# Molecular Characterization of HIV-1 Subtype C Strains from KwaZulu-Natal, South Africa with a Special Emphasis on Viral Fitness and Drug Resistance

**Submitted in partial fulfilment of the requirements for the degree of Doctor of Philisophy in the Faculty of Medicine, University of KwaZulu-Natal, 2004**

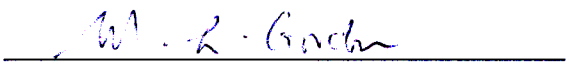
**By**

**Michelle Lucille Gordon**

**Promoter: Professor Sharon Cassol**

## Declaration

**I hereby declare that the whole thesis, unless specifically indicated to the contrary, is my own original work and has not been submitted for a degree at any other university.**



**Michelle Lucille Gordon**
**2004**

## Acknowledgements

I would like to acknowledge the following people for their assistance and support:

1. Professor Sharon Cassol, my supervisor and mentor
2. Professor Alan Smith, head of the Department of Virology, University of KwaZulu-Natal
3. Mr Tulio de Oliveira for introducing me to the field of bioinformatics and for his assistance, guidance and encouragement
4. Miss Karen Bishop, Miss Taryn Page and Miss Siva Danaviah, who were with me from the beginning of this project
5. Mr George Irving for the data capturing
6. Ms Zainub Bhorat for her efficiency in the office
7. The staff of the Molecular Virology and Bioinformatics Unit, Africa Centre
8. Mr Ryan Doherty for his assistance and encouragement and for being my tour-buddy in Belgium
9. Mrs Lorna Madurai for her continued encouragement
10. Mr Luiz-Felipe Gonzales for his assistance and phenotpying training
11. Mr Alex Calazens for his assistance and phenotpying training
12. Miss Kristel van Laethem for assisting with the interpretation of the resistance profiles
13. Ms Annemike van Damme for allowing me to spend time in her laboratory
14. Professor Amilcar Tanuri for allowing me to spend time in his laboratory and for his encouragement and support
15. Conceicao and family
16. The staff at the Sinikithemba Clinic, McCords Hospital, especially Dr Jane Hampton and Dr Janet Giddy.
17. Miss Deidre La Douce for always being there to help
18. Mrs Giselle La Douce for her critique of this manuscript
19. My family for their continued support
20. Mr Kevin Gordon, my loving and patient husband

**This thesis is lovingly dedicated to my family - past, present and future.**

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| 3TC | lamivudine; Epivir® |
| ABC | abacavir; Ziagen® |
| AIDS | acquired immune defiiciency syndrome |
| APV | amprenavir; Agenerase™ |
| ART | antiretroviral therapy |
| ARV | antiretroviral |
| AZT | zidovudine; Retrovir® |
| AZT | zidovudine |
| CA | capsid |
| bp | base pair |
| CBV | Combivir® (AZT+3TC) |
| cDNA | copy of deoxyribonucleic acid |
| cpm | copies per mililitre |
| CT | Cape Town |
| d4T | stavudine; Zerit® |
| ddC | zalcitabine; HIVID® |
| ddI | didanosine; Videx® |
| DLV | delavirdine; Rescriptor® |
| DNA | deoxribonucleic acid |
| dNTP | deoxyribonucleotide triphosphate |
| EFV | efavirenz; Sustiva™ |
| *env* | envelope |
| *gag* | group antigen |
| gp | glycoprotein |
| HAART | highly active antiretroviral therapy |
| HIV | Human immunodeficiency virus |
| HLA | human leukocyte antigen |
| IDV | indinavir; Crixivan® |
| INV | invirase; Fortovase™ |
| KS | Kaposis Sarcoma |
| KZN | KwaZulu-Natal |
| LPV/r | lopinavir; Kaletra™; ABT378/r |
| LRT | likelihood ratio test |
| LTR | Long terminal repeat |
| MA | Matrix |
| ML | maximum likelihood |
| MRCA | Most Recent Common Ancestor |
| MTCT | mother-to-child transmission |
| MTT | 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide |
| NC | nucleic acid binding protein |
| NFV | nelfinavir; Viracept® |
| NJ | neighbour joining |
| NNRTIs | non-nucleoside reverse transcriptase inhibitors |
| NRTIs | nucleoside/nucleotide reverse transcriptase inhibitors |
| nt | nucleotides |
| NVP | nevirapine; Viramune® |
| PCR | polymerase chain reaction |
| PIs | protease inhibitors |
| pMTCT | prevention of mother-to-child transmission |
| *pol* | polymerase |
| PR | protease |

| | |
|---|---|
| RVA | Recombinant Virus Assay |
| RT | reverse transcriptase |
| RTIs | reverse transcriptase inhibitors |
| RTV | ritonavir; Norvir™ |
| SIV | simian immunodeficiency virus |
| SQV | saquinavir; |
| TAMS | Thymidine-analogue Mutations |
| TB | Tuberculosis |
| TDF | tenofovir; Viread® |
| TDM | Therapeutic Drug Monitoring |
| TZV | Trizivir® (ABC+AZT+3TC) |
| US | United States of America |
| *vif* | virion infectivity factor |
| *vpr* | virion protein regulatory |
| *vpu* | viral protein unknown |
| WHO | World Health Organisation |

# Presentations and Publications Arising from this Study

## Publications


1. De Oliveira T, Salemi M, Gordon M, Vandamme A, Janse van Rensburg E,' Engelbrecht S, and Cassol S. Codon-Substitutions Models as Indicators of Positive Selection and Amino Acid Diversification: An Alternative Approach to HIV-1Vaccine Development? Genetics. 2004, in press.
2. Gonzales LJM, Brindeiro RM, Tarin M, Calazans A, Soares MA, Cassol S, Tanuri A. *In vitro* hyper-susceptibility of human immunodeficiency virus type 1 subtype C protease to Lopinavir. Antimicrob Agents Chemother 2003; 47:2817-22.
3. De Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, Zur Megede J, Barnett SW, Cassol S. Variability at HIV-1 subtype C protease cleavage sites: An indication of viral fitness? J Virol 2003; 77:9422-30.
4. Chelule PK, Mosam A, Gordon M, Palanee T, Page T, Coovadia HM, Cassol S. MDR1 and CYP3A4 polymorphisms among African, Indian and white populations in KwaZulu-Natal, South Africa. Clinical Pharmacology Therapeutics 2003; 74:195-6.
5. Gordon M, De Oliveira T, Bishop K, Coovadia HM, Madurai L, Engelbrecht S, Janse van Rensburg E, Mosam A, Smith A, Cassol S.Molecular characteristics of HIV Type-1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. J Virol. 2003; 77:2587-99.
6. De Oliveira T, Miller R, Tarin M, Cassol S. An integrated genetic data environment (GDE)-based LINUX interface for the analysis of HIV-1 and other microbial sequences. Bioinformatics. 2002; 19:1-2.

**Abstracts presented or accepted**

1. Gordon M, Graham N, Ruth Bland, Nigel Rollins, Claassen M, Coovadia HM, Bennish M, Cassol S. Resistance patterns in mother-infant pairs following single dose nevirapine (NVP) for the prevention of mother-to-child transmission (MTCT) of HIV-1. XIV International AIDS Conference, Bangkok, Thailand, July 12-16, 2004.
2. Gordon M, Graham N, Van Laethem K, Giddy J, Hampton J, Bishop K and Cassol S Surveillance of Antiretroviral Drug Resistance in a single HIV Clinic in KwaZulu-Natal (KZN) South Africa. XIV International AIDS Conference, Bangkok, Thailand, July 12-16, 2004.
3. Gordon M, Graham N, Bland R, Rollins N, De Oliveira T, Monosi B, Van Laethem K, Vandamme A, Cassol S. Surveillance of Resistance in KZN South Africa, including mother-infant pairs 6 weeks after single dose NVP. XIII International Drug Resistance Workshop, Tenerife, Spain, June 8-12, 2004.
4. De Oliveira T, Salemi M, Gordon M, Vandamme A, Cassol S. Evolutionary evidence that HIV-1 Tat and Rev are targets of strong positive selection: new insights for designing an AIDS vaccine? Oral Presentation at 10th International Workshop on HIV Dynamics and Evolution. UCLA Conference Center, Lake Arrowhead, California, April 13th-16th, 2003.
5. De Oliveira T, Miller R, Gordon M, Cassol S. An integrated Genetic Data Environment (GDE)-based LINUX interface for analysis of HIV-1 and other microbial sequences. Poster presentation at 10th International Workshop on HIV Dynamics and Evolution. UCLA Conference Center, Lake Arrowhead, California, April 13-16, 2003.
6. Chelule PK, Mosam A, Gordon M, Palanee T, Page T, Coovadia HM, Cassol S. Preparing for HIV-1 therapy in South Africa: Will host polymorphisms in MDR1 and CYP3A4 influence therapeutic outcome? 2nd IAS Conference on HIV Pathogenesis and Treatment. Paris, July 13th-16th, 2003.
7. Gonzalez LMF, Brindeiro R"M, Tarin M, Calazans A, Soares MA, Cassol S, Tanuri A. In vitro susceptibility to protease inhibitor from subtype C of human immunodeficiency virus type 1 isolates. 2nd IAS Conference on HIV Pathogenesis and Treatment. Paris, France, July 13th-16th, 2003.
8. Gordon M, De Oliveira T, Darder M, Bishop K, Cassol S. Resistance patterns of HIV-1 subtype C in South African patients failing highly active antiretroviral therapy (HAART). South Africa AIDS Conference, International Convention Centre, Durban, South Africa, August 3-6, 2003.

9. De Oliveira T, Salemi M, Tarin M, VanDamme AM, Cassol S. Detecting positive selection sites in the HIV subtype C complete genome. Annals of the 9th HIV Dynamics and Molecular Evolution Conference, Lake Arrowhead, CA, United States, 2002.
10. Tarin M, Oliveiro T, Bishop K, Smith A, Madurai L, Cassol S. Characterization of the reverse transcriptase (PR) and protease (PR) genes of treatment-naïve patients from KwaZulu-Natal (KZN) South Africa. XIV International AIDS Conference, Barcelona, Spain, July 7-12, 2002.
11. Tarin M, De Oliveira T, Bishop K, Cassol S. Africa Centre for Health and Population Studies, Nelson R. Mandela School of Medicine, University of Natal, Durban, South Africa. Comparison of Retrospective and Recent Reverse Transcriptase (RT) and Protease (PR) Sequences from Treatment-Naïve Patients in KwaZulu-Natal (KZN) and Cape Town, South Africa. XI International Drug Resistance Workshop, Seville, Spain, July 2-5, 2002.
12. De Oliveira T, Bishop K, Tarin M, Engelbert S, Cassol S. Detecting adaptive molecular evolution in HIV-1 subtype C complete genomes. 9th International Workshop on HIV Dynamics and Evolution, UCLA, Los Angeles, March 17-20.

# Abstract

As South Africa begins its National HIV-1 treatment program, it is urgent that we collect data that will help define the phylogenetic relationships, transmissibility and drug responsiveness of C viruses. In this thesis, data is presented on the genetic diversity of locally circulating drug naive subtype C strains, as an indication of their natural susceptibility to antiretroviral drugs, prior to the national roll-out of antiretroviral therapy. At the time this thesis was initiated, antiretroviral therapy was only available in South Africa in a few clinical trials and in the private sector, and it was therefore difficult to obtain large numbers of samples from treatment-experienced patients. Nevertheless, valuable information on the prevalence and patterns of resistance mutations in subtype C infected patients was obtained from small studies on patients receiving HAART, concomitant HAART and TB treatment, HAART and treatment for Kaposi Sarcoma, and single dose nevirapine for the prevention of mother-to-child transmission of HIV-1 infection.

The results show that the general antiretroviral drug naive population do not harbour any major resistance-associated mutations to the currently available protease and reverse transcriptase inhibitors, with no differences in genetic variation between the different ethnic groups infected with subtype C. Phenotyping of some of these isolates showed that they were susceptible to the available protease and reverse transcriptase inhibitors, and hyper-susceptible to the protease inhibitor, Lopinavir. Phylogenetic analysis of recent and retrospective subtype C isolates showed that there are multiple lineages of subtype C viruses circulating in South Africa, indicative of multiple introductions of subtype C across its many borders. Polymorphisms in the protease, reverse transcriptase and C2-V5 region of envelope in these drug naive samples lead to significant variation in the number, type and location of potential phosphorylation sites. There was also variation in the cleavage sites controlling the initiation and rate of Gag and Gag-Pol processing (p2/NC) and the activation of protease (TFP/p6*gag*) suggesting that there may be important differences in the way that B and C viruses regulate polyprocessing and virion assembly. Similar to studies on subtype B, 10 to 18% of the patients on HAART developed drug resistance. However, those on concomitant HAART and TB treatment developed resistance as early as one month after starting treatment. Generally, the resistance mutations that were seen were consistent with those seen in treatment experienced subtype B isolates. Of note was the high level of resistance to the entire class of NNRTIs. This could be reflective of the predominant use of NNRTI-based regimens, as well as the low

genetic barrier in this class of drugs. The NNRTI mutations included the V106M mutation that is considered a signature mutation of EFV experienced subtype C isolates. Resistance was high (40%) in mothers and infants 6 weeks after each received a single dose of NVP. K103N was most common mutation in the mothers, while Y181C was most common in the infants. Of note were the changes in functional properties caused by these mutations, by the introduction or alteration of putative myristoylation and phosphorylation sites in the RT.

Taken together, these data suggests that the pattern of resistance in African patients will be similar to that observed for the treatment of subtype B infection. However, patients should be closely monitored for viral rebound very early on in treatment. Also, given the high rate of resistance in mothers and infants after single dose NVP, the search for safer regimens to prevent MTCT should be intensified. Although the mechanisms are unknown, our results indicate that several of the phosphorylation-related substitutions in the *pol* and *env* genes of KZN and other C viruses are highly conserved and positively selected. It will be important to determine whether these sites play an important role in the replicative capacity and proteolytic processing of C viruses, and in viral entry. These data provide important benefits for public health policy and planning and for future patient treatment management.

# Chapter 1

## Review of the Literature

## 1.1 HIV-1 Discovery and Classification

The acquired immunodeficiency syndrome (AIDS) was first recognized in 1981 among homosexual men in the United States of America. It was only in 1983 that the organism later termed the human immunodeficiency virus (HIV) by the International Committee of Viral Taxonomy, was isolated by Barre-Sinoussi et al (1983). Soon after the discovery of HIV, the significant genetic variability of the virus evoked intense interest. An increase in the number of isolates from around the world, coupled with the vaccine initiative in the early 1990s, led to the discovery of equidistant clades (or subtypes), which have since been classified as subtypes A, A2, B, C, D, F1, F2, G, H, J, and K in addition to circulating recombinant forms (CRFs) (Novitsky et al, 2001; Janssens et al 1997; Myers et al, 1997; Louwagie et al, 1995). These subtypes form part of Group M (Major) viruses and are responsible for the majority of HIV-1 infections in the world. Other HIV-1 sequences can be classified either as O (Outlier), and N (Non-M and Non-O) groups (Gurtler et al, 1994; Myers et al, 1997; Robertson et al, 2000; Simon et al, 1998). It is thought that HIV entered the human population in at least seven different simian zoonosis events in West Central Africa (Hahn, 2000). HIV-1 shares the closest phylogenetic relationship with the SIVcpz strain found in a West Africa chimpanzee subspecies called *Pan troglodytes troglodytes* (Rambaut et al, 2004). The countries in West Central Africa (Gabone, Cameroon and Equatorial Guinea) had little contact with the outside world until after World War II when economic changes brought more travellers to Africa. This was probably when the disease started its slow spread around the world with the earliest evidence of HIV-1 infection found in the Congo in 1959 (Zhu et al, 1998).

## 1.2 Global Subtype Distribution

HIV-1 subtype B is the most prevalent subtype in North, Central and South America along with subtype F, B/F recombinants, and to a lesser extent -but ever increasing -, subtype C (Figure 1) (Spira et al, 2003). In Europe, subtype B is found in the west, while Central Europe is dominated by subtype G and A/G recombinants (Korber et al, 2000; Esu-Williams et al, 1997; Janssens et al, 1997). In Eastern Europe, subtype A featured early in the epidemic (1995/1996), while subtype B spread among the injecting drug users (IDUs). Both have now re-combined into a new CRF (A/B), which is spreading eastwards in that region (UNAIDS, 2002). The Middle East, India, China and Nepal are dominated by subtype C, while A/E recombinants appear in China, Thailand and the Philippines. Africa as a whole has all known HIV-1 subtypes.

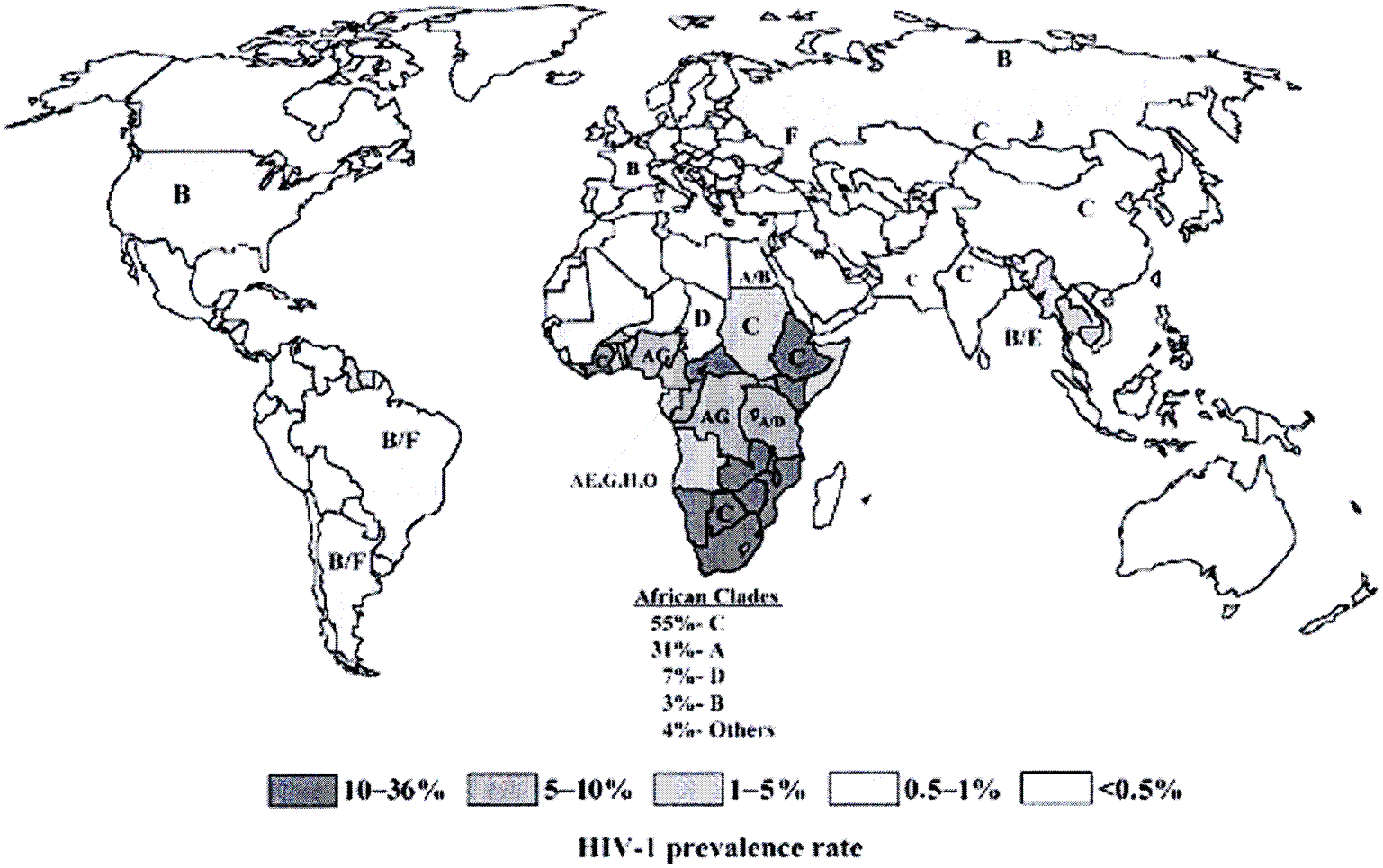



**Figure 1. Subtype diversity of HIV-1 infections prevalent world-wide (Spira et al, 2003)**

Central Africa has the largest variety of subtypes (A, D, G, F, H, A/G, A/E), which is indicative of a long-standing epidemic. In eastern Africa, subtypes A and D predominate, although subtype C is also spreading upward from the south and could impact on the incidence in the east in the future (Spira et al, 2003). West Africa is dominated by subtype A, G and more recently, A/G recombinants (over 30% of new infections), with sporadic occurrences of subtype D. Western Africa has a less severe epidemic, with a relatively modest presence in Senegal, Mauritania and Cape Verde. North Africa has very little HIV-1 infection, mostly subtype B, with prevalences below 1% in Morocco, Algeria, Libya and Egypt. Southern Africa is dominated by subtype C (Janssens et al, 1997), with sporadic cases of subtype A, B and D.

### 1.3 The Spread of Subtype C

Since subtype C began its devastating spread across southern Africa in the late 1980s, major outbreaks have now occurred in every country in this region. It is also responsible for the rapidly expanding epidemic in India, and is increasing in frequency in China and Brazil. In fact, of the 5 million new HIV-1 infections in 2002, subtype C accounted for >56% of these infections (Esparza et al, 2000). This is also influenced by the fact that most new infections are from heterosexual contact, which is the main route of transmission of HIV-1 subtype C.

#### 1.3.1 Africa

Earlier in the epidemic (late 1970s), HIV was not as well established in southern Africa as it was in central and eastern Africa, with the rate of spread being most rapid in east Africa (Hawlan, 2000). This began to change in the 1980s, when subtype C began its spread across southern Africa, starting in Tanzania, Zambia and Zimbabwe (Figure 2). In Tanzania, the first cases of HIV-1 were reported in 1983 (Mhalu, 1989). Early studies reported subtypes A and D as the predominant subtypes in this country (Holm-Hansen et al, 1996), while later reports have found equal proportions of A, C and D (Vidal et al, 2000). Since then there has been a dramatic increase in the HIV prevalence, with a concomitant increase in the prevalence of subtype C (12.7% in 2000 to 20% in 2003), particularly in the regions that border near Malawi and Zambia (Kapiga et al, 2002; Hoelscher et al, 2001; Renjifo et al, 1998). The interaction between Tanzania, Zambia and Malawi has played a role in the spread of subtype C in this region, with >93% of infections in Zambia due to subtype C (Handema et al, 2001; Morison et al, 2001; Trask et al, 2002). The earliest sequenced isolate of subtype C in the public database originates from Malawi. Malawi is one of the

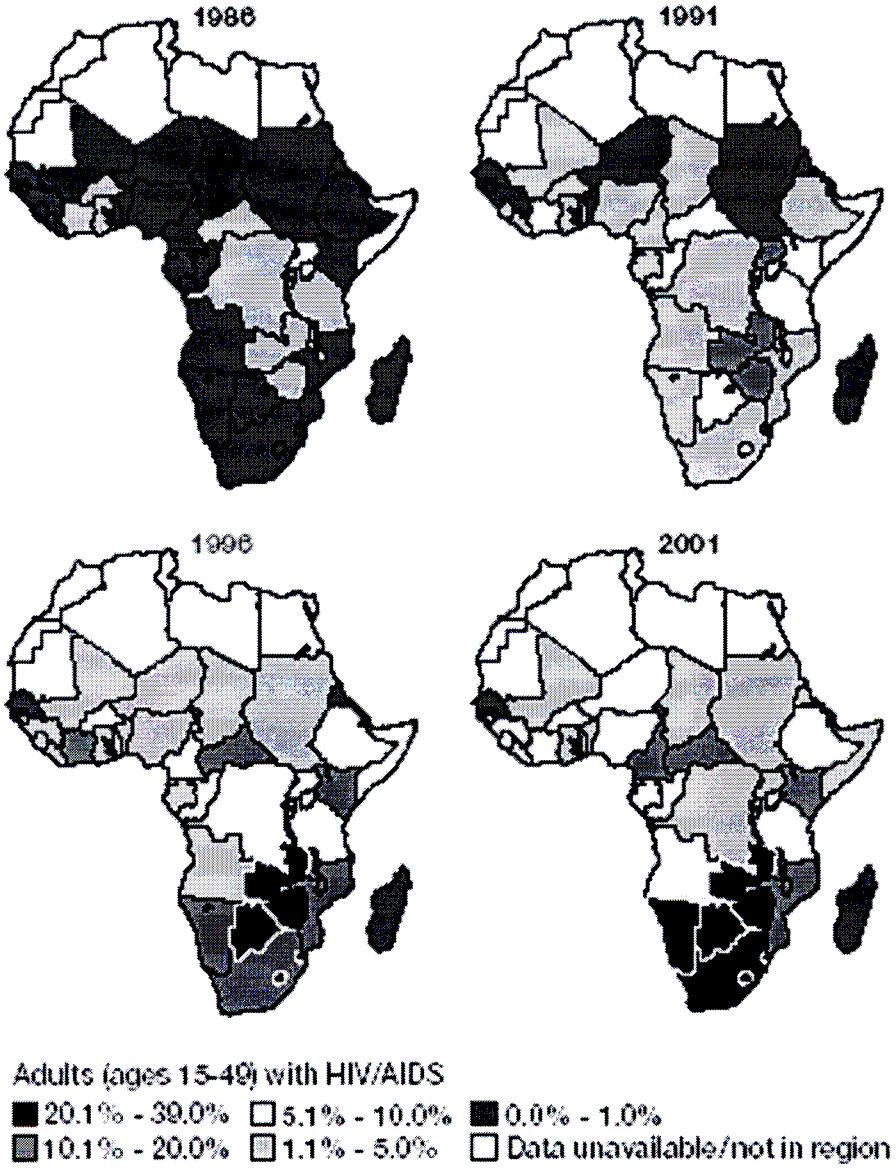



**Figure 2. HIV prevalence in Sub-Saharan Africa (UNAIDS, 2002)**

few regions that have population-based data derived from two detailed surveys carried out in northern Malawi in 1981 to 1984 and in 1986 to 1989 as part of a study on leprosy and other mycobacterial infections (Ponninghaus et al, 1993 and 1987). McCormack et al (2002) were able to subtype these samples. Their findings show that in 1982 to 1984, subtypes A, D and C were all present. Overall, subtype C accounted for 55% of the early samples (6/11). The next wave of sampling (1986 to 1989) showed an increase in the number of subtype C isolates, which corresponded to 90% of the cases, with non-C subtypes being more common in the north, bordering on Tanzania. They also found four recombinants: two AC, one AD and one DC. It is thought that the construction of a main road during the 1980s between Malawi and Tanzania could have influenced the spread of subtypes between the two countries (McCormack et al, 2002).

By the early to mid 1990s, the rapid expansion of the HIV-1 epidemic caused by subtype C had reached Botswana. HIV prevalence among antenatal clinic attendees tested in the major urban areas of Botswana (Gaborone, Francistown, and Selebi-Phikwe) increased from 6% in 1990 to 39% in 1997 (range of 34 to 43%) (Hensle, 1998). Outside of the major urban areas, median HIV prevalence increased from no evidence of infection in 1985 to 34% in 1997. To date, all isolates from Botswana have been identified as subtype C (Rodenburg et al, 2001).

Along with Botswana, South Africa, Zimbabwe, Namibia, Swaziland, Mozambique and Lesotho have also borne the brunt of the subtype C epidemic (Essex 1999; Bredell et al, 1998; Janssens et al, 1997). Prevalences have increased in Namibia from 7% in 1994 to 31% in 2000. Information on HIV prevalence among antenatal clinic attendees has shown a dramatic increase from 3.9% in 1992 to 34.2% in 2000 for Swaziland, and from 5.1% in 1991 to 42.2% in 2000 for Lesotho (UNAIDS, 2000). Similarly, prevalences in Mozambique have increased from less than 1% in 1988 to 13.2 % in 2000. In South Africa, only 166 cases of AIDS had been reported between 1982 and 1988 (Sher et al, 1989). However, by 1997 there was already a dramatic increase in HIV prevalence although at unequal rates throughout the country. The highest prevalence was found in KwaZulu Natal (northeast region) with a prevalence of 26.92%. In contrast, the prevalence rates in the Western Cape were much lower with a prevalence of only 6.29%. Of the isolates circulating in South Africa, 91% were subtype C (Van Harmelen, 1999). More recent reports from South Africa have shown the prevalence rates to be as high as 36.2% in KwaZulu Natal, while still a moderate 8.7% in the Western Cape (Rollins et al, 2002). There has also been an increase in the number of non-C subtypes detected in South Africa (subtypes A, D, G and A/G), mostly in individuals from other countries, however, some of these were found in combination with C. This could also be indicative of local

recombination (Bredell et al, 2002). Subtype C is the predominant subtype in KZN, with only one C/D recombinant reported (Gordon et al, 2003).

In addition to its spread across the southern parts of Africa, subtype C is also prevalent in Ethiopia, East Africa. Subtype C has been the predominant strain in the Ethiopian epidemic from the beginning, with 98% of samples isolated harbouring subtype C viruses (Abebe et al, 2001; Sherefa et al, 1994). The first case of HIV-1 infection in Ethiopia was registered in 1984 in the capital city Addis Ababa (Tsega et al, 1988; Zewdie et al, 1988). Prevalences began to increase from 11-13% in 1991, to 18% in 1996 (Fontanet et al, 1998; National AIDS program 1994). However, this figure is starting to decline as the prevalence rates of 15.1% in 2001 show. There are few studies describing the epidemiology of AIDS in Burundi, also in the east of Africa. Subtype C is also the dominant subtype in this region with a prevalence of 89% (Koch et al, in 2001). In Kenya, HIV-1 infection was first reported in the mid-1980s in a group of Nairobi sex workers (Kreiss, 1986; Obel 1984). Kenya also experienced a dramatic increase in HIV-1 prevalence during the 1990s, from 8.3% among women attending antenatal clinics in 1994, to 11% in 1997. Although subtype A is the major circulating subtype in Kenya, subtype C is also present in a significant minority (6.9% among breast-feeding women) (Neilson et al, 1999; Yang et al, 2004).

In Uganda, the first HIV/AIDS case was identified in 1982 (UNAIDS, 2001). Uganda originally had one of the highest prevalences of HIV-1 infection in the world. In rural regions of southwestern Uganda, overall seroprevalence rates had been reported to be 8.2% (1989) and 13% (1991) and were even higher in urban areas: 28.1% (1989) 40.4% (1991). However, Uganda was the first African country to have substantially decreased the effects of the HIV pandemic, with prevalences decreasing to 11.25% in 2000 (UNAIDS, 2001). Several studies have shown subtype A and D to be in similar proportion in Uganda and comprise the major components of the epidemic in that region (Hu et al, 1999; Brennan et al, 1997; Rayfield et al, 1998). Subtype C is still playing only a minor role in the epidemic in Uganda, accounting for only 3-8% of the isolates (Downing et al, 2000; Brennan et al, 1997). D/C recombinants have recently been isolated from Uganda (Wilbe et al, 2003).

