7. Discussion

The findings embodied in this thesis have been obtained from 101 ART naïve individuals and 19 treated individuals who showed drug resistance as indicated by detectable (least viral load was $\log_{10} 3.51$ RNA copies/ml) viral load. The individual showing ART failure were on treatment for at least 8 months. These individuals were identified at routine follow up when CD4 and viral loads were monitored. In the case of ART naïve individuals the time of HIV infection was not ascertained. Majority of these individuals were asymptomatic. These individuals were detected as HIV infected either in our hospital or referred from outside. In CMC any HIV infected individual detected in the health care setting is referred to the infectious disease clinic (ID) which functions under the department of medicine. Referrals of HIV infected individuals from outside are also managed by the ID physicians. The senior ID physicians of CMC were trained at US institutions and the juniors train as part of the NACO physicians program. The graduate student (Abraham Joseph Kandathil) was trained in drug resistance genotypic assays at Brown University, Providence, RI, USA under Dr Bharat Ramratnam. The advanced *in-silico* work was carried out by the graduate student at the Indian Institute of Science, Bangalore, India under Prof N.Srinivasan of the Department of Biophysics.

7.1 HIV-1 *pol* gene sequence analysis

7.1.1 In ART naïve individuals (Group 1)

In the absence of any drug exposure, protease sequences from B and non-B HIV-1 subtypes were shown to be polymorphic at 30% of the protease gene coded amino acid positions (148). In the RT about 40% of the first 240 amino acids were polymorphic(148). Some of these amino acid substitutions may occur at high rates in non-subtype B viruses at positions associated with drug resistance in subtype B. Such positions are 10, 20, 36, 63, 71, 77 and 93
for Pr and 69, 75, 98,106,118 and 179 for RT(148). It is not clear whether this may reduce the activity of NRTIs, NNRTIs and PIs or if this is a basis for the evolution of different primary resistance and cross-resistance pattern (318).

All the 101 RT sequences were found to have amino acid substitutions at potions 35, 39,245 and 272. The M41L amino acid substitution was seen in the RT of one of the strains and is known to cause high level resistance when it occurs along with T215Y (272). When present alone it causes low level resistance to zidovudine and stavudine according to the Stanford HIV drug resistance database. The combination of mutations M41L, L210W and T215F are known as thymidine analog mutations (TAMs) (16). The strain which showed the M41L substitution in the RT also showed T74S substitution in its protease gene. This finding in the strain from a treatment naïve person indicates either spontaneous mutation resulting in substitution in the amino acids of the pol gene which has been previously reported to indicate resistance to two classes of drugs, an unusual finding and difficult to explain. One of our patients had the Y181C mutations which is a mutation that is most frequently selected by nevirapine (16). The Y181C mutation has been shown to reduce resistance to Zidovudine (319). Two other NNRTI resistance conferring mutations were seen in one patient each. One of them was A98G, which is known to confer low level of resistance to NNRTI (24). The strain with the A98G substitution also had L210F/L substitution. The L210W substitution is a part of the TAMs. The other was V179E which causes potential low level resistance to all the NNRTIs based on the scores given by the Stanford HIV drug resistance database. Six strains had amino acid substitution at position 101 of which two had K101K/N and two strains each showed K101Q and K101R. K101N is known to cause resistance to nevirapine and delavirdine (24) . However, K101Q/R does not affect susceptibility to NNRTIs based on the report from the Stanford drug resistance database.
Amino acid substitutions were also seen at positions conferring resistance. These substitutions were however different from those known to confer drug resistance in clade B e.g., L100FL (n=1), K103R (n=2), V179I (n=5). The L100I substitution confers intermediate resistance to the NNRTIs(16). The codons for phenylalanine are UUU, UUC while for isoleucine one of the codons is AUU. Thus, a change in one nucleotide in a specific codon (U → A) can lead to a phenylalanine by isoleucine substitution. This subsequent change in amino acid can lead to resistance to the NNRTIs. L100F could be an intermediate to L100I substitution in RT of this particular strain. The situation is analogous to the M184I intermediate change which appears before the emergence of M184V strains (279). The K103R substitution does not confer resistance by itself but in association with V179D/E can cause resistance to NNRTIs (320). Five percent of the reverse transcriptase sequenced showed substitutions at position 179 and 20% of this had V179E substitution. The K103R mutation was not seen in combination with V179D/E. We could cautiously assume that the requirement for multiple mutations required for the development of resistance (genetic barrier) would be lesser in the strain with these mutations.

