites of mutations seen in Clade-C strains depicted on Atazanavir bound protease structure [2AQU]. The main chain atoms of the sites of mutations are shown in dotted representation using Pymol (DeLano Scientific LLC, San Carlos, California). The flap elbows and the lower half of the dimer interface are shown in green, fulcrum in orange and cantilever in white. The drug binding site is coloured blue and the helix (residues 86-94) is coloured red. Sites of the mutations analyzed in this study are labeled

FIG. 15(A) The hydrogen bond network in the region involving D30 and N88, seen in Atazanavir bound protease [2AQU]. The hydrogen bond interactions (here and in the rest of the figures) are shown as black dotted lines. The distance between the interacting atoms are also shown. Sites of mutations (here and in the rest of figures) are shown in yellow.

(B) The changes in the interaction pattern (only the major changes are shown) seen in the model for the strain EU030406 with the mutations D30N, K45R, N88D and L89V. Mutated residues (here and in the rest of figures) are shown in green. Note that the model has other mutations like R41K, Q61D and T74S (in this part of the structure). The backbone (here and the rest of the figures) is coloured as shown in Fig 13B.

(C) van der Waals contacts (red lines) between Met36 & Ile15 and neighbours, in Atazanavir bound protease [2AQU]

(D) Model of the strain EU030409 showing changes in the contact pattern as a result of mutations M36I and I15V

(E) The local hydrogen bond network involving Glu35 seen in Atazanavir bound protease structure [2AQU]

(F) Model of the strain EU030415 showing the changes due to E35D mutation Figures 12a, 12b, 12e and 12f were prepared using Pymol (DeLano Scientific LLC, San Carlos, California) while 12c and 12d were prepared using the Chimera software.

FIG. 16(A) Local interaction network in the region involving T12, L63 and H69 in the Atazanavir bound protease [2AQU]
(B) The changes in the interactions (only the major changes are shown) due to the mutations L63P, T12S, E21D and H69K. Note that the model has L19T mutation (in this part of the structure) which is not discussed in the text.

(C) Local hydrogen bond interactions involving Lys 20 in the Atazanavir bound HIV-1 protease [2AQU]

(D) Changes due to the K20T mutation. Note that the model also has M36I mutation in this part of the structure.

(E) Leu90 and the neighbouring residues with which it forms van der Waals contacts in the Atazanavir bound protease [2AQU]

(F) Changes in the contact pattern due to L90M mutation, observed in the protease structure 2F81

Figures 14A, 14B, 14C and 14D were prepared using PyMol (DeLano Scientific LLC, San Carlos, California) while 3F and 3F were made using the Chimera software
iii) \textit{G190A}

The Gly at position 190 in NNRTI bound RT structures makes van der Waals contacts with the drugs nevirapine and efavirenz. When Ala replaces Gly as a result of the mutation (G190A), the $\beta$-carbon in the Ala side chain, is at a distance of 2.46 Å and 2.58 Å from one of the fluorine atoms of efavirenz and the cyclo propyl group of nevirapine, respectively. Thus, the distance between the drugs and Ala are small and this would result in steric clash which can drastically affect the binding of the drug and confer resistance to the drug.

iv) \textit{Y318F}

For the initial samples that were sequenced during the period of standardization of the drug resistance assay, samples were sequenced to different RT positions to determine what length could be sequenced consistently. Hence, although the drug resistance assay was standardized to obtain mutation/polymorphism data for amino acid position till 300 in RT, we have RT amino acid sequences to positions 320 for some samples. One such sample which had a change from Tyr to Phe at position 318 was also studied using 3D models for its possible effects on drug resistance. Only one of the five samples in which the RT was sequenced successfully to an amino acid length of 320 amino acids showed this mutation. The mutation at position 318 in RT where Tyr is replaced by Phe (Y318F) in an efavirenz bound RT (1IKW) is shown in Fig.12. The O-H group in the side chain of Tyr 318 in the NNRTI bound complex (1IKW) interacts with the main chain carbonyl of His 235 through O-H…O hydrogen bond. This interaction helps to hold the side chain of Tyr 318 away from the NNRTI binding site (Fig. 12A). The mutation Y318F results in the loss of this interaction as Phe lacks side chains capable of hydrogen bonding. This will make Phe 318 side chain more flexible, thereby affecting inhibitor binding. The side chain hydroxyl of Tyr 318 is at a
distance of 3.3Å from efavirenz and favour van der Waals contacts between them. These interactions are also lost as a result of the Y318F mutation (Fig. 12B).