### 1.3.2 Southeast Asia

The HIV-1 epidemic first appeared in Southeast Asia in the late 1980s. In India, the epidemic appeared to begin in 1986, when six female sex workers were found positive for HIV antibodies in Tamil Nadu, in south India. (Seth and Sharma, 1991). Since then the epidemic has spread rapidly through every region of the country (Srikanth et al, 1997). Early studies in India identified subtype C as the major circulating subtype among female

sex workers and blood donors, along with the sporadic presence of subtypes A and ThaiB among the IDUs of Manipur in the northeastern part of India (Chakrabarti et al, 2000; Maitra et al, 1999; Gadkari et al, 1998; Jameel et al, 1995; Tsuscie et al, 1995; Baskar et al, 1994; Grez et al, 1994). The ThaiB strain may have been introduced independently into this region through drug-trafficking links with southeast Asian countries such as Myanmar and Thailand (Maitra et al, 1999; Sarkar et al, 1993). It is not known exactly when subtype C was introduced into this population of IDUs, however, there was a definite increase in prevalence relative to the ThaiB strain (from 46% to 68%) (Mandal et al, 2002). Subtype C probably arrived in China from India, via the over-land heroin trafficking routes (Beyrer et al, 2000; Yu et al, 1998; Neild et al, 1997). Beyer and colleagues (2000) suggested that subtypes B, and later C spread among the IDUs from Burma (which produces 60% of the world's heroin) to China across its eastern border. Phylogenetic analysis by Rodenburg et al (2001) have supported the suggestion that subtype C viruses now circulating in China may have been introduced from India, showing that viruses from China were more closely related to certain viruses from India (Rodenburg et al, 2001). There is now evidence that B and C have recombined and are spreading northwards in China (UNAIDS 2002; Yu et al, 1998; Neild et al, 1997; Beyrer et al, 2000)

### 1.3.3 Brazil

Brazil is the largest country in South America and is one of the most affected by the HIV-1 epidemic. The HIV-1 subtype distribution is complex when compared to other South American countries. Subtype B is the main circulating subtype, but other subtypes such as F, B, C and now B/C and B/F recombinants are also currently circulating (Bongertz et al, 2000; Caride et al, 2000; Couto-Fernandez et al, 1999; Brindeiro et al, 1999; Gao et al, 1998; Sabino et al, 1994; Csillag et al, 1994). Subtype C was already present in the southern regions of Brazil from as early as 1992, although the initial incidence was small (3%) (Bongertz et al, 2000; Brindeiro et al, 1999; Gao et al, 1998; Csillag et al, 14). More recent studies have shown a dramatic increase from the previous figures, with a prevalence of 30% for subtype C in the south and southeast regions, although the increase was less as one moved further north (Soares et al, 2003a; Soares et al, 2003b). The high incidence of subtype C in this region coupled with the equally high incidence of subtype B, facilitated a number of recombination events demonstrated by the presence of B/C and C/B recombinants Soares et al, 2003b).

## 1.4 Differences Between Subtypes B and C

At the molecular level, there are some characteristic differences between subtype B and C. These were first noted by Johansson et al, (1995), who compared the LTR sequences from

the available Ethiopian HIV-1 strains with Swedish HIV-1 subtype B strains and earlier published data. In all the Ethiopian HIV-1 subtype C strains, the first five nucleotides, (G/A)CAGA, were different in the Swedish subtype B strains. The most striking difference in the LTR region was the presence of putative third NF-KB site in the Ethiopian strains located upstream of the usual sites. At the same time, the core enhancer sequence GGGACTTTCC at site I was modified by a deletion of the A nucleotide and a change of the first T to a G (Johansson et al, 1995). The presence of an extra NF-KB site could be related to an increased fitness in subtype C since subtype C LTR sequences have a higher transcriptional activation relative to subtype, B and an increased response to the cytokine tumour necrosis factor alpha (Montano et al, 1997; Jeeninga et al, 2000; Montano et al, 2000). However, Rodenburg et al, (2001) found that the putative third NF-KB site frequently contained deletions or mutations that changed its consensus sequence, or were not found in some sequences, suggesting that this site may not be functional after all.

Another difference seen between the two subtypes is the presence of acidic amino acids at positions 11 and 25 in the V3 *env* sequence of subtype C viruses, which is a feature of viruses that use the CCR5 co-receptor for cell entry (Ping et al, 1999). CXCR4 co-receptor utilization and a syncytium-inducing phenotype are rare among subtype C isolates (Bjorndal et al, 1999; Cecilia, 2000; Tscherning et al, 1998). A switch from CCR5 to CXCR4 co-receptor tropism and NSI to SI phenotype has been associated with progression to AIDS in individuals infected with subtype B viruses. This has not been the case with most subtype C strains, which remain CCR5 tropic and NSI in primary lymphocyte cultures (Abebe et al, 1999; Bjorndal et al, 1999; Morris et al, 2001). Studies on animal models have suggested that there are different pathogenic sequelae for infection with CCR5 or CXCR4 utilizing strains (Berkowitz et al, 1998 and 1999; Harouse et al, 1999). This is strong evidence to suggest that the mechanism of pathogenesis and transmission is different in subtypes C and B. However, there have also been reports from South Africa and Zimbabwe on co-utilization of CCR5 and CXCR4 in subtype C isolates (Cilliers et al, 2003; Janse van Rensburg et al, 2002; Johnston et al, 2003). Janse van Rensburg et al, (2002) reported a change in the phenotypic characteristics of recently isolated subtype C strains, as compared to their previous study using older subtype C samples, and suggest that this is a result of a maturing of the subtype C epidemic in South Africa. Meanwhile, Johnston et al, (2003) suggest that antiretroviral treatment could create an environment facilitating the emergence of CXCR4 tropism in subtype C viruses. In their study of 28 patients from Zimbabwe, viruses from 50% of those on treatment induced syncytium formation when cultured with MT2 cells and the majority of those viruses were capable of using both CCR5 and CXCR4 as co-receptors for viral entry.

It has also been suggested that subtype C viruses grow more readily in Langerhans cells than subtype B (Soto-Ramirez et al, 1996). These cells line the cervix, vagina and foreskin of the penis and could possibly facilitate the spread of the virus through vaginal intercourse. This is discussed further in section 1.5. Other differences between the two subtypes include a truncated *rev* protein (Gao et al, 1998; Rodenburg et al, 2001) and a five-amino acid insertion in the *vpu* transmembrane in subtype C (Gao et al, 1998; Novitsky et al, 1999).

## 1.5 Transmission of Subtype C

HIV is transmitted through blood, genital secretions and breast-milk. It remains uncertain whether the existence of different subtypes has important implications for the transmission of HIV.

### 1.5.1 Horizontal Transmission

Many studies have found subtype B to predominate in HIV-1 infections acquired homosexually and by IDUs. Non-B infections are mostly seen in heterosexually acquired HIV-1 infections, and are usually among individuals with an epidemiological link with Africa or Asia (Herring et al, 2003; Liitsola et al, 2000; Op de Coul et al, 2001; Reinis et al, 2001; Lukashov et al, 1995). In fact, heterosexual transmission is responsible for the majority of new infections in southern Africa and India, a region where subtype C predominates (Mandal et al, 2000; Tien et al, 1999). As already mentioned, subtype C may replicate in Langerhans cells, which are found in genital mucosal epithelium and are thought to be the cells through which vaginal infection occurs. Langerhans cells are antigen-presenting cells found in the epidermis, and in oral and genital mucosal epithelium. They are, however, absent from rectal mucosa (Lalvani and Shastri, 1996). Therefore, one could argue that subtype C appears to be better adapted to penile-vaginal transmission, while subtype B may be more efficiently transmitted through blood (Mastro et al, 1997). Reports of subtype C transmission in China and India show that subtype C is transmitted to a lesser extent via intravenous drug use (Saraswathy et al, 2000). The spread of subtype C in these regions may also be influenced by heterosexual transmission from the IDUs to their partners.

### 1.5.2 Mother-to-child Transmission (MTCT)

In KZN, South Africa, MTCT accounts for 34% of HIV infections (Rollins et al, 2002). Transmission of HIV-1 from mother to infant may occur *in utero*, intra-partum, or postpartum through breast-feeding. Factors associated with perinatal transmission include

### 1.6 Prevention of Mother-to-Child Transmission (pMTCT)

Preventative strategies include routine antenatal voluntary counselling and testing (VCT), various combinations of a shortened period (3-6 months) of exclusive breast-feeding, perinatal administration of ARVs, and provision of safe and affordable replacement feeds.

#### 1.6.1 Use of Nevirapine (NVP) in pMTCT

Although long course anti-retrovirals (ARVs) and caesarian section substantially reduces vertical transmission, they are not practically achievable in most developing countries. The low cost of Nevirapine (NVP) (US$2) and simplicity of administration offers a great advantage for its use in MTCT prevention treatments, particularly for low-income countries (Saloojee, 2001). NVP has been widely used in adults as one of a combination of drugs to treat established HIV infection, but is also effective when given alone as a single dose to the mother at the beginning of labour and one dose administered to the baby within 72 hours of birth. NVP given to HIV-positive pregnant women rapidly crosses the placenta into the fetus with its effects lasting through the first week of life. The pharmacokinetics of NVP are characterised by rapid absorption and distribution, followed by prolonged elimination. The primary route of NVP elimination is through metabolism by the cytochrome P450 enzyme system. NVP elimination accelerates during long term administration because of autoinduction of the enzymes involved in its elimination pathway. A regimen of a single 200mg oral dose administered to the mother during labour, and a single 2 mg/kg dose administered to the newborn at 48 to 72 hours after birth maintains serum NVP concentrations above 100 μg/L (10 times the *in vitro* 50% inhibitory concentration against wild-type HIV-1) throughout the first week of life (Mirochinick et al, 2000). Clinical trials have confirmed the efficacy of NVP in preventing mother-to-child transmission (MTCT) of HIV. The HIVNET 012 trial in Uganda showed that NVP was able to reduce MTCT, with only 8.1% of infants exposed to NVP acquiring HIV at birth (Guay et al, 1999). Almost all the babies in the trial were breast fed, resulting in ongoing exposure to HIV. However, at 14 weeks, the rate of HIV infection was only 13%, while at 12 months the transmission rate was 15.7%, confirming that the reduction in the risk of transmission associated with NVP prophylaxis persists for at least the first year of life, despite the ongoing risk posed by breast-feeding (Guay et al, 1999). The Petra study conducted in Tanzania, Uganda and South Africa using various short-course regimens, found that Zidovudine (AZT) plus Lamivudine (3TC) given at birth and for 7 days postpartum to both mother and babes was effective in reducing transmission (Petra Study Team, 2002). In the South African Intrapartum NVP Trial (SAINT), 1307 mother-infant

pairs were randomised to either NVP during labour and post-delivery, or multiple doses of AZT/3TC during labour and for one week after delivery to mother and baby. In both treatment arms, about 40% of infants were breast-fed. Eight weeks after birth there was no significant difference observed between the rate of HIV infection or death across the two treatment arms, with a rate of 12.3% in the simpler NVP arm, and 9.3% in the more involved and expensive dual therapy arm (Moodley et al, 2003).

### 1.6.2 Resistance after Single Dose NVP

Evidence of drug resistance among women given NVP monotherapy during labour has raised concerns about this intervention. Single mutations in HIV-1 RT that can cause high level NVP resistance are likely to exist in most HIV-1 infected patients at low levels prior to ARV drug exposure (Richman et al, 1994). This favours the emergence of NVP resistance following NVP exposure (Eshleman et al, 2002). The emergence of NVP resistance after only a single dose is also a consequence of the long half-life of NVP upon initial dosing, before the induction of hepatic enzymes that speed up its metabolism (Katzenstein, 2000). Therefore, women receiving NVP for pMTCT experience a lengthy exposure (up to 2 weeks) of active but diminishing concentrations of NVP, creating selective pressure similar to those associated with NVP monotherapy (Havlir et al, 1996; Richman et al, 1994; Wei et al, 1995). The NVP-resistant viruses usually fade from detection in women and infants over time as wildtype viruses repopulate the viral quasispecies and become the predominant strain (Eshleman et al, 2001b). Therefore, it is likely that, in most cases, the re-introduction of NVP at delivery during a second pregnancy should result in the suppression of HIV replication, since most of the viral variants would be sensitive. If resistance mutations were to re-emerge, it would likely occur after delivery, and with the addition of two new ARVs, one would expect that HIV replication could be maximally suppressed and transmission prevented. However, there is growing evidence that the K103N mutation can persist as a minor population of the viral quasispecies for many years, which can very quickly re-emerge with the reintroduction of drug selection pressure (Little et al, 2004; Mellors et al, 2004; Palmer et al, 2004). Mellors et al, (2004) have shown that low-frequency NNRTI–resistant variants contribute to the failure of EFV-based regimens.

Recent reports suggest that risk of NVP resistance may depend on viral subtype (Eshleman et al, 2001a and 2004a). Eshleman and colleagues (2004a) have shown that the emergence of NVP resistance was higher in subtype D versus subtype A, and suggested that the rate of NVP resistance would vary from region to region, depending on which subtypes were prevalent. In the studies in Uganda on the emergence of resistance after single dose NVP,

resistance was found in 19% of women at 8 weeks post dose, with the K103N mutation most common. The Y181C mutation occurred most commonly in the infants. Further analysis of cloned variants showed that the Y181C mutation was detected as early as 7 days after a single dose of NVP, but faded in most women by 6-8 weeks, while the K103N mutation was more likely detected at 6-8 weeks than 7 days (Eshleman et al, 2004b). Eshleman and colleagues (2004b) suggested that the more rapid emergence and fading of Y181C versus K103N mutations was linked to differences in the NVP susceptibility and fitness of HIV-1 with these mutations.

A recent study conducted in Zimbabwe described a higher rate of NVP resistance in subtype C infected women, with 75% of the women showing resistance at 2 weeks and 34% at 8 weeks after NVP prophylaxis (Kantor et al, 2003b). This has led to concerns about the efficacy of NVP in South Africa and other resource-poor settings where subtype C predominates. Even more unsettling is the report of HIV shedding of resistant virus in the breast-milk after NVP prophylaxis, where 65% of the breast-milk samples tested showed NVP mutations (Lee et al, 2003).

### 1.7 Increased Fitness of Subtype C

Viral fitness is defined as the *in vivo* capacity of the naturally occurring viral population to expand in a specified ecosystem, such as human individuals, in which target cells may be scarce, and immune responses, competing viruses, and pharmacological agents are typically present. Within a given viral quasispecies, each clone has a fitness representative of its viral properties (e.g. activity and stability) undergoing selection in that particular environment. During viral replication within a defined environment, different genomes encode viruses that replicate at high rates and mutate, but generally remain under the same selective pressure (Domingo et al, 1999). Therefore, positive (Darwinian) selection implies that one or more members of the quasispecies are better suited to a given environment, while negative selection eliminates unfit variants (Domingo et al, 1999; Domingo and Holland, 1997).

*In vivo* assays provide the best estimate of viral fitness, using either sequencing (De Ronde et al, 2001; Devereux et al, 2001; Goudsmit et al, 1997; Goudsmit et al, 1996; Lukashov et al, 2001), differential hybridization (Eastman et al, 1998; Eastman et al, 1995) or primer-guided nucleotide incorporation assays (Frost et al, 2000). However, most *in vivo* fitness studies have been performed on blood samples (Eastman et al, 1998; Goudsmit et al, 1997; Goudsmit et al, 1996), and do not take into account the fitness of that strain in other body compartments. Host-host comparisons are also difficult due to the differences in host

genetics and immune response. Therefore, *in vivo* fitness studies are limited to the emergence of specific quasispecies or drug resistant mutants, and cannot determine the impact of specific substitutions on replicative capacity (Quinones-Mateu and Arts, 2002).

Multiple methods have been employed to measure HIV-1 replication capacity *in vitro* (Clavel et al, 2000; Nijhuis et al, 2001). Viral fitness is best defined in this context by replicative capacity during growth competition experiments (Quinones-Mateu and Arts, 2002; Holland et al, 1991). However, many studies have equated viral fitness with the catalytic activity of HIV-1 enzymes (Back et al, 1996; Nijhuis et al, 1999); virus turnover on HIV-1 infected individuals (Devereux et al, 2001; Goudsmit et al, 1997; Goudsmit et al, 1996); virus production in monoinfected cultures (Croteau et al, 1997; Doyon et al, 1996; Martinez-Picado et al, 1999; Sharma et al, 1997); using virus particle ratios (Bleiber et al, 2001), and an HIV-induced event in a single-cycle infection assay (Bjorndal et al, 1997; Maeda et al, 1998; Zennou et al, 1998). The replicative capacity of HIV-1 strains can provide information on their potential impact on viral load, drug resistance, and disease progression (Quinones-Mateu and Arts, 2002). For example, the "fittest" isolate that survives in growth competition assays may also lead to increased virulence within the host. However, many of these methods have shortcomings. For example, the methods of measuring viral fitness *in vitro* is by its replication capacity over a single cell replication cycle, which may not be a true representation of *in vivo* fitness as it is replicating under ideal circumstances including an abundance of target cells, and the absence of inhibitors and competitors (Mammano et al, 1998).

Differences with regard to replication capacity or fitness may exist among various HIV subtypes (Spira et al, 2003). Some studies have found that higher rates of viraemia coupled with lowest CD4 counts have led to a more rapid disease progression in subtype C infected individuals (Hu et al, 1999; Nielson et al, 1999; Kanki et al, 1999; Spira et al, 2003). Recent outgrowth of subtype C in southern Africa, India, and China has fuelled speculation that subtype C isolates may be more fit *in vivo*. However, Ball et al (2003) have shown subtype C HIV-1 isolates to be significantly less fit than subtype B isolates *in vitro*. Increased fitness of subtype B over subtype C was also observed in primary CD4+ T cells and macrophages from different human donors, but not in skin-derived human Langerhans cells. Detailed analysis of the retroviral life cycle during several B and C virus competitions indicated that the efficiency of host cell entry may have a significant impact on relative fitness. Furthermore, phylogenetic analyses of fitness differences suggested that, for a recombined subtype B/C HIV-1 isolate, higher fitness mapped to the subtype B *env* gene rather than the subtype C *gag* and *pol* genes. These results suggest that subtype B and C HIV-1 may be transmitted with equal efficiency (Langerhans cell data), but that

subtype C isolates may be less fit following initial infection (T-cell and macrophage data) and may lead to slower disease progression (Ball et al, 2003).

Velazquez-Campoy et al (2001) reported that proteases from the C and A subtypes exhibit higher biochemical fitness in the presence of widely prescribed protease inhibitors. Naturally occurring amino acid polymorphisms found in proteases from the C and A subtypes lower the binding affinities of existing clinical inhibitors up to a factor of 7.5, which by themselves are not enough to cause drug resistance, but instead significantly amplifies the effects of the drug-resistant mutation. By including the enzyme kinetic properties in the analysis, the biochemical fitness of the C and A subtype drug-resistant mutants can be up to 1000-fold higher than that of the wild-type B subtype protease in the presence of the studied inhibitors. They concluded that this might have important consequences on the long-term viability of current HIV-1 protease inhibitors (Velazques-Campoy, 2002).

## 1.8 HIV-1 Antiretroviral Therapy

There are currently four classes of approved ARVs available. The HIV-1 reverse transcriptase (RT) and protease (PR) enzymes are the molecular targets for 3 classes of ARV drugs: the protease inhibitors (PIs), nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), and non-nucleoside reverse transcriptase inhibitors (NNRTIs). The fourth class of ARVs is the recently approved fusion inhibitor, T20.

### 1.8.1 HIV-1 Protease

HIV-1 protease is an aspartic protease composed of two identical, non-covalently connected subunits, ninety-nine amino acids in length (Figure 3) (Loeb et al, 1989). The subunits form a substrate-binding cleft that is covered with a mobile flap. The hydrophobic substrate-binding cleft recognizes, and cleaves, the viral Gag, GagPol polyprotein and Nef precursors to yield the structural proteins and enzymes of the virus (Baldwin et al, 1995, Chen et al, 1995; Gulnik et al, 1995; Mahalingam et al, 2001; Olsen et al, 1999).

#### 1.8.1.1 Protease Cleavage

The structural and enzymatic proteins that comprise the virus core are initially translated as part of the Gag and GagPol polyprotein precursors. Protease cleaves at 12 sites: 5 in Gag (p17/p24, p24/p2, p2/NC, p7/p1 and p1/p6$^{gag}$); 6 sites in GagPol (NC/TFP, TFP/p6$^{pol}$, p6$^{pol}$/PR, PR/RT, RT/p66 and p66/IN and a single site in Nef.

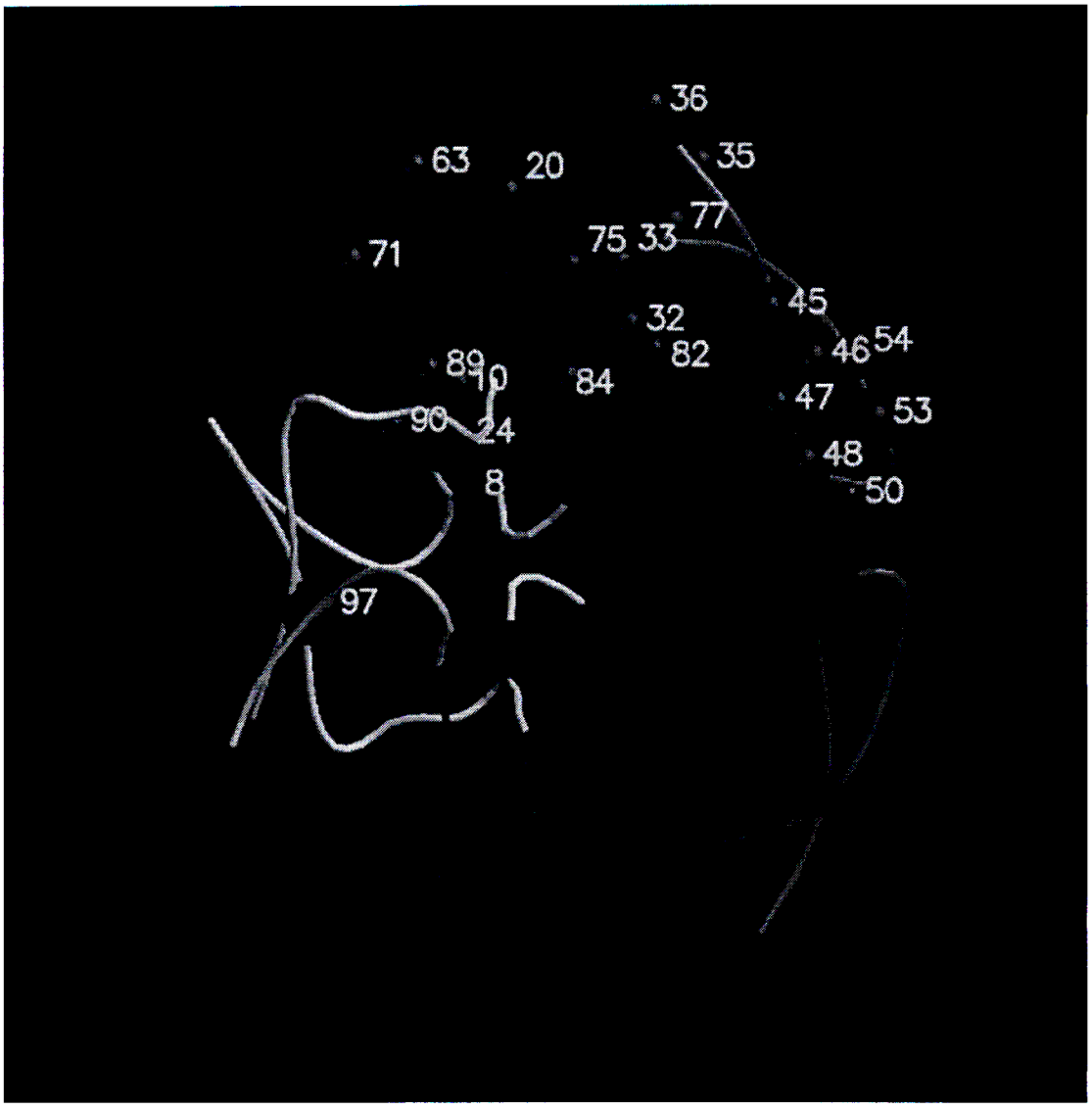



**Figure 3. HIV-1 protease structure showing the position of resistance associated mutations.**

Accurate and ordered processing of these precursors is an essential step in the production of infectious viral particles (Kaplan et al, 1994; Mervis et al, 1988). For the GagPol cleavage process, the first site to be cleaved is the p2/NC site to give 2 products, one with the MA, CA and p2 and the other with NC, TFP, $p6^{pol}$, PR, RTp51, RTp66 and IN. This is followed by cleavage at TFP/$p6^{pol}$ to once again give two products, this time with NC and TFP and the other product with the remaining sites. Finally, the remaining products are cleaved to give their individual proteins. A similar process occurs for the Gag cleavages. Firstly, the precursor is cleaved at p2/NC to give two products consisting of the MA, CA and p2, and the second product consisting of the NC, p1 and $p6^{gag}$ proteins. This is followed by cleavage at the MA/CA site and simultaneous cleavage at p1/ $p6^{gag}$.

Finally, the remaining sites CA/p2 and NC/p1 are cleaved (Cote et al, 2001; Pettit et al, 2003; De Oliveira et al, 2003). The timing of the initiation of these cleavages is regulated such that the cleavage products remain in the virus particle, even though processing appears to initiate while the polyprotein precursors are still cell associated (Kaplan et al, 1994; Karacostas et al, 1993).

Sequences at the protease cleavage sites have been studied to determine protease-substrate interactions and regulation of Gag processing, particularly the amino acids immediately upstream of the scissile bond (P1) and the amino acid immediately downstream of the scissile bond (P1'). The initial cleavages are carried out by the precursor-associated immature protease, including the cleavages that release the mature protease itself. The mature protease is responsible for the later cleavages. Thus, during virus assembly, the active dimeric enzyme originates as the result of the dimerization of two GagPol precursors. Once the protease domain is liberated from the precursors by cleavage at its N and C termini, a free and mature dimer, consisting of two protease monomers, is produced (Tessmer et al, 1998). Pettit et al (2002) showed that the rate of cleavage by the viral protease could be improved with the introduction of specific alternative amino acids at the P1 position. Substitutions at certain sites in Gag (MA/CA, CA/p2, NC/p1) mediated enhanced rates of cleavage, while for the other two sites (p2/NC, p1/p6); the wild-type amino acid conferred optimal cleavage (Pettit et al, 2002). The size of the amino acid at P1' also affected the rate of cleavage. PI resistant isolates with replication impaired Gag or GagPol processing can partially compensate by acquiring amino acid substitutions at *gag* cleavage sites (Prado et al, 2002; Mammano et al, 1998; Zennou et al, 1998). This is supported by changes seen in the protease cleavage sites within Gag (p2/ NC, NC/p1 and NC/TFP) after PI therapy in PI resistant individuals (Cote et al, 2001).

### 1.8.1.2 Protease inhibitors

There are seven FDA-approved protease inhibitors (PIs): Amprenavir (APV), Indinavir (IDV), Lopinavir (LPV - manufactured in combination with Ritonavir as Kaletra), Nelfinavir (NFV), Ritonavir (RTV), Saquinavir (SQV) and Atazanavir (ATZ). The function of the PIs is to prevent Gag-Pol polyproteins from successfully entering the active site and being cleaved. This consequently prevents the assembly of new, fully functional HIV virions.

### 1.8.1.3 Resistance to Protease Inhibitors

Most PR resistance mutations prevent inhibitor binding by altering the structure of the substrate cleft (Baldwin et al, 1995; Chen et al, 1995; Gulnik et al, 1995; Mahalingam et al, 2001; Olsen et al, 1999). Primary resistance mutations line the substrate cleft and occur at the following codon positions: D30N, V32I, G48V, I50V, V82A/F/T/S, I84V and L90M. The protease flaps (residues 33-62) extend over the substrate-binding cleft and are flexible to allow entry and exit of the polypeptide substrates and products (Shao et al, 1997, Scott et al, 2000). The tips of the flap, residues 46-54, are mobile and are also the site of many drug resistance mutations. As these mutations are outside of the substrate-binding region, they do not provide measurable resistance on their own. Instead, these mutations (at codons 46, 47, 53, and 54) either compensate for the decreased kinetics of enzymes with active site mutations or cause resistance by altering enzyme catalysis, dimer stability, inhibitor binding kinetics, or active site re-shaping through long-range structural perturbations (Erickson et al, 1999). Mutations associated with resistance to the currently available PIs are shown in Figure 4. Many PIs are now co-administered with a sub-therapeutic dose of Ritonovir (a P450 enzyme inhibitor) to inhibit the body's drug clearance mechanism, thereby increasing the plasma levels of the PI (Hurst et al, 2000). Since PIs are competitive inhibitors, the increased PI concentrations relative to the normal substrate are partly responsible for the enhanced ARV activity of these boosted PI regimens. Boosted PI levels can overcome small reductions in susceptibility conferred by early PI mutations, thus prolonging viral suppression (Condra et al, 2000; Kempf et al, 2001). Therefore viral rebound requires the acquisition of a greater number of mutations that produce even greater reductions in susceptibility.

#### 1.8.1.3.1 L24I

This mutation has been reported in patients receiving IDV and has not been shown to confer cross-resistance to other PIs, except possibly LPV (Condra et al, 2000; Kempf et al, 2001).

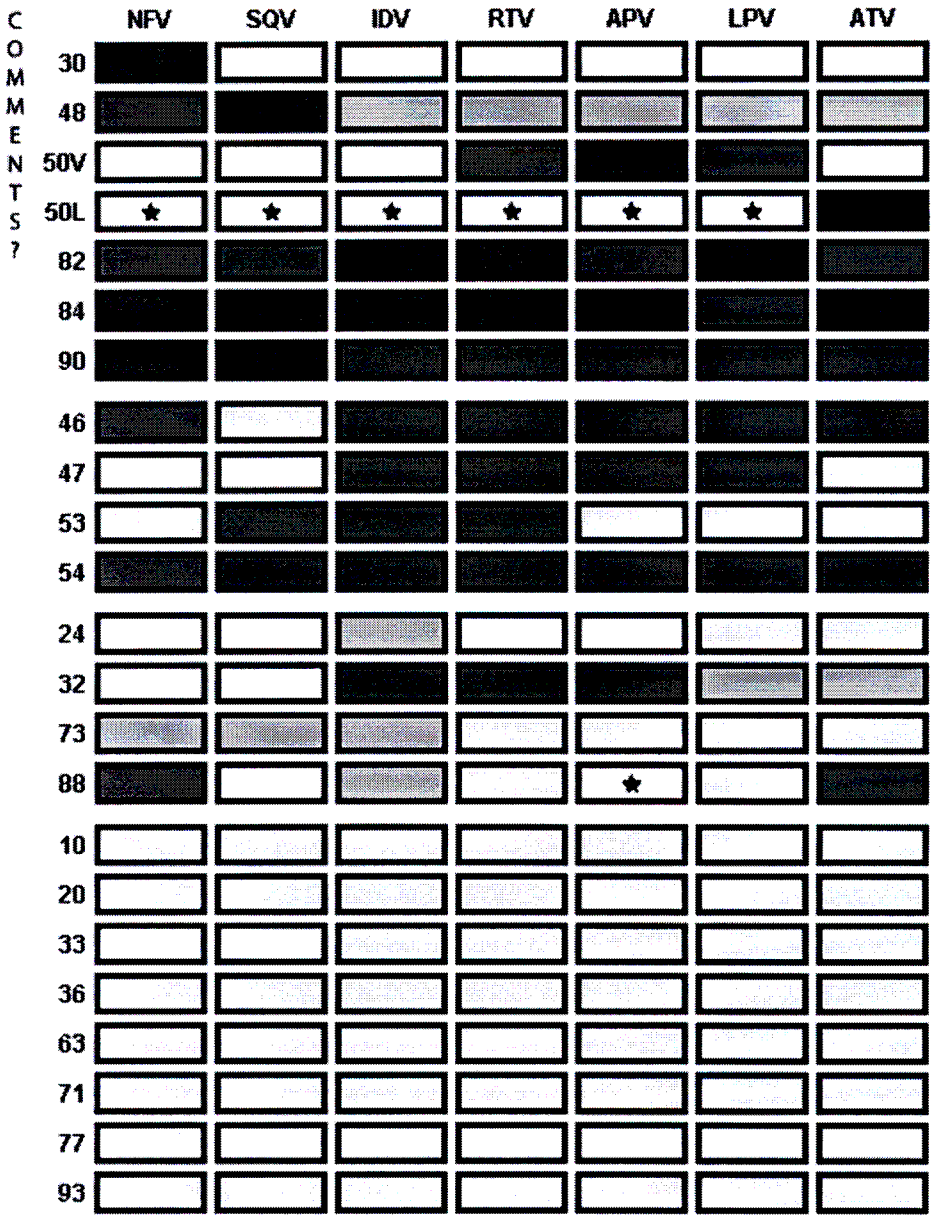



**Figure 4. Mutations associated with resistance to the currently available PIs (Shafer, 2003; http://hivdb.stanford.edu/)**

#### 1.8.1.3.2 D30N

D30N only occurs in patients receiving NFV. It does not confer *in vitro* or clinical cross-resistance to the other PIs on its own (Patick et al, 1996; Markowitz et al, 1998; Winters et al, 1998a; Zolopa et al, 1999). However, cross-resistance to IDV, RTV, and SQV has been observed in isolates that have D30N along with mutations at protease codon positions 88 and 90 (Parkin et al, 2001).