The most common substitutions in the amino acid sequence of the HIV-1 protease molecule (99 amino acids) that were observed were at position 69 (100%) followed by positions 19 (96%), 36 (99%) and 93 (96%). In the case of reverse transcriptase (first 300 amino acids), the amino acid substitutions were observed at positions 35, 39, 245 and 272 in all the hundred and one strains. Ninety seven (96%) strains showed amino acid substitutions at positions 173 and 207. Six strains had mutation at position 101 (6%), two strains showed mutations at position 103 (2%), seven at 179 (7%) and one strain at position 181. When compared to subtype B, the protease sequences of the study strains varied at 28 positions i.e., frequency of
variation was 28%. The reverse transcriptase sequences in comparison had variations at 73 positions i.e., frequency of variation was 24%.

In our study, over 90% of the strains showed substitutions at positions 36, 63 and 93 in the protease amino acid sequence. It has been reported in clade B that the prevalence of these substitution increases in patients harboring strains with multiple PI resistance mutations (266). At position 36 of the protease gene 80% showed the M36I substitution and 16% showed the M36V substitution. Ode et al (2007) have shown that the M36I which is a non-active site mutation decreases the volume of the binding cavity of the protease enzyme (321). The M36V substitution also reduces the volume of the binding cavity but not to the extent that M36I causes. It is shown that when the M36I occurs along with the D30N in subtype B it reduces the interaction with nelfinavir (321). Similar studies on the significance of mutations at positions 63 and 93 have not been published. At this point of time we do not know which of these are likely to be responsible for treatment failure subsequently.

Reduced susceptibility to more than one PI is most likely to be associated with amino acid substitutions at six positions: 10, 46, 54, 82, 84 and 90 (322). Among our patients 11% had an amino acid change at position 82 (V82I). The V82I change has not been associated with any level of resistance to PIs (323), in contrast to V82 A/T/F/S. Reports from other centers in India have shown similar level of prevalence of this mutation (30, 31, 300). Amino acid change at position 10 of the protease was seen in 3% of the study sequence and this included the sequence from the clade A1 strain. According to the Stanford HIV drug resistance database the L10I/V, which were seen in our strains, is associated with resistance to each of the PIs when present with other mutations.
Data obtained by genotypic analyses of strains from different centers in India have shown several similar mutations (30, 31, 300, 301). In one cross sectional observational study from India (western India), wherein only phenotypic analysis was carried out on treatment-naïve strains, primary resistance to protease inhibitors was 2.5% and reverse transcriptase inhibitors was 6.7% (299). The authors have commented upon lack of information on the individuals which could give a clue to the appearance of resistance derived from them.

The well recognized E35D substitution was seen in 11% of the protease gene sequences in our study. This E35D substitution affects the conformational equilibrium between the closed and semi-open conformations of the free protease and also causes a significant reduction in its binding free energy for its substrate and amprenavir (47). It has also been shown that the E35D mutation reduces interaction with the HLA B44 molecule impeding cellular immune response (47). It is thus postulated to favor escape from the immune system in addition to conferring drug resistance (47). Based on the report from the Stanford HIV drug resistance database one of our strains had a E35G amino acid change, which according to the database is slightly more common in viruses from PI-treated (particularly nelfinavir-treated) compared with untreated persons. This particular strain also had a T74S amino acid change which is associated with reduced susceptibility to nelfinavir. In total, 3% of the sequenced strains had this amino acid change at position 74.

Based on data obtained from the Stanford HIV genotypic drug resistance database and REGA HIV-1 subtyping tool the 100 sequences belonged to clade C and one belonged to A1. Genotypic data for the strains from different regions of India including south have shown the predominance of subtype C (324).
While doing the genotyping based drug resistance assay we could postulate that the predominant circulating strain in a given patient (quasispecies) would have been amplified and sequenced (16, 325). It is reported that the drug resistance mutations could lower replication fitness of the virus (326). Hence, there is a biological possibility that a quasispecies in a mixed population in a given individual which has different mutations but with a lower degree of fitness may be not amplified. If certain mutations confer replicative advantage to the virus in the presence of the drug to which the virus has become resistance that particular quasispecies will outgrow the others (16). It has been reported that amplification is of the predominant quasispecies which constitutes 30% or more of the circulating virus population (327).

In countries where the standard initial and alternate antiretroviral regimens are restricted it is important to evaluate if transmitted resistance has reached a population level that could affect the effectiveness of ART. Studies from India have not revealed any high prevalence of drug resistance conferring mutation (30, 31, 34, 300, 301). These are still initial days in the treatment of HIV infected individuals in India. But with plans to scale up ART in India it is necessary to have a HIV drug resistance surveillance network as per WHO protocol. As per WHO protocol the drug resistance prevalence in a geographical area can be categorized into three categories <5%, 5-15% and >15% (328). Surveillance for drug resistance and use of efficacious combination prevent the transmission of drug resistant HIV-1 strains.

