5. **Materials and Methods**

5.1 **Study group**

The individuals recruited for the different investigations were primarily drawn from HIV-1 infected individuals who were referred to the clinical virology department for CD4 and/or HIV-1 viral load determination as part of patient management by physicians of the departments of internal medicine, infectious disease, paediatrics and dermatology. The study was explained to all the subjects of the study group and a written consent was obtained. The written consent used in the study is shown in Appendix IV. An institutional review board clearance was given to use archived samples and prospectively collected samples for all the analysis. Information regarding treatment was obtained from the patients and their medical charts were reviewed by the collaborating physicians for other clinical information. Data on modes of transmission and time of infection were not elicited for these individuals due to privacy concerns. The sample size used in our study is adequate. Using Epi Info v6.04c we calculated a sample size for the study to estimate a prevalence of 7% at a precision of ±5% (95% CI). The 7% prevalence was approximated based on more recent studies published from India (30, 299). The estimates sample size was 101. This is a cross sectional study and the individuals recruited were by ‘convenient sampling’. We took a total of 120 HIV-1 infected individuals of which 101 were treatment naive.

The ART regimen initiated on CMC patients include a combination of NRTI (either AZT 300 mg twice daily or d4T 30mg twice daily with 3TC 150 mg twice daily) and a NNRTI (either NVP 200 mg once daily for 14 days followed by 200 mg twice daily or EFV 600mg once daily at bed time. All the drugs are generic fixed-dose combinations (AZT 300 mg+ 3TC 150 mg+ D4T 30 mg).
mg, d4T 30 mg + 3TC 150 mg, AZT 300 mg + 3TC 150 mg + NVP 200 mg, d4T 30 mg + 3TC 150 mg + NVP 200 mg).

The participants in the study could be broadly classified into two groups 1 and 2:

1) **ART naive group.**

   This group consisted of HIV-1 infected individuals who were not on ART regimens at time of recruitment. The individuals did not give a history of ART or seeking medical attention for the condition prior to meeting the ID physician at CMC. A total of 101 individuals were included in this study as a whole. Among these, 100 individuals had never been exposed to any ART regimen and one individual had taken ART before delivery of her child.

2) **ART failure group**

   This group consisted of HIV-1 infected individuals who had greater than 1,000 HIV-1 RNA copies/ml more than 8 months after initiation of ART by the ID physicians at CMC or referred by treating physicians outside of CMC. All the subjects in this group were on ART at the time of recruitment. A total of 19 individuals were included in this study as a whole.

**Establishment of HIV-1 infection status for study groups 1 and 2**

In the department of Clinical Virology individuals suspected to have HIV are tested with WHO/UNAIDS certified commercial assays. The algorithm of testing is shown in Fig. 5 and is in compliance with WHO/NACO strategy III for diagnosis of HIV infection. This system of testing is in vogue since 2005.
Fig. 5. The algorithm of testing is in compliance with WHO/NACO strategy III for diagnosis of HIV infection in the study.

Any samples which show discrepant results in assay 1 and assay 2 will be tested by two more assays (3rd/4th generation). If three of the four assays are negative the sample will be declared as negative. If two or more assays are positive, another sample (fresh sample) is tested for western blot/HIV-1 RNA estimation.

Assays 1 and 2 are ELISA based while Assay 3 is a rapid test based on flow through technology.
5.2 Collection and processing of samples:

Blood samples were collected from 101 HIV-1 infected individuals not on ART. Samples were also collected from 19 HIV-1 infected individuals with detectable HIV-1 viral load while on ART. The HIV-1 infected individuals had come to the Clinical Virology department of Christian Medical College Vellore, south India for CD4+ T cell estimation and or HIV-1 viral load during the year 2005-2007. Two ml of blood was collected in K3EDTA vacutainer tubes for CD4+T cell estimation. For HIV-1 viral load estimation and extraction of HIV-1 RNA for other studies, blood was collected in Na EDTA tubes. Plasma was separated from these tubes after they were centrifuged at 1500 rpm for 10 minutes at 4°C. The plasma was stored as multiple aliquot at -80°C until the time of testing.

The Guava® Easy CD4™ System (Guava Technologies, Hayward, CA, USA), as described earlier, was used for CD4 estimation (213). Briefly, the instrument uses two fluorescence parameters in combination with forward scatter (FSC) to identify cells. The reagents consists of a monoclonal anti-human CD3 antibody conjugated to the tandem dye phycoerythrin (PE)-Cy5 and monoclonal anti-human CD4 antibodies conjugated to PE. The antibody staining solutions were used to stain 10-µl portions of the blood samples. The CD4+ T-cell count was measured after staining using CD3 monoclonal antibody labelled with PE-Cy5 and CD4 monoclonal labelled with PE were added. The 1x lysing solution was then added after a 15-min incubation period at room temperature (20 to 25°C). After another 15 min of incubation
at room temperature, the samples were screened and the data were acquired on the Guava Personal Cell Analysis (PCA) instrument by using CytoSoft version 2.2 software.

For consistencies of results all blood samples for CD4 testing are collected from HIV infected individuals only on the day of testing between 8:00 -10:00 AM. Prior to each run a Guava®Check™ is run to verify the performance of the run. This step consists of checking fluorescent bead suspension of known concentration. The bead samples are run in triplicates. The count results, forward scatter (FSC) signal intensity and mean fluorescence intensities are displayed along with the coefficient of variation (CV). Only if the values and CV of the measured parameters are within the stipulated ranges the CD4 counts of the samples are measured. The stipulated range for the FSC intensity is between 150 to 300, for the photomultiplier intensity 1 mean fluorescence intensity it is between 850 to 1200 and photomultiplier intensity 2 mean fluorescence intensity it is between 200 to 300. The laboratory here is also a participant of two External Quality Assessment Scheme (EQAS) programs under the auspicious of NARI/NACO. Under this program, QC samples sent by National AIDS Research Institute, Pune, India (NARI)/NACO are measured for their CD3, CD4 values using the Guava® Easy CD4™ System and results dispatched to NARI for evaluation. Our laboratories always obtained values in acceptable range and passed the QC.