\(v)\) **K238T**

The interaction network involving Lys 238 is shown in Fig. 8C and it could play a role in maintaining the geometry of the NNRTI binding pocket. The mutation K238T would disrupt this interaction network and cause changes in the binding pocket (Fig. 12 D). The structure of RT structure in an open conformation (1RTJ, green) superposed on efavirenz bound RT (1IKW, yellow) is shown in Fig. 12 E. In the NNRTI bound RT structures, Lys 238 interacts with His 235 through van der Waals contacts. Lys 238 side chain makes van der Waals contacts also with His 315. The side chain amino group forms a salt bridge with the side chain carboxyl of Asp 237 which is weak in the case of nevirapine bound RT (1VRT, Fig. 12F). The change in His 235 side chain conformation, as a result of the mutation, could also affect Tyr 318 - His 235 interaction.

\(vi)\) **V75M**

The location of Val 75 with respect to dNTP binding site is shown in Fig. 13A. In the RNA template /primer bound RT (1RTD), Val 75 interacts with Phe 77, Val 60 and Lys 73 through van der Waals contacts. Phe 77 makes extensive van der Waals contacts with the neighbouring residues (Fig. 13B). These interactions help to hold the anti-parallel beta sheet (\(\beta3\) - \(\beta4\)) and the loop involving residues 64-70 over the bound dNTP. The van der Waals contacts between Phe 77 and Pro 150 help to position Gln 151 that interacts directly with the dNTP.
The structural model for V75M mutation shows (Fig. 13C) that the terminal methyl group of Met 75 side chain is close to Phe 77. The side chain of Phe 77 also moves to avoid steric hindrance and to facilitate formation of sulphur-π interaction between these two residues. This affects the structural environment of the neighbourhood. The hydrogen bond interaction between Lys 73 and Tyr 146 that help to hold the β3-β4 loop in position might be affected. The residues that are in the vicinity of Val 75 and Phe 77 such as Phe 61, Asp 76 and Arg 78 (Fig. 13D) interact with the template strand. These interactions could also be affected as a result of the mutation and the position of template bases might be altered.

vii) V118I

The position of Val 118 has tight binding in the template/primer bound RT structural region with a hydrophobic environment. In the template/primer and thymidine triphosphate (TTP) bound RT (1RTD), Val 118 interacts with Leu 149 and Phe 160 through van der Waals contacts (Fig. S1A in Appendix III). In the V118I mutant model (Fig. S1B in Appendix III), we noted changes in the side chain conformation of Tyr 115, Leu 149 and Ser 156 owing to the bulkier -CH2-CH3 group in the side-chain of Ile compared to Val which has only –CH3. Tyr 115 interacts directly with the dNTP. Ser 156 forms O-H---O hydrogen bond with the main chain carbonyl of Pro 150 and this interaction aids in positioning Gln 151 for dNTP binding. Thus the side chain conformational changes of these residues triggered by the presence of the bulkier Ile in the place of Val 118 would affect dNTP binding interfering with the transcriptional role of RT.
viii) \textbf{Q151M}

In the template/primer and thymidine triphosphate (TTP) bound RT structure (Fig. S1C in Appendix III), the side chain NH of Gln 151 interacts with the main chain carbonyl at Lys 43 through a N-H…O hydrogen bond. The side chain carbonyl of Gln 151 also forms hydrogen bonds with side chain N-H of Arg 72 and O (1B) P of dTTP. These interaction help directly or indirectly in TTP binding. The side chain of Gln 151 also forms van der Waals contacts with TTP and the side chain benzene rings of Tyr 115 and Phe 116. As a result of the Q151M mutation, these hydrogen bond interactions would be lost due to the apolar nature of Met side-chain (Fig. S1D in Appendix III). The dramatic loss of these interactions in the mutant could affect the closing of $\beta$3-$\beta$4 loop upon dNTP binding. Though the side chain sulphur of Met 151 can interact with the $\pi$ cloud of Phe 116 the terminal carbon of Met 151 side chain is at a distance of 2.27A from one of the phosphate oxygens of dTTP. This would lead to steric hindrance and can hamper the binding of dNTP.