Based on our findings, we have observed no alarming presence of drug resistance mutations or prevalence of known amino acid substitution known to confer drug resistance on HIV strains in the treatment naïve population from India, reported earlier (30, 31, 34, 300, 301).
We however suggest the implementation of resistance testing to monitor treatment naïve population for early detection of transmitted drug resistance (Patient being infected with a previously resistant strain). In this study, we observed a Pr sequence with T74S and E35G substitution and a RT sequence with A98G, K101R and L210F/L substitutions. If strains with major mutations associated with resistance to drugs used routinely in the ART regimen of a country are found in two consecutive years with no evidence of previous treatment, sentinel surveillance should be considered as per the WHO recommendation (327). To the best of the candidate’s knowledge and the survey of the literature, India yet has no report of this problem. This will help to keep a tab on drug resistance. Further studies are needed to confirm if mutation causing resistance in clade B have the same effects/mechanism in clade C HIV-1 strains.

7.1.2 In ART experienced individuals (Group 2)

The RT amino acid sequence was found to have more drug resistance conferring mutations than the Pr amino acid sequence. This could be explained as the outcome of all patients having received at least one NRTI/NNRTIs containing regimen. The most common NRTI conferring mutation was M184V seen in 90% of the strains while 45% of the strains had the NNRTI conferring mutation K103N. Whereas, only 4 of the 19 patients had PI mutations as PI use is still limited in India. The M46I mutation was seen in 20% of the protease amino acid sequence.

Nineteen of the twenty one strains sequenced in this group had the M184V mutation. This mutation is selected by lamividune (3TC), emtricitabine (FTC) or abacavir (ABC) and confers high level resistance to 3TC (28). HIV-1 strains in patients receiving incompletely
suppressive 3TC regimens usually develop M184V as their first mutation (28). The other mutations observed included nucleotide excision mutations (NEM)/TAMS 41, 67, 70, 210, 215 and 219 (24). The most commonly occurring NEM was at position 215. The T215Y mutation was observed in 25% of the sequences while T215F was seen in 20%. These NEMs cause drastic reduction in susceptibility to azidothymidine (AZT). They do not however drastically affect the action of 3TC although they can affect the synergism of 3TC along with AZT (28).

The mutations conferring resistance to NNRTIs were also seen with one strain showing resistance only to NNRTIs. The NNRTI resistance conferring mutations included K103N (45%), Y181C (15%) and Y188L (20%). Additionally, the Y181I mutation conferring resistance to NNRTIs was seen in 5%. Only 5% of the sequences showed the Q151M mutation. The strain showing the Q151M substitution did not show any of the TAMs except for K70R. The strain also showed secondary mutations at positions 75, 77 and 116. The Q151M mutation is seen in patients receiving didanosine (ddI) in combination with AZT and stavudine (d4T) and is rarely seen in 3TC containing NRTI regimens (275, 284). As most first line ART regimens in India contain 3TC this may explain the low prevalence of Q151M in the study sequences. Previous publications from India have the absence or low prevalence of the Q151M mutation (33, 34). Based on data (n=100 subjects) from the West from previously published studies, M184V was the most commonly observed mutation in patients on NRTIs (28). Our strains from 19 patients have shown mutational patterns observed appear similar to published data (28, 33, 34). Minor mutations conferring resistance to NNRTIs were also observed and this included V90I, K101E and V106I. The associations between major mutations like TAMs (NEMs), M184V with minor mutations were looked for. We were unable to discern any identifiable pattern of association between the major and minor
mutations. Furthermore, the RT mutations were not restricted to any particular site domain of the RT molecule relative to its function.

Four patients showed resistance to PIs. With the exception of L90M the other mutations were observed near or in the flap region of the protease as deduced from 3D modelling. The flaps must be flexible in order to allow movement of the substrate and product (329). D30N was observed in one of the patients showing failure to PIs. D30N confers resistance to NFV only and does not show cross resistance to other PIs (24). The patient who had the D30N mutation also had M46I. The M46I mutation was however not observed in the sequence of the sample collected on the second visit which was two years later. The viral load on the second visit showed a log_{10} 0.43 reduction. This could be indicative of reduced viral fitness of the strain as indirectly indicated by reduced viral load. Alternately, there could have been amplification of a different quasispecies in the second sample. The M46I mutation contributes to reduced susceptibility to nelfinavir and fosamprenavir (330).

L90M mutation was observed in two sequences and was present along with M46I. L90M has been reported from patients treated with saquinavir, nelfinavir, indinavir and ritonavir (24). L90M has been implicated in clinical cross resistance to each of the PIs (24). Based on previously published literature, L90M was the most commonly seen mutations among infected individuals failing a PI regimen. In India PIs are not normally included in the first line ART because of its cost. The L89V mutation was seen in our treatment naïve population. Based on the Stanford drug resistance database the mutation contributes to darunavir (DRV) resistance which is however not used in the ART regimen in India. This
may possibly explain the lack of selection of this mutation in the treatment experienced group.

Based on clinical data available for the individuals in this group, majority of the individuals were classified in WHO clinical stage 1. This underlines the importance of periodic laboratory monitoring of individuals on ART using both CD4 counts and HIV-1 viral load estimation. This will benefit the individuals as it would help in regimen shifting before more deleterious mutations like the TAMs develop in the infecting HIV-1 strains.