#### 1.8.1.3.3 V32I

The V32I mutation occurs in patients receiving IDV, RTV or APV. By itself, it appears to cause minimal resistance to any one PI. It causes high-level resistance when present with other PI resistance mutations M46I, I47V, V82A and I84V (Parkin et al, 2003a).

#### 1.8.1.3.4 M46I/L

The M46L mutation has been reported in patients treated with IDV, RTV, APV and NFV and contributes resistance to each of the PIs except SQV (Condra et al, 2000; Molla et al, 1996; Schapiro et al, 1996; Patick et al, 1998).

#### 1.8.1.3.5 I47V

Mutations at codon 47 have been reported in patients receiving IDV, RTV, APV, and often occur in conjunction with the V32I mutation (Markland et al, 2000; Parkin et al, 2003a).

#### 1.8.1.3.6 G48V

This mutation occurs primarily in patients receiving SQV and rarely in patients receiving IDV (Jacobsen et al, 1995; Patick et al, 1996; Hertogs et al 1998; Winters et al, 1998a). Its affect on APV and LPV is not known. It often occurs with mutations at codon positions 54 and 82 and isolates with all three mutations are phenotypically resistant to each of the PIs (Shafer et al, 1998; Palmer et al, 1999).

#### 1.8.1.3.7 I50V/L

The I50V mutation has been reported in patients receiving APV as their first PI (Maguire et al 2002). It causes reduced APV susceptibility, and has been shown to cause cross-resistance to RTV and LPV *in vitro* (Partaledis et al, 1995; Tisdale et al, 1995; Colonno et al, 2000; Molla et al, 2001; Parkin et al, 2001; Prado et al, 2002). I50L occurs in patients

receiving ATZ as their first PI (Colonno et al, 2002). In addition to causing reduced ATZ susceptibility, it causes hypersusceptibility to each of the remaining PIs (Colonno et al, 2002).

#### 1.8.1.3.8 F53L

This mutation has been in more than 10% of patients treated with multiple PIs and has been associated with phenotypic resistance to LPV (Kantor et al, 2002a; Kempf et al, 2001).

#### 1.8.1.3.9 I54V/T/L/M

Mutations at this position have been reported in patients treated with IDV, RTV, APV, SQV, and LPV and contribute resistance to each of the available PIs (Condra et al, 2000; Molla et al, 1996; Schapiro et al, 1996; Patick et al, 1998).

#### 1.8.1.3.10 G73C/S/T

These mutations have been reported in patients receiving IDV, SQV or occasionally NFV monotherapy, but usually are found in patients failing multiple PIs (Shafer et al, 1999; Wu et al, 2003). It usually occurs in conjunction with the L90M mutation (Kantor et al, 2002a).

#### 1.8.1.3.11 V82A/T/F/S

V82A/T/F/S occurs in HIV-1 isolates from patients receiving IDV or RTV (Condra et al, 1996; Molla et al, 1996). V82A is the most common mutation at this position, while V82S is the least common. Alone, these mutations confer decreased susceptibility to IDV, RTV and LPV however, when present with other PI mutations they contribute phenotypic and clinical resistance to each of the PIs (Shafer et al, 1998; Sham et al, 1998; Winters et al, 1998a; Falloon et al, 2000; Kempf et al, 2001). V82I occurs more frequently in untreated individuals with non-B isolates (1% of subtype B versus 5-10% of non-B subtypes), and confers minimal or no resistance to the available PIs (Gonzales et al, 2001; Brown et al, 2001; Descamps et al, 1998; King et al, 1995).

#### 1.8.1.3.12 I84V/A/C

Mutations at this codon position have been reported in patients receiving only one PI (either IDV, RTV, SQV or APV) and cause cross-resistance to each of the PIs (Condra et al, 1996; Molla et al, 1996; Craig et al, 1998; Hertogs et al, 2000; Sevin et al, 2000). I84V

tends to develop in isolates that already have the L90M mutation (Kantor et al, 2002a). I84A and I84C occur rarely, but are also associated with resistance to multiple PIs when present with other PI mutations (Mo et al, 2003).

#### 1.8.1.3.13 N88D/S

Mutations at position 88 have been reported in patients receiving NFV or occasionally IDV. Alone, these mutations cause low-level NFV resistance. When they occur in conjunction with D30N or M46I, they cause high-level resistance to NFV (Colonno et al, 2000; Petropoulos et al, 2000; Ziermann et al, 2000). N88S causes hypersensitivity to APV but the clinical significance of this is not known (Ziermann et al, 2000).

#### 1.8.1.3.14 L90M

The L90M mutation has been reported in patients treated with SQV, NFV, IDV and RTV. And causes cross-resistance to each of the available PIs (Lawrence et al, 1999; Zolopa et al, 1999; Falloon et al, 2000; Hertogs et al, 2000; Para et al 2000; Dronda et al, 2001; Kempf et al, 2001). The mechanism by which L90M causes PI resistance is not known.

#### 1.8.1.3.15 Polymorphisms in PR and Resistance

Amino acid variants at codons 10, 20, 36, 63, 71, 77, and 93 also make frequent contributions to drug resistance when present with other protease mutations (Figure 4) (Condra et al, 1995; Rose et al, 1996; Martinez-Picado et al, 1999; Nijhuis et al, 1999; Mammano et al, 2000). Codon 63 is the most polymorphic of these sites. Mutations at these positions increase in heavily treated patients (Wu et al, 2003; Hertogs et al, 2000; Yahi et al, 1999). In some HIV-1 subtypes, mutations at codons 20, 36 and 93 occur at higher rates than they do in subtype B isolates (Cornelissen et al, 1997; Shafer et al, 1999; Pieniazek et al, 2000; Gonzales et al, 2001). It has been suggested that individuals harbouring isolates containing multiple accessory mutations may be at a greater risk of virologic failure during PI therapy (Perez-Alvarez et al, 2001; Perno et al, 2001). However, most studies to date have not supported this hypothesis (Frater et al, 2001; Perez-Alvarez et al, 2001; Perno et al, 2001; Servais et al, 2001; Kuritzkes et al, 2000; Bossi et al, 1999; Harrigan et al, 1999). In addition, mutations at position 17, 22, 23, 45, 58, 66, 74, 75, 76, 83 and 85 in PR have recently been associated with treatment, but the phenotypic and clinical impact of these mutations are not yet known (Parkin et al, 2003a; Wang and Larder et al, 2003; Wu et al, 2003).

### 1.8.2 HIV-1 Reverse Transcriptase (RT)

The RT enzyme is both an RNA-dependent DNA polymerase and DNA-dependent DNA polymerase. RT is a heterodimer consisting of p66 and p51 subunits (Figure 5). The p66 subunit consists of 560 amino acids containing a polymerase domain (the DNA-binding groove and the enzyme's active site), an RNase H subdomain and connection subdomain (Larder and Stammers, 1999; Sarafianos et al, 1999a). The shape of the polymerase domain of the p66 subunit can be likened to a human hand with subdomains referred to as fingers, palm, and thumb. Most RT inhibitor resistance mutations are in the 5' polymerase coding regions, particularly in the "fingers" and "palm" subdomains (Shafer et al, 2003). The p51 consists of the first 440 amino acids of the p66 subunit, but displays no enzymatic activity and functions as a scaffold for the enzymatically active p66 subunit.

#### 1.8.2.1 Nucleoside/nucleotide RT Inhibitors (NRTIs)

There are currently seven nucleoside analogues, and one nucleotide analogue approved by the FDA, including Zidovudine (AZT), Didanosine (ddI), Zalcitibine (ddC), Stavudine (d4T), Lamivudine (3TC), Abacavir (ABC) and Emtricitibine (FTC) (Figure 6). Tenofovir disoproxil fumarate (TDF) is the only FDA-approved nucleotide analogue. The NRTIs are prodrugs that must first be phosphorylated by host cellular enzymes before they can be used by the RT enzyme. Nucleosides must be tri-phosphorylated and nucleotides must be di-phosphorylated. The phosphorylated NRTIs compete with natural deoxynucleoside triphosphates (dNTPs) for incorporation into the newly synthesized DNA chains where they cause chain termination and block further extension of the proviral DNA during RT (Shafer et al, 2003).

#### 1.8.2.2 NRTI Resistance

NRTI resistance can occur by two biochemical mechanisms. In the first, certain mutations allow the RT enzyme to discriminate against NRTIs during synthesis and prevent their addition to the growing DNA chain (Larder and Stammers, 1999; Sarafianos et al, 1999a; Huang et al, 1998). The second mechanism is mediated by nucleotide excision mutations (NEMs) that increase the rate of ATP-dependent hydrolytic removal of the chain-terminating NRTI and enable continued DNA synthesis (Shafer et al, 2003; Arion et al, 2000; Meyer et al, 1999; Arion et al, 1998). This mechanism of resistance has also been referred to as pyrophosphorolysis and primer unblocking.

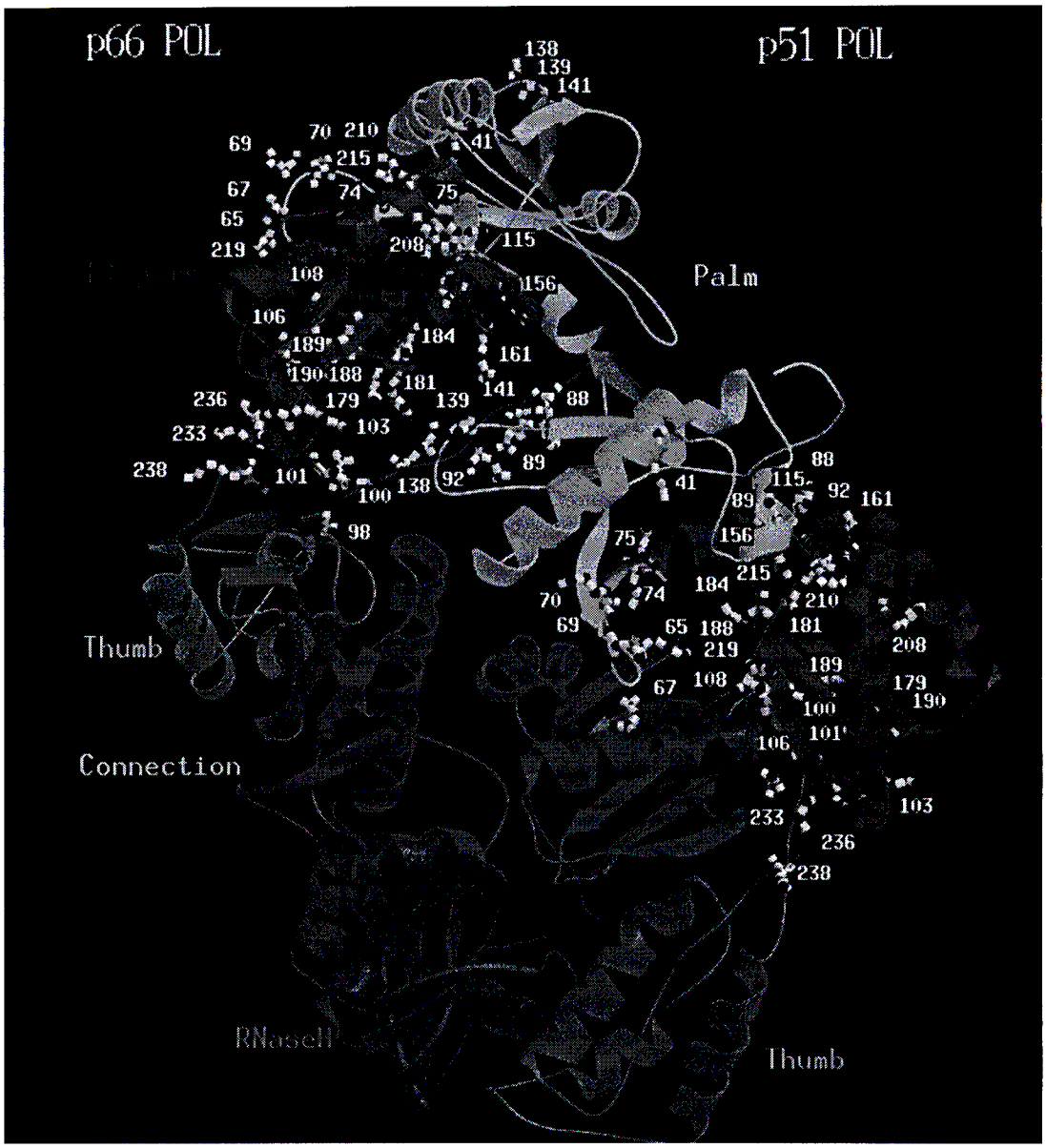



**Figure 5. HIV-1 RT structure. The active site (positions 110, 185 and 186) is shown in red sticks in the p66 subunit**

#### 1.8.2.2.1 Nucleotide Excision Mutations (NEMs)

The nucleotide excision mutations mediate the ATP-dependent hydrolytic removal of a dideoxy nucleotide monophosphate (ddNMP) from a terminated cDNA chain (Meyer et al, 2000; Arion et al, 1998). They were originally identified for their role in causing AZT resistance, but several studies have since shown that the NEMs are associated with clinical resistance to AZT, d4T, ABC, and to a lesser extent, to ddI, ddC and TDF (Miller et al, 2001; Miller and Larder, 2001; Wainberg et al, 2001) (Figure 6). The interaction of the incoming dNTP with a ddNMP-terminated primer results in the formation of a stable "dead-end" catalytic complex between RT, primer, template, and dNTP (Sarafianos et al, 2003; Miller et al, 2001; Lennerstrand et al, 2001; Boyer et al, 2001; Tong et al, 1997). The formation of such a dead-end complex interferes with the ability of NEMs to facilitate the resumption of virus DNA chain elongation. The bulky azido group of AZT may interfere with the formation of such a dead-end catalytic complex by sterically preventing the addition of the next dNTP (Boyer et al, 2001; Lennerstrand et al, 2001).

Various combinations of mutations at codons 41, 67, 70, 210, 215, and 219 have been implicated in the removal of the ddNMP (Meyer et al, 2000; Arion et al, 1998; Huang et al, 1998; Harrigan et al, 1996; Kellam et al, 1992; Larder et al, 1989). Biochemical studies suggest that D67N and K70R are the mutations most responsible for rescue of chain-terminated primers (Meyer et al, 1999; Arion et al, 1998). Some studies suggest that the T215Y/F mutations lead to increased RT processivity (Caliendo' et al, 1996, Arion et al, 1998). Mutations at positions 41 and 210 usually occur with mutations at position 215 and appear to stabilize the interaction of 215Y/F with the dNTP binding pocket (Yahi et al, 2000; Huang et al 1998; Harrigan et al, 1996; Hooker et al, 1996). Mutations at positions 67 and 219 may occur with mutations at position 70 or with mutations at position 215, suggesting alternate pathways for the development of resistance to the NRTIs (Shafer, 2003). In summary, the favoured pathway for the development of resistance is: **M41L/T215Y→M41L+T215Y→M41L+K210W+T215Y→M41L+D67N+K210W+T215Y** with **D67N/K70R→K219Q/E/N→D67N+K70R/K70R+K219Q→T215Y+K219Q** as the alternate pathway.

The NEMs interfere with the clinical response to AZT, d4T, ddI and ABC (Molina et al, 2003; Kozal et al, 1993). A combination of three or more NEMs with M184V results in complete loss of ABC activity (Lanier et al, 2004; Lanier et al, 1999). M184V also appears to reverse the effect of the NEMs on resistance to d4T and TDF (Naeger et al, 2001; Shulman et al, 2001; Miller et al, 2002; Palmer et al, 1999). Preliminary data

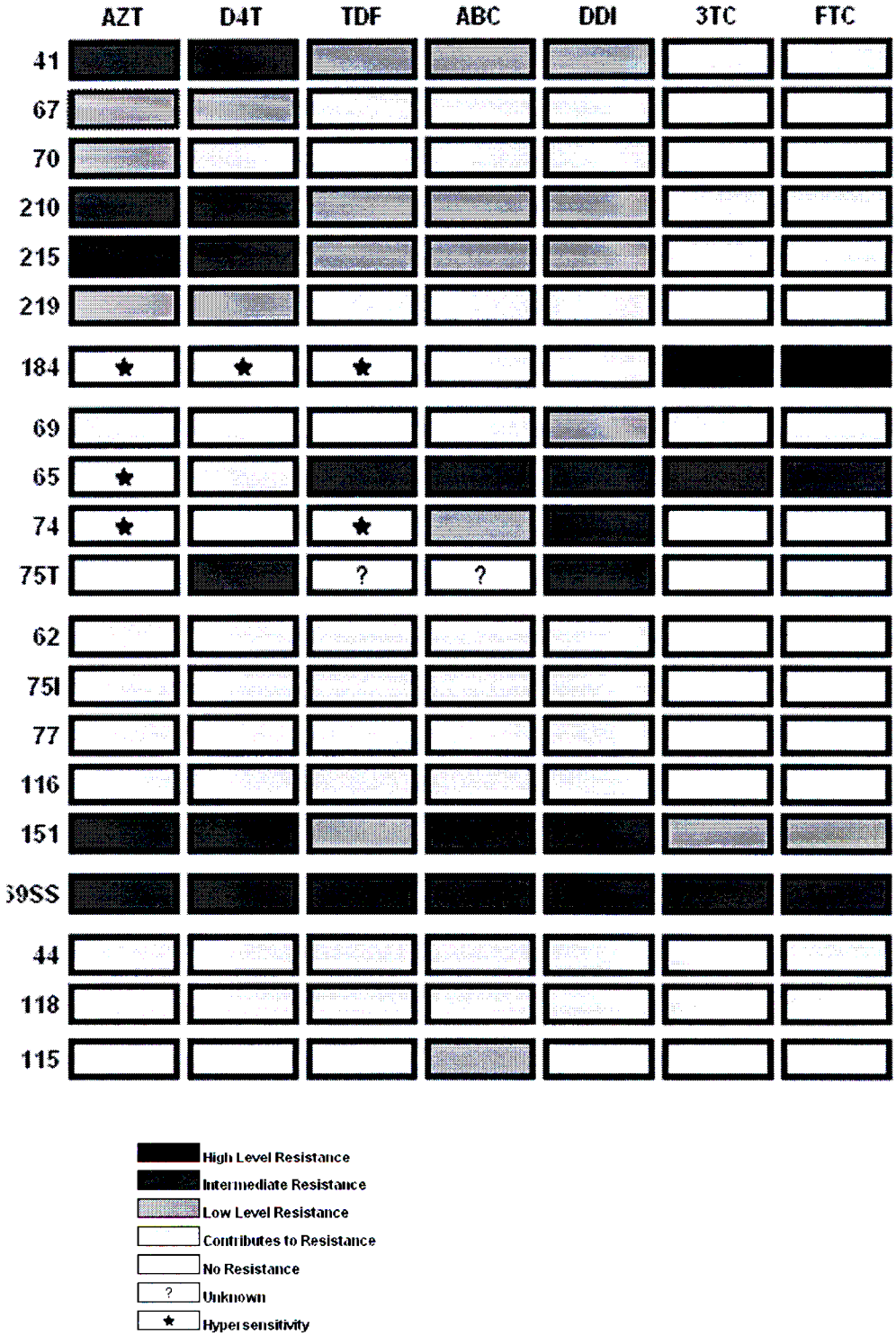



**Figure 6. Mutations associated with resistance to the currently available NRTIs (Shafer, 2003; http://hivdb.stanford.edu/). ddC is not currently used and is therefore not shown in the figure.**

suggest that TDF usually retains antiviral activity even in patients with extensive previous NRTI therapy (Miller et al 2001; Schooley et al, 2001).

#### 1.8.2.2.2 K65R

K65R is seen mainly in patients receiving triple NRTI regimens lacking AZT (Gallant et al, 2003; Roge et al, 2003). It confers intermediate levels of resistance to ddI, ABC, ddC, 3TC, FTC and TDF, and low-level resistance to d4T (Parkin et al, 2003b; Miller et al, 2000; Petropoulos et al, 2000; Wainberg et al, 1999; Tisdale et al, 1997; Gu et al, 1994, Zhang et al, 1994). However, it also causes hypersensitivty to AZT (Rhee et al, 2003). Position 65 interacts with the γ-phosphate of the bound dNTP and improves discrimination between the dNTPs and most NRTIs (Lanier et al, 2003; Parkin et al, 2003). The K65R mutation has been shown to increase RT replication fidelity and processivity *in vitro*, but has also been associated with a decrease in replicative capacity (Miller et al, 2003; Shah et al, 2000; Arion et al, 1996). K65R generally occurs in association with other mutations (such as M184V and Q151M) that discriminate between the NRTIs and the natural dNTP substrates (Gonzales et al, 2003).

#### 1.8.2.2.3 T69D/N/SA Ins

Mutations at position 69 include T69D/N/S/A, as well as single and double amino acid insertions. T69D has been reported in patients treated with either of the available NRTIs (Shafer, 2003; Fitzgibbon et al, 1992). In site-directed mutagenesis studies, other mutations at this position including T69N, T69S, and T69A have been shown to confer resistance to AZT, ddI, ddC, and d4T (Winters et al, 2001). The 69Ins occurs in approximately 2% of heavily treated patients and when alone, causes low level resistance to each of the NRTIs. When it occurs together with T215Y/F and other AZT-resistance mutations it causes high-level resistance to each of the NRTIs, including a 20-fold resistance to TDF which is the highest reported level of resistance to this drug (Masquelier et al, 2001; Larder et al, 1999; Winters et al, 1998b; de Jong et al, 1999). It has been suggested that this mutation may act in a similar manner to the NEMs by causing ATP-mediated primer unblocking (Lennerstrand et al, 2001).

#### 1.8.2.2.4 L74V/I

L74V is usually seen in patients treated with ddI or ABC monotherapy and causes intermediate resistance to ddI and ddC and low-level resistance to ABC (Harrigan et al, 2000; Miller et al, 2000; Tisdale et al, 1997; Winters et al, 1997; Kozal et al, 1994; Shafer

et al, 1994; St. Clair et al, 1991). It also causes hypersensitivity to AZT and TDF (Parkin et al, 2003b; Shafer, 2003; Miller et al, 2001; St. Clair et al, 1991). L74V plays a role in the repositioning of the template/primer complex (Sturmer et al, 2003). This mutation has also been shown to be the cause of decreased RT processivity and decreased replication in cell culture (Sharma et al, 1999). Recent evidence suggests that L74V causes a reduction in the efficiency of the excision reaction associated with RT (Frankel et al, 2004). L74I causes intermediate resistance to ddI (Kantor et al, 2002a).

#### 1.8.2.2.5 V75T/I/M/A

The V75T mutation is rarely found *in vivo*. It develops in isolates cultured in the presence of increasing concentrations of d4T and causes intermediate resistance to d4T, ddI and ddC (Lacey et al, 1994). Mutations at this position cause drug resistance through nucleotide discrimination as well as through a non-ATP-mediated mechanism of primer unblocking (Lennerstrand et al 2001; Selmi et al, 2003). V75I generally occurs in isolates that also have the multinucleoside resistance mutation, Q151M. The V75M/A mutations contribute to d4T resistance (Rhee et al, 2003; Bloor et al, 1998).

#### 1.8.2.2.6 M184V/I

M184V sterically hinders inhibitor binding by changing the geometry of the dNTP binding pocket by direct interaction of the altered residue with the incoming dNTP, thus leading to a decrease in 3TC incorporation (Sarafianos et al, 1999b; Drosopoulos et al, 1998). Several *in vitro* studies have shown that RT enzymes with M184V displayed increased fidelity and decreased processivity (Drosopoulos et al, 1998; Sharma et al, 1999; Oude et al, 1997; Wainberg et al, 1996; Boyer et al, 1995; Back et al, 1997). M184V is usually the first mutation to develop in isolates from patients receiving incompletely suppressive 3TC-containing regimens (Staszewski et al, 2001; Descamps et al, 2000; Havlir et al, 2000; Maguire et al, 2000). It is also selected during therapy with ABC, ddC, ddI and FTC (Quinn et al, 2003; Harrigan et al, 2000; Miller et al, 2000; Winters et al, 1997; Gu et al, 1992). M184V, in combination with the NEMS or in combination with mutations at positions 65, 74, or 115 in RT, leads to both *in vitro* and *in vivo* ABC resistance (Harrigan et al, 2000; Katlama et al, 2000; Lanier et al, 1999; Palmer et al, 1999). The M184V mutation also reverses T215Y-mediated AZT resistance by its ability to impair the rescue of chain-terminated DNA synthesis (Gotte et al, 2000; Shafer et al, 1998; Larder et al, 1995; Boucher et al, 1993; Tisdale et al, 1993). However, this can be overcome by the presence of four or more AZT resistance mutations (Whitcomb et al, 2003; Shafer et al, 1998; Tisdale et al, 1993). Re-sensitisation does not appear to apply to AZT resistance

caused by Q151M (Shafer et al, 1996). M184I also causes high-level resistance to 3TC. It usually develops before M184V in patients receiving 3TC because HIV-1 RT is more prone to G to A mutations than to A to G mutations (i.e. ATG to GTA) (Keulen et al, 1997; Ji et al, 1994). However, the enzymatic efficiency of M184I is less than that of M184V and nearly all patients with mutations at this position eventually also develop M184V (Frost et al, 2000).

#### 1.8.2.2.7 Q151M

Q151 is in the B2 strand in the fingers domain of p66 subunit of HIV-RT and are in direct contact with the incoming dNTP. Its role is to participate in the positioning of the incoming dNTP in the HIV-RT active site (Courcambeck et al, 2002). Alone, the Q151M mutation causes intermediate levels of resistance to AZT, ddI, ddC, d4T and ABC (Van Laethem et al, 2000; Iversen et al, 1996; Shafer et al, 1996; Shirasaka et al, 1995). It causes high-level resistance to each of these NRTIs and low-level resistance to 3TC and TDF when present with mutations at positions 62, 75, 77, and 116 (Miller et al, 2000; Palmer et al, 1999). This mutation was found to develop in up to 50% of patients who had received dual therapy with ddI and AZT/d4T (Cowley et al, 2001; Shafer et al, 1994). Q151M confers ability of RT to discriminate between an analogue and its natural counterpart and has no effect on repair of analogue termination (Deval et al, 2002). MultiNRTI resistance via the Q151M pathway is more common in non-B subtypes (Kantor et al, 2003a).

### 1.8.2.3 Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs)

There are currently three FDA-approved NNRTIs, namely, NVP, delavirdine (DLV), and efavirenz (EFV) (Figure 7). The NNRTIs bind to a hydrophobic pocket close to the RT active site (Hsiou et al, 2001). They inhibit HIV-1 replication allosterically by displacing the catalytic aspartate residues relative to the polymerase binding site (Esnouf et al, 1995; Spence et al, 1995; Kohlstaedt et al, 1992).

### 1.8.2.4 NNRTI resistance

Resistance usually emerges rapidly when NNRTIs are administered as monotherapy or in the presence of incomplete virus suppression, with a single mutation in the NNRTI-binding pocket resulting in high-level resistance to one or more NNRTIs (Figure 7). Many researchers suggest that this rapid appearance of resistance may be caused by the selection of a pre-existing population of mutant viruses within an individual (Conway et al, 2001;

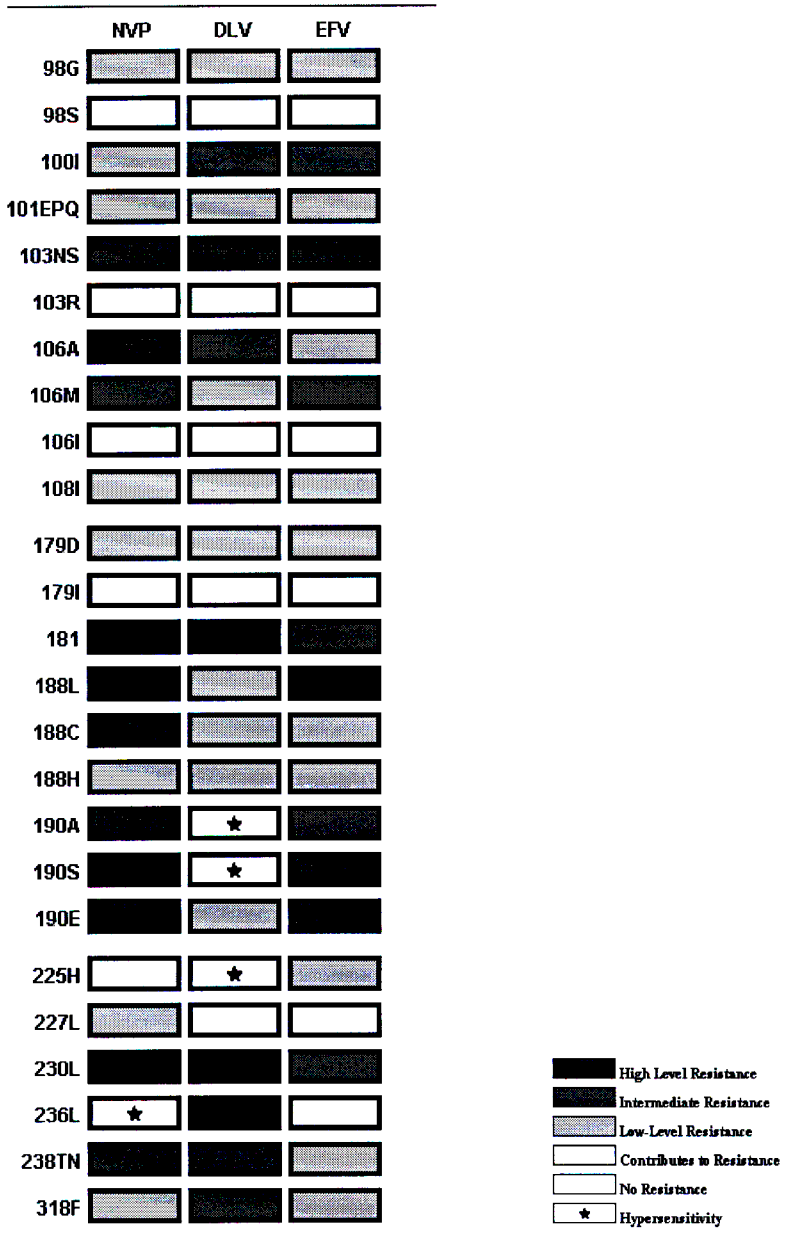



**Figure 7. Mutations associated with resistance to the currently available NNRTIs (Shafer, 2003; http://hivdb.stanford.edu/).**

Jackson et al, 2000; Havlir et al, 1996; Wei et al, 1995). NNRTI resistance mutations may compromise virus replication by changes in the conformation of the dNTP binding pocket and changes in RNaseH activity (Van Laethem et al, 2000; Archer et al, 2000; Kleim et al, 1994).