\textbf{6.4 Construction of HIV-1 Protease 3D models}

The protease 3D models were constructed for the selected strains of Vellore origin and reference amino acid sequences from the HIV-1 Los Alamos database. Nucleotide mutations causing non-synonymous aa substitutions were examined. The criteria for strain selections are as mentioned in section 5.6 of materials and methods. Fig. 14a represents superposition of the structures with PDB IDs 2AQU (red), 1HHP (white) and 1TW7 (green) which have closed semi-open and open conformations respectively. The sites of the mutations analyzed in this study are labelled in Fig 14b.
i) **D30N, K45R, N88D and L89V**

The dynamic changes in the 3D structure of the protease molecule when bound to an inhibitor as shown in Fig. 15A. The carboxyl side chain of Asp 30 interacts with the amino group side chain of Lys 45 is established through a salt bridge interaction. The side chain amino group of Lys 45 also interacts with the main chain carbonyl at Met 46 by forming water mediated hydrogen bonds. However, this interaction is not seen in all drug bound complexes. Lys 45 interacts with Gln 58 forming N_H---O hydrogen bond. This interaction is often water mediated.

The side chain amino group of Asn 88 interacts with both the main chain carbonyl and the amino group of Thr 74. The hydrogen bond interaction with the amino group is water mediated and this interaction is conserved in all the drug bound structures analyzed. The side chain carboxyl interacts with the side chain hydroxyl of Thr 31 through an O_H---O hydrogen bond. The side chain amino group of Gln 92 also forms a N_H---O hydrogen bond with the main chain carbonyl at Asn 88.

The side chain of Leu 89 interacts with the side chains of Asn 88, Thr 31, Ile 85 and Ile 64 through van der Waals contacts. The interaction with Asn 88 and Thr 31 can stabilize their side chain conformations. The mutations D30N, K45R, N88D and L89V were observed in a strain sequenced from an individual showing treatment failure (EU030406). The mutation D30N results in the loss of salt bridge interaction with Lys 45 (Fig. 15B). The D30N mutation confers high resistance to nelfinavir because of the loss of the O_H---O hydrogen bond interaction between the drug and the side chain carboxyl of Asp 30. The mutation L89V affects van der Waals contacts with Asn 88 and Thr 31. This relaxes the hold on the side chains of Asn 88 and Thr 31. When Asn 88 is mutated to Asp, the hydrogen bond interactions
with the main chain groups of Thr 74 are lost (Fig. 15B). The loss of Asn 88 hydrogen bond interactions with Thr 74 and the salt bridge interaction between Lys 45 and Asp 30 affect the hydrogen bond network (Fig. 15A) seen in most of the drug bound complexes.

In the crystal structures of unliganded protease (e.g., PDB IDs: 1HHP, 3PHV) that were analyzed, Asn 88 does not interact with the main chain carbonyl and amino group of Thr 74 (Fig. S2A & B in Appendix III). The interaction between Lys 45 and Gln 58 is weak or absent. As shown in the Fig. 15B, we see similar changes in the protease model which contained the following mutations D30N, K45R, N88D and L89V.

ii) M36I & I15V

These two mutations observed in most of the study strains sequenced. In the drug bound protease structures of strains without these mutations, the side chain of Met 36 and Ile 15 (from the β carbon - CB) forms van der Waals contacts with the side chain of Leu 38, Lys 20, Leu 33, Glu 34, Ile 62 and Val 75 (Fig. 15C). When Met 36 is mutated to Ile and Ile 15 to Val, the side chains becomes a bit shorter, which reduces the extent and strength of van der Waals contacts (Fig. 15D). The van der Waals contact with V75 and Ile 62 are lost and the interactions with Leu33 and Leu38 are weakened. There is also a significant reduction in the extent of contact between Met 36 and Ile 15 (not shown in the figure). Both the mutations result in shorter side chains that favour the lowering of the elbow region, which is observed during the transition from the closed to the open state (Fig. 14A). In some of the structures of protease in the unliganded state (Fig. S3 in Appendix III), the side chain of Met 36 is ‘curled in’. This would reduce the steric hindrance when the elbow is lowered. The flap elbow of the protease is constituted by amino acids located at positions 35-42. When the protease is in an
open state for the substrate to enter the active site the flap elbow is lowered to aid in opening of the flap and exposing the active site.