Based on data obtained from the Stanford HIV genotypic drug resistance database all the submitted sequences belonged to clade C. This was also supplemented with phylogenetic analysis and results obtained from the REGA HIV-1 subtyping tool. This is consistent with genotypic data for the strains from different regions of India including south which have shown the predominant subtype to be clade C (324). The prevalence of major mutations observed in reverse transcriptase and protease sequences of ART naïve and ART experienced HIV-1 infected individuals were harmonized for those observed based on surveillance drug resistance mutations 2008 (SDRM) and international AIDS society 2007 (IAS-USA) respectively. The RT amino acid sequence in the treatment experienced group in our study had a higher prevalence of D67N (38%), K103N (43%), M184V (90%) and G190A (24%) than those seen in the HIV drug resistance Stanford database. This pattern of mutations may be an outcome of the ART regimen followed in India.
Based on our findings, we suggest the use of resistance testing at multiple centres in India to check the prevalence of the drug resistance mutation that arises following failure of the first line regimen and establish a national database like the Stanford HIV-1 drug resistance database. This will help in establishing national guidelines for second line regimens in India by NACO. Further studies are needed to confirm if mutation patterns that arise following failure of therapy for clade C is the same for clade B strains. As alluded to before when compared to subtype B, the protease sequences of the study strains (group 1) the frequency of variation was 28% and for the reverse transcriptase frequency of variation was 24%. This would help us know if the polymorphisms present in the RT and Pr of HIV-1 in clade C develop the same mutational pattern when compared to clade B under selective pressure of the ART regimen.

7.2 Comparison of HIV-1 drug resistance algorithms

The Stanford HIV Drug Resistance Database (SHDB) and the Geno2Pheno (G2P) databases differed in the sub-typing results. While the SHDB denoted all but one of the strains as clade C, the G2P sub-typed them as clade C (n=86), CRF 08_BC (n=29), clade D (n=4) clade H (n=2) and CRF 03_A0 (n=1). The sub-typing results of the SHDB showed 100% concordance with the Rega HIV-1 sub-typing tool and hence can be considered a reliable tool for sub-typing of HIV-1 strains based on pol sequence (Protease:1-99 amino acids and Reverse Transcriptase: 1-300 amino acids). The SHDB compares the sequence to a list of reference sequences for each of the HIV-1 M group and then assigns the submitted sequence the subtype of the closest reference amino acid sequence. The poor performance of G2P may be due to lack of experience with clade C strains. This could possibly be due to the smaller
number of non-clade B sequences that were used to develop the G2P software as compared to SHDB.

The databases varied in the amino acid change detected at positions 3 and 37 for the Pr and four amino acid positions 122, 214, 272 and 277 for RT. A predominant number of sequences that were submitted to G2P and DS showed a variation at amino acid positions 3 (V3I) and 37 (S37N) in the protease and amino acid position 214 (L214F) in the RT. Strains submitted to SHDB primarily showed variations at positions 122 (K122E), 272 (A272P) and 277 (K277R). These differences arise because of the sequence to which the three algorithms compare the submitted sequences to determine the mutation and polymorphisms present in the query sequence. The SHDB compares the submitted sequence to a consensus clade B sequence whereas G2P uses HXB2 (Genbank Id: K03455) as the reference strain for alignment of the submitted sequence. This could explain the consistent difference at positions 3 and 37 for the PR and amino acid positions 122,214,272 and 277 for RT. The difference in alignment strain used by the two databases however did not lead to non-recognition of any major drug resistance mutation. There are no published reports on the association of sites 3 and 37 for the Pr and positions 122, 214, 272 and 277 for RT with susceptibility to ART drugs and hence variation at these positions does not pose a problem in interpretation of results in the present scenario.

There were differences in drug resistance interpretation not only between databases but also within the same database in the case of G2P and DR_SEQAN (DS). The SHDB generated only one output while G2P and DS had 3 and 2 outputs, respectively. The G2P estimates phenotypic drug resistance data based on the submitted sequence and also uses two machine learning approaches viz., decision trees and SVM. The G2P output thus contains
interpretation based on these two systems. In addition G2P 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$_{50}$ based on reports from published literature in addition to genotypic resistance. The interpretation of IC$_{50}$ based on sequence data by DS hampered the mean percentage concordance (MPC) of DS when compared to SDHB. When the IC$_{50}$ values were however omitted and only genotypic output was used to calculate, the MPC was found to better between SDHB and G2P for all groups of drugs. The IC$_{50}$ interpretations are based on published literature associating mutations with phenotypic data. An update of this with more recent reports from published literature may increase the MPC of DS when compared to SHDB.