Extraction of RNA for use in the study was done using QIAamp® viral RNA extraction reagents (Qiagen GmbH, Hilden, Germany). Before extraction the sample was first lysed using the lyses solution (Buffer AVL) under highly denaturing conditions to inactivate RNases and to ensure isolation of intact viral RNA. Buffering conditions were also adjusted to provide optimum binding of the RNA to the QIAamp membrane, which is a silica-gel-
based membrane. The sample was then loaded onto a QIAamp spin column. The RNA get bound to the membrane and the contaminants are washed away in two steps using two different wash buffers (AW1 and AW2). RNA is finally eluted out in a special RNase-free buffer (AVE). AVE is RNase-free water that contains 0.04% sodium azide to prevent microbial growth and subsequent contamination with RNases. The steps involved in RNA extraction are as follows:

1) Add 560 µL of buffer AVL into a 1.5 mL microcentrifuge tube.
2) Add 140 µL of plasma to the buffer AVL in the microcentrifuge tube. Mix by pulse-vortexing for 15 seconds and spin down at 8000 rpm for 1 minute.
3) Incubate at room temperature (15°C-25°C) for 11 minutes.
4) Add 560µL of ethanol (96-100%) to the sample. Mix by pulse-vortexing for 15 seconds and spin down at 9000 rpm for 30 seconds.
5) Transfer 630µL of the solution from step 4 to the QIAamp spin column without wetting the rim. Close the cap and centrifuge at 8,000 rpm for 1 minute.
6) Place the QIAamp spin column into a clean 2-ml collection tube and discard the tube containing the filtrate.
7) Transfer the remaining 630µL to the QIAamp spin column. Spin at 8000 rpm for 1 minute. Place the QIAamp spin column into a clean 2-ml collection tube and discard the tube containing the filtrate.
8) Add 500 µL of AW1 and spin at 8000 rpm for 1 minute. Place the QIAamp spin column into a clean 2-ml collection tube and discard the tube containing the filtrate.
9) Add 500 µL of AW2 and spin at 14000 rpm for 4 minutes. Place the QIAamp spin column into a clean 2-ml collection tube and discard the tube containing the filtrate.
10) Place the QIAamp spin column into a clean 2-ml collection tube (not provided) and discard the tube containing the filtrate. Spin at 14000 rpm for 2 minutes.
11) Place the QIAamp spin column into a clean 1.5-ml collection tube (not provided).
12) Add 50µL elution buffer that has been equilibrated to room temperature. Incubate at room temperature for 1 minute and then centrifuge at 8,000 rpm for 30 minutes.

The extracted RNA (20 µL) was subjected to HIV-1 Viral load estimation using Artus RealArt™ HIV-1 reagents (Qiagen Hamburg Gmbh, Germany) using Rotor- Gene™ 3000 (Corbett research, Australia). The HIV-1 Artus Kits are intended for in-vitro diagnostic use in Europe (CE0197) as per the product info available in the company website. (http://www1.qiagen.com/Products/artusHIVirusRT-PCRKitCE.aspx?ShowInfo=1). The cycling conditions were as follows: 50°C for 10 min, 95°C for 10 min followed by 45 cycles of 95°C for 20 sec, 50°C for 30 sec, 71°C for 30 sec. The Artus RealArt™ is a real-time quantitative PCR assay which is based on TaqMan® principle. In this assay there is detection and measurement of products generated during each cycle of the polymerase chain reaction process. The PCR products are measured while they are accumulating in the reaction mix i.e., in ‘real time’. Quantitation is achieved with a standard curve using plasmid standards of known copy numbers fed in the onboard software provided with the real time PCR machine (Corbett Research, Australia). There are 4 standards (HIV 1 RG QS 1-4) which represent 1 x 10^4 IU/µL- 1 x 10^1 IU/µL. Each run also included an internal control whose copy number was known. The internal control was made making a 1/100 dilution of a sample with a known copy number in HIV negative plasma.

The reverse transcription and specific amplification of a 73 bp region of the LTR region in the HIV-1 genome is measured. This assay makes use of two channels viz., Cycling A FAM and Cycling A JOE. The former is used for the direct detection of the specific amplicon and the latter to identify possible PCR inhibition. Samples that are negative are checked for
possible PCR inhibition by checking for any signals in the Cycling A JOE channel. When PCR inhibition is indicated, the assay is repeated for the inhibited sample. The assay is repeated for the sample following re-extraction after dilution to a factor of 10.