The mutation K20T (discussed in detail below) which occurred along with these mutations in most of the strains, also result in a shorter side chain that can aid lowering of flap elbow.

iii) E35D (E35N and E35K)

In the inhibitor bound protease structures, the side chain carboxyl of Glu 35 forms a salt bridge with the side chain NH2 of Arg 57 (Fig.15E). The mutation E35D results in weakening of this interaction since Asp has a shorter side chain (Fig.15F). Thus the elbow becomes slightly flexible upon mutation, which favors its lowering. The E35D mutation has been shown to increase flap flexibility. E35N and E35K mutation, which were seen in our strains, will have more profound effect as the salt bridge is lost as a consequence of this mutation. Increase in the flexibility of flap elbow as a result of M36I and E35D mutations can also affect the conformations of residues in the drug binding pocket (e.g. Val 32, Fig. 15C). Val 32 interacts with most of the HIV-1 protease inhibitors and so these mutations can affect inhibitor binding.

iv) T12S, E21D, L63P, H69K

In the drug bound structures of protease, Glu 21 is seen to have a flexible side chain (Fig.16A). However, in some of the semi-open structures [3PHV, 2HB4] and in the open form [1TW7], the flexibility is restricted by the side chain O_H---O hydrogen bond formed with Thr 12. Following the T12S and E21D mutations, Ser 12 is found to interact with Asp 21 through O_H---O hydrogen bond (Fig.16B). In one of the unliganded protease structures, the side chain carboxyl of Glu 21 forms water mediated hydrogen bonds with the side chain
amino group of Arg 8 which is at the dimer interface (Fig. S4 in Appendix III). This interaction affects the salt bridge formed between Arg 8 and Asp 29 of the other monomer during dimerization. These mutations restrict the side chain flexibility of Glu 21 and could indirectly aid in dimer stabilization. In some of the drug bound protease structures analyzed [e.g. 2AQU] Glu 21 forms van der Waals contacts with Leu 10. This interaction would be lost as a result of the E21D mutation.

In most of the drug bound protease structures, the side chain of Leu 63 interacts with the side chain of Leu 72 through weak van der Waals forces. The main chain carbonyl of Leu 63 is seen to interact with the main chain amino group of Gly 16 either directly or through water mediated hydrogen bonds. The mutation L63P results in stronger van der Waals contacts with Leu 72. Also, new contacts are formed with the side chain of Ile 62. These contacts can alter the side chain conformations of Leu 72 and Ile 62. A change in the Leu 72 side chain conformation might affect the two hydrogen bonds that Gln 92 forms with the main chain carbonyl and amino groups of Leu 72 (Fig. S5 in Appendix III). In most of the unliganded structures of protease [PDB ID: 1HHP, 3PHV, 1HVP, 2G69], this interaction between Gln 92 and Leu 72 is absent. A change in Ile 62 side chain conformation can affect the hydrophobic contacts in the region involving Met 36, Leu 33, Leu 38, Ile 15, Val 75 and Ile 62. His 69 is protonated in the drug bound protease structures and the side chain NE1 interacts with the free C-terminal carboxyl through a salt bridge (Fig.16A). The side chain ND1 makes a hydrogen bond to the main chain carbonyl of Cys 67. The above interactions are however present in only one subunit of most of the drug bound structures studied. In both subunits, the side chain interacts with the aromatic ring of Phe 99 through π-stacking interaction. In the structure of multi-drug resistant protease [PDB ID: 1TW7] in the open state, His 69 is seen to move close to the C-terminal carboxyl. Following H69K mutation, the side chain
amino group of Lys 69 can form a salt bridge with the C-terminal carboxyl (Fig. 14B). The interactions with Cys 67 and Phe 99 would be lost as a result of the mutation. This might aid in the lowering of the cantilever beta-hairpin.