The number of drugs and drugs for which susceptibility patterns were interpreted were different for the three algorithms. While SHDB and G2P had interpretation patterns for 19 drugs G2P had this for 17 drugs. Susceptibility profiles to etravirine (ETR), tipranavir (TPV) and darunavir (DRV) were shown by SHDB only. The SHDB however does not interpret sensitivity to zalcitabine (ddC) and ritonavir (RTV). G2P does not give susceptibility profile for emtricitabine (EMT). In terms of susceptibility patterns of the most relevant drugs used in ART, SHDB was the most relevant followed by DS.

The three sequences which showed an error message of ‘overflow’ on submission to DS was sent to Prof Luis Menéndez-Arias, the curator of the DS software. In a personal communication received from him, he mentioned it was a possible problem with the software which had been rectified in the newer version of the software. On submission of the sequence to the newer version by Prof Luis Menéndez-Arias, to DS a result was obtained.
This software has not been released to the existing DR_SEQAN website. Hence, reanalysis was not done for all the study sequences except for these three sequences.

The presence of resistance mutations can greatly affect the efficacy of certain ART regimens (331). Drug resistance testing is strongly recommended following a regimen failure and also in the treatment of primary infection especially in areas where the transmission of drug resistance strains is high (36, 332). Hence, this study was done to compare the interpretation of the different algorithms to assess the suitability for clade C strains from India. A similar comparative study on different genotypic algorithms with non-clade B strains showed a higher inter algorithm discordance for predictions of protease inhibitor susceptibility than for reverse transcriptase inhibitor susceptibility among treatment naïve HIV-1 infected individuals(333). The SHDB database also makes use of non clade B strains to generate its output as per the release notes on the SHDB website. Hence, we have used the SHDB as a ‘gold standard’ for this comparison. We do not however have phenotypic data for the strains included in this study and hence not in a position to assign a particular algorithm as superior. The algorithms need to be updated regularly so as to increase their utility. These results also show a possible need for subtype specific algorithm for interpretation of HIV-1 genotypic drug resistance.

7.3 Construction of HIV-1 Reverse Transcriptase 3D models

The gene mutations resulting in amino acid substitutions studied for their possible roles in affecting drug susceptibilities were those that conferred resistance to NRTIs (V75M, V118I and Q151M) and NNRTIs (G190A, M230L K238T and Y318F). The mutations that affected
NRTI activity were located in the finger sub-domain of the RT as deduced in the 3D models. The finger sub-domain of the RT is composed of amino acids at position 1-85 and 118-155 (37). The NNRTIs resistance conferring mutations were not restricted to any particular sub-domain of the molecule. Three NNRTI mutations occurred in the thumb sub-domain (M230L, K238T and Y318F) and one mutation occurred in the palm sub-domain (K101H). The palm sub-domain of the RT is composed of amino acids at positions 86-117 and 156-237 and the thumb sub-domain is composed of amino acids at positions 238-318 (37).

The NRTIs bring about inhibition of HIV-1 replication by chain termination of the growing viral DNA strand (24). During extension of the viral strand by HIV-1 RT, the dNTPs that are to be added to the viral DNA have to interact with the finger sub-domains of the RT (102). These interactions help in appropriate positioning of the dNTPs and they also act as clamps to hold the dNTPs in place. Hence, mutations that confer resistance primarily occur in the finger sub-domain of the RT affect interactions with the dNTPs/drug (NRTIs). The mutations at positions 75,118 and 151 of the RT contributes to drug resistance due to template repositioning and discrimination between substrate and drug.

The NNRTIs like nevirapine inhibit RT action by binding to a hydrophobic pocket in RT (232). The inhibition is hence allostERIC in nature. The hydrophobic pocket of the RT to which the NNRTIs bind is located between the β6-β10-β9 and β12-β13-β14 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). Inhibition of RT by nevirapine occurs by two possible mechanisms. The drug has been postulated to either prevent the translocation of the template-primer complex following incorporation of a new nucleotide by restricting the movement of the
thumb or by altering the orientation of the carboxylate side chains involved in the catalysis by altering the conformation of $\beta6-\beta10-\beta9$ (37).

Mutations like K101H, M230L and K238T that were studied were postulated to affect NNRTI activity by affecting drug entry into the binding pocket and by altering the geometry of the pocket. These mutations cause a loss of interactions which plays a role in stabilizing the structure of RT under different conformations. The loss of these interactions can make the RT more flexible and/or alter the geometry of the hydrophobic pocket. Mutations like G190A and Y318F were seen to directly affect the drug binding into the hydrophobic pocket. In the case G190A mutation, the new amino acid that is present in the RT after the mutation prevents the entry of the NNRTIs into the pocket because of steric hindrances while in the case of Y318F there is a loss of van der Waals bond between the drug and RT.