#### 1.8.2.4.1 K103N/S/R/Q/T

K103N occurs more commonly than any other mutation in patients receiving NNRTIs and causes high-level resistance to each of the available NNRTIs (Bacheler et al, 2000; Demeter et al, 2000; Hanna et al, 2000; Conway et al, 2001; Deeks et al, 2001; Torti et al, 2001; Petropoulos et a; 2000; Huang et al, 1999; Young et al, 1995). Alone, the K103N mutation causes resistance to each of the available NNRTIs (Casado et al, 2000; Demeter et al, 2000; Joly et al, 2000). K103S appears to have the same effect as the K103N mutation (Harrigan et al, 2003). K103R does not appear to cause NNRTI resistance on its own, but can cause cross-resistance to all of the NNRTIs when present with V179D (Petropoulos et al, 2003). The K103Q/T mutations do not appear to cause NNRTI resistance.

Position 103 is located in the vicinity of an entrance to the binding pocket. The structure of the RT enzyme is only minimally changed by the K103N mutation. Unliganded, it forms a network of hydrogen bonds that are not present in the wildtype enzyme which could stabilize the closed pocket form of the enzyme and interfere with the ability of inhibitors to bind to the enzyme (Hsiou et al, 2001).

#### 1.8.2.4.2 V106A/M/I

V106A causes high-level resistance to NVP, intermediate levels of resistance to DLV, and low-level resistance to EFV (Bacheler et al, 2001; Petropoulos et a; 2000; Balzarini et al, 1998, Fujiwara et al, 1998; Young et al, 1995; Byrnes et al, 1993; Emini et al, 1993; Larder et al, 1993). V106M is a putative subtype C specific mutation occurring in EFV treated patients and causes high-level resistance to each of the NNRTIs but does not arise under pressure by NVP or DLV (Brenner et al, 2003). Subtype C tends to get the V106M rather than the V106A because of the GTG to ATG mutation at amino acid position 106. Subtype B usually has a GTA to GCA mutation. When on EFV treatment, the V106M mutation is followed by mutations at codons 181 and 188 (Brenner et al, 2003). V106I is a treatment-associated polymorphism that has a minimal effect on susceptibility to the currently available NNRTIs (Shafer, 2003).

#### 1.8.2.4.3 A98G/S

A98G causes low-level resistance to each of the NNRTIs (Petropoulos et al, 2000; Young et al, 1995; Byrnes et al, 1994). A98S is a polymorphism that does not appear to be associated with resistance (Shafer, 2003).

#### 1.8.2.4.4 L100I

L100I causes intermediate resistance to EFV and DLV and low-level resistance to NVP (Petropoulos et al, 2000; Fujiwara et al, 1998; Winslow et al, 1996; Young et al, 1995; Byrnes et al, 1994; Byrnes et al, 1993). L100I usually occurs with K103N in patients receiving EFV (Bacheler et al, 2000).

#### 1.8.2.4.5 K101E/R/Q/P

K101E causes low-level resistance to each of the NNRTIs (Petropoulos et al, 2000; Young et al, 1995; Byrnes et al, 1994). K101R/Q do not appear to be associated with drug resistance. Although K101P occurs frequently in heavily treated patients, its phenotypic impact has not yet been reported (Shafer, 2003).

#### 1.8.2.4.6 V108I

V108I causes low-level resistance to each of the NNRTI (Petropoulos et al, 2000; Young et al, 1995; Byrnes et al, 1994).

#### 1.8.2.4.7 V179D/I

V179D causes low-level resistance to each of the NNRTIs (Winslow et al, 1996; Young et al, 1995; Byrnes et al, 1994; Byrnes et al, 1993). V179I is a common polymorphism that is not associated with drug resistance (Shafer, 2003).

#### 1.8.2.4.8 Y181C/I

Y181C/I occurs commonly in patients treated with NVP and causes high-level resistance to NVP and DLV and low-level resistance to EFV (Petropoulos et al, 2000; Fujiwara et al, 1998; Young et al, 1995; Byrnes et al, 1994; Byrnes et al, 1993).

#### 1.8.2.4.9 Y188C/L/H

Mutations at this position occur in patients treated with NVP and EFV and cause high-level resistance to NVP and EFV and intermediate resistance to DLV (Petropoulos et al, 2000; Fujiwara et al, 1998; Young et al, 1995; Byrnes et al, 1994; Byrnes et al, 1993).

#### 1.8.2.4.10 G190A/S/E/C/Q/V/T

The most common mutations at this position, G190A and S, cause high-level resistance to NVP and intermediate (G190A) or high (G190S) levels of resistance to EFV (Petropoulos et al, 2000; Fujiwara et al, 1998). The other mutations at this position are also markedly less susceptible to NVP and EFV (Huang et al, 2003). Most G190 mutations cause hypersusceptibility to DLV, except G190E, which causes low-level resistance to DLV (Huang et al, 2003). Certain 190 substitutions (C, Q, V, T, and E have been shown to have markedly reduced fitness *in vitro* (Huang et al, 2003; Kleim et al, 1994). This was correlated with reduced virion-associated RT activity and incomplete PR processing of the viral p55 (gag) polyprotein (Huang et al, 2003).

### 1.9 Inhibitors of HIV-1 Fusion and Entry

Since the development of resistance to any ARV treatment regimen is unavoidable, there is a constant need for the development of new classes of drugs. T-20 (by Trimeris and Roche) is the first agent in the new class drug called fusion inhibitors that target the entry stage of the virus life cycle. The HIV-1 *env* glycoprotein consists of a surface glycoprotein (gp120) and a transmembrane glycoprotein (gp41). During HIV-1 infection, the virus's gp120 binds to both a CD4 and chemokine receptor on the target cells, followed by a conformational change in gp41 that promotes the fusion of the viral and cellular membranes (Eckert et al, 1999). For successful fusion to occur, the heptapeptide repeat domains (HR-1 and HR-2) of gp41 must form a hairpin structure (Chan et al, 1998). T-20 is a 36-amino-acid peptide that corresponds to a part of the HR-2 region (residues 127-162) and binds to HR1 near the fusion domain, preventing interactions with the HR-2 region and keeping the virus and cell membranes from fusing (Kilby et al, 1998). It has been shown to be effective for patients with viruses resistant to RT and PR inhibitors when given by means of intravenous infusion or twice-daily subcutaneous injection (Lalezari et al, 2003; Kilby et al, 1998). It has also been shown to be effective subtype C isolates *in vitro* (Cilliers et al, 2004).

In a study of treatment-experienced patients in which T-20 was added to an optimised ARV regimen, 32% of subjects achieved either a greater than 1.0 $\log_{10}$ copies/ml reduction

in plasma HIV-1 RNA levels from baseline or a viral load less than 400 copies/ml. Thirteen percent of subjects achieved a viral load less than 50 copies/ml (Lalezari et al, 2003). In another study, a regimen of APV/RTV/ABC/EFV, with or without T-20 (at 1 of 3 doses), was evaluated in PI-experienced NNRTI-naive patients. Again, T-20 recipients were more likely to achieve a plasma HIV-1 RNA level less than 400 copies/ml, with the greatest reduction in viral load observed in those subjects receiving a T-20 dose of 100 mg twice daily (Lalezari et al, 2003). However, resistance was observed to evolve in studies of T-20 monotherapy and should therefore be used as part of a regimen consisting of several active agents (Sista et al, 2002). Resistance is associated with changes in the HR1 region, particularly in the GIV motif (Wei et al, 2002; Kilby et al, 2002; Hanna et al, 2002). This motif is highly conserved among HIV-1 isolates (Hanna et al, 2002; Xu et al 2002).

### 1.10 HIV-1 and the Mechanisms of Drug Resistance

During uncontrolled HIV infection, the high HIV replication rate and RT-induced mutation rate generates every possible mutation in the HIV genome each day (Richman et al, 1994). This results in a large pool of genetically related but distinct HIV strains called quasispecies, each of which has the potential to develop into the dominant strain. During treatment with ARVs, the likelihood that resistant mutants will emerge is not only a function of the fidelity of the viral RT enzyme and rate of virus replication, but is also determined by the ability of the viral target site to mutate. For example, AZT selects for mutations in the RT more readily *in vitro* and *in vivo* than d4T, probably because of the relatively large 3'-azido group on AZT, making it even more different to the physiological thymidine than d4T. Therefore, the RT molecule can more readily mutate to interact differently with this unphysiologic sugar moiety on AZT, thus selecting for mutations more easily than against d4T. Similarly, 3TC has a distinctive sugar moiety that probably accounts for the rapidly emergence of resistance to this drug (Gao et al, 1993; Schinazi et al, 1993; Tisdale et al, 1993).

Figure 8 demonstrates the development of drug resistance during incomplete viral suppression. With increasing drug exposure, the selective pressure on the replicating virus population increases to promote the more rapid emergence of drug-resistant mutants. As long as significant levels of virus replication are sustained, the likelihood of such mutants increases. As antiviral activity increases still more, the amount of virus diminishes to the point where the likelihood of resistance emerging begins to diminish.

This likelihood becomes nil when replication is completely inhibited (Richman, 1994). However, with continuing viral replication in the absence of completely suppressive antiviral drug activity, the cumulative acquisition of multiple mutations can occur over

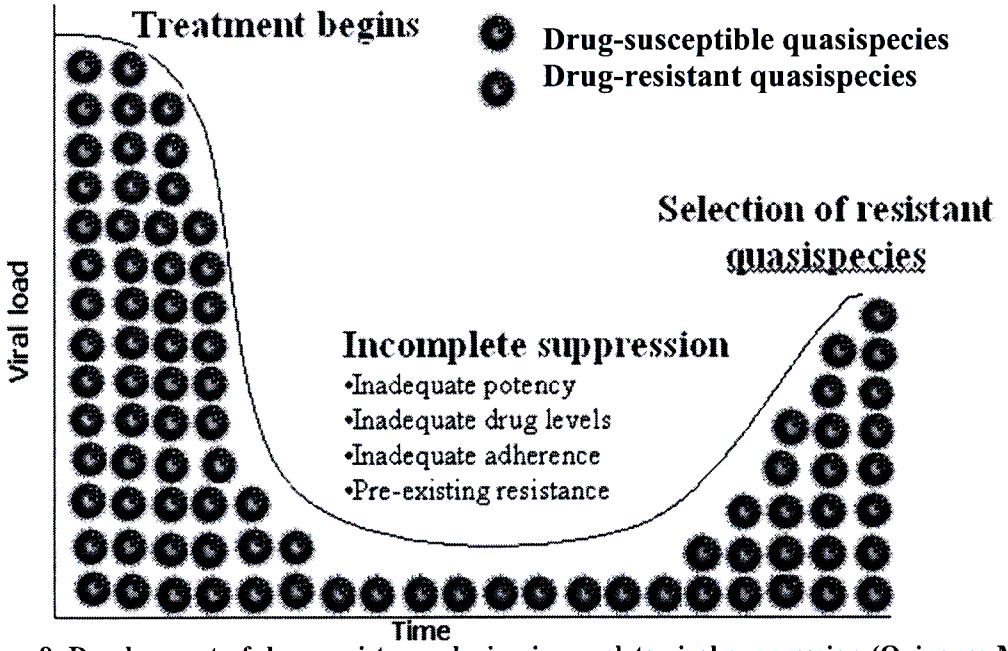



**Figure 8. Development of drug resistance during incomplete viral suppression (Quinones-Mateu and Arts, 2002)**

time (Condra et al, 1995; Larder et al, 1989). The selective pressure of drug treatment permits the outgrowth of these pre-existing mutants (Coffin et al, 1995). After the failure of a particular regimen, it is difficult to dissect out the component of the drug regimen that is attributable to the acquisition of drug resistance. This is possibly due to the fact that patients who are more likely to develop drug resistance are more likely to have confounding factors that will predispose to a poor prognosis (Richman, 1994). Also, patients may contain mixtures of certain viruses with different susceptibilities, and these populations may be represented differently in different organs of the patient under treatment. Therefore the key to ensure antiviral efficacy is total viral suppression with a potent combination ARV regimen. This is because resistant mutants do not emerge if a regimen can prevent the outgrowth of pre-existing mutants and block ongoing replication that permits the emergence of new mutants. A second approach is to switch to an alternate therapy against which the population has no cross-resistance. However, when a patient changes to a new regimen, the rate of viral replication and selection of mutations is influenced by the extent to which the mutations selected by the previous regimen also confer resistance to the new drugs. If the virus has some degree of cross-resistance, mutation-generating replication events continue to occur at a rate determined by the overall effectiveness of the new regimen, resulting in the more rapid development of resistance.

It therefore seems that the development of drug resistance is an inevitability, and therapies must be designed to optimise treatment in the face of resistant populations. One approach to sustain antiretroviral drug activity is to generate plasma drug levels that exceed the susceptibility of drug-resistant viruses (Havlir et al, 1996). Another approach is to use multiple drugs targeted to the same viral protein (convergent therapy), whereby mutations induced by drug 1 may sensitise the virus to drug 2, or may prevent the emergence of viable mutants to drug 2 (Chow et al, 1993).

### 1.11 Rate of Selection of Resistance

Besides the effect of the potency of the drug regimen and the adherence to the regimen, the inherent replication rate of a resistant virus also influences the rate of its emergence (Richman, 1994). Some resistance mutations do not appear to substantially affect the viral replication rate and will appear as the dominant quasispecies much faster, while those that compromise viral replication may take longer to emerge, or may acquire additional compensatory mutations that help to restore the viral replication rate. Also, some mutations occur as single nucleotide changes (eg, NNRTIs) and will appear faster than those that require two base changes (eg, T215Y, T69D).

### 1.12 Disappearance of Resistance Mutations

The same factors that influence the development of a resistant population also affect the "disappearance" of mutations from the dominant species (Quinones-Mateu and Arts. 2002; Richman, 1994). An example of when this occurs is during treatment interruption, where there is no longer any drug selection pressure and the most efficiently replicating strain (which could be the wild-type strain) becomes the dominant quasispecies. This "reversion to wild-type" can occur within a few weeks of therapy cessation. Although the drug resistance mutations would no longer be detectable by conventional genotypic assays, these strains still persist in the minority species. The minor populations can re-emerge rapidly if selective pressure is reapplied. Similarly, resistance mutations may be "lost" when a different drug regimen is used against a drug-resistant variant without any cross-resistance to the new regimen. Quasispecies that carry previously selected mutations that are no longer advantageous and may be associated with a reduced rate of replication, will tend to be out competed by virus strains that are able to replicate more efficiently without these mutations, and may disappear from genotypic reports (Shafer, 2003; Quinones-Mateu and Arts. 2002).

### 1.13 Hypersusceptibility

This is defined as an increased susceptibility to a particular drug, i.e. virus replication is inhibited by lower amounts of drug than those required to inhibit a reference wild-type strain. It has been shown that a number of NRTI-resistance mutations cause hypersensitivity to the NNRTIs (Whitcomb et al, 2002). Patients with NNRTI hypersusceptibility have better outcomes on NNRTIs than do non-hypersusceptible patients (Shulman et al, 2001; Haubrich et al, 2002; Hammer et al, 2002; Katzenstein et al, 2003).

### 1.14 Drug Resistance in Non-B Subtypes

With a growing demand for treatment and access to antiretroviral therapy in resource-limited settings, the susceptibility and resistance patterns of non-subtype B viruses, particularly subtype C, to antiretroviral drugs is an important question. Initial reports suggest that non-B subtypes respond equally as well as subtype B to therapy (Turner et al, 2004; Laurent et al, 2002; Shafer et al, 1997). A study of 72 drug naive subtype C isolates from South Africa found no naturally occurring primary RT or protease drug resistance mutations (Gordon et al, 2003). Another study from Zimbabwe found that subtype C

viruses were as susceptible as subtype B to commonly used nucleoside and non-nucleoside RT inhibitors (Shafer et al, 1997). Turner et al (2004) found that although polymorphisms in non-B subtypes reflect different patterns of codon usage, this may only rarely affect the patterns of drug resistance-associated mutations. On the other hand, a study by Dumans et al (2004) found that a polymorphism at position 210 in subtype F resulted in a lower prevalence of the L210W mutation in subtype F patients from Brazil. In addition, some studies have reported differences in the response of non-B isolates to therapy (Gonzalez et al, 2004; Grossman et al, 2002a; Navratne et al, 2002; Loemba et al, 2002; Loveday et al, 2002). Loveday et al (2002) found a significantly inferior virologic response in subtype C versus subtype B infected patients at 24 and 48 weeks of follow-up.

In the PR region, the D30N appears at very low rates in NFV treated non-B isolates (Gonzalez et al, 2004; Grossman et al, 2004; Navratne et al, 2002), while common in subtype B isolates treated with NFV. Gonzalez et al (2004) found that the D30N mutation drastically reduced subtype C fitness *in vitro*, which could account for the lack of prevalence of this mutation in subtype C treated isolates. Non-B isolates failing NFV treatment more commonly developed the L90M mutation (Gonzalez et al, 2004; Grossman et al, 2002a; Navratne et al, 2002). The N88S mutation was more common in CRF01_AE strains in Japan, with D30N and N88D not seen at all. N88S is also frequently associated with L90M in subtype G isolates (Camancho et al, 2004). The rate of some protease resistance mutations was higher in subtype B sequences (positions 10, 30 and 90), while others were higher in subtype C sequences (positions 36, 41, 89 and 93) (Auerbuck et al, 2002). Similarly, Qari et al (2004) found distinct patterns of resistance with PI treatment in subtype F and A/G isolates. They also found that L10V and V77I amplified PI resistance in their subtype F and A/G isolates. Within the RT, subtype C viruses were found to develop the key nucleoside excision mutations (NEMS) at positions 41, 67, 70, 210, 215, and 219, but to a lesser extent than subtype B viruses (Bollyky et al, 2002; Grossman et al, 2002b). Pillay et al (2002a) from South Africa, found that a high percentage (6/12) of subtype C positive infants treated with ddI plus d4T developed resistance, including the multi-drug resistant Q151M mutation. Novel mutations present at higher rates in treated persons infected with subtype C include mutations at RT positions 20, 36, 48 53, 123, 174 and 214 (Bollyky et al, 2002; Kantor et al, 2002b). Spira et al (2003) also report a cluster of silent mutations in subtype C isolates at amino acid positions RT 65, 138 and 161 linked to NRTIs resistance *in vitro*.

Differences also occur for the NNRTI class of drugs. The most notable difference is the V106M resistance mutation seen in subtype C patients treated with EFV, which does not arise under pressure by NVP or DLV (Brenner et al 2003). The V106M mutation is said to confer high-level multi non-nucleoside RT inhibitors (NNRTI) resistance. This mutation

occurs because the sequence for valine in subtype C isolates is GTG, while it is GTA in subtype B. The change from V to M (ATG) is therefore easier in subtype C isolates as it only requires a single nucleotide substitution, while the V to M mutation in subtype B isolates would require two substitution events. The V106A (GTA to GCA) mutation occurs more commonly in subtype B isolates, and causes NVP resistance without any cross-resistance to the other NNRTIs. Eshleman et al (2004b) found that subtype influences the selection and fading of the NVP resistance mutations K103N and Y181C. In a study on the pattern of NNRTI resistance in non-B subtypes, Cane et al (2004) found that the G190A mutation occurred more frequently in subtype A and C isolates than subtype B. The impact of other signature sequences in subtype C and other non-B subtypes on response to ARV therapy needs to be further investigated (Kantor et al, 2003a; Loemba et al, 2002; Grossman et al, 2001).

### 1.15 Transmission of ARV Drug Resistance

The transmission of HIV-1 viruses that are resistant to ARVs has serious implications both for the treatment of individuals and for public health. Individuals infected with a resistant strain of HIV often have fewer drugs from which to choose as many of the drugs are cross-resistant and this lowers the chances of that individual benefiting from long-term therapy. This may also have grave implications for perinatal transmission prevention programs, as drug resistant virus circulating in these areas may cause these programs to fail. Poor adherence because of stigma, side effects and cost, is one of the main causes of the development of drug resistance (Little, 2001). There is a growing concern that drug-resistant strains will spread as more people take ARVs without full viral suppression. On the other hand, some researchers feel that while the spread of drug-resistant virus is a threat, it is unlikely to have an impact at epidemic levels since people with resistant HIV strains also have wild-type strains in their pool of viruses, which are possibly more transmissible than resistant strains (Boden et al, 1999; Loveday et al, 1999). There is emerging evidence that resistant virus is less infectious, with a recent study suggesting that resistant virus has only a 25% capacity to infect others compared to wild-type virus (Leigh Brown et al, 2003). However, another study has found that in individuals who were infected with resistant viruses, all the resistant mutations detected in the index case persisted in the infected patients, even after they had disappeared from the index case after treatment interruption (Ravaux et al, 2003). Another study has shown that transmission-acquired resistance does not appear to influence the rate of disease progression, as it does in individuals that develop resistance (Pillay et al, 2002b).

There is currently an upward trend in the transmission of drug resistance, probably as a result of the increased use of ARV therapy. Studies from the 1990s found less than 10% of new infections involved drug-resistant virus, but more recent studies in Europe and the US have shown a dramatic increase of up to 27% (Wensig et al, 2003; Little et al, 2002; UK Collaborative Group 2001). However, a Swiss study found that the transmission of drug-resistant viruses peaked in 1997, contradicting the notion of a rise in transmission of drug-resistant virus (Yerly et al, 2001). They suggested that better control of viremia over time and greater attention to the need for adherence might have reduced the risk of transmission of resistant virus (Yerly et al, 2001).

Little surveillance data exists on the prevalence of drug resistant HIV in resource-limited settings. The only country in which resistance has been systematically studied in a treatment naïve population is Brazil. A study by Dumans et al (2002) conducted on sample collected in 1998, after the introduction of free access to HAART in Brazil in 1996, found that there were no primary resistance mutations in either the PR or RT region. A national survey conducted on samples collected in 2001 found a slight increase, with around 2% showing genotypic resistance to each of the three drug classes and 8% showing resistance to at least one drug (Brindeiro et al, 2003). A smaller pilot survey among 71 treatment naïve patients in Rio de Janeiro found only one case of drug resistance, associated with 3TC (Dias Tavares et al, 2003). To increase the data in these regions, the WHO and the International AIDS Society have established a joint surveillance project to track prevalence of genotypic resistance in treatment naïve individuals after the introduction of HAART.

### 1.16 Resistance Testing

HIV drug resistance testing is an invaluable asset in HIV treatment management (Kuritzkes, 2004; Katzenstein, 2003). It is designed to identify gene mutations or viral growth characteristics that suggest reduced drug susceptibility. Current methods of HIV resistance testing include *in vitro* genotypic assays and phenotyping (drug-susceptibility assays) that detect mutations known to confer drug resistance. The HIV genotype refers to the actual DNA sequence of the virus, while the phenotype reflects the physical traits expressed by the genotype. There are limitations to both genotypic and phenotypic assays: they are both unable to detect minor drug resistant populations and there is limited knowledge about the clinical significance of certain combinations of mutations and of certain levels of phenotypic drug resistance. (Hirsch et al, 2003; Vandamme et al, 2001; Shafer et al, 2000).

Genotypic assays are more commonly used in clinical settings because of their wider availability, lower cost and quicker turnaround. Generally they utilize a two-step procedure of PCR to amplify a specific region of the HIV genome (RT and PR) and a specific mutation detection methodology that distinguishes each type of genotyping assay. The three most prevalent mutation detection methods are DNA sequencing, gene chip arrays and the line probe assay (Wilson, 2003). DNA sequencing has been the gold standard genotyping technology and has the advantage of identifying potentially every drug resistance associated mutation (Richman, 2000). There are variety of sequencing assays available: the HIV-1 GenotypR PLUS (Specialty Laboratories); TRUGENE HIV-1 genotyping test (Visible Genetics); VircoGEN II (Virco), Viroseq HIV genotyping system (Applied Biosystems); GeneSeq (ViroLogic), HIV-1 Mutation Analysis (Focus Technologies), HIV ViroTYPE (Rheumatology Diagnostoc Laboratory); GenoSure (LabCorp and Virco); and HIV-1 Genotype (Quest Diagnostics) (Tobin and Frenkel, 2002; Wilson, 2003). The value of genotyping depends on an understanding of the relationships between identified mutations and specific drug susceptibility, as well as viral fitness (Wilson, 2003).

There are a number of on-line databases to assist in genotype interpretation: the Los Alamos HIV Sequence Database (http://hiv-web.lanl.gov); the Stanford HIV RT and Protease Sequence Database (http://hivdb.stanford.edu/hiv); the HIVResistanceWEB (http://www.hivresistanceweb.com); and The International AIDS Society-USA (http://www.iasusa.org).

Phenotypic tests measure virus drug susceptibility, resulting from known or unknown resistance-related mutations and their interactions. Various biological methods have been developed to assess the drug susceptibility of HIV-1 (Brun-Vezinet et al, 1992; Japour et al, 1993; Larder et al, 1990). In most cases these assays require the cocultivation of patient peripheral blood mononuclear cells (PBMCs) with donor PBMCs to obtain a viral stock. Kellam et al, (1994) and later Maschera et al, (1995) developed an innovative method for generating viruses for drug susceptibility testing, called the recombinant virus assay (RVA). RVA involves the production of viable virus *in vitro* by homologous recombination of RT-PCR products from plasma virus with an infectious PR or RT deleted cloned HIV-1 provirus. The resulting recombinant viruses derive all their biological properties from the subtype B molecular clone, except for RT and PR, which are encoded by the genes from the patient isolates. Boucher et al (1996) improved on this method, using the cell-killing assay described by Pauwels et al (1988), instead of the limiting HeLa CD4+ plaque reduction assay to assess the drug susceptibility of the viruses. The cell-killing assay measures the capacity of a virus to induce lysis of target

**Table 1. Recommendations from expert panels for resistance testing.**

| Presentation | International AIDS Society-USA (Hirsch et al, 2003) | US Department of Health and Human Services (2003) | Euro-Guidelines Group (2001) |
|---|---|---|---|
| Failure of first drug regimens | Recommended | Recommended | Recommended |
| Failure of subsequent drug regimen | Recommended | Recommended | Recommended |
| Primary/acute HIV-infection | Recommended | Recommended | Consider testing |
| Established (untreated) HIV infection | Recommended | Recommended | Consider testing |
| Pregnancy | Recommended | Recommended | Recommended |
| Post-exposure prophylaxis | Recommended | Recommended | Recommended index case |
| After discontinuation of drugs | -- | Not recommended | -- |
| Paediatrics | -- | -- | Recommended |

cells. Drug susceptibility results are reported as the amount of drug required to inhibit viral replication by 50% ($IC_{50}$) or 90% ($IC_{90}$). The result is reported as the fold-change in viral susceptibility compared with a wild-type reference strain. HIV strains deemed "resistant" to a particular drug might still be inhibited by that drug, although higher concentrations of drug are required. Whether that concentration of drug is pharmacologically achievable *in vivo* determines the efficacy of that drug against those less susceptible strains.

The widespread use of resistance testing has led to the development of formal guidelines by expert panels (Table 1). The International AIDS Society-USA Consensus Panel on Resistance Testing, the Panel on Clinical Practices for Treatment of HIV infection of the US Department of Health and Human Services (DHHS) and the EuroGuidelines Group for HIV Resistance have all published recommendations for the appropriate use of drug resistance testing (US Department of Health and Human Services, 2003; Hirsch et al, 2003; The EuroGuidelines Group for HIV Resistance, 2001). All three panels recommend that information obtained through resistance testing be used in conjunction with a patient's treatment history, viral load and immunologic status, as well as medication tolerance and compliance. All panels also recommend drug resistance testing in first and multiple drug treatment failures, as well as prior to starting therapy in patients with acute or established infection (Shafer, 2003).

## 1.17 Objectives

The objectives of this thesis are:

To define the spectrum of naturally occurring resistance-associated polymorphisms and mutations in RT and protease of subtype C viruses;

To determine how these polymorphisms and mutations impact on the structure-function relationships of the RT and protease;

To determine how these genetic substitutions impact on response to antiretroviral therapy, focusing on drug regimens that are affordable, safe and effective for use in Africa.

# Chapter 2

## Sequence Analysis Methods and Bioinformatics Tools

## 2.1 Sequence Analysis Methods

Bioinformatics is a rapidly developing branch of biology that uses techniques and concepts from informatics, statistics, mathematics, chemistry, biochemistry, physics and linguistics. Bioinformatics tools can be used to assemble sequences into a contiguous consensus sequence (contig), perform sequence alignments, perform phylogenetic analyses, translate DNA sequences into protein sequences, as well as predict protein structure and function. This thesis involves extensive analysis of HIV-1 *env*, RT and *pol* sequences using a broad usage of different bioinformatics tools.

To rule out contamination between samples before beginning any sequence analysis, each new sequence was compared to other sequences amplified at the same time, as well as to other sequences previously amplified in our laboratory. Newly-acquired sequences were analysed by constructing a phylogenetic tree, and analysing the tree for extreme inter-patient clustering or extreme intra-patient divergence. Sequences were also compared to published sequences by performing a BLAST search (Altschul et al, 1997) at the Los Alamos website (http://hiv-web.lanl.gov/content/hiv-db). Blast is a program that finds sequences with very high similarity to the query sequence, which could indicate contamination. Sequences datasets were aligned with CLUSTALW (Thompson et al, 1994), a program for multiple sequence alignment, and manually edited based on the codon alignment (aligning the nucleotide sequence using the amino acid code). This was done using the GDE (Genetic Data Environment) v 2.2 (Smith et al, 1994) -based Linux interface (De Oliveira et al, 2002) which is discussed in more detail below. Phylogenetic trees were initially generated with the F84 model of substitution and the neighbour-joining method (version 4.0b2a/4.0b10) of PAUP* (Phylogenetic Analysis Using Parsimony) (Swofford, 1999) (Sinauer Associates, Sunderland, Mass.). For subtype analysis, new sequences were compared to subtype reference strains available at the Los Alamos subtype database (http://hiv.lanl.gov/content/hivdb/SUBTYPE_REF/align.html). For these analyses, the phylogenetic trees were rooted with a homologous region of HIV-1 group O (OCM_MP5180). To investigate whether the sequences were recombinant forms of HIV-1, recombination analyses were performed with Simplot (Lole et al, 1999), a method that uses a sliding-window approach to calculate bootstrap plots for constructing neighbor-joining trees with the DNADIST, NEIGHBOR, or CONSENSE programs of the PHYLIP package (Felsenstein, 1989). To further examine inter- and intra-patient relationships, subtype C specific trees were constructed with selected reference sequences obtained from the GeneBank and Los Alamos public databases. To ensure that the best model of evolution was used for these analyses, the respective datasets were tested using Model

Test v 3.0 (Posada and Crandall, 1998). This program runs within PAUP*. The best model was then pasted into the sequence file for that particular dataset (in nexus format) and used for subsequent analysis.

To investigate sites under positive selection, the codeml option in the PAML (Phylogenetic Analysis by Maximum Likelihood) (Rambaut, 2000) software package was used to calculate the ratio of nonsynonymous to synonymous amino acid substitutions (*dn/ds*) or ω. This value is a measure of natural selection pressure at the protein level and provides a powerful tool for understanding the mechanisms of DNA sequence evolution. An individual amino acid was considered to be positively selected if the *dn/ds* ratio was significantly greater than 1.0.