v) K20T

In the inhibitor bound closed conformation of protease, the side chain amino group of Lys 20 interacts with the side chain carboxyl of Glu 34 and the side chain carbonyl of Asn 83 either directly or through water mediated hydrogen bonds. In addition to this it also forms a N—H—O hydrogen bond with the side chain carbonyl of Gln 18 (Fig. 16C). K20T mutation results in the loss or weakening of the interactions with Glu 34 and Gln 83 (Fig. 16D). The interaction with Gln 18 is also lost as a result of this mutation. A similar change is also seen in some of the protease structures in the unliganded state (Fig. S6 in Appendix III). The side chain carboxyl of Glu 34 forms a salt bridge with the side chain amino group of Lys 20 in the multidrug resistant protease structure [1TW7] which is in the open conformation (Fig. S7 in Appendix III). This interaction could be lost as a result of this mutation. As mentioned above, the shortening of the side chain as a result of the mutation can also aid in the lowering of flap elbow.

vi) L90M

In the wild type drug bound protease structures, the side chain of Leu 90 interacts with Thr 26, Ile 85, Gly 86, Leu 24, Cys 95 and Asp 25 (main chain carbonyl 3.86Å) through van der Waals forces (Fig. 16E). These interactions help to maintain the geometry of the active site. It also interacts with the side chain of Leu 5 of the other protomer through van der Waals forces. In one of the subunits of the L90M mutant structure available in PDB [2F81], Met 90 is observed to have two side chain conformations (Fig. 14F). In the side chain conformation
in one of the occupancy states (68% populated), Met 90 forms a number of van der Waals contacts with the main chain atoms of Asp 25 and Thr 26. The terminal methyl group is also close to the main chain carbonyl of catalytic Asp 25 (3Å). In the other conformation, Met 90 can interact with the side chains of Leu 5 and Leu 97 of the other protomer. The interaction pattern is similar in the other subunit also. The mutation can result in change in the geometry of the active site due to changes in the extent of interactions with Asp 25 and Thr 26. The mutation can also strengthen the dimer interface because of the newly formed interaction with Leu 97 in the extended conformation. The dimer interface is formed by amino acid residues 1-4 and 96-99 as a result of the overlapping β strands below the active site of the enzyme previously referred to as ‘interleaving’ of the β strands.

6.5 HIV-1 env (V3) gene amplification, sequencing, co-receptor prediction and susceptibility to co-receptor antagonists.

6.5.1 HIV-1 V3 sequences quality control

The REGA HIV-1 Subtyping Tool - v2.0 identified 47 strains as clade C and one strain as clade A1 based on pol and env sequence data. In the phylogenetic tree the 47 study sequences were seen to cluster with the clade C V3 sequences (Fig. 17). One strain (EU781864) was seen to closely match with the clade A sequence. None of the strains were identical (not showing similarity greater than 99%) ruling out cross contamination.
Fig 17 Phylogenetic tree constructed using nucleotide sequences of 48 strains from Group 1 and the consensus sequences of clade C strains from India (D13427, D13424, DQ325336, DQ325332, DQ325314, DQ149152, DQ149147, AY128271, AY128263, AJ292013, AJ292009) clade C CXCR4 strains from Africa (L22956, DQ382379, DQ382378, DQ382373, DQ382372, DQ382362) clade A strain (AY350379) and clade B CXCR4 strains (U08444, U72495, AF075720)

For the construction of the phylogenetic tree, the evolutionary history was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Phylogenetic analyses were conducted in MEGA4.
The extracts of ten strains obtained using QIAamp® viral RNA extraction reagents in addition to amplification using a one step RT-PCR was also subjected to amplification of the V3 region without cDNA synthesis. This was done to ensure that only RNA extracts were being amplified and not possible DNA contamination in the RNA extracts. No amplification was observed from samples not subjected to cDNA synthesis. This was indicated by the absence of any amplification bands when the amplified products were run on a 2% agarose gel containing ethidium bromide as shown in the Figure. The desired amplification bands were however seen in all the ten samples whose V3 regions were amplified using the one step RT-PCR.