The mutation studied were seen to affect drug efficacy by a number of ways such as altering the geometry of the drug binding pocket, perturbing the entry of the drug into the active site, repositioning the template bases and by discriminating the inhibitors from their natural substrates. These are similar to the mechanisms that are known for clade B strains(39, 334). The mutations that were analyzed in this study conferred resistance to NRTIs and NNRTIs. Drugs belonging to these two classes have different mechanisms of actions and hence the sites of mutations on RT would also be expected to be distinct. NRTI resistance analyzed in this study was mediated either by discrimination between inhibitors and natural substrate or by template repositioning, whereas, NNRTI resistance affected the geometry of the active site, drug binding and entry.
These plausible mechanisms of resistance that we have elucidated in clade C strains from India are similar to the mechanisms that are known for clade B strains (39, 334). Hence, although we have looked at mutation that have not been looked before in clade B there are no new mechanisms of resistance based on molecular modelling studies. There have been no reports of similar studies from India to date on HIV-1 RT (Pubmed search done on 5th Sept 2008). It is important to do further studies to look into mechanisms of resistance as this will aid in development of drugs which could play a role to control infections by strains resistant to presently used groups of NRTI/NNRTIs. The protease inhibitors (PIs) are examples of compounds that have been developed by structure-based rational design (44). This is more difficult in the case of NRTIs as these drugs have to be converted into their active form before they can inhibit the activity of HIV-1 RT (24). The HIV-1 RT is inhibited by the phosphorylated form of the NRTIs. Phosphorylation of NRTIs is brought about by cellular kinases present in the host cells. Hence, a structure-based design would be difficult for development of newer NRTIs. It could however be attempted to develop newer NNRTIs as these group of drugs bind directly to the hydrophobic pocket present in HIV-1 RT. Newer NNRTIs like TMC 125 have employed a structure based approach as part of its development strategy (335). The groups of drugs targeting RT are widely used in India and hence would need significant research effort to develop newer drugs.

Similar to HIV-1 protease, RT is also a dynamic enzyme but unlike in the case of the former the mutations do not greatly hamper the dynamics of the enzyme (42, 46). Resistance in RT can broadly be seen as changes in interactions at the active site of the enzyme which can in turn affect its behavior to the inhibitors. Newer groups of drugs in this class would need to take into account these mutations to be effective. RT does present itself as a potential target for controlling HIV infection (39). There are two copies of RT in a HIV virion and each
catalyzes about 20,000 nucleotide incorporations. Thus, there are approximately 5,000 chances for chain termination. We have studied only at a part of the protein and this may not reflect the behaviour of the entire enzyme. However, the study looked particularly at changes in interactions of the amino acids in and around the drug binding site. It is likely that our approach though useful has its limitation.

### 7.4 Construction of HIV-1 Protease 3D models

In this study we looked at the changes in interaction patterns between amino acid residues in HIV-1 protease. The effects of mutations occurring together and single mutations on the network of interactions were studied. Most of the polar amino acid residues in HIV-1 protease have flexible side chains. These residues can act in a concerted fashion by forming hydrogen bonds to stabilize the closed, semi-open and open conformations. Studies from a few centers in India have shown the protease subtype C to be highly polymorphic when compared to the subtype B protease amino acid sequence (30, 31).

Based on data from the Stanford HIV Drug Resistance Database (SDHB) the L90M is associated with cross resistance to PIs like NFV and SQV. The L90M mutation was found to disrupt the extent interactions that are observed between Lys at position 90 with the active site residues Asp at position 25 and Thr at position 26. This disruption leads to a change in the geometry of the active site and this may be responsible for the cross resistance. Ode et al has also described a disruption in the binding cavity of protease (336). The HIV-1 clade C strains appear to more often have the L90M mutation than D30N conferring resistance to NFV(337).
The D30N which is known to cause high level resistance to NFV (ref) was studied in conjunction with other mutations K45R, N88D and L89V as present in the patients sequenced strain (Gen Bank ID: EU030406). These changes cause a disruption in a network of hydrogen bonds and van der Waals contact as a result of which the protease is unable to remain in a closed conformation. The D30N mutation also causes a loss of the hydrogen bond that is present between Asp at position 30 and NFV. This results in D30N conferring high level resistance to NFV. This is as previously described by Ode et al (336).

Another set of mutations whose effects were looked together was the combination of T12S, E21D, L63P and H69K which was seen in a strain resistant to PIs. However, polymorphisms at positions 12, 63 and 69 occurred commonly among the individuals in study group 1 also. The prevalence of these polymorphisms among the study group 1 is shown in Table 9 of the results chapter. These changes in interactions that these amino acid substitutions cause within a protease molecule are similar to interactions observed in an unliganded protease molecule. As a result of these changes the PIs will be unable to bind optimally to the binding site on the protease molecule.