Ancestral sequences were reconstructed using the baseml application in the PAML package. This method involved the use of maximum-likelihood methods and a nucleotide substitution model to identify nucleotide substitutions along each branch of the tree (Yang, 2000; Yang et al, 2000). Genetic diversity was measured for individual datasets using a Kimura 2-parameter (K2P) model implemented in the MEGA (Molecular Evolutionary Genetics Analysis) program version 2.0 (Kumar et al, 2001) (Arizona State University, Tempe). Nucleotide sequences were also translated and analysed for the presence of signature patterns using VESPA (Viral Epidemiology Signature Pattern Analysis) (Korber and Myers, 1992). Biologically important sites were predicted using PROSITE, a database of protein families and domains, which was run in GENEDOC, a windows-based multiple sequence alignment editor. The different software packages and their applications are briefly discussed below.

### 2.1.1 Likelihood Ratio Test (Anisimova et al, 2001)

All models in the Likelihood Ratio Test (LRT) are nested i.e. all models are a special case of the General Time Reversible (GTR) model. For example, the K2P model is a special case of the more complex HKY85 model because HKY85 assumes that base frequencies can vary, while the K2P model assumes equal base frequencies (see Table 2). Therefore, the K2P model uses less parameters and is a simpler model. Simpler models are quicker to run and have less error associated with model parameters. However, more complex models account take additional variables into account (Table 3) and, thus are more likely to lead to the true tree.

**Table 2. The most commonly used evolutionary models are described. The models increase in complexity from the Jukes-Cantor (simplest) to General Time Reversible (most complex) model.**

| Model | Substitution rate/base frequency | Number of free parameters |
|---|---|---|
| Jukes- Cantor (JC) | One rate of substitution<br>Equal base frequencies | No free parameters |
| F81 | One rate of substitution<br>Unequal base frequencies | Three free parameters for base frequencies |
| Kimura's Two Parameter (K2P) | Two types of substitution (tratio)<br>Equal base frequencies | One free parameter for tratio |
| HKY85 | Two types of substitution (tratio)<br>Unequal base frequencies | One free parameter for tratio<br>Three free parameters for base frequencies |
| General Time Reversible (GTR) | Six types of substitution<br>Unequal base frequencies | Five free parameters for substitution rate matrix (rmatrix)<br>Three free parameters for base frequencies |

**Table 3**

The addition of among-site variation to the above models increases their complexity as follows:

| | |
|---|---|
| A proportion of sites invariant (I) | One additional free parameter for *pinvar* |
| Gamma distribution (G) | One additional free parameter for gamma rate parameter |
| Both combined (G+I) | Two additional parameters |

The LRT can test whether the more complex model of evolution is significantly better than a similar, but simpler model. The most commonly used models of evolution are listed in Table 2. The LRT uses the following formula:

LRT = 2([-lnL*null*] – [-lnL*alternative*])

Where L is the maximum likelihood of the tree topology under the hypothesis and data. The degrees of freedom (equal to the difference in the number of free parameters) is used to calculate the Chi Square Critical Value, which determines whether the null hypothesis is significantly better than the alternative model. For example, to test if the more complex GTR model (8 free parameters) is better suited to the data than the simpler HKY85 model (4 free paramters), the –lnL of that topology and data for each model is calculated. Eg: GTR –lnL=1784.84185; HKY85 –lnL=1787.08478. Because HKY85 is a special case of the GTR model, it is the *null* hypothesis. Therefore:

LRT = 2(1787.08478 – 1784.82185) = 4.52586

Degrees of Freedom = 8-4 = 4

Chi Square Critical Value = 9.488

Since the LRT value is less than 9.488, the null hypothesis for two types of substitution is not significantly different to 6 types of substitution, and the simpler model can be used in the analysis.

### 2.1.2 Model Test (Posada et al, 1998)

Modeltest 3.0 evaluates 56 models of evolution/variations within the framework of a nested hierarchical analysis. Briefly, the data file is opened in PAUP* and the control file (modelblock3) is executed. This file contains a set of PAUP* commands that calculates a neighbour-joining tree, evaluates the tree under the entire suite of models, and saves the likelihood scores to a separate file. The models evaluated in this step include the simplest (JC) to the most complex, or parameter-rich (GTR) model, with the sequential addition of a proportion of invariant sites and gamma-distributed rates. The output (model.scores) is then run through modeltest. This program runs a series of log-likelihood tests on nested model comparisons using the LRT described above. Modeltest also makes another comparison of the models, with the Akaike Information Criterion (AIC), which allows for comparison of different models without the nested requirement. The AIC rewards models for good fit but penalizes extra parameters. Each model gets an AIC score. Smaller values of AIC indicate better models and the one with the best AIC score is selected.

$AIC = -2 \ln L + 2n$ where n = # of independently adjusted parameters in model

### 2.1.3 PAUP* (Swofford, 1999)

PAUP* is used to infer and interpret phylogenetic trees. It provides full support for tree searching under DNA maximum-likelihood and distance-based optimality criteria (in addition to parsimony). A wide range of DNA substitution models are available, including the GTR model and its submodels (Jukes-Cantor, Kimura 2P, HKY85, F84 used in DNAML and fastDNAml, Tamura-Nei). PAUP* also allows the investigator to choose the structure of any model for likelihood, with the option of either having the program estimate the parameters during the analysis (increased computational time), or letting the investigator set the parameters to values obtained otherwise (for example, from Modeltest).

### 2.1.4 SimPlot (Lole et al, 1999)

SimPlot identifies recombinant genomes. Although SimPlot reads most sequence file formats, for our purposes, all files were saved in fasta format. To run the program, the query sequence (the suspected mosaic) was aligned with a set of reference sequences. Note that SimPlot can analyse no more than 26 groups of sequences (or individual sequences. The program generates a graph consisting of a set of lines (or

optionally strings of points) that reflect the similarity (or distance) of each Reference sequence (or Group) to the query sequence. In order to generate this plot a sliding window is passed across the alignment in small steps (the window size and step size are selectable). For the analyses, a sliding window of 400 bp and a step-size of 50 bp was used . The BootScanning option in the SimPlot software package was used to identify the breakpoints in the recombinants (Salminen et al, 1995). To perform BootScanning, the PHYLIP software suite (Felsenstein) is needed, which is available for download from ftp://evolution.genetics.washington.edu/pub/phylip. In this application, a minimum number of three reference strains are compared with the query sequence. Again, a window size of 400 bp and a step-size of 50 bp was used to calculate bootstrap plots for the construction of neighbour-joining trees with DNADIST, NEIGHBOR or CONSENSE programs of the PHYLIP package.

### 2.1.5 PAML (Rambaut, 2000; Yang et al, 2000)

PAML is a package of programs for phylogenetic analyses of DNA or protein sequences using maximum likelihood (ML). It is maintained and distributed for academic use free of charge by Ziheng Yang. It has many capabilities including, estimating the branch lengths in a phylogenetic tree and assessing parameters of evolution including the transition/transversion rate ratio, the shape parameter of the gamma distribution for variable evolutionary rates among sites, and rate parameters for different genes. It can also be used for testing hypotheses concerning sequence evolution, such as rate constancy and independence among nucleotide or amino acid sites, rate constancy among lineages (the molecular clock), and homogeneity of evolutionary process in multiple genes. Finally, it can also be used for calculating substitution rates at each site and reconstructing ancestral nucleotide or amino acid sequences, as well as phylogenetic tree reconstruction by maximum likelihood and Bayesian methods.

In this thesis, two different applications of the PAML software package were used – codeml program to identify sites under positive selection, and the baseml program for ancestral reconstruction and estimation of the mutation rate.

#### 2.1.5.1 Baseml

Baseml can perform ML analysis of nucleotide sequences: estimation of tree topology, branch lengths, and substitution parameters under a variety of nucleotide substitution models (JC69, K80, F81, F84, HKY85, TN93, REV); constant or gamma rates for sites; molecular clock (rate constancy among lineages) or no clock, among-gene and within-

gene variation of substitution rates; models for combined analyses of multiple sequence data sets; calculation of substitution rates at sites; reconstruction of ancestral nucleotides.

The default control file for baseml is baseml.ctl. An example and the description of each of the parameters is shown in the appendix.

#### 2.1.5.1.1 Estimation of Mutation Rate

The option clock=3 in the baseml.ctl file implements Andrew Rambaut's TipDate models (Rambaut 2000). For this application, the dates are specified at the end of the sequence names. To run this analysis, a ML tree was constructed in PAUP* using the appropriate model of evolution. The same dataset was exported in PHYLIP v3.2 format. The file was then edited (in a text editor) to remove the "YF" found at the beginning of the file. The name of the sequence file and tree file were entered into the baseml.ctl control file. After running the analysis, the mutation rate was obtained from the output "rst" file.

#### 2.1.5.1.2 Reconstruction of Ancestral Sequences

Nucleotides of extinct ancestors can be reconstructed based on contemporary sequence data. This was originally done using Parsimony methods based on the criterion that the number of changes along the tree at each site is minimized. Algorithms based on this criterion were developed by Fitch (1971) and Hartigan (1973). Baseml uses a likelihood approach to determine the branch lengths and the substitution pattern for ancestral reconstruction and assigns a nucleotide to an interior node at a site according to its posterior probabilities (Yang et al, 1995). To run the reconstruction, a ML tree was constructed in PAUP* using the appropriate model of evolution for that dataset. The same dataset was exported in PHYLIP v3.2 format. The file was then edited (in a text editor) to remove the "YF" found at the beginning of the file. The name of the sequence file and tree file were entered into the ancestor.ctl control file. After running the analysis, the results were collected in the "rst" file. The tree file was extracted from the "rst" file and viewed in TREETOOL. The respective ancestral nodes were identified for each patient. The "rst" file also contained the reconstructed nucleotide sequences for each ancestral node. These were saved in fasta format and translated into the amino acid sequences in GDE. Patient sequences were then compared to the generated ancestral sequences.

### 2.1.5.2 Codeml

Codeml performs ML analysis of protein-coding DNA sequences using codon substitution models (Goldman and Yang 1994). The program allows for the calculation of the codon-usage table; estimation of synonymous and nonsynonymous substitution rates; likelihood

ratio test analysis of positive selection or relaxed selective constraints along lineages based on the $d_n/d_s$ rate ratios; identification of amino acid sites or evolutionary lineages potentially under positive selection; reconstruction of ancestral codon sequences. The format of the sequence files and the requirement of a ML tree from the same dataset was the same as for baseml. The variables in the program files were also used in the same way as the baseml.ctl file, except that the option for codon sequences was used. The number of substitution sites (nssites) were varied according to the model for the dN/dS ratio (ω) (Nielsen and Yang, 1998; Yang et al, 2000). Nssites = m corresponds to model Mm in Yang et al (2000). Three models were used: neutral (1), selection (2) and discrete (3). The posterior probabilities for site classes as well as the expected ω values for sites were listed in the "rst" file, used to pinpoint sites under positive selection.

### 2.1.6 VESPA (Korber and Meyer, 1992)

The VESPA program detects signature patterns (distinctive amino acid or nucleotide residues) in a set of query sequences relative to a set of reference sequences. VESPA calculates the frequency of each amino acid (or nucleotide) at each position (column) in an alignment for the query and the reference set, and selects the positions for which the most common character in the query set differs from that in the background set. For this application, nucleotide sequences were translated into amino acid sequences and exported in NEXUS, non-interleaved format, using GDE2.2. The NEXUS file was then opened in a text editor and all other information besides the sequence names and amino acid sequences were removed. VESPA was run using a "query set" and a consensus subtype A, B, C or D sequence as the "background" set. The output file was opened in a text editor to extract the results.

### 2.1.7 PROSITE (Hulo et al, 2004)

PROSITE is a curated database of protein families and domains hosted by the Swiss Institute of Bioinformatics. It is based on the observation that most proteins can be grouped, on the basis of similarities in their sequences, into families that share functional attributes. By analyzing the constant and variable properties of such groups of similar sequences, it is possible to derive a signature for a protein family or domain, which distinguishes its members from all other unrelated proteins. These protein signatures, which are detailed in the PROSITE database and listed in Table 4, were used to assign sequences to a specific family of proteins and consequently, function.

**Table 4. Protein signature sequences**

| Function | Signature Sequence |
|---|---|
| N-glycosylation site | N-{P}-[ST]-{P} |
| cAMP- and cGMP-dependent protein kinase phosphorylation site | [RK](2)-x-[ST] |
| Protein kinase C phosphorylation site | [ST]-x-[RK] |
| Casein kinase II phosphorylation site | [ST]-x(2)-[DE] |
| Tyrosine kinase phosphorylation site | [RK]-x(2,3)-[DE]-x(2,3)-Y |
| N-myristoylation site | G-{EDRKHPFYW}-x(2)-[STAGCN]-{P} |
| Amidation site | x-G-[RK]-[RK] |

### 2.1.8 SWISS-MODEL (Schwede et al, 2003)

Swiss-Model (http://www.expasy.org/swissmod/SWISS-MODEL.html) is a site for 3D homology modelling of protein structure. The Swiss-Model process is detailed as follows: The software program BLASTP2 finds all similarities of the target sequence with sequences of known structure from the ExNRL-3D database. The program SIM then selects all templates with sequence identities above 25% and projected model size larger than 20 residues. Input files are then generated for ProModII, which generates all the models found in ExPDB. Gromos96 performs energy minimisation of all models. Models can be visualised with Deep View Swiss-pdbViewer, an interface allowing the analysis of several proteins at the same time (Guex and Peitsch, 1997).

### 2.1.9 PhD (Rost et al, 1994)

Secondary structure was predicted using the PhD software program via the Predict protein server (http://www.embl-heidelberg.de/predictprotein/ predictprotein.html). PhD uses a standard feed-forward neural network. The first network predicts the secondary structure of the central residue in a 13-residue window. Afterwards the predicted secondary structure is fed into the second network which also predicts the secondary structure using the output from the first net as its input. Finally several different networks are created, using slightly different parameters for the training and a jury network is trained using the output from the two earlier layers. After the predictions, if a helix is greater than three residues long, it is kept as a helix, and if it is three or less, it is changed into a loop. Prediction of secondary structures is a step towards the prediction of the three-dimensional structure of a protein by providing a rough estimate of structural features, and with at least one known homologue available, the PhD method has an expected overall accuracy of 71.4% for proteins (Rost et al, 1994).

### 2.1.10 HIV-1 Genetic Data Environment (GDE)-Based LINUX Interface (De Oliveira et al, 2002)

This was another tool developed in our laboratory (De Oliveira et al, 2002), which was available for the analysis of the sequence data. GDE is a "front-end" sequence analysis program originally developed for Sun UNIX™ systems with an OpenView X Window manager (Smith et al, 1994). GDE was subsequently adapted to Seqlab, a graphical user interface (GUI) that incorporates most of the software distributed in the GCG toolset (Womble, 2000). LINUX, a re-implementation of UNIX™, is one of the most frequently-used operating systems. The combined GDE-LINUX approach, GDE-L, reduces the complexity and repetitive nature of input/output formatting and facilitates the development of user-defined local databases that are population-, pathogen- and/or disease-specific. New software is integrated into GDE-L by editing a unique menu file that controls the menu appearance (.GDEmenus). This same menu file is then used to send the parameter and input files to applications such as BLAST, CLUSTALW, PAUP* and READSEQ. All formats supported by the READSEQ sequence conversion software are readily accepted in GDE-L. Newly-created .GDEmenus files are then copied to a home directory. Accessory PERL scripts permit the automatic integration of sequence and BLAST-formatted datasets without editing the menu control file. All of the sequence-specific databases, phylogenetic datasets and programs needed to perform these analysis have been integrated into a single HIV-1 GDE-Linux interface. This interface was used for sequence location, (numbering according to the HXB2 reference strain), contamination detection, multiple alignment using CLUSTALW (Thompson, 1994), and genetic subtyping using PAUP* (Swofford, 2000).

# Chapter 3

## Molecular Characteristics of Human Immunodeficiency Virus Type 1 Subtype C Viruses from KwaZulu-Natal, South Africa: Implications for Vaccine and Antiretroviral Control Strategies

### 3.1 Introduction

One of the most dramatic changes in the global AIDS pandemic has been the rapid emergence and devastating spread of human immunodeficiency virus type 1 (HIV-1) subtype C (De Oliveira et al, 2001; Kuiken et al, 1999; Rollins et al, 2002; Sharp et al, 1994; UNAIDS/W.H.O. Working Group on Global HIV/AIDS and STD surveillance, 2000). As a result of this rapid escalation, HIV-1 C viruses now account for more than 56% of all global infections (Esparza et al, 2000). First identified in retrospective specimens from Ethiopia and South Africa (Johansson et al, 1995; Salminen et al, 1996; Zacharova et al, 1997), subtype C began a devastating spread across southern Africa in the late 1980s (De Oliveira et al, 2001). Major outbreaks have now occurred in every country of southern Africa, with some regions reporting adult prevalence rates as high as 40% (Department of Health/Directorate Health Systems Research; Rollins et al, 2002; Van Harmelen et al, 1999). Recent studies suggest that subtype C is spreading northward into the Congo, Tanzania, Burundi, and Kenya, where it is becoming increasingly predominant relative to other subtypes (Janessens and Nkengasong, 1997; Koch et al, 2001; Rodenburg et al, 2001). C viruses also dominate the rapidly expanding epidemic in India (Shankarappa et al 2001) and are increasing in frequency in China (Gao et al, 1998; Rodenburg et al, 2001; Yu et al, 1998) and Brazil (Brindeiro et al, 1999; Soares et al, 2003). C/D recombinants have been identified in several countries, including Tanzania, Kenya, and India (Hoelscher et al, 2001; Koulinska et al, 2001; Renjifo et al 1998), and C/B recombinants have been detected in China (Yang et al, 2002). The reasons for the increase in HIV-1 C are not known but may be related to host, viral, or socioeconomic factors. At the viral level, it has been suggested that an extra NF-KB binding site in the long terminal repeat may enhance gene expression, altering the transmissibility and pathogenesis of C viruses (Tatt et al, 2001). Others have suggested that C viruses may be more stable and that their protease genes may have increased catalytic activity relative to other subtypes (Velazquez-Campoy et al, 2001). Additional features of subtype C include a five-amino-acid insertion in the transmembrane domain of Vpu (McCormick-Davis et al, 2000), a prematurely truncated second exon of *rev* (Gao et al, 1998; Rodenburg et al, 2001; zur Megede et al, 2002), and an increase in amino acid variation at protease cleavage sites (de Oliveira et al, 2003). Recent advances in sequencing and bioinformatics (De Oliveira et al, 2003; Posada and Crandall, 1998; Pybus et al, 2001; Yang et al, 2000) make it easier to analyze full-length HIV-1 sequences and correlate the genetic information with the immunological and biological properties of the virus. These advances, combined with the development of promising vaccine candidates and

simplified, more affordable drug regimens, are paving the way for enhanced prevention and treatment efforts in southern Africa. As with HIV-1 B, it is expected that safe and efficacious treatment of C infections will not only reduce the morbidity and premature death associated with HIV-1 and AIDS (Garcia et al, 2002; Inter-Agency Task Team on Mother-to-Child Transmission of HIV; Kazatchkine et al 2000; Palella et al, 1998) but will also play a role in reducing transmission (Jackson et al, 2001). Since we are on the brink of implementing intervention strategies in a region of the world where subtype C infections predominate, it is urgent that we collect information that will help define the phylogenetic relationships, transmissibility, and drug responsiveness of C viruses. In this study, we analyzed the C2V5 and *pol* subgenomic regions of 72 contemporary viruses from KwaZulu-Natal and compared the results with those for 18 retrospective C isolates from South Africa.

## 3.2 Methods

### 3.2.1 Specimen Collection and Processing

A total of 72 treatment-naive HIV-1-infected children ($n$=16) and adults ($n$=56) representing different ethnicities, genders, age groups, and stages of disease were selected for study. Samples were obtained in Durban and surrounding areas, including Ulundi and the Hlabisa region of northern Kwazulu-Natal and Tongaat and Phoenix in the coastal region north of Durban. Participants were recruited from among symptomatic and asymptomatic adult patients, tuberculosis patients, women and children attending district health clinics, and children being treated for pneumonia. After obtaining informed consent, blood samples were collected in EDTA anticoagulant tubes (most adult patients) or as dried blood spots (most pediatric patients). Plasma was isolated within 6 h of collection; dried blood spots were stored with desiccant at -20°C until analyzed.

### 3.2.2 Viral Load and CD4+ T-cell counts

RNA was extracted from plasma and dried blood spots with a guanidinium-silica method (Nuclisens isolation kit; Organon Teknika) and an automated extractor (Organon-Teknika). Virus levels were measured with the Nuclisens HIV-1 QT kit, an assay with a quantitative range of 40 to >500,000 copies of HIV-1 RNA/ml of plasma. When applied to 50 µl of dried blood, the lower limit of detection is 1,600 HIV-1 RNA copies/ml of blood. Specificity of the method has been previously assessed and shown to be greater than

98.9% (Cassol et al, 1997). CD4+ cell counts in venous blood were determined according to a standard FACSCount method.

### 3.2.3 Sequencing of the Envelope C2V5 Region

Sequencing of *env* was performed directly on a 621-bp PCR product generated from the C2V5 region (nucleotides 7026 to 7646, relative to HXB2) (Korber et al, 1998). RNA was extracted from plasma with the ViroSeq method (Applied Biosystems). Plasma RNA and Nuclisens-extracted dried blood spot RNA were reverse transcribed to cDNA with Superscript II and random hexamer primers (Invitrogen Corp., San Diego, Calif.). The RNA template and random primers (100 ng) were heated to 70°C for 10 min, chilled on ice, and reverse transcribed at room temperature in a 20 µl reaction volume containing 1X reaction buffer, 10 mM dithiothreitol, 0.5 mM each deoxynucleoside triphosphate, and 200 U of Superscript reverse transcriptase (Invitrogen) at 42°C for 50 min, followed by 15 min at 70°C. The C2V5 *env* region was amplified from the cDNA with MK605 (5'-AATG TCAGCACAGTACAATGTACAC-3'; positions 6945 to 6969) and CD4R2 (5'-TATAATTCACTTGTCCAATTGTCC-3'; positions 7652 to 7675) as outer primers (Cassol et al, 1996) and (M13F)-ES7 (5'-tgtaaaacgacggccagtCTGTTAAATGGCAGTC TAGC-3'; positions 7002 to 7021) and (M13R)-ES8 (5'-caggaaacagctatgaccCA CTTCTCCAATTGTCCCTCA-3'; positions 7648 to 7668) as inner primers. The first and second PCR steps were carried out in final volumes of 25 µl and 50 µl, respectively, containing 1X PCR buffer, 2.0 mM $MgCl_2$, 0.2 mM each deoxynucleoside triphosphate, 2.5 pmol of each primer, and 1.25 U of Amplitaq Gold. The PCR conditions were 95°C for 13 min, followed by six cycles at 95°C for 30 s, 65°C for 45 s, and 72°C for 60 s, with a decrease of 1°C per cycle. This was followed by 29 cycles at 95°C for 30 s, 60°C for 45 s, and 72°C for 60 s, with an increase of 5 s for each extension cycle, and a final extension of 72°C for 10 min. Amplified DNA was visually quantified by agarose gel electrophoresis, purified on a Microcon (Amicon) spin column, and sequenced on an automated 3100 genetic analyzer (Applied Biosystems Inc., Foster City, Calif.) with M13 sequencing primers and a Big-Dye terminator cycle sequencing kit.

### 3.2.4 Sequencing of Reverse Transcriptase and Protease

Sequencing of *pol* (nucleotides 2253 to 3485, relative to HXB2) (Korber et al, 1998) was performed with the ViroSeq HIV-1 genotyping system (Applied Biosystems). Plasma and dried blood spot RNAs were reverse transcribed with Moloney murine leukemia virus reverse transcriptase. A 1.8-kb fragment containing the protease (amino acids 1 to 99) and

reverse transcriptase (amino acids 1 to 312) regions was then amplified in a 40-cycle PCR with Amplitaq Gold DNA polymerase and AmpErase dUTP/uracil-*N*-glycosidase to minimize the risk of cross-contamination. PCR products were visually quantified by agarose gel electrophoresis. Following purification, the products were sequenced with six of the seven kit primers (primer D was not used) and Big-Dye terminator reagents and run on a 3100 genetic analyzer as described above. Sequences were assembled, translated, and analyzed for the presence of amino acid polymorphisms. A report was generated for each sequence, with mixtures of wild-type and mutant bases being classified as mutant.

### 3.2.5 Genetic Subtyping and Phylogenetic Analysis

To rule out contamination between samples, each new sequence was compared to other sequences amplified at the same time, as well as to other sequences previously amplified in our laboratory and published sequences in the Los Alamos BLAST search database (Altschul et al, 1997). The sequences were aligned with CLUSTAL W (Thompson et al, 1994) and manually edited with the codon alignment of the Genetic Data Environment (GDE 2.2) program (Smith et al, 1994). New sequences were then compared to subtype reference strains in the Los Alamos subtype database (http://hiv-web.lanl.gov/content/hiv-db/SUBTYPE_REF /align.html). Following degapping with the degapped option in PAUP*, phylogenetic trees were generated on a Linux computer with the F84 model of substitution and the neighbor-joining method (version 4.0b2a) of PAUP* (Swofford, 1999). Trees were rooted with a homologous region of HIV-1 group O (OCM_MP5180). To examine intrasubtype relationships, each KwaZulu-Natal sequence was analyzed against a subset of published C sequences from Zimbabwe, South Africa, Brazil, Tanzania, Zambia, Ethiopia, Israel, and eastern India. Appropriate evolutionary models were selected with the Akaike identification system (Akaike, 1997), implemented in MODELTEST 3.0 (Posada and Crandall, 1998). With this method, a pairwise distance matrix was calculated and used to construct neighbor-joining maximum likelihood trees. Parameters of the reverse transcriptase/protease model, TVM + I + G, were: $f_A = 0.3986$, $f_C = 0.1653$, $f_G = 0.2033$, and $f_T = 0.2328$; $R$ matrix values, $R_{A\rightarrow C} = 2.7534$, $R_{A\rightarrow G} = 10.1383$, $R_{A\rightarrow T} = 0.9138$, $R_{C\rightarrow G} = 1.3684$, $R_{C\rightarrow T} = 13.5383$, and $R_{G\rightarrow T} = 1.0000$; proportion of invariable sites = 0.4263; and heterogeneous variable site distribution (gamma) with alpha shape = 0.8233. Parameters of the *env* model, GTR + I + G, were: $f_A = 0.3801$, $f_C = 0.1838$, $f_G = 0.2890$, $f_T = 0.1472$; $R$ matrix values, $R_{A\rightarrow C} = 3.3002$, $R_{A\rightarrow G} = 8.3576$, $R_{A\rightarrow T} = 3.7717$, $R_{C\rightarrow G} = 1.9646$, $R_{C\rightarrow T} = 23.3707$, $R_{G\rightarrow T} = 1.0000$; proportion of invariable sites

= 0.1534; and heterogeneous variable site distribution (γ) with alpha shape (α) = 0.7332. Trees were viewed with Treetool and Treeview.

### 3.2.6 Genetic Diversity and Intersubtype Recombination Analysis

Mean genetic distances were measured with the Kimura-2 parameter model implemented in MEGA (Kumar et al, 2001). To investigate whether the sequences were recombinant forms of subtype C, recombination analyses were performed with the recombination identification program (Siepel et al, 1995), Bootscanning (Salminen et al, 1995), recombination detection program (Robertson et al, 1995), and Simplot (Lole et al, 1999), a method that uses a sliding-window approach to calculate bootstrap plots for constructing neighbor-joining trees with the DNADIST, NEIGHBOR, or CONSENSE programs of the PHYLIP package (Felsenstein, 1998).

### 3.2.7 Nucleotide and Amino Acid Sequence Analysis

Nucleic acid sequences were also analyzed with SNAP (http://hiv-web.lanl.gov) (Korber and Myers, 1992) and Codeml, a program from the PAML software package (Rambaut, 2000). Various software programs were then used to calculate the ratio of synonymous to nonsynonymous amino acid substitutions as a measure of natural selection pressure at the protein level. Programs included SNAP and MEGA (Kumar et al, 2001), which calculate a synonymous-to-nonsynonymous (*ds*/*dn*) substitution ratio, and Codeml, which calculates a *w* (*dn*/*ds*) value. High rates of synonymous mutation are indicative of conservation and a strict requirement for biological function, while high rates of nonsynonymous substitution are indicative of adaptive change, presumably in response to host selection pressure. To identify amino acid patterns that are characteristic of KwaZulu-Natal viruses, nucleotide sequences were translated and aligned and the consensus KwaZulu-Natal sequence was analyzed by viral epidemiology signature pattern analysis (Korber and Myers, 1992). Consensus sequences were screened for the presence of biologically important sites with Prosite, a database of protein families and domains.

### 3.2.8 Identification of Resistance Mutations and Correlation with Phenotype

The Stanford HIV-SEQ and ß-test programs were used to identify and assess the impact of resistance-associated mutations and polymorphisms on phenotypic resistance. Each reverse transcriptase and protease sequence was compared to that of a subtype B reference strain, HXB2, in the Stanford HIV reverse transcriptase and protease sequence database (http://hivdb.Stanford.Edu/hiv/). Mutations associated with reduced sensitivity to

antiretroviral drugs were assigned a drug penalty score based on genotypic-phenotypic correlative data.

### 3.2.9 Nucleotide Sequence Accession Numbers

GenBank accession numbers for sequences obtained in this study including information on the year of specimen collection and risk category are provided in Table 5.

## 3.3 Results

### 3.3.1 Study Population

Demographic and laboratory results for the 72 KwaZulu-Natal patients are summarized in Table 6. Six specimens came from three sets of epidemiologically linked sex partners. After recording the genotype, the male partner of each couple was excluded from further analysis. Many adult patients had HIV-1-related symptoms; 10 had tuberculosis. Two children were asymptomatic; the remaining 14 children had a variety of symptoms, ranging from pneumonia to weight loss, hepatomegaly, splenomegaly, and diarrhea. All of the children were black. To investigate genetic change over time, 18 retrospective samples (8 from KwaZulu-Natal and 10 from Cape Town) were sequenced and included in the analysis.