6.5.2 HIV-1 viral load and CD4 counts

The HIV-1 viral load and CD4 counts in the study population were as follows.

**Study group 1 viral load and CD4 counts in the study population.**

The mean HIV-1 viral load in the treatment naïve group was log\(_{10}\) 5.85 copies/ml and with a median value of log\(_{10}\) 5.45 copies/ml. The viral load in this group ranged from log\(_{10}\) 3.04 – 7.0 copies/ml. The mean CD4 counts in this group was 235 cells/mm\(^3\) with a median value of 202 cells/mm\(^3\). The CD4 counts in this group ranged from 10 – 676 cells/mm\(^3\).

**Study group 2 viral load**

The mean HIV-1 viral load in this group was log\(_{10}\) 5.33 copies/ml with a median value of log\(_{10}\) 4.17 copies/ml. The viral load in this group ranged from log\(_{10}\) 3.83 – 6.12 copies/ml. The mean CD4 counts in this group available for 7 individuals was 274 cell/mm\(^3\) with a median value of 375 cells/mm\(^3\). The CD4 counts in this group ranged from 53 – 436.
cells/mm³. The mean viral load and CD4 counts for the R5 and X4 strains as classified by the C matrix of the Web PSSM were also calculated in order to evaluate differences amongst R5 and X4 strains. The data on viral load and CD4 were analysed as shown below:

**Group with R5 utilizing strains**

The mean HIV-1 viral load in the R5 group (n=45) was log₁₀ 5.80 copies/ml and ranged from log₁₀ 3.04 copies/ml to log₁₀ 7.0 copies/ml. The CD4 counts in this group ranged from 34 – 676 cells/mm³ with a median value of 217 cells/mm³. CD4 counts were available for only 44 individuals in this group. The group harbouring R5 strains consisted of 29 (64%) males and 16 females who had an age range of 12-61 years.

**Group with X4 utilizing strains**

The X4 study group had 1 female and 1 male whose ages were 20 and 30 years, respectively. The mean viral load and CD4 counts in this group were log₁₀ 5.70 copies/ml and 135 cell/mm³, respectively. Both these individuals were asymptomatic.

A scatter plot of the HIV-1 viral load and CD4 counts in group 1 and 2 are shown in Fig 18

### 6.5.3 HIV-1 co-receptor usage prediction based on V3 region sequencing

Based on output from Geno2Pheno, three strains were identified as CXCR4 co-receptor users (X4). One strain belonged to the treatment failure group. Using the C matrix of the Web PSSM, two strains were identified as CXCR4 utilizing. The strains identified as X4 belonged to the treatment naïve group. For one of the CXCR4 utilizing strain when the ambiguous base-calls were expanded, there were 4 possible amino acid sequences. Among the four, three sequences were predicted as X4. The sequence predicted as R5 had a PSSM score of -21.93
Fig 18 HIV-1 coreceptor usage of test strains shown in the context of CD4 and viral load of the individuals