We also studied the effect of the M36I mutation in conjunction with I15V on the protease molecule. Previous analysis of molecular dynamics in non-clade C protease molecules have shown that M36I increased the flap flexibility and widened the binding cavity of the protease. The presence of both M36I and the I15V mutation results in the loss of interactions which favours the lowering the elbow region of the protease, which is observed during the transition from the closed to the open state.
The E35D mutation has been shown to affect the conformational ability of the free protease between the closed and the semi-open conformation and also causes a significant reduction in the binding free energy of the protease for its substrate and amprenavir. We studied E35N and E35K in addition to E35D. The E35N and E35K mutation was also found to increase the flap flexibility because of the loss of a salt bridge which would have existed between Glu at position 35 and Arg at position 57. We postulate that the E35K would further increase flap flexibility because of the repulsion between Lys at position 35 and Arg at position 57.

The K20T mutation was also associated with changes that are observed in an unliganded protease structure. The shortening of the side chain as a result of the change from Lys to Thr was postulated to aid in the lowering of the flap elbow resulting in an open conformation of the protease.

The drug resistance mutations are found to alter the interaction network that holds the protease in closed conformation. Water mediated interactions also help to stabilize the different side chain conformations. Mutations were seen to confer resistance by a number of ways. These changes included alterations in the geometry of the active site, weakening of interactions that stabilize the closed conformation and also affecting dimer interactions. Although clade C HIV-1 proteases are polymorphic when compared to clade B proteases our studies suggests a similar mechanism of resistance in clade C strains.

A number of mutations that were analyzed were found to increase the flexibility of the flaps, which in a drug bound protease acts as clamps to hold the inhibitor in place. These mutations
thus favour the partially open conformation by increasing the flexibility of the flaps. Thus, inhibitors with a very low dissociation constant will probably be more effective in treatment of HIV-1 individuals harbouring PI resistant strains. Residues like Gln 58, Asn 83, Asn 88 and Gln 92 along with their interactions with the neighbours are important for the transition from closed to the open conformations. These conformational structures dependent on certain amino acids can be potential targets for newer PIs. Drugs that interfere with dimer formation if administered with other inhibitors can make ART very effective since some of the resistance mutations also stabilize the interactions at the interface, which is required for the dimer stability when the flexibility is increased. These observations tailored with molecular dynamics simulations can help in the development of newer inhibitors to combat strains which have multiple mutations.

7.5 HIV-1 \textit{env} (V3) gene amplification, sequencing, co-receptor prediction and susceptibility to co-receptor antagonists.

Forty seven strains drawn from study groups 1 and 2 were identified as clade C and one strain as clade A1 based on \textit{pol} and \textit{env} sequence data. The majority of the study population strains were found to utilize CCR5 with discrepancies in the strains found to utilize CXCR4 between Geno2Pheno and Web PSSM. While Geno2Pheno identified 3 (7\%) strains of forty six (clade C) were predicted to utilize CXCR4 while the Web PSSM identified 2 such strains. One strain was identified as CXCR4 utilizing by both the methods. There was one CXCR4 strain from the ART failure group using Geno2Pheno which was however not detected using Web PSSM C matrix. One strain gave ambiguous results with Web PSSM because of the position of the amino acid deletion. The clade A strain also gave ambiguous results based on the matrix that was used. The clade A strain was not taken for analysis. Based on Web
PSSM, the mean positive charge of the CCR5 strains was 6 and the net charge 4.06 while for CXCR4 strains the mean positive charge was 6 and the mean net charge was 3.1. The X4 strains were found to have a lower net charge as compared to the clade C X4 strains from Africa (54).

The crown motif sequence was the same for both the R5 and the X4 strains as shown in Fig. 19 in the results section. Previous studies have shown that the crown motif (GPGQ) varies between X4 and R5 strains among clade C strains (338).

There are eight currently functional sequence based methods available to detect co-receptor usage (339). The Geno2Pheno is a support vector machine based method. In addition to sequence data it also makes use of additional parameters like CD4 counts and HIV-1 viral load to make the predictions. Based on published literature the Web PSSM was found to be one of the better tools to predict CXCR4 utilization for non clade B and clade C strains (340). The discordance with a phenotypic assay was 22.5% for determining CXCR4 utilization in clade C strains by Geno2Pheno and 6.3% using Web PSSM_simi (340). Literature defines discordance as strains classified as R5 by phenotypic assay and X4 by the bioinformatics tool (340). However, for the Web PSSM only the clade B matrix was used to determine co-receptor utilization of the non clade B strain in the study (340). The Web PSSM’s C matrix has shown good concordance with phenotypic assays for detecting X4 strains among clade C strains (338). Previous report from Africa has stated a 100% concordance of Web PSSM with a phenotypic assay in detecting CXCR4 strains among clade C strains (54). Hence, it would be appropriate to state that the strain reported is CXCR4 utilizing. This is the first report of clade C X4 strains from India. The X4 strains were identified in 1 female and 1
male mean viral load and CD4 counts in this group were $\log_{10} 5.70$ copies/ml and 135 cell/mm$^3$, respectively. Both these individuals were asymptomatic. The significance of the lower CD4 counts would be difficult to comment on as there were no follow samples from these patients. From this preliminary finding there is clear indication that more work needs to be done on identifying the role of X4 strains emerging among subtype C and its role in pathogenesis of the disease. This is an indication of the two decades long epidemic with subtype C resulting in the emergence of a more virulent strain (X4) in the HIV-1 infected population. In a study from the Africa there is evidence for a similar phenomenon (54). One of the limitations of this study is that we have not tested these strains with a phenotypic assay. We have used only genotypic tools due to lack of appropriate culture facilities for HIV propagation.