### 3.3.2 Genetic Divergence, Subtyping, and Phylogenetic Tree Analysis

As shown in Table 7, the average intersequence divergence among KwaZulu-Natal sequences was significantly higher than among subtype C sequences from Brazil and India, but comparable to values observed for Botswana and other countries in southern Africa. There was no measurable difference in diversity between Indian and black or between adult and pediatric subgroups. Eleven *env* samples carried insertions and deletions and could not be sequenced directly from the PCR product. Overall, KwaZulu-Natal *env* sequences differed from the reference sequences of subtypes A, B, and D by 30.4%, 29.3%, and 32.2%, respectively. Pol sequences differed from subtype A, B, and D reference strains by 11.6%, 11.08%, and 11.01%, respectively. Maximum-likelihood and neighbor-joining distance methods were used to determine subtype. As expected, 60 of 61 (98.4%) matched *env-pol* sequence pairs and all of the retrospective sequences grouped as subtype C. These phylogenetic relationships were supported by bootstrap values of >95%. One sample, ZA021p01, had different *env* and *pol* subtypes, suggesting recombination between these two regions. Further analysis by recombination identification program,

## Table 5. GenBank accession numbers and year of sampling[a]

| *pol* sequence name | Genebank accession no. | Yr of sampling | Transmission[b] | *env* sequence name[c] | Genebank accession no. |
|---|---|---|---|---|---|
| ZA004p01 | AY136957 | 2001 | P | NA | |
| ZA005p01 | AY136958 | 2001 | A | ZA005e01 | AY137011 |
| ZA006p01 | AY136959 | 2001 | P | NA | |
| ZA007p01 | AY136960 | 2001 | A | ZA007e01 | AY137012 |
| ZA008p01 | AY136961 | 2001 | A | NA | |
| ZA009p01 | AY136962 | 2001 | A | ZA009e01 | AY137013 |
| ZA010p01 | AY136963 | 2001 | A | ZA010e01 | AY137014 |
| ZA011p01 | AY136964 | 2001 | A | ZA011e01 | AY137015 |
| ZA012p01 | AY136965 | 2001 | A | NA | |
| ZA013p01 | AY136966 | 2001 | A | ZA013e01 | AY137016 |
| ZA014p01 | AY136967 | 2001 | A | ZA014e01 | AY137017 |
| ZA015p01 | AY136968 | 2001 | A | NA | |
| ZA016p01 | AY136969 | 2001 | A | ZA016e01 | AY137018 |
| ZA017p01 | AY136970 | 2001 | A | ZA017e01 | AY137019 |
| ZA018p01 | AY136971 | 2001 | A | ZA018e01 | AY137020 |
| ZA019p01 | AY136972 | 2001 | A | ZA019e01 | AY137021 |
| ZA020p01 | AY136973 | 2001 | A | ZA020e01 | AY137022 |
| ZA021p01 | AY136974 | 2001 | A | ZA021e01 | AY137023 |
| ZA022p01 | AY136975 | 2001 | A | ZA022e01 | AY137024 |
| ZA025p01 | AY136978 | 2001 | A | NA | |
| ZA024p01 | AY136977 | 2001 | A | ZA024e01 | AY137026 |
| ZA023p01 | AY136976 | 2001 | A | ZA023e01 | AY137025 |
| ZA026p01 | AY136979 | 2001 | P | NA | |
| ZA027p01 | AY136980 | 2001 | P | ZA027e01 | AY137027 |
| ZA028p01 | AY136981 | 2001 | A | ZA028e01 | AY137028 |
| ZA029p01 | AY136982 | 2001 | A | NA | |
| ZA030p01 | AY136983 | 2001 | A | ZA030e01 | AY137029 |
| ZA031p01 | AY136984 | 2001 | A | ZA031e01 | AY137030 |
| ZA032p01 | AY136985 | 2001 | A | ZA032e01 | AY137031 |
| ZA033p01 | AY136986 | 2001 | P | ZA033e01 | AY137032 |
| ZA034p01 | AY136987 | 2001 | P | ZA034e01 | AY137033 |
| ZA035p01 | AY136988 | 2001 | P | ZA035e01 | to be submitted |
| ZA036p01 | AY136989 | 2001 | P | ZA036e01 | AY137034 |
| ZA037p01 | AY136990 | 2001 | P | ZA037e01 | AY137035 |
| ZA038p01 | AY136991 | 2001 | P | ZA038e01 | AY137036 |
| ZA039p01 | AY136992 | 2001 | P | ZA039e01 | AY137037 |
| ZA040p01 | AY136993 | 2001 | P | ZA040e01 | AY137038 |
| ZA041p01 | AY136994 | 2001 | P | NA | |
| ZA042p01 | AY136995 | 2001 | P | ZA042e01 | AY137039 |
| ZA043p01 | AY136996 | 2001 | P | ZA043e01 | AY137040 |
| ZA044p01 | AY136997 | 2001 | P | ZA044e01 | AY137041 |
| ZA045p01 | AY136998 | 2001 | A | ZA045e01 | AY137042 |
| ZA046p01 | AY136999 | 2001 | A | NA | |
| ZA047p01 | AY196498 | 2001 | A | ZA047e01 | AY137043 |
| ZA048p01 | AY196499 | 2001 | A | ZA048e01 | AY137044 |
| ZA049p01 | AY196500 | 2001 | A | ZA049e01 | AY137045 |
| ZA050p01 | AY196501 | 2001 | A | ZA050e01 | AY137046 |
| ZA051p01 | AY196502 | 2001 | A | ZA051e01 | AY137047 |
| ZA052p01 | AY196503 | 2001 | A | ZA052e01 | AY137048 |
| ZA053p01 | AY196504 | 2001 | A | ZA053e01 | AY137049 |
| ZA054p01 | AY196505 | 2001 | A | ZA054e01 | AY137050 |
| ZA055p01 | AY196506 | 2001 | A | ZA055e01 | AY137051 |
| ZA057p01 | AY196507 | 2001 | A | ZA057e01 | AY137053 |
| ZA058p01 | AY196508 | 2001 | A | ZA058e01 | AY137054 |
| ZA059p01 | AY196509 | 2001 | A | ZA059e01 | AY137055 |
| ZA060p01 | AY196510 | 2001 | A | ZA060e01 | AY137056 |
| ZA061p01 | AY196511 | 2001 | A | ZA061e01 | AY137057 |
| ZA062e01 | AY196512 | 2001 | A | ZA062e01 | AY137058 |
| ZA063p01 | AY137008 | 2001 | A | ZA063e01 | AY137059 |
| ZA064p01 | AY137006 | 2001 | A | ZA064e01 | AY137060 |
| ZA065p01 | AY137007 | 2001 | A | ZA065e01 | AY137061 |
| ZA066p01 | AY137004 | 2001 | A | ZA066e01 | AY137062 |
| ZA068p01 | AY196513 | 2001 | A | ZA068e01 | AY137064 |
| ZA069p01 | AY196514 | 2001 | A | ZA069e01 | AY137065 |
| ZA071p02 | AY137000 | 2001 | A | ZA071e01 | AY137067 |
| ZA073p01 | AY196515 | 2001 | A | ZA073e01 | AY137070 |
| ZA074p01 | AY196516 | 2001 | A | ZA074e01 | AY137071 |
| Za075p01 | AY196517 | 2001 | A | ZA075e01 | AY137069 |
| ZA077p02 | AY137001 | 2001 | A | ZA077e01 | AY137073 |
| ZA078p02 | AY137003 | 2001 | A | ZA078GRe02 | AY137072 |
| ZA079p02 | AY137002 | 2002 | A | ZA079GRe02 | to be submitted |
| ZA080p01 | AY137005 | 2001 | A | NA | |

[a] *pol* and *env* sequences from the same virus and individual are shown on the same line. All samples were collected in Durban, South Africa, and surrounding regions.
[b] A, adult (heterosexual) transmission; P, pediatric (perinatal) transmission.
[c] NA, not available.

recombination detection program, Bootscanning, and Simplot confirmed that ZA021p01 was an intersubtype recombinant that typed as subtype C in *env* and as a C/D recombinant in *pol*, with the breakpoint for recombination occurring approximately twothirds of the way along the reverse transcriptase gene.

To investigate within-subtype clustering, trees were constructed with published C sequences from eight different countries (Figures 9 and 10). Full-length reference sequences were selected because these strains contained both the *env* and *pol* genes. Unlike sequences from India, where seven out of nine (77.7%) samples grouped as a single monophyletic group, KwaZulu-Natal sequences were widely dispersed across multiple clusters, or sublineages. The topology of samples within these maximum-likelihood and neighbor-joining trees was similar for both *env* and *pol* and for retrospective specimens collected prior to 1992. As shown in Figure 9, many of the retrospective sequences were localized internally, closer to the root of the sublineage. For most trees, the bootstrap support for the delineated sublineages was higher than 70%. Overall, 58 of the 69 (84.0%) KwaZulu-Natal sequences grouped within sublineages. The number of samples within each sublineage ranged from 3 to 14 for *env* and from 4 to 12 for *pol*. One of the largest sublineages consisted of only KwaZulu-Natal sequences. The remaining sublineages contained sequences from other countries in southern Africa (primarily Botswana, but also Tanzania and Zambia), in addition to those from KwaZulu-Natal. One Kwa- Zulu-Natal sample clustered with sequences from Brazil, Ethiopia, Zambia, and Israel. Sequences from different patient groups were distributed across the phylogenetic tree and showed no obvious evidence of geographic or subgroup clustering based on ethnicity, age, or sex (Figure 10).

### 3.3.3 KwaZulu-Natal and Subtype-specific Signature Motifs

The KwaZulu-Natal protease consensus sequence was identical to the consensus sequence of subtype C at 100% of 99 amino acids, but differed from the consensus of subtypes A, B, and D at seven, eight, and six positions, respectively (Figure 11). Compared to the B consensus, amino acid substitutions were identified at 32 different positions. The mean number of substitutions was nine, with 65 (94.2%) isolates having eight or more substitutions relative to subtype B. The reverse transcriptase consensus differed from the subtype C consensus at only one position, codon V60I. Forty-three (62.3%) of the KwaZulu-Natal sequences had an isoleucine at reverse transcriptase codon 60 rather than the valine residue that is typical of B and C subtypes. This polymorphism was not present in the consensus sequence of retrospective samples. Comparison of the KwaZulu-Natal

**Table 6. Characteristics of and laboratory results for children and adults in the study**

| Variable | All patients (*n*=72) | Adults (*n*=56) | Children (*n*=16) |
|---|---|---|---|
| Mean age ±SD (range) | | 38.4 yrs | 15.72 mo |
| Sex, no. of subjects (%) | | | |
| Male | | 19 (34) | NA[a] |
| Female | | 37 (66) | NA |
| Ethnicity, no. of subjects/total (%) | | | |
| Black | 46/72 (64) | 30/56 (54) | 16/16 (100) |
| White | 2/72 (3) | 2/56 (4) | 0/16 (0) |
| Coloured | 2/72 (3) | 2/56 (4) | 0/16 (0) |
| Indian | 22/72 (31) | 22/56 (39) | 0/16 (0) |
| CD4 cell count, no. of subjects/total (%) | | | |
| ≤200 cells/mm$^3$ | | 12/40 (30) | NA |
| 201 to 500 cells/mm$^3$ | | 17/40 (43) | NA |
| ≥ 501 cells/mm$^3$ | | 11/40 (28) | NA |
| Average CD4 cell count cells/mm$^3$ | | 366 | NA |
| Plasma HIV RNA, no. of subjects/total (%) | | | |
| ≤400 copies/ml | 0/50 (0) | 0/36 (0) | 0/14 (0) |
| 401 to ≤$10^4$ copies/ml | 7/50 (14) | 7/36 (19) | 0/14 (0) |
| >$10^4$ to ≤$10^5$ copies/ml | 16/50 (31) | 14/36 (39) | 2/14 (14) |
| >$10^5$ copies/ml | 27/50 (53) | 15/36 (42) | 12/14 (86) |
| Average Plasma HIV RNA (copies/ml) | 248,260 | 113,020 | 383,500 |
| HIV-1 subtype, no. of subjects/total (%) | | | |
| Subtype C | 60/61 (98) | 48/49 (98) | 12/12(100) |
| D/C recombinant | 1/61 (2) | 1/49(2) | 0/12 (0) |

[a]NA = not available

**Table 7. DNA distances between subtype C sequences from different population groups.**

| Country and ethnic group (no. of viruses) | Mean % distance (SE) | | | |
|---|---|---|---|---|
| | *pol* | *pol* 1st and 2nd positions | *env* | *env* 1st and 2nd positions |
| South Africa (73 ) | 4.93 (0.27) | 2.39 (0.24) | 19.18(1.0) | 19.3 (1.1) |
| Indian(from KZN) (19) | 5.07 (0.34) | | 19.99 (1.1) | |
| Black (from KZN) (44) | 4.9 (0.3) | | 19.38 (1.1) | |
| Botswana (51) | 5.92 (0.30) | 2.86 (0.24) | 19.25 (0.99) | 18.7 (1.0) |
| India (9) | 3.44 (0.29) | 2.22 (2.28) | 11.78 (0.84) | 11.7 (0.9) |
| Tanzania (4) | 4.86 (0.47) | 2.08 (0.43) | 17.81 (1.5) | 17.8 (1.6) |
| Zambia (2) | 5.16 (0.6) | 2.23 (0.5) | 20.3 (2.0) | 20.9 (2.4) |
| Brazil (2) | 2.65 (0.45) | 1.85 (0.47) | 12.65 (1.4) | 12.1 (1.6) |

consensus to that of subtype B revealed 19 different amino acid substitutions. Thirty-six (52.2%) had 21 to 25 substitutions and another 33 (47.8%) had 14 to 20 substitutions relative to HIV-1 B. The most frequent substitutions are shown in Table 8. The characteristic GPGQ motif at the tip of the V3 loop was conserved in 98.4% of KwaZulu-Natal samples (Figure 12). The RIGPGQTFYATG dodecapeptide (amino acids 13 to 24 of V3), previously identified in 69.1% of V3 sequences from Calcutta, India (CIN), was detected in only 28.6% of KwaZulu-Natal specimens. One of the most variable amino acids was the C-terminal glycine (G). Overall, 34.9% of KwaZulu-Natal sequences had an asparagine substitution at this position; 19% had a deletion mutation. Most (87.7%) of the substitutions and deletions were present within the black subgroup. The deletion mutation, present in 24.2% of Black and 5.3% of Indian sequences, caused a decrease in length of the V3 loop from 35 to 34 amino acids. With the exception of a single D25K mutation, no basic amino acid substitutions were detected at V3 loop positions 11, 24, or 25.

### 3.3.4 Amino Acid Substitutions Associated with Drug Resistance

The amino acid sequence of each KwaZulu-Natal sequence was compared to sequences in the Stanford University HIV reverse transcriptase and protease sequence database in order to identify polymorphisms and mutations previously associated with drug resistance in HIV-1 B infections. No primary resistance mutations to protease inhibitors were detected in any of the KwaZulu-Natal samples. However, a substantial number of accessory (secondary) mutations were found at the following positions, in order of decreasing frequency: I93L (97.1%), M36I (85.5%), M63P/S/I/V/H (37.7%), K20R (13.0%), V77I (7.2%), and L10I (1.4%). Similarly, no primary or accessory mutations to resistance against nucleoside reverse transcriptase inhibitors were identified. However, three patients were found to harbor resistance mutations to nonnucleoside reverse transcriptase inhibitors: patient ZA024p01 had a K103N mutation, and her male partner, ZA023p01, carried a G190A mutation in addition to K103N. A third patient, ZA010p01, had a single A98G mutation. Table 9 summarizes the frequency and pattern of these mutations.

### 3.3.5 Amino Acid Substitution and Selection Pressure

KwaZulu-Natal sequences were then compared internally to assess the mutational behavior of reverse transcriptase and protease in the absence of drug therapy. Analysis by the likelihood ratio method of Yang (Yang, 2000) indicated that both genes were under strong purifying (negative) selection pressure (*dn/ds* or $w < 1$), with >95% of sites having $w1 = 0.019$ and $w2 = 0.395$. In contrast, only 5 (5.1%) amino acids in protease (codons 12,

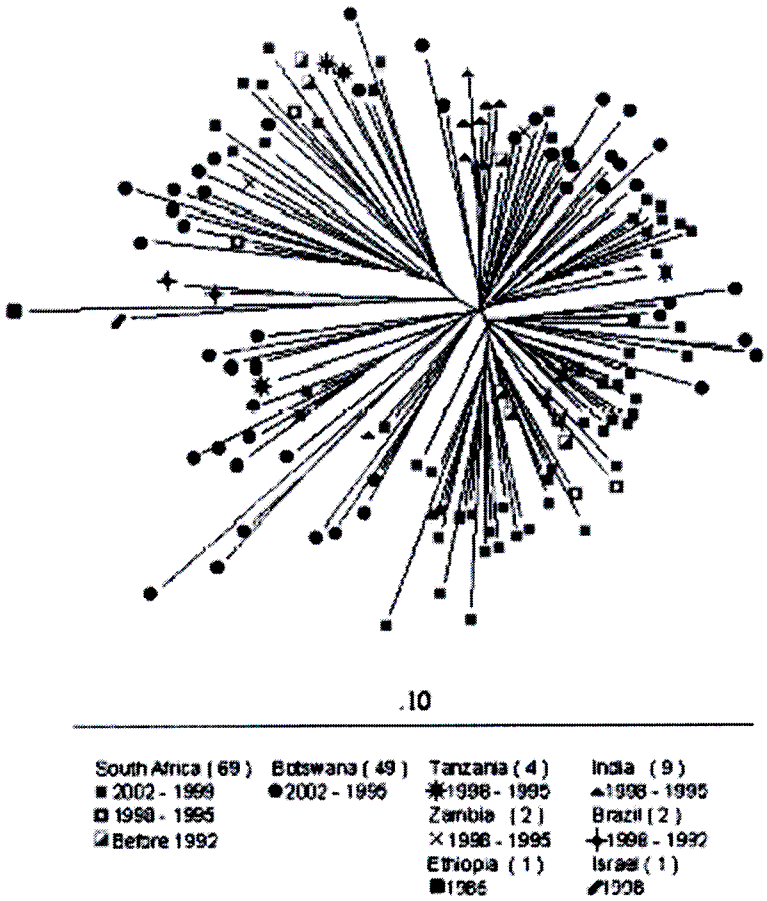



**Figure 9.** Representative *pol* tree showing the relationships between retrospective and contemporary sequences from South Africa, Botswana, and other countries affected by the subtype C epidemic. The sequences are coded by the country of origin and year of isolation. The following sequences were included in the analysis: 49 previously described isolates from Botswana (accession numbers AF110960, AF110963, AF110967, AF110970, AF110972, AF110973, AF110978, and AF443074 to AF443115), 9 sequences from India (accession numbers AF286232, AF286223, AF286231, AB023804, AF067159, AF067155, AF067154, AF067157, and AF067158), 4 sequences from Tanzania (accession numbers AF286234, AF286235, AF361874, and AF361875), 2 sequences from Zambia (AF286224 and AF286225), 2 sequences from Brazil (U52853 and AF2862228), 1 sequence from Ethiopia (U46016), 1 sequence from Israel (AF286233), and 69 sequences from South Africa, including 5 previously described sequences (AF286227, AY043173, AY043174, AY043175, and AY043176), 3 sequences from another study (Van Harmelen et al, 2001), and 61 sequences newly generated from this study (14 retrospective and 47 contemporary strains).

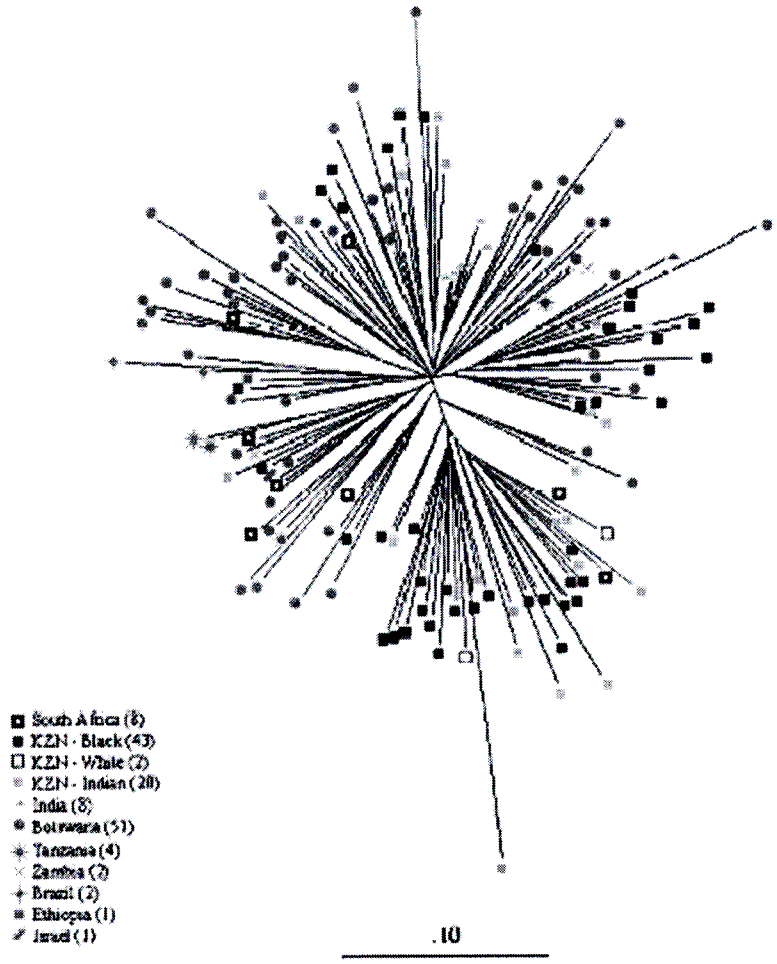



**Figure 10.** Phylogenetic relationship of C2V5 envelope sequences from KwaZulu-Natal, Botswana, Zambia, and Tanzania.

19, 35, 37, and 63) and 15 (4.8%) amino acids in reverse transcriptase (codons 36, 39, 123, 135, 162, 166, 174, 196, 207, 211, 214, 245, 272, 277, and 286) were found to be under strong positive Darwinian (*d*) selection ($w = 2.055$). As shown in Figure 11, these amino acids were not randomly distributed but were located at discrete loci along the reverse transcriptase and protease genes. Seven (35%) amino acids that were under positive (diversifying) selection pressure (protease positions 12S and 19I; reverse transcriptase positions 36A, 39E, 123G, 211K, and 245Q) were present in both the KwaZulu-Natal and subtype C consensus sequence but not in the consensus sequences of subtypes A, B, and D, suggesting that these signature residues may offer a subtype-specific fitness advantage to C viruses.

### 3.3.6 Impact of Substitution on Functional Motifs

Naturally occurring polymorphisms also resulted in significant variation in the number and type of phosphorylation sites. Overall, 17 potential phosphorylation sites were identified in the *pol* gene, 3 in the protease and 14 in the reverse transcriptase. Twelve of the *pol* sites were conserved among KwaZulu-Natal patients and in the consensus sequences for subtypes A, B, C, and D (Figure 11). These included the predicted protein kinase C site at codons 12 to 14 near the N terminus of the protease and the two casein kinase II phosphorylation motifs at the active site. Most KwaZulu-Natal sequences had an S-X-K rather than a T-X-K motif at protease codons 12 to 14. Conserved phosphorylation sites in reverse transcriptase included protein kinase C codons 68 to 70; tyrosine kinase codon 49 to 56; cyclic AMP/cyclic GMP-dependent codons 65 to 67, 102 to 105, and 125 to 128; and CKII codons 3 to 6, 107 to 110, 191 to 194, 215 to 218, and 253 to 256. Two KwaZulu-Natal patients lacked a cyclic AMP phosphorylation site at reverse transcriptase codons 102 to 105 due to the presence of a K103N mutation. Some phosphorylation sites, such as the CKII sites at reverse transcriptase positions 39 to 41 and 200 to 203, were present in subtypes A, B, and D but absent from most of the KwaZulu-Natal and subtype C sequences. Other differences included the absence of an internal myristoylation site (Maurer-Stroh et al, 2002) at reverse transcriptase codons 196 to 201 in nine patients and the presence of an amidation site at protease codons 67 to 70 in subtype A, subtype C, and all but two of the KwaZulu-Natal sequences. With a single exception, all of the natural reverse transcriptase mutations were embedded within cytotoxic T-lymphocyte, T-helper, or overlapping cytotoxic T-lymphocyte/T-helper epitopes, as defined for B viruses. Of particular interest, with respect to the *env* gene, was a cluster of substitutions located at or in close proximity to the bottom of the V3 loop, a region known to play a major role in viral tropism and coreceptor usage. This cluster included amino acid -1, immediately

**Table 8. Frequency of the most common amino acid substitutions in the *pol* gene compared to subtype B.**

| Protein | Amino acids | % of strains |
|---|---|---|
| Protease | H69K/Q/Y | 100 |
| | I93L | 97.1 |
| | L19I/V/E/T/A | 97.1 |
| | I15V | 92.8 |
| | M36I/L/T | 91.3 |
| | L89M | 89.8 |
| | R41K | 89.8 |
| | T12S/A/P | 81.2 |
| Reverse transcriptase | V35T/I/K/M/Q | 100 |
| | Q207E/D/G/N/S/R/K | 100 |
| | V245Q/K/L/H | 98.6 |
| | T39E/D/A/K/N | 98.5 |
| | I293V | 97.1 |
| | V292I | 97.1 |
| | T200A/I/E | 95.7 |
| | E291D | 95.7 |
| | K173A/T/V/G/I | 94.2 |
| | S48T/E | 92.7 |
| | K122E/Q | 93.2 |
| | D177E/G/N | 89.9 |
| | A272P/Q/S/R | 89.9 |
| | T286A/V | 81.1 |
| | E36A/T/V | 78.3 |
| | D123G/N/S | 73.9 |
| | K277R/S | 66.7 |
| | V60I | 62.3 |
| | R211K | 56.5 |

**Table 9. Amino acid substitutions at codons associated with drug resistance[a]**

| Patient no. | Protease substitution(s) | Reverse transcriptase substitution(s) |
|---|---|---|
| 21 | M36I | |
| 36, 29, 62 | I93L | |
| 7, 8, 11, 13, 14, 15, 16, 19, 27, 30, 31, 32, 33, 34, 37, 39, 40, 41, 42, 47, 51, 53, 54, 55, 57, 59, 61, 73, 78, 79, 80 | M36I, I93L | |
| 20 | V771, I93L | |
| 10 | M36I, I93L | A98G |
| 35, 66, 74 | L63P, I93L | |
| 49 | K20R, V77I, I93L | |
| 5, 9, 22, 38, 58, 68, 69 | K20R, M36I, I93L | |
| 12, 17, 18, 25, 26, 28, 45, 46, 48, 50, 52, 63, 64, 65, 71, 75, 77 | M36I, L63P/T/S/H, I93L | |
| 6 | M36I, V77I, I93L | |
| 23 | L63T, V771, I93L | K103N*, G190A |
| 24 | L63T, V771, I93L | K103N* |
| 44 | L10I, M36I, L63P, I93L | |
| 60 | M36I, L63P, V77I, I93L | |
| 43 | K20R, M36I, L63V, I93L | |

[a,b]Amino acid substitutions relative to the North American/European HIV-1 B subtype
*Primary mutation associated with resistance to nevirapine, delavirdine and efavirenz

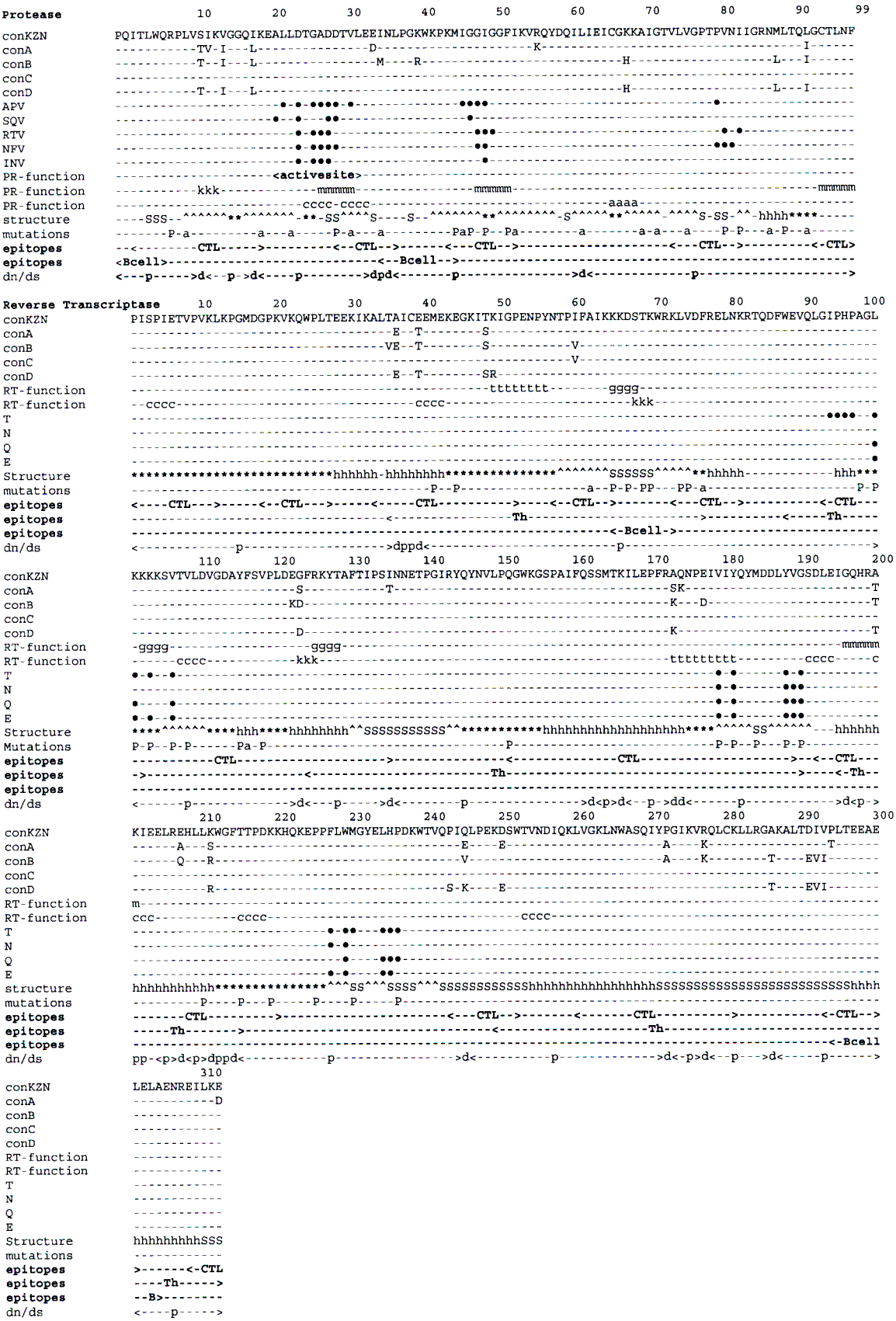

Protease
10 20 30 40 50 60 70 80 90 99
conKZN PQITLWQRPLVSIKVGGQIKEALLDTGADDTVLEEINLPGKWKPKMIGGIGGFIKVRQYDQILIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNF
conA
conB
conC
conD
APV
SQV
RTV
NFV
INV
PR-function <activesite>
PR-function
PR-function
structure
mutations
epitopes <CTL>
epitopes <Bcell>
dn/ds
Reverse Transcriptase
10 20 30 40 50 60 70 80 90 100
conKZN PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
conA
conB
conC
conD
RT-function
RT-function
T
N
Q
E
Structure
mutations
epitopes <CTL>
epitopes <Th>
epitopes <Bcell>
dn/ds
110 120 130 140 150 160 170 180 190 200
conKZN KKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRAQNPEIVIYQYMDDLYVGSDLEIGQHRA
conA
conB
conC
conD
RT-function
RT-function
T
N
Q
E
Structure
Mutations
epitopes <CTL>
epitopes <Th>
epitopes
dn/ds
210 220 230 240 250 260 270 280 290 300
conKZN KIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEEAE
conA
conB
conC
conD
RT-function
RT-function
T
N
Q
E
structure
mutations
epitopes <CTL>
epitopes <Th>
epitopes <Bcell>
dn/ds
310
conKZN LELAENREILKE
conA
conB
conC
conD
RT-function
RT-function
T
N
Q
E
Structure
mutations
epitopes <CTL>
epitopes <Th>
epitopes <B>
dn/ds

**Figure 11. Correlation of signature patterns with structure and function for protease and reverse transcriptase.** conKZN, KwaZulu-Natal consensus; conA, conB, conC, and conD, consensus sequences for subtypes A, B, C, and D, respectively; APV, SQV, RTV, NFV, INV, drug binding sites for amprenavir, saquinavir, ritonavir, nelfinavir, and indinavir, respectively; functn, RT, reverse transcriptase; CTL, cytotoxic T-lymphocyte epitope; F, drug-binding site; k, protein kinase C phosphorylation site; c, casein kinase phosphorylation site; m, myristoylation site; aaaa, amidation site; t, tyrosine kinase phosphorylation site; g, cyclic AMP- and cyclic GMP-dependent protein kinase site; T, thiocarboxanilide UC-781; N, nevirapine; Q, quinoxaline HBY 097; E, efivirenz; a, accessory mutation; P, primary mutation; caret, extended _-strand; S, bend; star, hydrogen-bonded turn; h, helix; p, purifying selection pressure; d, Darwinian (positive) selection pressure.