- ● - CXCR4 receptor utilizing HIV-1 strain present in the individual
- ◊ - CCR5 receptor utilizing HIV-1 strain present in the individual
which is very near the prediction cut-off of -21.64. The other scores were well above the cut-off with their values being -19.03, -18.37 and -15.47. Since at least two of the four possible sequences must have been present in the population, and assuming that ambiguous reads were the result of polymorphism in the bulk-sequenced sample, at least one X4 must exist in the sample. One clade C strain had a deletion in the V3 amino acid sequence based on the nucleotide sequence submitted. Web PSSM generated the output taking the deletion in the 24th and 25th position. When the deletion was in the 24th position it was classified as a R5 strain but when it was accounted for in the 25th position it was classified as X4. The scores for this strain using Web PSSM were -23.10 for deletion at position 24 and -21.24 for deletion at position 25. There were no ambiguous base calls in this sample. The ambiguity in prediction arose due to 2 equal-scoring best alignments emitted by the Web PSSM tool. The clade A1 strain was also analyzed using both the B and C matrices available using Web PSSM. When analyzed using clade B matrices the strain was identified as CCR5 utilizing (R5). This strain was not taken for analyses. The positive and the net charges of the V3 sequences were also calculated by Web PSSM. The mean positive charge of the R5 strains was 6 and the net charge 4.06 with mean difference between the positive charge and the net charge being +1.83. For the X4 strain the mean positive charge was 6 and the mean net charge was 3. The net charge of the R5 strains ranged from 1 to 6 while for the X4 strains it was 3 to 4. The crown motif sequence (GPGQ) was unaltered in both the R5 and X4 strains as shown in Fig. 19a and b at positions 15-18. While interpreting Fig. 17 a and b, it should be noted that we really don’t know the frequency of the different combinations of nucleotides that are actually present in the viral population. Probably not all of the possible nucleotide combinations were present and hence not all the amino acids were present either. The results are summarized in table 12. The results displayed by G2P on submission of the sequences are shown in Appendix VII E and F.
Fig. 19: The amino acid characters for all the R5 (n=45) Fig. 19a and X4 (n=1) Fig. 19b strains as determined by Web PSSM are shown. The size depicts the relative proportion of the amino acid at a site. Taller the logo, lesser is the variability at the amino acid position. The x axis depicts the amino acid position. The seq logo was developed at http://weblogo.berkeley.edu/logo.cgi
Table 12: Classification of the strains into R5 and X4 based on the C matrix of Web PSSM

<table>
<thead>
<tr>
<th>Predicted Co-receptor Usage</th>
<th>No*</th>
<th>CD 4 cell counts (mean ± SD)</th>
<th>HIV-1 RNA copies/ml (mean ± SD)</th>
<th>Charge of the amino acid sequence of the env V3 region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Net Charge (mean ± SD)</td>
<td>Pos. Charge (mean ± SD)</td>
</tr>
<tr>
<td>CCR5</td>
<td>45</td>
<td>251.2 ± 162.7#</td>
<td>Log 5.80 ± 6.20</td>
<td>6 ± 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.06 ± 1.01</td>
</tr>
<tr>
<td>CXCR4</td>
<td>2</td>
<td>135±72.12</td>
<td>Log 5.70 ± 5.42</td>
<td>3.1 ± 0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6± 0</td>
</tr>
</tbody>
</table>

* The strain with belonging to clade A was removed from analysis

# CD4 counts were available only for 44 individuals

6.6 HIV-1 drug resistance mutations in ART experienced individuals (study group 2) in relation to viral loads and CD4 counts.

The most common mutation observed in the RT amino acid sequences of individuals in study group 2 was the M184V mutation. This was not seen in individuals belonging to group 1 (n=101). Individuals with M184V were associated with a mean viral load of log_{10} 5.54 copies/ml with a median value of log_{10} 4.22 copies/ml. The CD4 counts ranged in this group from 32 to 436 cells/mm³ and the mean was 289 cells/mm³. The other strains had M184V in association with TAMS and these individuals had a mean viral load of log_{10} 5.40 copies/ml with a median value of log_{10} 4.9 copies/ml. The CD4 counts ranged in this group from 26 to
580 cells/mm³ and the mean was 225 cells/mm³. The results are summarized in Table 13. HIV-1 strains from two individuals with Gen bank ID Nos.: EU 781841 and 78184 had M184V as the only NRTI resistance conferring mutation. The plasma from these two individuals had higher viral load compared to other individuals in this group. These two individuals who were concordant couples were on incompletely suppressive therapy of d4T and 3TC at the time of genotypic drug resistance testing. Cross contamination for these two strains was ruled out as shown in the phylogentic tree (Fig.9) where they formed two distinct branches with different branch lengths. One patient with the Q151M mutation had a viral load of log₁₀ 6.19 copies/ml.