At the end of each run the threshold and Ct is also calculated by the onboard software. The Ct is defined as the cycle at which the sample fluorescence crosses the threshold. The run is validated only if the CV of the 4 standards is within the accepted range and the negative controls (water) are negative. For every three samples a negative control is included. This assay has been previously validated with an ‘in-house’ HIV-1 RNA real time PCR assay in our laboratory using NIB standards (UK)

5.3 HIV-1 pol gene amplification and sequencing

5.3.1 HIV-1 pol gene amplification

The extracted RNA (20µl) was subjected to a one step RT-PCR (Qiagen GmbH, Hilden, Germany) along with specific forward and reverse primers, HIV-1 out 1 and HIV-1 out 2 respectively (custom synthesized at Invitrogen, CA, USA). The list of primers and their sequences used are shown in Table 1. (The sequences for all the primers were provided by the kind courtesy of Dr. Deenan Pillay, HPA, Collindale, UK). In the department, the appropriate flow through while doing PCR is practiced which includes the use of dedicated suite of room, filter block tip and dedicated micro-pipettes and use of adequate disposable gloves while adding the sample. This practise is similar to the earlier prescribed precautions prescribed by Kwok and Higuchi (302). The Qiagen one tube RT-PCR reagents are a mixture of Omniscript and Sensicript Reverse transcriptase which helps in transcription of any RNA quantity between 1pg to 2 µg. It also contains HotStart Taq DNA polymerase for
amplification of the cDNA. The cycling conditions were as follows: 50°C for 30 min, 95°C for 15 min followed by 40 cycles of 94°C for 30 sec, 60°C for 30 sec, 72°C for 2 min and a final extension of 72°C for 7 min. The amplification reactions were carried out on PTC-100 (MJ research, California, USA) or MyCycler™ (BioRad California, USA).

The first round products were further amplified using Qiagen Hot Start Master Mix (Qiagen GmbH, Hilden, Germany) with forward and reverse primers, PCR res 1 and PCR res 2 respectively. The cycling conditions were as follows: 95°C for 15 min followed by 23 cycles of 94°C for 30 sec, 58°C for 30 sec, 72°C for 2 min and a final extension of 72°C for 7 min. The amplified products were run on a 2% agarose (Sigma Aldrich Inc, St Louis, MO, USA) gel containing ethidium bromide to check for the specific 1800 bp size amplicon as shown in Fig. 6. The agarose gel was visualized using the gel documentation system Gel Doc 2000 (BioRad, California, USA) using the software Quantity One version 4.1.1 (BioRad, California, USA).

5.3.2 HIV-1 pol gene sequencing

The amplified products contained in their original amplification tubes were then purified with Millipore (Billerica, MA, USA) plates by vacuum drying. The purification before setting up the sequencing reaction helps to remove unused dNTPs and primers. The following steps were followed:

1. The amplified product was made up to 100 µL using sterile distilled water
Fig 6. Gel picture showing the detection of the amplified pol gene of HIV-1.

Lane 1 Molecular Weight Marker (Φ X 174, Hae III digest)
Lane 6, 11 Negative control (PCR grade water)
Lane 2, 3, 4, 5, 7, 8, 9, 10, 12 HIV-1 pol gene (1800 bp)
2 The diluted product was then transferred to the Millipore micotiter plate and attached to a Millipore vacuum manifold.

3 Vacuum pressure was applied for approximately 10 minutes or until the well dried.

4 Upon drying of the wells 100 µL of sterile distilled water was added and step 3 was repeated.

5 Upon complete drying of the wells, 20 µl of sterile distilled water was added and mixed 25 times with a pipette and the contents transferred to PCR tubes.

The purified DNA was collected from the wells of the plate using a micropipette and transferred to PCR tubes which were then subjected to sequencing using the Big Dye terminator assay.

The sequencing reaction contained the following: 1 µl of Ready Reaction mix (ABI PRISM Big Dye terminator cycle sequencing reagent, Applied Biosystems, California, USA), 2 µl of Ready Reaction buffer (ABI PRISM Big Dye terminator cycle sequencing reagent, Applied Biosystems, California, USA), 3.2 µl of primers at a concentration of 1 picomole / µl and 1 µl of the purified product. The Ready Reaction mix contains the Ampli Taq polymerase FS, deoxynucleotides and dideoxynucleotides. The dideoxynucleotides are responsible for chain termination which was the principle of the sequencing reaction employed. Ampli Taq polymerase FS contains a point mutation in the active site (F667Y) which leads to less discrimination against dideoxynucleotides. The enzyme also has a G46A mutation which removes the 5’-3’ nuclease activity. The reaction was made up to a final volume by addition of 10 µl of PCR grade water. The reaction mixture was then subjected to 25 cycles of 96°C for 15 sec, 50°C for 20 sec and 60°C for 4 minutes.
The sequenced products were then subjected to purification before being analyzed on the ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems, California, USA). The sequenced samples were subjected to post sequencing clean up using a Millipore vacuum system. The following steps were followed.

1. The sequenced sample product was made up to 40 µL using injection solution.
2. The diluted product was then transferred to the Millipore micotiter palte and attached to a Millipore vaccum manifold.
3. Vacuum pressure was applied for approximately 10 minutes or until the well dried.
4. Upon drying of the wells 40 µL of injection solution was added and step 3 was repeated.
5. Upon complete drying of the wells, 30 µl of sterile distilled water was added and mixed 25 times with a pipette and the contents transferred to PCR tubes for analysis.

The ABI PRISM 310 Genetic Analyzer contains a cathode, anode and a capillary. The capillary tube and the cathode touch the sample while the anode is placed in a buffer. A current is then passed from the cathode to anode and this is known as electrokinetic injection. The sample forms a tight band in the capillary during injection following which the end of the capillary near the cathode is immersed in the buffer and the current is still passed. When the sample enters a detector window in the capillary a laser excites the fluorescent dye labels and the reading is taken. The output is in the form of an electropherogram. Sequencing reaction readouts were taken only if the electropherogram indicated no mixtures. This was indicated by the absence of more than one peak throughout the electropherogram. The sequencing was done at our institutional facility using sequencing primers. We used six sequencing primers A, B, C, E, F and G but not D. Primer D was not used because the sequence electropherogram obtained upon using D was not readable on repeat testing. The
Table 1: Primer sequences used for HIV-1 genotypic drug resistance assay