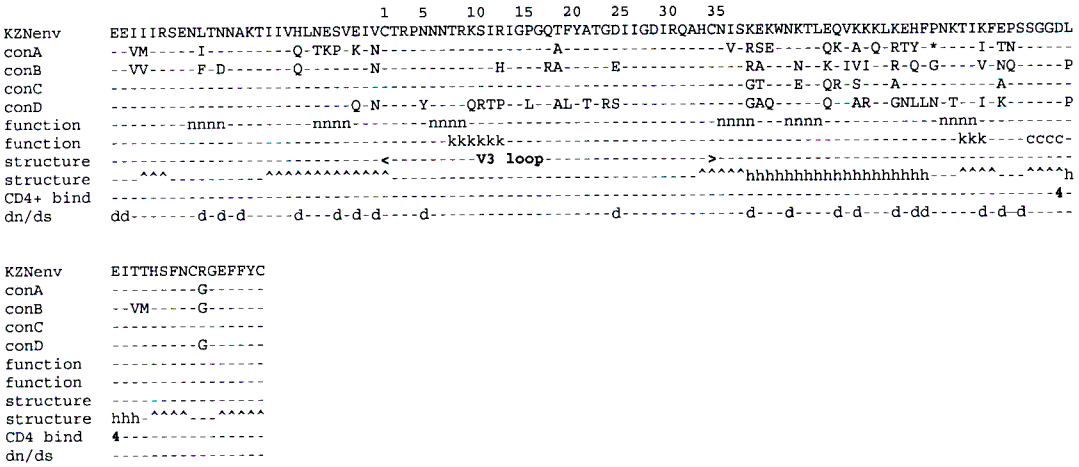



**Figure 12. Correlation of signature patterns with structure and function of V3 loop.** KNZenv, KwaZulu-Natal consensus; conA, conB, conC, and conD, consensus sequences for subtypes A, B, C, and D, respectively; k, protein kinase C phosphorylation site; c, casein kinase phosphorylation site; n, N-linked glycosylation site; caret, extended β-strand; h, helix; 4, CD4+ binding site; d, Darwinian (positive) selection pressure.

upstream from the cysteine residue at the beginning of V3, and amino acid positions 11 and 13 within the V3 loop itself. In common with other C viruses, strains from 89.0% of KwaZulu-Natal patients had amino acid substitutions that resulted in elimination of the N-linked glycosylation site at position - 1 (amino acid 301 according to the numbering of Korber at al, 1998). In 91.0% of patients, loss of glycosylation was associated with a serine (S) substitution at position 11 and the presence of a positively charged arginine (R) residue at V3 position 13. The resultant S-X-R motif gave rise to a second, alternative protein kinase C site immediately adjacent to the phosphorylation site at amino acids 8 to 10. These findings suggest a potential linkage between deglycosylation and phosphorylation in the V3 loop of C viruses. Most A variants also carried the extra protein kinase C site at position 11 to 13 but lacked the N-linked glycan at position -1. Instead, a more distal N-X-S glycosylation site (positions-7 to -5) was frequently absent in A viruses. Another protein kinase C site, located downstream from the C terminus of V3 at positions 45 to 47 (relative to V3), was missing in most KwaZulu-Natal viruses. This site is highly conserved among subtype B viruses. In common with subtype B, KwaZulu-Natal and other C viruses contained a highly conserved CKII site at amino acids 68 to 71.

### 3.4 Discussion

Despite the dramatic impact of HIV-1 and AIDS on the KwaZulu-Natal region of South Africa, few studies have examined the genetic diversity and molecular phylogeny of KwaZulu-Natal viruses. To date, only eight full-length South African sequences have been published (Van Harmelen et al, 2001; zur Megede et al, 2002). The primary goals of this study were to identify regions of high variability, characterize amino acids that are unique to local strains, and identify sites that are highly conserved and thus likely to be important for vaccine development and the assessment of antiretroviral therapy. Our results indicate that C viruses in KwaZulu-Natal have a higher level of nucleotide diversity than previously reported (Van Harmelen et al, 1999; Van Harmelen et al, 2001) and that the epidemic, in its explosive phase, is characterized by multiple circulating sublineages in both the Indian and black communities. The restricted distribution of subtype C viruses from India compared to the multilineage pattern of Indian viruses from Africa indicates that the two Indian epidemics have different origins and different evolutionary histories. The presence of retrospective samples (collected prior to 1990) at internal (basal) branches in three of the sublineages suggests that each lineage is derived from a different founder variant and that these variants have been cocirculating in South Africa for at least 10

years. Of significant note was the cosegregation and close relatedness of sequences from Kwa-Zulu-Natal black and Indian inhabitants, not only to each other, but also to published sequences from Botswana. This close relationship with sequences from Botswana was not observed in a previous study (Novitsky et al, 2002), presumably because of the small number of samples included from South Africa ($n$ = 5). Taken together, our findings confirm the existence of multiple HIV-1 C sublineages in southern Africa and demonstrate that the spread of these different lineages has been substantial. The finding that C viruses from KwaZulu-Natal are substantially more diverse than those in India and Brazil is consistent with other studies and has been attributed to the longer duration of the AIDS epidemic in Africa (Brindeiro et al 1999; Shankarappa et al, 2001). The overall evolutionary rate of *pol* and *env* sequences, as measured by a dated-tip likelihood method (Rambaut, 2000), was 35% and 68% higher than that of subtype B. Despite the high level of diversity, KwaZulu-Natal viruses were remarkably well conserved at the amino acid level, both within subtype C and among different individuals. This is due to the fact that a large number of the nucleotide substitutions are silent (synonymous) mutations that cause no change in the amino acid sequence. As a result, the consensus sequence for the KwaZulu-Natal protease was identical to the consensus sequence for subtype C, while the reverse transcriptase consensus sequences differed from the C consensus at a single amino acid, codon 60. High rates of synonymous-to-nonsynonymous nucleotide change have also been observed among subtype C isolates from Zimbabwe (Shafer et al, 1997) and Ethiopia (Loemba et al, 2002). This inherent property of African subtype C viruses is a reflection of the differential pressure exerted on the three positions of the amino acid code or the KwaZulu-Natal reverse transcriptase gene, the mutation rate for the third position of the codon was four times higher than that observed for the second position and 30 times higher than for the first codon position (data not shown). The conservation of subtype C at the amino acid level offers considerable promise for the development of a consensus- or ancestor-based "supervaccine" (Novitsky et al, 2002). Recent primate studies suggest that it may be possible to overcome diversity and achieve cross-protection against different HIV-1 variants (Dunn et al, 1997; Shibata et al, 1997). However, it should be stressed that the long-term impact of silent mutations on vaccine efficacy is not known. In the context of antiretroviral therapy, one recent study found that, despite numerous naturally occurring mutations in reverse transcriptase, C viruses from Zimbabwe were as susceptible as subtype B viruses to commonly used nucleoside and nonnucleoside reverse transcriptase inhibitors (Shafer et al, 1997). However, another recent study found that, although C viruses in Ethiopia were susceptible to reverse transcriptase inhibitors, the presence of

silent mutations led to a more rapid emergence of resistance (Loemba et al, 2002). These data emphasize the need for carefully designed prospective trials to determine whether existing polymorphisms influence the development of resistance in C-infected patients.

With the exception of two primary resistance mutations, K103N and G190A, which occurred in a single husband-wife pair, none of the reverse transcriptase or protease polymorphisms occurred at drug-binding sites or at active sites of the enzymes. Both mutations are known to cause high-level resistance to nevirapine in persons infected with subtype B (Raffi et al 2000). Although believed to be naturally occurring, the possibility that these mutations represent treatment-induced changes cannot be excluded. As many as 15% of patients in the private sector in South Africa have received or are currently receiving some form of antiretroviral therapy. Many protocols include nevirapine because of its low cost and long half-life. Nevirapine is also being increasingly used for the prevention of mother-to child HIV-1 transmission in KwaZulu-Natal and other regions of Africa (Jackson et al, 2001).

All of the remaining *pol* polymorphisms occurred in regions involved in the three-dimensional configuration of reverse transcriptase and protease. One such polymorphism, which occurred in a single patient, was A98G in the reverse transcriptase. This mutation was also detected in a treatment-naïve patient from Ethiopia (Loemba et al, 2002). In persons infected with subtype B, A98G has been associated with low-level resistance to nonnucleoside reverse transcriptase inhibitors. Other polymorphisms were localized within the hinge region of protease, a region that induces conformational changes during drug binding. A subset of these mutations, M36I/R41K/H69/L89 M, has been linked to increased catalytic activity in subtypes A and C (Velazquez-Campoy et al, 2001). Another series of polymorphisms, at codons 12, 15, 19, and 93, occurred in >80% of KwaZulu-Natal viruses and formed a KwaZulu-Natal/subtype C signature motif. The first three amino acids of this motif are located near the N terminus of protease, in an extended β-strand; the fourth, I93L, is located in a hydrogen-bonded turn, immediately upstream of the protease/reverse transcriptase cleavage site. The marked dominance of I93L among C viruses, its close proximity to the protease/reverse transcriptase cleavage site, and its linkage to the T12S/T15V/L19I signature warrant further investigation. Studies of HIV-1 B have reported that mutations in the protease and Gag-Pol cleavage sites contribute to drug resistance, are specifically selected during therapy, and can lead to improved enzyme kinetics (Cote et al, 2001; Doyon et al, 1996c These conserved codons were concentrated within active sites and at drug-binding sites in reverse transcriptase and protease and at nucleoside triphosphate binding sites in reverse transcriptase. The remaining 5% of amino

acids were under strong positive selection pressure and were concentrated in regions associated with maintaining the tertiary structure and facilitating conformational changes. Some positively selected codons, such as protease 63 and reverse transcriptase 123 and 174, showed extensive interpatient and intersubtype variation. Other codons (such as protease 12S and reverse transcriptase 39E, 245Q, 272P, and 277R) were highly conserved among Kwa-Zulu-Natal and subtype C sequences and formed part of an HIV-1 C signature sequence. The conservation of codons in the face of strong diversifying pressure suggests that they may play an important role in the evolutionary, structural, and phenotypic properties of C viruses. A few positively selected codons were conserved across several subtypes, suggesting that they may contribute to the evolutionary history of group M viruses. Although many factors contribute to the generation of new variants, one of the most important is related to cytotoxic T lymphocytes and the role they play in recognizing epitopes presented by major histocompatibility complex class I molecules. With a single exception, all of the naturally occurring reverse transcriptase mutations were embedded within cytotoxic T-lymphocyte, T-helper, or overlapping cytotoxic T-lymphocyte/T-helper epitopes as previously defined for B viruses (Korber et al, 2000). Several signature sequences in *env* also mapped to known subtype B cytotoxic T-lymphocyte epitopes, including the heavily glycosylated regions at the bottom of V3 and the associated protein kinase C phosphorylation site at V3 position 11. Information on subtype C epitopes is just beginning to emerge and, when combined with novel methods of analysis, may lead to new insights into the immune selection pressures occurring during seroconversion and in response to therapy. By examining sites under positive selection pressure, we may be able to identify targets of the host immune system and select appropriate epitopes for inclusion in a subtype C vaccine. Although it is well known that most C viruses lack a V3 glycosylation site and a basic amino acid residue at position 11, the biological significance of these findings remains unclear. Disruption of V3 glycosylation has also been reported to occur in 52%, 34%, and 20% of subtype G, A, and D viruses, respectively. Studies of subtype B have suggested that this N-linked glycan may play a role in the interaction of gp120 with its coreceptors (Li et al, 2001) and in perinatal transmission. Nakayama et al, (1998) found that absence of this V3 glycan caused a marked reduction in CXCR4-dependent but not CCR5-dependent viral entry. Others have suggested that the V3 glycan is not necessary for CXCR4 usage (Losman et al, 1991) and that its absence leads to enhanced infectivity of CXCR4-expressing cells (Polzer et al, 2001). Li et al. (2001) found that multiple factors contribute to coreceptor usage and that the effects exerted by the V3 glycan are both isolated and context dependent. Similarly,

the absence of a basic amino acid at position 11 of V3 and at positions 24 and 25 has been associated with a non-syncytium-inducing phenotype and CCR5 coreceptor-using properties, while the presence of basic charge has been correlated with CXCR4 and syncytium-inducing phenotypes (Hoffman et al, 1999; Hoffman et al, 2002; Hwang et al, 1991; Kato et al, 1999; Nakayama et al, 1998). As with deglycosylation, these correlations have been imprecise. Our findings, showing a potential linkage between V3 deglycosylation and the presence of a serine phosphorylation site at position 11, suggest that factors other than glycosylation and charge may have to be taken into account when assessing the function of V3. Based on the knowledge that C viruses are almost exclusively non-syncytium inducing and CCR5 using, it is tempting to speculate that deglycosylation may allow better access to the CCR5 coreceptor, while phosphorylation may alter the conformation of gp120, exposing retroviral sites that are needed for efficient CCR5-mediated viral entry. Although highly speculative, this possibility warrants further study given the critical importance of V3 for host cell recognition and viral entry. Differences were also observed in the number and position of phosphorylation sites in reverse transcriptase and protease. Phosphorylation is known to modulate the activity of many proteins that interact with nucleic acids, including DNA and RNA polymerase. It is also known that, in addition to reverse transcriptase and protease, several protein kinases are incorporated into mature HIV-1 virions (Tozer et al, 1999), where they are available not only to regulate the activity of reverse transcriptase and protease, but also to participate in interactions with the host cell. Phosphorylation of threonine residue at reverse transcriptase codon 215 has been shown to increase discrimination against azidothymidine, leading to drug resistance (Lazaro et al, 2000), and phosphorylation of protease substrates can lead to impaired proteolytic cleavage (Tozer et al, 1999). Our data indicate that several phosphorylation sites in the *pol* gene of KwaZulu-Natal and subtype C viruses are highly conserved and positively selected. It will be important to determine whether these sites play a significant role in the replicative capacity and proteolytic processing of C viruses.

# Chapter 4

## Variability at Human Immunodeficiency Virus Type 1 Subtype C Protease Cleavage Sites: an Indication of Viral Fitness?

## 4.1 Introduction

One of the most dramatic changes in the human immunodeficiency virus type 1 (HIV-1)-AIDS epidemic has been the rapid emergence and devastating spread of subtype C viruses (Kuiken et al, 2002; http://www.unaids.org/epidemicupdatedec01/report /index html). HIV-1 C now accounts for >56% of all circulating viruses and is the most commonly transmitted subtype worldwide (Esparza and Bhamarapravati, 2000). Subtype C predominates in southern Africa (Gordon et al, 2003; Novitsky et al, 1999; Van Harmelen et al, 1999) and India (Shankarappa et al, 2001) and is increasing in frequency in China Gao et al, 1998, Rodenburg et al, 2001) and Brazil (Brindeiro et al, 1999; Gonzales et al, 2003; Soares et al, 2003). The disproportionate increase in C viruses relative to other HIV-1 strains suggests that subtype C may be more easily transmitted or that it has a higher level of "fitness" at the population level. One possible explanation is that founder effects relating to the ongoing introduction of subtype C into new population groups with different host factors, or different social and sexual practices, may be responsible for the rapid spread. However, founder and host effects cannot account for the fact that C viruses are overtaking preexisting virus subtypes in several different geographical regions, including Yunnan Province in China and the southern region of Brazil (Brindeiro et al, 1999; Gao et al, 1998; Rodenburg et al, 2001; Soares et al, 2003). It is increasingly evident that additional (non-host) viral factors are also contributing to the rapid spread of HIV-1 C. Viral studies indicate that subtype C has distinct genetic and phenotypic properties that differentiate it from other HIV-1 subtypes. Various studies have postulated that an extra NF-KB binding site in the long terminal repeat (Rodenburg et al, 2001), a prematurely truncated Rev protein (Gao et al, 1998, Rodenburg et al, 2001), or a 5-amino-acid insertion in Vpu (McCormick-Davis et al, 2000) may influence viral gene expression, altering the transmissibility and pathogenesis of C viruses (Tatt et al, 2001). Factors related to viral entry and pathogenesis, such as the CCR5 and non-syncytium-inducing properties of C isolates (Ball et al, 2003; Peeters et al, 1999; Ping et al, 1999), may also contribute to the increased spread of C viruses. One area of research that is receiving consideration is the possibility that C viruses have a more active, catalytically efficient protease (Velazquez-Campoy et al, 2001). The C protease is highly conserved at the amino acid level and has a distinct signature sequence that differentiates it from those of subtypes A, B, and D (Gordon et al, 2003; Velazquez-Campoy et al, 2001). A subset of these signature residues, present in the hinge (M36I/R41K/H69K) and α-helix (L89M) of C (and A) proteases, has been linked to increased catalytic activity (Velazquez-Campoy et al, 2001). Another signature pattern, identified in >80% of C viruses from South Africa, is

T12S/ T15V/L19I/I93L (Gordon et al, 2003). The 12S, 15V, and 19I residues of this motif are located near the N terminus of the protease in an extended β-chain. The 93L polymorphism is located within a hydrogen-bonded turn immediately upstream from the protease-reverse transcriptase (RT) cleavage site, in close proximity to 12S/15V/19I and the dimerization domain (Pettit et al, 2003; Shehu-Xhilaga et al, 2001; Swanstrom and Wills, 1997; Tessmer et al, 1998). All of these polymorphisms lie outside the catalytic site of the protease, in regions that would be expected to alter the enzyme's activity toward its natural cleavage sites, leading to adaptive (compensatory) changes in the cleavage site itself. Since protease inhibitors (PIs) are currently the most active antiretroviral drugs used for the treatment of HIV-1 (Zennou et al, 1998), it is important to collect information, not only on the C protease but also on its drug responsiveness, substrate-inhibitor specificities, and cleavage site characteristics. This information is needed to design PIs that are maximally effective against C viruses and to obtain new insights into the mechanism of drug resistance. Studies have shown that resistance mutations in the B protease are associated with impaired proteolytic processing, decreased enzymatic activity, and a failure to produce mature infectious virions (Croteau et al, 1997; Doyon et al, 1996; Zennou et al, 1998). Compensatory cleavage site mutations can partially compensate for these defects (Zybarth and Carter, 1995). In this report, we describe the natural variability of subtype C cleavage sites in viruses from Africa, India, and Brazil and compare the results to cleavage site patterns in representative B and group M viruses.

## 4.2 Methods

### 4.2.1 Cleavage Site Characteristics

The HIV-1 protease is a small, 99-amino-acid aspartic enzyme that mediates the cleavage of Gag, Gag-Pol, and Nef precursor polyproteins. These reactions occur late in the viral life cycle, during virion assembly and maturation at the cell surface. The process is highly specific, temporally regulated, and essential for the production of infectious viral particles (Jacks et al, 1998; Kaplan et al, 1994; Krausslich et al, 1989; Swanstrom and Wills, 1997). As shown in Figure 13, the main structural proteins are formed by cleavage of the Pr55*gag* polyprotein into matrix (MA; p17), capsid (CA; p24), nucleocapsid (NC; p7), p6*gag*, and two spacer peptides, p2 and p1. The viral enzymes are formed by cleavage of Pr160*gag-pol*, a fusion protein derived by ribosomal frame shifting (Jacks et al, 1998). Although Pr160*gag-pol* also contains p17, p24, and p2, its C-terminal cleavage products are NC, a transframe protein (TFP), p6*pol*, protease (PR), reverse transcriptase (RTp51), RNase H (RTp66), and integrase (IN) Ikuta et al, 2000, Tessmer et al, 1998). In total, 12 proteolytic

reactions are required to generate a mature infectious virion. Each reaction occurs at a unique cleavage site that differs in amino acid composition (Billich et al, 1988). Some cleavage sites contain phosphorylated Ser/Thr or Tyr residues that alter the sites' susceptibilities to cleavage (Tozer et al, 1999). P6*gag*, the major phosphoprotein of HIV-1, plays an essential role in the release of virus from the membranes of infected cells (Muller et al, 2002).

### 4.2.2 Sequence Data and Construction of Cleavage Site Fragments

A total of 84 full-length nucleotide sequences were selected for analysis. These sequences included two C isolates from South Africa, TV001 and TV002 (Zur Megede et al, 2002), in addition to another 25 subtype C, 30 subtype B, and 27 representative group M reference sequences (including A [$n$=3], B [$n$=4], C [$n$=5], D [$n$=3], F1 and F2 [$n$=4], G [$n$=2], H [$n$=2], J [$n$=2], and K [$n$=2] subtypes) extracted from the Los Alamos database (Table 10) (Kuiken et al, 2002). Sequences were selected based on the patient being treatment naive. Since the prevalence of drug resistance in untreated patients has been reported to range from 1 to 11% (Little, 2000), sequences were also screened and excluded from the study if they were found to contain primary resistance mutations. The majority of the sequences were obtained by direct DNA PCR amplification and cloning of peripheral blood mononuclear cells. Nucleotide sequences were aligned by CLUSTAL W (Thompson et al, 1994) and manually edited with the codon alignment of the Genetic Data Environment (GDE version 2.2) program (Smith et al, 1994). Calculation of the pairwise distance matrix, phylogenetic inference, and tree construction were performed on a dual-processor Linux computer by using the PAUP version 4.0b2a program (Sinauer Associates, Sunderland, Mass.) and a GDE for Linux HIV-1 interface (De Oliveira et al, 2002). Thirty-base-pair segments, consisting of 15 nucleotides (5 amino acids) on each side of the 12 cleavage sites, were extracted and concatenated into a 360-bp nucleotide sequence.

### 4.2.3 Reconstruction of Ancestral Cleavage Site Sequences

To examine the evolutionary histories of individual cleavage sites, Phylogenetic Analysis under Maximum Likelihood (PAML) software (Rambaut, 2000) was used to identify amino acid and nucleotide substitutions along each branch of the tree. Branch lengths were estimated using a nucleotide substitution model; amino acid sequences were deduced from the reconstructed nucleotide triplets. The analyses involved the use of maximum-likelihood methods and a time-reversible model which assume different substitution rates,

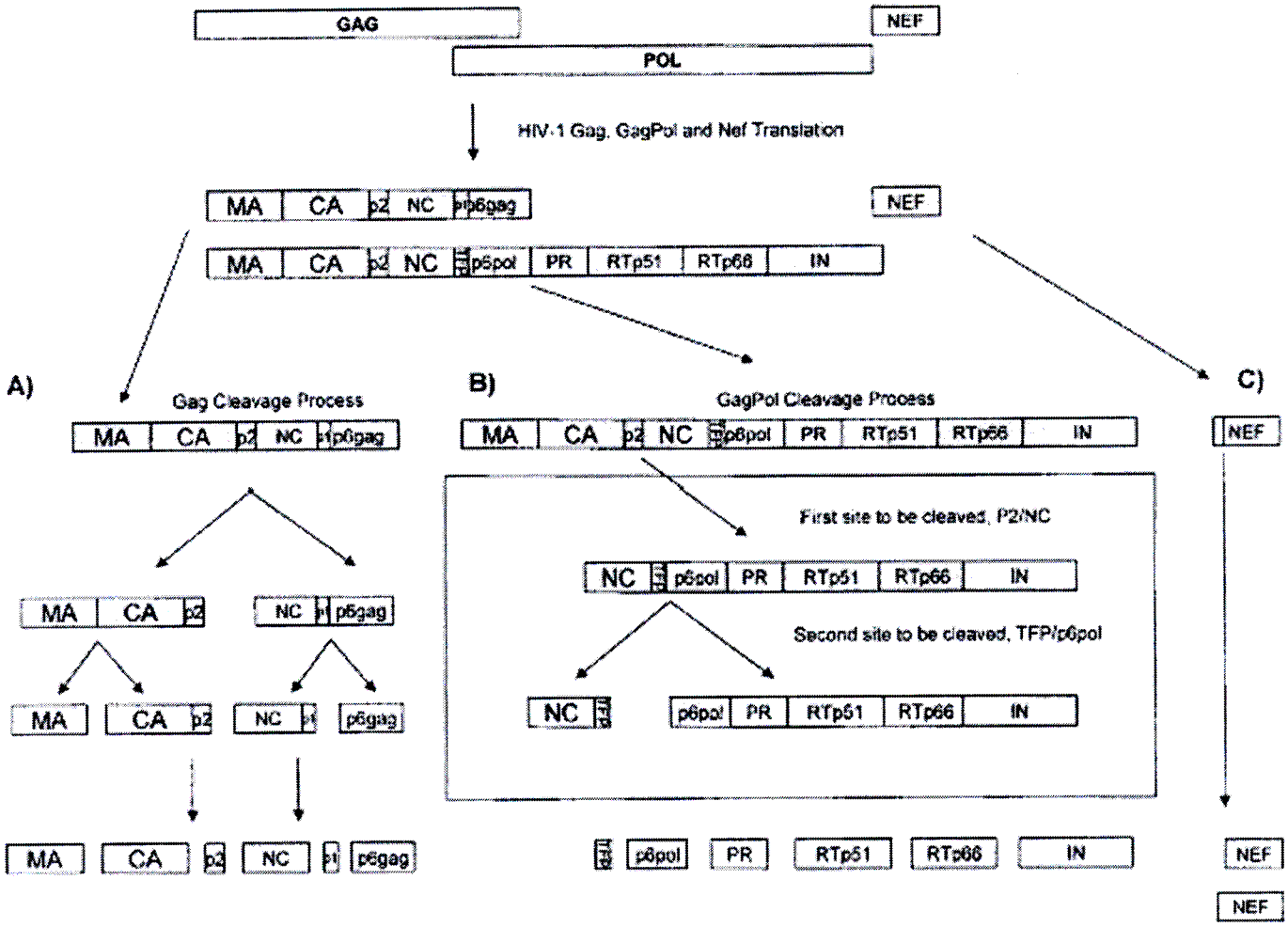



**Figure 13. Schematic of the Gag and Gag-Pol processing sites showing the 12 individual protease cleavage sites: 5 cleavage sites in Gag (p17/p24, p24/p2, p2/NC, p7/p1, and p1/p6*gag*), 6 cleavage sites in Gag-Pol (NC/TFP, TFP/p6*pol*, p6*pol*/PR, PR/RT, RT/p66, and p66/IN), and a single site in Nef..**

**Table 10. Sequences used for the analysis of protease cleavage sites[a]**

| Group or subtype | Data set | Accession no. |
|---|---|---|
| M | A1.KE.93.Q23-17[a] | AF004885 |
| | A1.UG.85.U455 | M62320 |
| | A1.UG.92.92UG037 | U51190 |
| | B.FR.83.HXB2 | K03455 |
| | B.US.83.RF | M17451 |
| | B.US.86.JRFL | U63632 |
| | B.US.90.WEAU160 | U21135 |
| | C.BR.92.92BR025 | U52953 |
| | C.BW.96.96BW0502 | AF110967 |
| | C.ET.86.ETH2220 | U46016 |
| | D.CD.83.ELI | K03454 |
| | D.CD.83.NDK | M27323 |
| | D.CD.84.84ZR085 | U88822 |
| | F1.BE.93.VI850 | AF077336 |
| | F1.BR.93.93BR020.1 | AF005494 |
| | F2.CM.95.MP255 | AJ249236 |
| | F2.CM.95.MP257 | AJ249237 |
| | C.ZA.98.TV001c8.5 | AY16222 |
| | C.ZA.98.TV002c12 | AY16224 |
| | G.BE.96.DRCBL | AF084936 |
| | G.FI.93.HH8793-12.1 | AF061641 |
| | H.BE.93.VI991 | AF190127 |
| | H.BE.93.VI997 | AF190128 |
| | J.SE.93.SE7887 | AF082394 |
| | J.SE.94.SE7022 | AF082395 |
| | K.CD.97.EQTB11C | AJ249235 |
| | K.CM.96.MP535 | AJ249239 |
| B | B.US.84.NY5CG | M38431 |
| | B.US.-.AD8 | AF004394 |
| | B.CN.-.RL42 | U71182 |
| | B.US.90.WEAU160 | U21135 |
| | B.US.-.P896 | U39362 |
| | B.GA.-.OYI | M26727 |
| | B.US.-.DH123 | AF069140 |
| | B.GB.-.CAM1 | D10112 |
| | B.NL.86.3202A21 | U34604 |
| | B.AU.87.MBC925 | AF042101 |
| | B.ES.89.89SP061 | AJ006287 |
| | B.FR.83.HXB2 | K03455 |
| | B.US.83.SF2 | K02007 |
| | B.US.90.WCIPR9018 | U69591 |
| | B.US.84.MNCG | M17449 |
| | B.US.87.BC | L02317 |
| | B.TW.-.TWCYS | AF086817 |
| | B.US.86.JRFL | U63632 |
| | B.AU.86.MBC200 | AF042100 |
| | B.DE.86.HAN | U43141 |
| | B.US.83.RF | M17451 |
| | B.US.97.ARES2 | AB078005 |
| | B.AU.96.MBCC98 | AF042104 |
| | B.KO.97.WK | AF224507 |
| | B.AR.98.ARCH054 | AY037268 |
| | B.AR.00.ARMS008 | AY037269 |
| | B.AR.99.ARMA132 | AY037282 |
| | B.DE.86.D31 | U43096 |
| | B.US.88.WR27 | AF286365 |
| | B.US.86.YU2 | M93258 |
| C | C.ET.86.ETH2220 | U46016 |
| | C.BR.92.BR025 | U52953 |
| | C.IN.93.93IN999 | AF067154 |
| | C.IN.93.93IN904 | AF067157 |
| | C.IN.93.93IN905 | AF067158 |
| | C.IN.94.94IN11246 | AF067159 |
| | C.IN.94.94IN476 | AF286223 |
| | C.IN.95.95IN21068 | AF067155 |
| | C.BW.96.96BW01B03 | AF110959 |
| | C.BW.96.96BW0402 | AF110962 |
| | C.BW.96.96BW0502 | AF110967 |
| | C.BW.96.96BW1104 | AF110969 |
| | C.BW.96.96BW1210 | AF110972 |
| | C.BW.96.96BW17B03 | AF110980 |
| | C.BW.96.96BW1626 | AF110978 |
| | C.BW.96.MJ4 | AF321523 |
| | C.ZM.96.ZM651 | AF286224 |
| | C.ZM.96.ZM751 | AF286225 |
| | C.ZA.97.ZA012 | AF286227 |
| | C.BR.98.98BR004 | AF286228 |
| | C.IL.98.98IS002 | AF286233 |
| | C.IN.98.98IN012 | AF286231 |
| | C.IN.98.98IN022 | AF286232 |
| | C.TZ.98.98TZ013 | AF286234 |
| | C.TZ.98.98TZ017 | AF286235 |
| | C.ZA.98.TV001c8.5 | AY16222 |
| | C.ZA.98.TV002c12 | AY16224 |

[a] Identification information for the sequences is in the format: subtype.country. isolationyear.commonname. The country is represented by the two letter country code using the international naming convention from ISO 3166

base frequencies, and transition/transversion rate ratios (kappa) (Yang, 2000; Yang et al, 2000). Using this approach, we were able to reconstruct the ancestral sequences and internal nodes for each of the12 protease cleavage sites in the B, C, and M group data sets. The number of proximal ancestors for each data set was $n$ - 1, which translated into 29 ancestral sequences for subtype B, 26 sequences for subtype C, and 26 sequences for the group M viruses. The most recent common ancestor (MRCA) nucleotide sequence for each virus in the three data sets was saved and translated into its corresponding amino acids.

### 4.2.4 Diversity and Cleavage Site Polymorphisms

Nucleotide diversity at cleavage sites was measured using a Kimura 2-parameter model with a distance matrix implemented in the MEGA program version 2.0 (Arizona State University, Tempe). Amino acid diversity was measured using a Poisson distribution method implemented in the same MEGA package. *P* values for diversity measurements were calculated by applying the *t* test to the distance matrix of each data set. To determine whether the sequences had evolved over time, amino acid profiles for individual Gag, Gag-Pol, and Nef cleavage sites were compared to the inferred MRCA for that site.