Previous reports have shown CCR5 tropic viruses are present early in infection and have shown to decrease with exposure to antiretrovirals (50, 253). An increased frequency of X4 strains were seen in ART treated patients. The findings are important with the recent introduction of a CCR5 antagonist for the treatment of HIV-1 infections. Hence, in such scenarios it would be imperative to determine the co-receptor utilized by the strains before administration of these classes of drugs. Strains resistant to maraviroc have been detected in patients elsewhere in whom they have been administrated. These resistant strains were CXCR4 strains which had emerged from a preexisting population of HIV-1 quasispecies in the patient (341). It would be important to monitor the scenario when this group of drugs is introduced in India.
7.6 HIV-1 drug resistance mutations in ART experienced individuals (study group 2) in relation to viral loads and CD4 counts.

Among the 21 strains from the treatment failure group it was observed that individuals who had the M184V as the sole NRTI resistance conferring mutation, without TAMs had a lower median HIV-1 viral load compared to the group which harboured TAMs in addition to the M184V mutation. Studies have shown an association of lower viral fitness in strains having the M184V mutation. Strains were found to have similar fitness to wild type strains in T-cell lines where nucleotide concentration is high but reduced fitness in macrophages and PBMC cells which have a limited nucleotide concentration (342-344). It has been proposed that the decreased processing capacity of the mutant RT in polymerization is due to M184V substitution and is aggravated due to low nucleotide concentrations (343). The M184V mutant strains were found to have a reduced fitness as measured by parallel cultures using both single cycle and multiple cycles as well as by growth competition assays using the multiple cycle technique (343-346). Some TAMs have been observed associated with an increase in viral fitness though we did not observe the previously described such TAMs.

The Q151M complex which consists of A62V, V75I, F77L, F116Y and Q151M confers resistance to all the drugs in the NRTI group except tenofovir based on data from SHDB. Multiple cycle growth competition assays have shown HIV-1 strains with the Q151M complex to have an improved viral fitness compared to the wild type strain (347). The Q151M mutation was found in only one individual in study group 2. It has been suggested that the reason behind the less frequent development of Q151M may be due to the lower fitness of the intermediate mutants (347, 348).
We have attempted to correlate fitness of the viral strains harbouring certain drug resistance conferring mutations with viral load levels and CD4 counts in these individuals but have not employed any fitness assays. The group with M184V mutation had a mean CD4 count of 289 cells/mm$^3$ with a median count of 329 cells/mm$^3$. Previous publications have shown an inverse relation between certain resistance mutations and CD4 counts. It was observed in a study on 273 patients from Spain that more than 75% of patients with multidrug resistance mutations had CD4 counts greater than 200 cells/µl and nearly 50% had viral load less than 10,000 RNA copies/ml. The authors infer that low viral load and high CD4 counts were possibly high due to lower viral fitness of the strains harbouring the mutations. They hence, suggest that the failing regimen should be maintained till a newer regimen can be administered (349).

Strains with reduced viral fitness compared to the wild type strains may not be transmitted effectively. One study observed that patients with primary HIV infection (PHI) had lower prevalence of M184V mutation in comparison to their prevalence in the potential transmitter population (PT) (350). The observed ratio of M184V mutation between PHI and PT was 0.14. The study suggested that decreased viremia and viral fitness due to M184V could affect the transmission of strains harbouring this mutation. This could possibly be one of the reasons behind the absence of the M184V mutation in study group 1 even though it was the most prevalent mutation in study group 2.

Though viral fitness assays were not done by us as alluded to earlier, the stable CD4 counts and low HIV-1 viral load in 5 patients with the RT showing M184V substitution in study
group 2 may be an indication of the same phenomenon that has been observed in the study from Spain (350).

Our findings have been discussed in the context of previously published literature. Interestingly we found many concordant findings between reports on clade B strains and clade C strains from Africa and India. There were certain unique features observed in clade C strains that we analyzed like: higher levels of polymorphisms in both RT and Pr amino acid sequence of clade C strains when compared to clade B strains, the first documentation of mutations in the protease gene conferring resistance to PIs in clade C from India, the first time description of the mechanism and significance of the E35K substitution in the Pr, predominant predicted CCR5 utilization by strains even in individuals with high viral loads and low CD4.