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1st round:</strong></td>
<td></td>
</tr>
<tr>
<td>HIVout1</td>
<td>5’ AAT GAT GAC AGC ATG Y(T+A)CA GGG AGT</td>
</tr>
<tr>
<td>HIVout2</td>
<td>5’ AGT CTT TCC CCA TAT TAC TAT GCT TTC</td>
</tr>
<tr>
<td><strong>2nd round:</strong></td>
<td></td>
</tr>
<tr>
<td>PCRres1</td>
<td>5’ GGA AAA AGG GCT GTT GGA AAT GTG</td>
</tr>
<tr>
<td>PCRres2</td>
<td>5’ GGC TCT TGA TAA ATT TGA TAT GTC CAT TG</td>
</tr>
<tr>
<td><strong>Sequence primers:</strong></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5’ AGC CAA CAG CCC CAC CAG</td>
</tr>
<tr>
<td>B</td>
<td>5’ GTT AAA CAA TGG CCA TTG ACA GAA GA</td>
</tr>
<tr>
<td>C</td>
<td>5’ TGG AAA GGA TCA CCA GCA ATA TTC CA</td>
</tr>
<tr>
<td>D</td>
<td>5’ GGA ACT GTA TCC TTT AGC TTC CC</td>
</tr>
<tr>
<td>E</td>
<td>5’ GGG CCA TCC ATT CCT GGC</td>
</tr>
<tr>
<td>F</td>
<td>5’ CCA TCC CTG TGG AAG CAC ATT G</td>
</tr>
<tr>
<td>G</td>
<td>5’ CTG TAT TTC TGC TAT TAA GTC TTT TGA TG</td>
</tr>
</tbody>
</table>
primer sequences are mentioned in Table 1. Not using D still allows amplification without loss of any sequence segment. There are two reports on a commercial assay which with a similar experience with their sequencing primers (303, 304). These published papers do not reveal the sequences. The complete protease gene was analyzed whereas only the first three hundred amino acids were analyzed for the reverse transcriptase. There were no stop codons and the nucleotide ambiguities were less than 5%. Consensus sequence was created using BioEdit sequence alignment editor version 7.0.4.1 and aligned using ClustalW (http://www.ebi.ac.uk/Tools/clustalw/) (305). The sequences were analyzed by Stanford HIV drug resistance database for genotypic drug resistance analysis (http://hivdb.stanford.edu). Sequences obtained were also submitted to REGA HIV-1 subtyping tool - v2.0 to identify the subtype of the strains (http://dbpartners.stanford.edu/RegaSubtyping/) (306). An electropherogram obtained from one of the sequence primers is shown in Fig.7.

5.3.3 HIV-1 pol gene sequence quality control

The phylogenetic tree was constructed to confirm the subtyping analysis and also as a contamination control (quality control) for the sequencing reactions as recommended. Phylogenetic analysis was performed for the sequences with Mega 4 software using a minimum evolution method with Kimura two-parameter and bootstrap value of 500 replicates (307). These were compared to the HIV-1 pol sequences obtained from the Los Alamos HIV sequence data base. The sequences included were consensus sequences 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
Fig. 7. Electropherogram showing nucleotide sequence of a part of the HIV-1 pol gene.
China (AY967806), Ethiopia (AY255823), Myanmar (AB097871) and South Africa (AF544009 and DQ093604). The M group consensus sequence was used to root the tree. Separate phylogenetic trees were constructed for using the nucleotide sequences of strains from the ART naive group and the ART failure group. Phylogenetic trees were also constructed using the translated amino acid sequence of the strains from the ART naive group and the ART failure group.

For quality control, we also randomly selected amplified products from ten strains (10% of the total treatment strains) and sent them to an external commercial centre along with the sequencing primers (1st Base, Singapore) for sequencing.

5.4 Comparison of HIV-1 drug resistance algorithms

The sequences obtained from amplification and sequencing of the pol gene was submitted to three different HIV-1 drug resistance algorithms - the Stanford HIV Drug Resistance database, Geno2Pheno and DR_SEQAN (290-292). A total of 122 sequences were compared. This included 101 HIV-1 strain sequences from as many ART naive individuals and 21 HIV-1 strain sequences from 19 individuals showing viral load while on ART. The sequences were edited using BioEdit sequence alignment editor version 7.0.4.1 and submitted to the Stanford HIV Drug Resistance database, Geno2Pheno and DR_SEQAN (305). In the SHDB analyzes each mutation is given a score. The score for each drug is totaled and based on the score, resistance classified as: susceptible, potential low-level resistance, low-level resistance, intermediate resistance and high level resistance. The scores for each mutation are assigned based on published studies associating mutations with antiretroviral drugs. G2P
estimates phenotypic drug resistance data based on the submitted sequence (291). G2P uses two machine learning approaches viz., decision trees and support vector machines (SVM). The output thus contains interpretation based on these two systems. This database also proposes a score for the probability of the submitted sequence to originate from a resistant subpopulation rather than a susceptible population (291). The DS reports the fold increase of the IC\textsubscript{50} based on reports from published literature while the genotypic resistance profile is done based on a series of rules (292). The SHDB thus generated only one output while G2P and DS had 3 and 2 outputs, respectively. The susceptibility patterns were obtained for nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) which included lamivudine (3TC), abacavir (ABC), zidovudine (AZT), stavudine (d4T), didanosine (ddI), tenofovir (TDF), non-nucleoside reverse transcriptase inhibitors (NNRTIs) which included delavirdine (DLV), efavirenz (EFV), nevaripine (NVP) and protease inhibitors (PIs) which included atazanavir (ATV), amprenavir (APV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV) and saquinavir (SQV). While submitting to G2P the cut off levels for the phenotypic assay was set at the default level. The SHDB algorithm defines five levels of drug resistance: susceptible, potential low level resistance, low level resistance, intermediate resistance and high level resistance. The G2P algorithm defined three levels - susceptible, intermediate and resistant. The DS algorithm defined the levels as susceptible, partial resistance, significant resistance and high level resistance. In addition, it denoted hypersusceptibility as antagonistic; e.g., the mutation M184V, which induces resistance to 3TC, is responsible for hypersusceptibility towards thymidine analogues like AZT and d4T (16). For convenience of analysis we classified these different outputs into three broad classes as shown in Table 2. Results were considered as concordant when all the three algorithms denoted the same level of resistance in all forms of its output. Results were considered as discordant when at least one of the algorithms used for analysis reported a level of resistance different from the other
Table 2: Classification of the different outputs of the algorithms into three broad classes