Five strains were found to harbour mutations conferring resistance to PIs. These individuals had a mean viral load of log₁₀ 5.0 copies/ml with a median value of log₁₀ 4.87 copies/ml. An individual infected with a strain showing the V82A mutation in the protease gene had a viral load of log₁₀ 4.35 copies/ml. An individual on whose strain genotypic drug resistance was done on two occasions while on the same regimen (ABC+ddI+NFV) was observed to show the loss of the M46I mutation in its protease amino acid sequence while on the second visit. The second visit was also associated with a log₁₀ 0.43 copies/ml of reduction in viral load. The CD4 counts were not available for this individual. The Gen bank IDs of the strains on the 1st and the 2nd visit were EU037772 and EU037775.
A: Phylogenetic tree constructed using nucleotide sequences of 101 strains from Group 1 and the consensus group M, for M group, A1, A2, B, C, D, AE, F1, F2, G, H; Indian subtype C sequences (AF286232, AF067155, AF067157, AY713414, AY049708, DQ826595 and DQ826669); non-Indian subtype C sequences from China (AY967806), Ethiopia (AY255823), Myanmar (AB097871) and South Africa (AF544009 and DQ093604). The M group consensus sequence was used to root the tree.

B: Phylogenetic tree constructed using nucleotide sequences of 21 strains from Group 2 and the consensus group M, for M group, A1, A2, B, C, D, AE, F1, F2, G, H; Indian subtype C sequences (AF286232, AF067155, AF067157, AY713414, AY049708, DQ826595 and DQ826669); non-Indian subtype C sequences from China (AY967806), Ethiopia (AY255823), Myanmar (AB097871) and South Africa (AF544009 and DQ093604). The M group consensus sequence was used to root the tree.

C: Phylogenetic tree constructed using nucleotide sequences of 101 strains from Group 1 and 21 strains from Group 2

For the construction of the phylogenetic tree the evolutionary history was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Phylogenetic analyses were conducted in MEGA4.
D: Phylogenetic tree constructed using amino acid sequences of 101 strains from Group 1 and the consensus group M, for M group, A1, A2, B, C, D, AE, F1, F2, G, H; Indian subtype C sequences (AF286232, AF067155, AF067157, AY713414, AY049708, DQ826595 and DQ826669); non-Indian subtype C sequences from China (AY967806), Ethiopia (AY255823), Myanmar (AB097871) and South Africa (AF544009 and DQ093604). The M group consensus sequence was used to root the tree.

E: Phylogenetic tree constructed using amino acid sequences of 21 strains from Group 2 and the consensus group M, for M group, A1, A2, B, C, D, AE, F1, F2, G, H; Indian subtype C sequences (AF286232, AF067155, AF067157, AY713414, AY049708, DQ826595 and DQ826669); non-Indian subtype C sequences from China (AY967806), Ethiopia (AY255823), Myanmar (AB097871) and South Africa (AF544009 and DQ093604). The M group consensus sequence was used to root the tree.

F: Phylogenetic tree constructed using amino acid sequences of 101 strains from Group 1 and 21 strains from Group 2

For the construction of the phylogenetic tree, the evolutionary history was inferred using the Minimum Evolution method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4.
Table 13: CD4 counts and HIV-1 viral loads of individuals harbouring strains with M184V mutations and individuals harbouring strain with M184V mutations in association with TAMS

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Individuals with strains harbouring M184V mutations</th>
<th>Individuals with strains harbouring M184V mutations in association with TAMS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CD4 cell count (cells/mm³)</td>
<td>HIV-1 viral load (log₁₀ RNA copies/ml)</td>
</tr>
<tr>
<td>1</td>
<td>375</td>
<td>3.98</td>
</tr>
<tr>
<td>2</td>
<td>436</td>
<td>4.30</td>
</tr>
<tr>
<td>3</td>
<td>171</td>
<td>5.80</td>
</tr>
<tr>
<td>4</td>
<td>397</td>
<td>3.83</td>
</tr>
<tr>
<td>5</td>
<td>340</td>
<td>3.51</td>
</tr>
<tr>
<td>6</td>
<td>245</td>
<td>6.37</td>
</tr>
<tr>
<td>7</td>
<td>318</td>
<td>4.60</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>4.23</td>
</tr>
<tr>
<td>9</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>11</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Mean</td>
<td>289</td>
<td>5.54</td>
</tr>
<tr>
<td>Median</td>
<td>329</td>
<td>4.22</td>
</tr>
</tbody>
</table>

ND : Not done