<table>
<thead>
<tr>
<th>Algorithm Interpretation</th>
<th>Study Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susceptible and antagonistic</td>
<td>Susceptible (S),</td>
</tr>
<tr>
<td>Potential low level resistance, low level</td>
<td>Intermediate (I),</td>
</tr>
<tr>
<td>resistance and partial resistance</td>
<td></td>
</tr>
<tr>
<td>Intermediate resistance, significant resistance and resistance</td>
<td>Resistant (R).</td>
</tr>
</tbody>
</table>
algorithms in any one of its outputs. The data was entered in MS Excel spreadsheet and Epi
Info v 6.03 was uses to calculate the percentage concordance (observed agreement) for each
drug was then calculated. The pol gene sequences obtained for all the study strains were
submitted to the REGA HIV-1 subtyping tool - v2.0 to identify the subtype of the strains
(http://dbpartners.stanford.edu/RegaSubtyping/) (306).

5.5 Construction of HIV-1 Reverse Transcriptase 3D models

Amino acid sequences of HIV-1 RT generated from RNA extracted from plasma of blood
samples were collected from HIV-1 infected ART naïve individuals and those failing ART
were taken to develop 3D models of RT. Extraction of HIV-1RNA, amplification and
sequencing were done as described in section 5.3. RT sequences from 10 strains (8 from
treatment failure and 2 from treatment naïve groups) and one reference clade C
(AAY23520.1) were selected to generate three dimensional models. The treatment naïve
group was selected based on the presence of unusual mutation/ mutations not seen in clade B
strains. The ten strains selected are shown in table 3. The data on mutational patterns were
based on information obtained from the Stanford HIV drug resistance database. The
nucleotide sequences were translated into the amino acid sequences using the ExPasy
translate tool (http://www.expasy.ch/tools/dna.html). The translated sequences were aligned
using clustalW (http://www.ebi.ac.uk/clustalw/). Three dimensional models were generated
by mutating the available crystal structures of drug bound and unbound RT using the
Table 3: GenBank ID of strains that were modelled for the Reverse Transcriptase along with their percentage identity to clade C strains and mutations observed

<table>
<thead>
<tr>
<th>GenBank ID</th>
<th>Subtype (% identity*)</th>
<th>NRTI Resistance Mutations</th>
<th>NNRTI Resistance Mutations</th>
<th>Other Mutations</th>
</tr>
</thead>
</table>

*GenBank ID (full name) indicates the strain's unique identifier in the GenBank database.

*Percentage identity* refers to the proportion of nucleotides or amino acids that are identical between the compared strain and the reference strain.
<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1IKW</td>
<td>Efavirenz</td>
</tr>
<tr>
<td>1VRT</td>
<td>Nevirapine</td>
</tr>
<tr>
<td>1DLO</td>
<td>Unliganded</td>
</tr>
<tr>
<td>1HMV</td>
<td>Unliganded</td>
</tr>
<tr>
<td>IRTJ</td>
<td>Unliganded</td>
</tr>
</tbody>
</table>

*based on information from the Stanford HIV Drug resistance database*
Biopolymer tool of Sybyl (Tripos Inc., St. Louis, MO). The protein databank codes of the crystal structures of bound and unbound forms of RT used in this study are 1IKW (308), 1VRT (309), 1DLO (233), 1HMV(234), 1RTJ(230) and 1RTD (102) and are shown in table 4. The drugs bound to these structures are shown in table. Energy minimizations were carried out by targeting the sites of mutations and neighbourhood using the Maximin tool of Sybyl package. All the atoms that were within a distance of 6Å (Angstrom) from the mutated residue (“HOT” region), were allowed to move during minimization. Kollman United force field available in Sybyl package was used for energy minimization (310). A cutoff of 9Å was used for calculating non-bonded interactions. The inhibitor molecule was not included during energy minimization since the sites of most of mutations were not located close to the active site. Mustang and DALI (Holm, et al. 1996 Holm, et al. 2000) were used for multiple and pairwise structure superposition, respectively (311-313). PyMOL (DeLano Scientific LLC, San Carlos, California) and Setor were used to visualize modelled protein structures and to analyze interactions between amino acid residues(314).

The following criteria were used for hydrogen bonding and van der Waals interactions:

*Hydrogen bond interaction:*

Distance cutoff: 4 Å

Donor-Acceptor bond angle: > 90°

*Van der Waals interaction:*

Distance cutoff: Sum of van der Waals radii ± 0.5 Å

The computers used for construction of the 3D models had Linux as the operating software. These computers belonged to the department of Molecular Biophysics, Indian Institute of Science, Bangalore and were used under the immediate supervision of Prof N Srinivasan.
5.6 Construction of HIV-1 Protease 3D models

This component of the study involved construction of 3D models using amino acid sequences of HIV-1 protease molecule. This was generated from RNA extracted from HIV-1 infected ART naïve individuals and those failing ART. Extraction of HIV-1 RNA, amplification and sequencing were done as described in section 5.3. Protease sequences from thirteen strains (three and ten strains from treatment failure and treatment naïve groups respectively) and one reference Indian clade C strain sequence (accession ID AF067155) from the Los Alamos HIV sequence database were selected for generating three-dimensional models. The thirteen strains selected are shown in table 5. The strains for modelling were selected from the treatment naïve group based on the presence of one or more unusual mutations not present in clade B strains. The data on mutational patterns were based on information got from the Stanford HIV drug resistance database. This system is capable of comparing the sequence of interest with the consensus clade B sequence. The nucleotide sequences were translated using the ExPasy translate tool (http://www.expasy.ch/tools/dna.html). The translated sequences were aligned using clustalw (http://www.ebi.ac.uk/clustalw/). Three dimensional models were built by mutating the drug /inhibitor bound protease structures available in Protein Data Bank using the Biopolymer tool of Sybyl 6.5 package (Tripos Inc., St. Louis, MO, USA). The lists of models used are shown in table 6. Localized energy minimization was carried out for the mutated residues using the Maximin tool of Sybyl package. The effect of minimization decreases with distance from the site of mutation. Kollman United force field available in Sybyl package was used for energy minimization (310). The number of steps of energy minimization was decided based on the number of mutations in the model, eg: 250 steps of energy minimization was carried out for a strain with Genbank ID EU030408 which had 22 mutations. Mustang v 0.3 (Konagurthu, et al.2006) and DALI v2.4.1 (Holm, et al. 1996 Holm, et al. 2000) were used for multiple and pairwise structure superposition,
respectively (311-313). Gromacs package was then used in carrying out energy minimization in the presence of a solvent (water). The mutated models were centred in a cubic box whose faces were 7.5Å away from the model. The models were then solvated in single point charge (SPC) water. The charge of the system was neutralised using the Genion tool. OPLS-AA force field implemented in the Gromacs package was used to minimize the solvated models for 1000 steps.

PyMOL (DeLano Scientific LLC, San Carlos, California), Chimera and Setor (314, 315) were used to visualize protease structures and also to analyze interactions present in the modelled structures. Inter-atomic distances were calculated using the distance measurement tools available in Chimera and Pymol (DeLano Scientific LLC, San Carlos, California) and various interactions such as hydrogen bonding and van der Waal’s were identified. The PDB structures bound to drugs used in ART are referred to as ‘drug bound’ while the protease molecules bound to investigational compounds are referred to as ‘inhibitor bound’

The following criteria were used for hydrogen bonding and van der Waals interactions:

*Hydrogen bond interaction:*

Distance cutoff: 4 Å

Donor-Acceptor bond angle: > 90°

*Van der Waals interaction:*

Distance cutoff: Sum of van der Waals radii ± 0.5 Å
Table 5: GenBank ID of strains that were modelled for the Protease along with their percentage identity to clade C strains and mutations observed

<table>
<thead>
<tr>
<th>GenBank ID</th>
<th>Subtype (% identity*)</th>
<th>Major Mutations</th>
<th>Minor Mutations</th>
<th>Other Polymorphisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU030407</td>
<td>C (90.9%)</td>
<td>M46I, L90M</td>
<td>K43T</td>
<td>T12S, I15V, L19I, K20T, E35D, M36I, R41K, L63P, H69K, T74S, I93L</td>
</tr>
<tr>
<td>EU030412</td>
<td>C (92.9%)</td>
<td>N88NK</td>
<td></td>
<td>T12S, I15V, L19T, M36I, R41K, L63T, I64M, H69K, L89M, I93L</td>
</tr>
<tr>
<td>EU030413</td>
<td>C (93.9%)</td>
<td></td>
<td>L89LMV</td>
<td>T12S, K14R, I15V, L19I, M36I, R41K, H69K, I93L</td>
</tr>
<tr>
<td>EU030414</td>
<td>C (92.9%)</td>
<td></td>
<td></td>
<td>I13V, E35N, M36I, N37D, R41K, L63P, H69K, L89M, I93L</td>
</tr>
<tr>
<td>EU030415</td>
<td>C (93.3%)</td>
<td></td>
<td></td>
<td>T12S, I15V, L19I, M36V, R41K, L63P, H69K</td>
</tr>
<tr>
<td>EU030416</td>
<td>C</td>
<td></td>
<td></td>
<td>T12S, I15V, L19I, E35D, R41K,</td>
</tr>
</tbody>
</table>
Table 6: The HIV-1 protease templates obtained from the Protein Data Bank for construction of the study 3D models

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Drug/Inhibitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AQU</td>
<td>Atazanavir</td>
</tr>
<tr>
<td>1T3R</td>
<td>TMC 114</td>
</tr>
<tr>
<td>1MUI</td>
<td>Lopinavir</td>
</tr>
<tr>
<td>1C6Z</td>
<td>Saquinavir</td>
</tr>
<tr>
<td>1OHR</td>
<td>Indinavir</td>
</tr>
</tbody>
</table>

*based on information from the Stanford HIV Drug resistance database*
5.7 HIV-1 *env* (V3) gene amplification, sequencing, co-receptor prediction and susceptibility to co-receptor antagonists.

Forty HIV-1 infected treatment naïve individuals based on their CD4 counts (convenient sampling) from the original set of 101 individuals were taken for sequencing and co-receptor usage prediction study. The forty individuals were stratified into two groups based on their CD4 counts: one group (n=20) had CD4 counts less than or equal to 200 cells/mm³ while the other group included individuals whose CD4 counts were more than 200 cells/mm³. We also included 8 individuals who were failing ART. The majority of the study population consisted of South Indians (n=45, 93%) followed by individuals from East (n=3, 7%) India.

5.7.1 HIV-1 *env* (V3) gene amplification

The V3 region was amplified with the extracted RNA which was subjected to a one step RT-PCR (Qiagen GmbH, Hilden, Germany) along with specific forward and reverse primers, ED5 and ED12, respectively(316). (Custom synthesized at Sigma Aldrich, Bangalore, India). The list of primers and their sequences used are shown in Table 7. The cycling conditions were as follows: 50°C for 30 min, 95°C for 15 min followed by 3 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min followed by 32 cycles of 94°C for 15 sec, 55°C for 45 sec, 72°C for 1 min and a final extension of 72°C for 5 min.

The first round products were further amplified using Qiagen Hot Start Master Mix (Qiagen GmbH, Hilden, Germany) with forward and reverse primers, ED 31 and ED 33, respectively (316). The list of primers and their sequences used are shown in Table 7. The cycling
Table 7: Primer sequences used for HIV-1 *env* (V3) gene amplification

<table>
<thead>
<tr>
<th>First round:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ED 5</strong></td>
<td>5’ ATG GGA TCA AAG CCT AAA GCC ATG TG</td>
</tr>
<tr>
<td><strong>ED 12</strong></td>
<td>5’ AGT GCT TCC TGC TGC TCC CAA GAA CCC AAG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second round:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ED 31</strong></td>
<td>5’ CCT CAG CCA TTA CAC AGG CCT GTC CAA AG</td>
</tr>
<tr>
<td><strong>ED 33</strong></td>
<td>5’ TTA CAG TAG AAA AAT TCC CCT C</td>
</tr>
</tbody>
</table>
conditions were as follows: 95°C for 15 min followed by 3 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min followed by 32 cycles of 94°C for 15 sec, 55°C for 45 sec, 72°C for 1 min and a final extension of 72°C for 5 min. The pol gene of these strains was also amplified and sequenced as described in section 5.3.

The amplified products were run on 2% agarose (Sigma Aldrich Inc, St Louis, MO, USA) gel containing ethidium bromide to check for the specific 500 bp size amplicon. The agarose gel was visualized using the gel documentation system Gel Doc 2000 (BioRad, California, USA) using the software Quantity One version 4.1.1 (BioRad, California, USA).

5.7.2 HIV-1 env (V3) gene sequencing

The amplified products were sequenced and subsequently analyzed on the ABI PRISM 310 Genetic Analyzer (PE Applied Biosystems, California, USA). The electropherogram obtained was analyzed and consensus sequence was created using Sequence Navigator software (v 1.0.1, PE Applied Biosystems, California, USA). The sites of nucleotide ambiguity were marked as prescribed by the nomenclature committee of the international union of biochemistry (317). The sequences were aligned using ClustalW (http://www.ebi.ac.uk/Tools/clustalw/) and translated into all possible combinations of amino acid sequences using custom PERL scripts (provided by the kind courtesy of Prof Mark Jensen, formerly at department of Genetics and Epidemiology, University of Georgia Athens, Atlanta, USA). A total of 135 amino acids sequences were obtained and used for analysis as mentioned in section 3.7.4.
5.7.3 HIV-1 *env* (V3) gene sequence quality control

Phylogenetic analysis was performed for the forty eight nucleotide V3 sequences using Mega 4 software by the minimum evolution method with Kimura two-parameter and a bootstrap value of 500 replicates (307). The phylogenetic tree was constructed to confirm the subtyping analysis and also as a contamination control (quality control) for the sequencing reactions. These were compared to the HIV-1 V3 sequences which included 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) obtained from the GenBank.

The *pol* and *env* gene sequences obtained for all the study strains were submitted to REGA HIV-1 subtyping tool - v2.0 to identify the subtype of the strains ([http://dbpartners.stanford.edu/RegaSubtyping/](http://dbpartners.stanford.edu/RegaSubtyping/)) (306). The REGA HIV-1 subtyping tool cannot do recombination analyses on sequences less than 800 bp and hence the contiguous *pol* gene sequences of each of the strains were submitted for subtyping.

5.7.4 Prediction of co-receptor usage and susceptibility to co-receptor antagonists.

The nucleotide and the translated amino acid sequence obtained in section 3.6 were used for this part of the study. The 48 nucleotide sequences were then submitted to Geno2pheno [coreceptor] ([http://coreceptor.bioinf.mpi-inf.mpg.de/](http://coreceptor.bioinf.mpi-inf.mpg.de/)) and all the 135 amino acid sequences were submitted to Web PSSM ([http://ubik.microbiol.washington.edu/ computing/pssm/](http://ubik.microbiol.washington.edu/computing/pssm/)) to
determine the co-receptor usage of the HIV-1 strains. While using Geno2Pheno a 10% false positive rate was assumed as prescribed (False positive rate was defined as the probability of classifying an R5-virus falsely as X4). Along with submission of the V3 sequence of the strain, CD4 and viral load data were also submitted to Geno2Pheno for prediction of the co-receptor used by the strain. For Web PSSM, the clade C matrix was used, while for the one clade A strain in the analysis set both B and C matrices were used. We also observed the V3 crown motifs of all the strains (Amino acid positions 15-18 in the V3 region constitute the crown motif). The sequences were visualized using Weblogo (http://weblogo.berkeley.edu/logo.cgi).

5.8 Detail of strains used to construct the phylogenetic trees

All the nucleotide sequences of the strains used for construction of the phylogentic trees were downloaded from the Los Alamos data base and the Genbank.

5.8.1 Strains used for construction of phylogenetic tree using pol gene sequences of the study strains

i) Clade C strains from India

- AF286232: HIV-1 strain 98IN022, complete genome submitted in 2000
- AF067155: HIV-1 isolate 21068, complete genome submitted in 1998
- AF067157: HIV-1 isolate 301904, complete genome submitted in 1998
• AY713414: HIV-1 isolate 94IN_20635-4, complete genome submitted in 2004

• AY049708: HIV-1 isolate 01IN565.10, complete genome submitted in 2001

• DQ826595: HIV-1 isolate NARI-IN-RT3, pol protein (pol) and gag protein (gag) gene, partial cds, submitted in 2006

• DQ826669: HIV-1 isolate NARI-IN-NPT106, pol protein (pol) and gag protein (gag) gene, partial cds, submitted in 2006

ii) Non Indian clade C strains

• **AY967806 (China):** HIV-1 isolate YNRL9840, partial genome, submitted in 2005

• **AY255823 (Ethiopia):** HIV-1 isolate 99ET1, gag polyprotein precursor (gag) and gag-pol fusion polyprotein precursor, genes, partial cds; and vif protein (vif), vpr protein (vpr), tat protein (tat), truncated rev protein (rev), vpu protein (vpu), envelope glycoprotein precursor (env), and nef protein (nef) genes, complete cds, submitted in 2003

• **AB097871 (Myanmar):** HIV-1 proviral DNA, nearly complete genome, isolate: mIDU101, submitted in 2002

• **AF544009 (South Africa):** HIV-1 isolate DU151 subtype C from South Africa pol protein (pol) gene, partial cds, submitted in 2002
• **DQ093604 (South Africa):** HIV-1 isolate 04ZASK173B1 from South Africa, complete genome, submitted in 2005

5.8.2 Nucleotide sequences of strains used for construction of the phylogenetic tree using V3 gene sequences of the study strains along with their Genbank IDs

i) Clade C strains from India

- **D13427**: HIV-1 isolate HIV-1GT8, env gene for envelope protein, partial cds, submitted in 1992

- **D13424**: HIV-1 isolate HIV-1GT5, env gene for envelope protein, partial cds, submitted in 1992

- **DQ325336**: HIV-1 isolate cal 341, non-functional env gene, partial sequence, submitted in 2005

- **DQ325332**: HIV-1 isolate cal 262 env gene, partial sequence, submitted in 2005

- **DQ325314**: HIV-1 isolate cal 136 ,env gene, partial sequence, submitted in 2005

- **DQ149152**: HIV-1 isolate Drj 040, env gene, partial cds, submitted in 2005

- **DQ149147**: HIV-1 isolate Drj 036, env gene, partial cds, submitted in 2005
• **AY128271**: HIV-1 isolate IN97005 *env* gene, partial cds, submitted in 2002

• **AY128263**: HIV-1 isolate IN98009 from India *env* gene, partial cds, submitted in 2002

• **AJ292013**: HIV-1 proviral partial *env* gene, isolate Ind.Cal 1008, submitted in 2001

• **AJ292009**: HIV-1 proviral partial *env* gene, isolate Ind.Cal 126, submitted in 2001

ii) **Clade C strains from Africa**

• **L22956**: HIV-1 isolate ZAM20, proviral *env* gene, submitted in 1995

• **DQ382379**: HIV-1 isolate 05ZAFV27 from South Africa, *env* gene, complete cds, submitted in 2006

• **DQ382378**: HIV-1 isolate 05ZAFV26 from South Africa, *env* gene, complete cds, submitted in 2006

• **DQ382373**: HIV-1 isolate 05ZAFV15 from South Africa, *env* gene, complete cds, submitted in 2006

• **DQ382372**: HIV-1 isolate 05ZAFV14 from South Africa, *env* gene, complete cds, submitted in 2006

• **DQ382362**: HIV-1 isolate 05ZAFV3 from South Africa, *env* gene, complete cds, submitted in 2006
iii) Clade A Strain

- **AY350379**: HIV-1 isolate 070 from Italy *env* gene, partial cds, submitted in 2003

iv) Clade B strains

- **U08444**: HIV-1 isolate 593 clone 1 from Haiti, *env* gene, partial cds, submitted in 1994
- **U72495**: HIV-1, isolate MN, from USA, *env* gene, V3 region, partial cds, submitted in 1987
- **AF075720**: HIV-1 isolate MN clone MNE4 from the USA, *env* gene, complete cds, submitted in 1998

The consensus sequences for M group, A1, A2, B, C, D, AE, F1, F2, G and H were downloaded for the Los Alamos HIV database.

### 5.9 Details of Genbank submission

All the strains that were sequenced in the study were submitted to the Gen Bank. The GenBank IDs of the submitted sequences are shown in Appendix VI. It is important to submit the sequences to Genbank as they the HIV drug resistance public databases access this information to improve their algorithms for non-B subtypes especially for Indian clade C. Till
date, Stanford HIV drug resistance database (accessed on 30\textsuperscript{th} September 2008) has included 10 of the submitted sequences for upgrading the subtype C information.

5.10 Statistical Analysis

The results obtained on submitting the sequences to the three different algorithms were entered in MS Excel spreadsheet. Results on comparison were classified as ‘concordant’ or ‘disconcordant’. Results were considered as concordant when all the three algorithms denoted the same level of resistance in all forms of its output. Results were considered as discordant when at least one of the algorithms reported a level of resistance different from the other algorithms in any one of its outputs. 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 and median CD4 and viral load values of individuals belonging to group 1 and group 2 were estimated using the tools available in the MS excel worksheet.