### 4.2.5 Assessment of Positive Selection Pressure

Nucleotide sequences were also analyzed with Codeml, a program from the PAML software package (Rambaut, 2000). The likelihood ratio test (Anisimova et al, 2001) and recently developed codon-based models (Yang, 2000, Yang et al, 2000) were used to assess natural selection and adaptive evolution at the amino acid level. These selection models use maximum-likelihood scores to account for variation in the *dn/ds* (nonsynonymous/synonymous) ratio (ώ) at individual codons along the length of the sequence. High rates of synonymous mutation are indicative of conservation and a strict requirement for biological function, while high rates of nonsynonymous substitution are indicative of adaptive change in response to host selection pressure. An individual amino acid was considered to be positively selected if the *dn/ds* ratio was significantly greater than 1.0.

## 4.3 Results

### 4.3.1 Viral Characteristics

No primary RT- or PI-resistant mutations were detected among the 84 full-length sequences selected for study. Although attempts were made to include only sequences

amplified directly from HIV-1 proviral DNA and to match these sequences based on duration of infection, plasma viral load, and CD4 count, this proved difficult. A surprisingly small amount of full-length sequence data was available from treatment-naive patients infected with subtype B, and when available, it was often poorly annotated. Despite these limitations, the frequency and pattern of naturally occurring polymorphisms observed in this study were remarkably similar to those reported for a control group of subtype B infections treated with nucleoside reverse transcriptase inhibitors but not with PIs or nonnucleoside reverse transcriptase inhibitors (Cote et al, 2001). Most of the non-C sequences came from regions of the world where treatment is not yet readily available.

### 4.3.2 Genetic Diversity and Patterns of Amino Acid Variability at Individual Cleavage Sites

Variation at the 12 cleavage sites of subtypes B and C and group M is shown in Table 11. Seven (58.3%) sites (p17/p24, p24/p2, NC/p1, NC/TFP, PR/RT, RT/p66, and p66/IN) were found to be relatively well conserved, both over time and between subtypes, with a mean intrasubtype distance ranging from 0.40 (0.20%) to 7.49 (5.62%). The remaining five (41.7%) sites exhibited moderate (p1/p6*gag*) to extensive (p2/NC, TFP/p6*pol*, p6*pol*/PR, and Nef) variation, with mean intrasubtype diversities reaching levels as high as 42.42 (15.16%). For the purposes of this study, we have referred to these three patterns as conserved, moderately variable, and variable. Polymorphisms were more common among C than B viruses ($P$=0.0001). Overall, six cleavage sites (p17/p24, p2/NC, NC/p1, PR/RT, RT/p66, and Nef) had significantly higher levels of diversity among C viruses ($P$=0.0001); five sites (p24/p2, p1/p6*gag*, NC/TFP, p6*pol*/PR, and p66/IN) had similar levels of diversity in both subtypes, and one site, TFP/p6*pol*, was highly variable among B and group M viruses (mean distances, 16.7 and 24.7%, respectively) but relatively conserved in C viruses (mean distance, 7.6%) ($P$=0.0001). Compared to the M data set, the level of polymorphism at C cleavage sites was as wide ranging as that observed for the entire M group, a data set containing nine different HIV-1 subtypes. One cleavage site, p17/p24, was significantly more diverse in subtype C (mean divergence, 5.15%) than in group M (mean divergence, 2.23%) and subtype B (mean divergence, 2.31%) viruses (0.0001 for both comparisons). Seven sites (p24/p2, p2/NC, NC/p1, p6*pol*/PR, RT/p66, p66/IN, and Nef) exhibited similar levels of diversity in both data sets ($P$=0.004-0.961). Only four group M cleavage sites (p1/p6*gag*, NC/TFP, TFP/p6*pol*, and PR/RT) had mean diversities that were significantly greater than that observed for subtype C ($P$=0.0001). For >50% of the sites, the variability of B viruses was significantly lower than that

observed for subtype C or group M ($P$=0.0001), despite the fact that the B viruses covered a broader time frame. The observed polymorphisms were not randomly distributed across the variable cleavage sites but were confined to specific amino acids, most of which were positively selected in C but not in B viruses (Figure 15). The least variable residues were the P1 positions of p1/p6*gag*, TFP/p6*pol*, and Nef and the P1' positions of p6*pol*/PR and Nef. As shown, these positions flank the scissile bond, with P1 located immediately upstream and P1' located immediately downstream of the cleavage junction. The most variable positions were the P1, P3, P4, P5, P3', and P5'residues of p2/NC; the P3' residue of p1/p6*gag*; the P1' and P4' residues of TFP/p6*pol*; the P1 to P4 residues of p6*pol*/PR; and the P3 to P5 and P2' residues of Nef.

### 4.3.3 Subtyping and Phylogenetic-tree Analysis

When subjected to phylogenetic analysis (Figure 14), the concatenated 360-bp fragments of the group M data set fell into eight subtype-specific clusters representing subtypes A, B and D, C, F, G, H, J, and K, with the cleavage sequences for subtypes B and D segregating together in the same subcluster. This pattern was supported by high bootstrap values, by high-score maximum-likelihood trees, and by phylogenetic analysis of the deduced amino acids. These findings reveal the subtype-specific nature of protease cleavage sites and suggest that the evolution of cleavage sites parallels that of the full-length genome. With the exception of C.98IN022, B.AR00.ARMS008, and B.US.P896, all of the cleavage site fragments in the B and C data sets segregated into two distinct monophyletic groups representing either subtype B or subtype C viruses (data not shown). The longer branch lengths in the C subcluster were reflective of the increased diversity of C viruses relative to subtype B.

### 4.3.4 Identification and Dating of Common Ancestors

Maximum likelihood methods were next used to reconstruct the internal nodes of the phylogenetic tree and to estimate the times of divergence of individual sequences from their MRCA. These estimates were determined by measuring the number of substitutions along each branch of the tree. MRCAs for the B, C, and group M data sets are shown in Figure 15. Two different patterns were observed based on the relationship between a given sequence and its MRCA. Conserved ($n$=7) and moderately variable ($n$=1) cleavage sites shared the same (identical) MRCA among all three data sets. The proximal location of the MRCA relative to the root of the tree suggests that, for these sequences, cleavage site

**Table 11. Inter- and intra-subtype diversity at 12 cleavage sites, expressed as amino acid distances between sequences**

| Protease site[a] | Mean % distance | | | *P* value | | |
|---|---|---|---|---|---|---|
| | Subtype B | Subtype C | Group M | B vs C | B vs M | C vs M |
| p17 - p24 | 2.31 | 5.15 | 2.23 | <0.0001 | 0.806 | <0.0001 |
| p24 - p2 | 2.41 | 3.07 | 2.67 | 0.077 | 0.454 | 0.267 |
| p2 - p7 | 18.66 | 42.42 | 39.22 | <0.0001 | <0.0001 | 0.062 |
| p7 - p1 | 0.40 | 5.42 | 5.22 | <0.0001 | <0.0001 | 0.7244 |
| p1 - $p6^{gag}$ | 8.93 | 9.81 | 14.47 | 0.318 | <0.0001 | <0.0001 |
| NC-TFN | 5.30 | 3.90 | 6.30 | 0.0012 | 0.0215 | <0.0001 |
| TFN-$P6^{pol}$ | 16.69 | 7.60 | 24.70 | <0.0001 | <0.0001 | <0.0001 |
| $p6^{pol}$ PR | 17.66 | 16.28 | 15.55 | 0.15 | 0.019 | 0.365 |
| PR - RT | 0.70 | 1.56 | 2.93 | <0.0001 | <0.0001 | <0.0001 |
| RT - p66 | 0.81 | 7.49 | 6.03 | <0.0001 | <0.0001 | 0.004 |
| p66 - IN | 2.44 | 2.22 | 2.24 | 0.524 | 0.61 | 0.961 |
| Nef | 11.89 | 24.52 | 25.68 | <0.0001 | <0.0001 | 0.33 |
| Gag (501 aa) | 6.99 | 9.75 | 15.72 | <0.001 | <0.001 | <0.001 |
| Pol (1004 aa) | 6.02 | 5.83 | 5.89 | 0.376 | 0.419 | 0.875 |
| Nef (207 aa) | 14.66 | 16.50 | 18.50 | 0.07 | <0.001 | <0.001 |
| 12 protease sites (360 bp -120 aa) | 4.80 | 10.10 | 12.10 | <0.001 | <0.001 | <0.001 |

[a]amino acids, aa

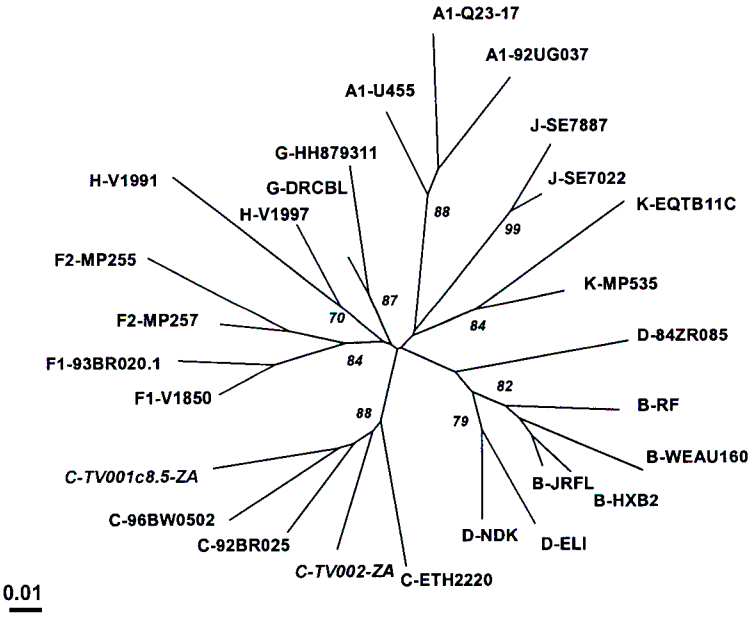



**Figure 14. Phylogenetic relationships of the South African Tygerberg virology (TV) cleavage site sequences relative to other subtypes in the group M data set. This representative maximum-likelihood tree is based on concatenation and analysis of the 12 protease site nucleotide sequences as a single segment of 360 bp. An indication of the degree of sequence dissimilarity is given by the distance from the central node. The percentage of bootstrap trees out of 1,000 replications supporting a particular phylogenetic group is shown alongside the node considered.**

p17/p24 (MA/CA) - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | V | S | Q | N | Y | / | P | I | V | Q | N |
| M group | I(1) | | | | | | | | L(1) | | |
| | S(1) | | | | | | | | | | |
| | D(1) | | | | | | | | | | |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | A(1) | | R(1) | | | | | | | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | I(3)* | | R(1) | | | | | | A(1) | | |
| | D(2) | | | | | | | | | | |

p24/p2 (CA/p2) - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | K | A | R | V | L | / | A | E | A | M | S |
| M group | | | | I(4) | | | | | | | |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | | | | I(2) | | | | | | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | | | K(1) | I(2) | | | G(1) | | | | |

p2/NC (p2/p7) - | Variable |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | S | T | A | I | M | / | M | Q | K | G | N |
| M group | N(11)* | A(10)* | N(7)* | V(4) | L(1) | | L(1) | | R(13) | S(13) | |
| | A(3) | S(3) | T(6) | A(3) | | | V(1) | | | | |
| | H(2) | P(1) | S(2) | M(1) | | | I(1) | | | | |
| | T(1) | V(1) | V(1) | | | | | | | | |
| | P(1) | | | | | | | | | | |
| | Q(1) | | | | | | | | | | |
| | G(1) | | | | | | | | | | |
| | V(1) | | | | | | | | | | |
| B MRCA | . | A | T | . | . | / | . | . | R | . | . |
| Subtype B | P(3) | N(2) | N(3) | V(3) | | | L(1) | | K(5) | S(1) | |
| | T(1) | T(2) | A(3) | M(1) | | | I(1) | | G(1) | | |
| | A(1) | V(1) | | | | | | | | | |
| C MRCA | N | . | N | . | . | / | . | . | . | S | . |
| Subtype C | S(13)* | A(6)* | S(2)* | V(2) | L(9)* | | I(2) | | R(16) | G(5)* | |
| | T(1) | N(4) | T(1) | | | | V(1) | | | N(1) | |
| | | V(2) | G(1) | | | | | | | | |
| | | M(1) | | | | | | | | | |
| | | S(1) | | | | | | | | | |

NC/p1 (p7/p1) - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | E | R | Q | A | N | / | F | L | G | K | I |
| M group | G(1) | | | | | | | | | R(2) | F(1) |
| | K(1) | | | | | | | | | | M(1) |
| | | | | | | | | | | | L(1) |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | K(1) | | | | | | | | | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | | G(1) | R(1) | | D(1) | | | | | R(3) | L(1) |

NC/TFP - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | E | R | Q | A | N | / | F | L | R | E | N |
| M group | G(1) | | | | | | | | | | D(6) |
| | | | | | | | | | | | T(1) |
| | K(1) | | | | | | | | | | V(1) |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | K(1) | | | | | | | | | | D(21) |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | | G(1) | R(1) | | D(1) | | | | | | T(1) |
| | | | | | | | | | | | D(1) |

p1/p6$^{Pol}$ - | Moderately Variable |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | R | P | G | N | F | / | L | Q | S | R | P |
| M group | | | | | L(1) | | P(5) | | N(10)* | | L(3)* |
| | | | | | | | I(1) | | K(1) | | S(1) |
| | | | | | | | | | | | T(1) |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | | R(1) | | | | | P(4) | | N(3) | | L(3)* |
| | | | | | | | | | R(1) | | T(1) |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | | | | | L(1) | | | | N(11)* | | T(2) |
| | | | | | | | | | G(1) | | S(1) |
| | | | | | | | | | | | L(1) |

TFP/p6$^{Pol}$ - | Variable |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | E | N | L | A | T | / | Q | Q | G | E | A |
| M group | | D(6) | | V(1) | S(1) | | P(10)* | K(2) | R(5) | K(9) | |
| | | T(1) | | | | | L(2) | | | | |
| | | V(1) | | | | | | | | | |
| B MRCA | . | . | . | . | . | / | P | | | K | . |
| Subtype B | | D(21) | M(1) | V(1) | | | L(8) | R(1) | R(2) | E(4) | |
| | | | | | | | Q(1) | | | | |
| C MRCA | . | . | . | . | . | / | P | . | . | . | G |
| | | D(1) | | | | | Q(1) | E(1) | | K(3)* | |
| Subtype C | | | | | | | | | | | |
| | | T(1) | | | | | | | | | |

p6$^{Pol}$/protease - | Variable |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | T | S | F | S | F | / | P | Q | I | T | C |
| M group | V(9)* | T(4)* | L(11)* | N(14)* | C(1) | | | | | | |
| | P(5) | A(1) | S(3) | D(3) | | | | | | | |
| | G(3) | G(1) | | C(1) | | | | | | | |
| | S(3) | | | R(1) | | | | | | | |
| | R(1) | | | | | | | | | | |
| | D(1) | | | | | | | | | | |
| | Q(1) | | | | | | | | | | |
| B MRCA | V | . | . | . | . | / | . | . | . | . | L |
| Subtype B | I(3) | | L(2) | N(7) | L(5) | | | R(1) | V(1) | | |
| | | | | D(2) | C(1) | | | H(1) | | | |
| C MRCA | G | T | L | K | . | / | . | . | . | . | L |
| Subtype C | | S(2)* | F(8)* | V(1)* | L(3)* | | L(1) | | | A(1) | |
| | | N(2) | | S(1) | C(3) | | | | | | |
| | | A(3) | | | | | | | | | |

protease/RT - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | C | T | L | N | F | / | P | I | S | P | I |
| M group | R(1) | | | H(1) | L(1) | | | V(1) | | | |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | | | | | | | | | C(1) | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | R(1) | | | D(1) | | | | | | | |

RT/p66 - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | G | A | E | T | I | / | Y | V | D | G | A |
| M group | | V(4) | D(1) | | Y(4) | | | | | | |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | F(3) | | | | | | | | | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | L(1) | V(10)* | | | Y(2)* | | | | | | |

p66/IN - | Conserved |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | I | R | K | V | L | / | F | L | D | G | I |
| M group | V(1) | | M(1) | | | | | | | | |
| B MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype B | V(1) | | R(1) | I(1) | | | | | N(1) | | |
| C MRCA | . | . | . | . | . | / | . | . | . | . | . |
| Subtype C | | | R(2) | | | | | | | | |
| | | | I(1) | | | | | | | | |

NEF - | Variable |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M MRCA | P | D | C | A | W | / | L | E | A | Q | E |
| M group | A(14)* | A(4)* | L(4)* | V(2) | | | V(1) | K(1) | | | T(2) |
| | D(2) | S(3) | S(1) | T(1) | | | | Q(1) | | | Q(1) |
| | E(1) | T(1) | V(1) | | | | | | | | |
| | | | Y(1) | | | | | | | | |
| B MRCA | A | . | . | . | . | / | . | . | . | . | . |
| Subtype B | P(1) | A(7) | I(1) | V(1) | | | Q(1) | K(1) | | H(1) | |
| | | N(1) | | T(1) | | | | | | | |
| | | T(1) | | | | | | | | | |
| C MRCA | A | . | . | . | . | / | . | . | . | . | . |
| Subtype C | P(3)* | A(7)* | L(1)* | | R(1) | | | R(5)* | P(2)* | | Q(1) |
| | T(1) | H(1) | G(1) | | | | | Q(4) | E(1) | | K(1) |
| | E(1) | E(1) | | | | | | K(2) | T(1) | | |

**Figure 15. Amino acid polymorphisms at Gag, Gag-Pol, and Nef cleavage sites. The letters refer to the amino acid substitutions; the numbers in parentheses refer to the number of times the substitution was observed. Each cleavage site sequence consists of the 5 amino acids upstream and the 5 amino acids downstream of the scissile bond, indicated by a shill. The labeling of amino acids is according to the convention of P1 to P5 going from the scissile bond toward the amino terminus and P1' to P5' going toward the carboxy terminus. Positively selected amino acids are marked with asterisks. Dots represent amino acids that are identical to those in the M MRCA.**

diversification occurred after subtype divergence. In contrast, variable cleavage sites ($n$=4) showed a high degree of divergence both from their subtypespecific MRCA and from the group M MRCA. Ancestral nodes for the variable sites were located closer to the tips of the tree (data available upon request).

### 4.3.5 Variability of Cleavage Sites Relative to Other Regions of the HIV-1 Genome

For C viruses, the average intersequence divergence among the concatenated cleavage site fragments was higher (10.1%) than those observed for the Gag (9.8%) and Pol (5.8%) proteins but lower (16.5%) than that observed for Nef. In contrast, B cleavage sites were significantly less diverse (4.8%; $P$=0.0001) than those of the Gag (7.0%), Pol (6.0%), and Nef (14.7%) proteins of subtype B. Among group M viruses, cleavage site diversity was significantly higher (12.1%) than that calculated for Pol (5.9%) but lower than that determined for Gag (15.7%) and Nef (18.5%). These results are presented in more detail in Table 11.

### 4.3.6 Physical-chemical Properties of Amino Acids at P1-P1'Cleavage Junctions

Overall, excluding the highly conserved asparagine (N) residue at the P1 position of NC/p1 and NC/TFP, >97.0% of P1-P1' amino acids in group M were nonpolar. Of these, 77.8% were hydrophobic, 21.5% were small amino acids (78.9% proline, 20.8% alanine, and 0.3% glycine), 0.3% were polar uncharged (one serine and three glutamine), 0.2% were polar charged (one arginine and one aspartic acid), and 0.3% were ungrouped (cysteine) residues. These amino acids were localized at specific positions within the cleavage sequence. The small amino acids were localized primarily to the P1' position of p17/p24, p24/p2, TFP/p6*pol*, p6*pol*/PR, and PR/RT. Hydrophobic amino acids were concentrated at the P1-P1' junction. As previously reported for HIV-1 B (23), the P1-P1' amino acids of subtype C fell into two different patterns defined by the size of the P1' amino acid: type I, represented by p2/NC and p1/p6*gag*, and type II, represented by p17/p24 and p24/p2. Both types carried large nonpolar, hydrophobic amino acids (leucine, tyrosine, phenylalanine, and methionine) at position P1 and either a large (type I) or small (type II) hydrophobic amino acid (proline, alanine, or glycine) at P1'.

### 4.3.7 MRCAs and Subtype-specific Signature Patterns

A summary of amino acid signature patterns relative to the subtype B and C and group M MRCAs is shown in Tables 12 and 13. Mutations at cleavage sites defining the enzymatic (PR/RT, RT/p66, and p66/IN) and structural (p17/24, p24/p2, and NC/p1) components of

HIV-1 were relatively uncommon and, when detected, were found at greater frequencies among C versus B viruses. In total, only 1.2, 0.8, and 1.1% of the 840 amino acids at each of the PR/RT, p66/IN, and p24/p2 cleavage sites carried substitutions. As a result, the majority of sequences at conserved sites (81.5 to 96.7%) were identical to both the subtype-specific MRCA and the common ancestor of the group M viruses. Several cleavage sites involved in the regulation of polyprotein processing and protease activation, p2/NC, TFP/p6*pol*, and p6*pol*/PR, were highly variable and differed significantly from both the subtype-specific and group M MRCAs. With the exception of the TFP/p6*pol* site, which was more variable in subtype B, divergence from the MRCA was greatest for C and M viruses. None of the p2/NC and p6*pol*/PR sequences in the group M data set was identical to the M MRCA signatures for p2/NC and p6pol/PR, STAIM/MQKGN and TSFSF/PQITC, respectively.

### 4.3.8 Positive Selection of Amino Acids at Protease Cleavage Sites

The concatenated 360-bp cleavage site fragments were next compared internally to assess the mutational behavior of P1-P5 and P1'-P5' sites in the absence of drug therapy. As described in Materials and Methods, the analyses were performed using codon-based maximum-likelihood methods that test for the variation in selection pressure (*dn/ds*) at individual amino acids along the length of the sequence. Application of the likelihood ratio test indicated that the best-fit model for subtype C and group M sequences was the positive-selection (discrete) model ($\chi^2$ = 54.47; $P < 0.0001$ and 62.34 and $P < 0.0001$, respectively), while for B viruses the neutral model performed as well as the positive model ($\chi^2$ = 1.2; $P > 0.05$). Overall, when analyzed as a single concatenated fragment, all three datasets were found to be under negative (purifying) selection, with *dn/ds* (ω) values ranging from 0.26 to 0.31 for all sites. Only 20 (16.6%) of the 120 amino acids within the 12 cleavage sites of subtype C were found to be under positive selection, with a ω3 value of 1.7. Group M and subtype B cleavage sites had fewer positively selected amino acids: 11.1 and 6.7%, respectively (Figure 15).

## 4.4 Discussion

The presence of polymorphisms in the protease of subtype C would be expected to result in adaptive (compensatory) changes in the natural cleavage sites that are recognized and cleaved by the C enzyme. To test this hypothesis, we examined the prevalences and patterns of cleavage site mutations in the Gag, Gag-Pol, and Nef proteins of subtype C compared to those of non-C viruses. Using phylogenetic and ancestral reconstruction

**Table 12. Relationship between cleavage site signature patterns and common ancestors[a]**

| Group or subtype | No. (%) of sequences with the same MRCA | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | p17/p24(MA/CA) VSQNY/PIVQN | p24/p2 (CA/p2) KARVL/AEAMS | NC/p1 (p7/p1) ERQAN/FLGKI | NC/TFP ERQAN/FLREN | p1/p6$^{gag}$ RPGNF/LQSRP | protease/RT CTLNF/PISPI | RT/p66 GAETF/YVDGA | p66/IN IRKVL/FLDGI |
| M | 23 (85.2) | 23 (85.2) | 21 (77.7) | 17 (63.0) | 13 (48.1) | 22 (81.5) | 19 (70.4) | 25 (92.6) |
| B | 28 (93.3) | 28 (93.3) | 29 (96.7) | 8 (26.7) | 23 (76.7) | 29 (96.7) | 29 (96.7) | 26 (86.7) |
| C | 21 (77.7) | 23 (85.2) | 21 (77.7) | 22 (81.5) | 15 (55.6) | 25 (92.6) | 14 (51.9) | 24 (88.9) |

a) Conserved and moderately variable cleavage sites (group M MRCA is identical to subtype B and C MRCAs)

**Table 13. Relationship between cleavage site signature patterns and common ancestors[a]**

| Group or subtype | p2/NC (p2/p7) | | TFP/p6$^{pol}$ | | p6$^{pol}$ /PR | | NEF | |
|---|---|---|---|---|---|---|---|---|
| | Sequence | No. (%) same | Sequence | No. (%) same | Sequence | No. (%) same | Sequence | No. (%) same |
| M | STAIM/MQKGN | 0 (0) | ENLAF/QQGEA | 6 (22.2) | TSFSF/PQITC | 0 (0) | PDCAW/LEAQE | 5 (18.5) |
| B | SATIM/MQRGN | 13 (43.3) | ENLAF/PQGKA | 5 (16.7) | VSFSF/PQITL | 13 (43.3) | ADCAW/LEAQE | 18 (60.0) |
| C | NTNIM/MQKSN | 3 (11.1) | ENLAF/PQGEG | 21 (77.8) | GTLNF/PQITL | 10 (37.0) | ADCAW/LEAQE | 6 (22.2) |

[a]Variable cleavage sites (subtype B and C MRCAs differ from the group M MRCA)

methods, we identified two groups of natural cleavage site sequences. The first group consisted of eight sequences, seven of which were highly conserved in all HIV-1 subtypes, and an eighth sequence which was moderately variable. Mutations at these sites were uncommon and, when present, were observed at relatively equivalent frequencies among different subtypes. These cleavage sites, which defined the main structural (MA, CA, and NC) and enzymatic (RT, RNase H, and integrase) proteins of HIV-1, were under strong negative (purifying) selection pressure, had a common ancestor, and showed little genetic evolution over time. The second group consisted of five cleavage sites that were under positive (diversifying) selection pressure, exhibited extensive inter- and intrasubtype variability, and showed little (or no) resemblance to the common ancestor of group M or to the subtype-specific MRCA. Our data showing that the majority (58.3%) of cleavage sites are highly conserved in all subtypes was not unexpected, given the narrow specificity of the HIV-1 protease relative to cellular proteases, such as pepsin (Ridky et al, 1996). The strong purifying selection pressure exerted on these sites is presumably a reflection of the need to maintain the spatial configuration of the enzyme-substrate complex, conserve the hydrophobic nature of the scissile bond, and retain the biological activity of functionally important sites, such as the P1' proline of p17/p24 and the P1 and P1'-P5' residues of NC/p1. Cleavage of the p17/p24 site is known to play an important role in virion maturation, while processing of NC/p1 is required for ribosomal frame shifting and Gag-Pol expression (Ericson-Viitanen et al, 1989, Jacks et al, 1998; Kaplan et al, 1994; Krausslich et al, 1989, Pettit et al, 1994). In B viruses, cleavage of p17/24, p24/p2, and NC/p1 has been shown to be suboptimal, with the NC/p1 site being rate limiting (Pettit et al, 2002). It has been suggested that the slow, regulated cleavage of these structural proteins may represent a common strategy to ensure that the assembled virions have the full complement of proteins needed to bud from the cell surface, bind to a new cell, and initiate a new round of viral replication (Tessmer et al, 1998). The carboxyl terminus of NC is particularly interesting. Unlike other cleavage sites, which carry an aromatic amino acid at P1 and either a leucine or proline residue at P1', the C termini of NC/p1 and NC/TFP carry an asparagine (N) residue at P1 opposite a phenylalanine (F) residue at P1' (Cote et al, 2001, Pettit et al, 2003; Pettit et al, 2002; Pettit et al, 1994). In this study of 84 untreated patients, no mutations were detected at the P2 or P1'-P3' positions of NC/p1 or NC/TFP and only a single N3C mutation was detected at P1. Taken together, these findings underscore the unique nature and limited mutability of the NC/p1 and NC/TFP cleavage junctions. Although these sites were strongly conserved in natural infection, recent studies have shown that an A3V substitution at the P2 positions of NC/p1 and

NC/TFP is a common adaptive change, occurring in 29% of PI-resistant patients taking indinavir, saquinavir, and/or ritonavir for the treatment of subtype B (Cote et al, 2001; Doyon et al, 1996; Zennou et al, 1998; Zhang et al, 1997). This valine substitution is frequently associated with an M46 I or L mutation (and possibly a V82 mutation) in the protease and leads to altered polyprocessing and improved growth of protease-mutated viruses. Whether similar "second locus" mutations will be observed during the treatment of non-B subtypes remains to be established. The identification of common patterns may facilitate the development of broad based inhibitors with increased specificity and improved binding to the mutated protease. These secondary inhibitors might preempt (or delay) the emergence of resistance. Our analyses also revealed important differences among HIV-1 subtypes. Particularly intriguing was the identification of five cleavage sites that exhibited extensive variability across all subtypes, with C viruses being significantly more variable than subtype B. Variation was restricted to a few specific amino acids, most of which were positively selected in C but not in B viruses. In contrast to conserved sites, variable cleavage sites tended to be those with regulatory rather than structural or enzymatic functions. At least four of the variable sites (p2/NC, p1/p6gag, TFP/p6pol, and p6pol/PR) are known to play major roles in the regulation of polyprotein processing and, in the case of TFP/p6*pol*, in the activation of the protease enzyme (Pettit et al, 2003; Pettit et al, 2002; Pettit et al, 1994; Ridky et al, 1996; Shehu-Xhilaga et al, 2001). Studies of subtype B have shown that p2/NC is the initial and most rapidly processed cleavage site, controlling both the rate and the order of Gag and Gag-Pol polyprocessing (Shehu-Xhilaga et al, 2001). Our results indicate that p2/NC is by far the most variable cleavage site, with intrasubtype diversity ranging from 18.7% in subtype B to levels of 42.4% in subtype C. The p1/p6*gag* cleavage product, p6*gag*, is a major phosphoprotein that is critical to the release of mature, infectious virions (Muller et al, 2002). Although not well studied, phosphorylation of Gag and Gag-Pol sequences has been shown to alter susceptibility to cleavage, attenuating or even preventing the proteolytic process (Tomasselli and Heinrickson, 1994). The TFP/p6*pol* cleavage site, defining the N terminus of p6*pol*, was the only site to have a significantly higher level of diversity among B than among C viruses. TFP/p6*pol* is a novel cleavage site located 8 amino acids downstream from NC in the TFP domain of Gag-Pol (Pettit et al, 2003, Tessmer et al, 1998). Although TFP/p6*pol* lies outside (and upstream) of the protease, the EDL tripeptide of this cleavage site (ENL in the case of C viruses) has been postulated to have a major influence on protease activation and on the timing and specificity of Gag-Pol cleavage, delaying the release of the protease until after the viral particle has budded from the cell membrane. Such a

mechanism may protect the cell from the cytotoxic effects of proteolysis (Thomas et al, 1996; Tomasselli and Heinrickson, 1994). The observed subtype variation in the cleavage sites controlling the initiation and rate of Gag and Gag-Pol processing (p2/NC) and the activation of protease (TFP/p6*gag*) suggests that there may be important differences in the way that B and C viruses regulate polyprocessing and virion assembly. Subtle cleavage site differences could, over time, have a major differential impact on the pathogenesis of HIV-1 subtypes and on response to therapy. Early treatment studies suggest that C viruses give an excellent initial response to highly active antiretroviral therapy but that the duration of the response may be less than that reported for B viruses.

In summary, our results point to important inter- and intrasubtype differences in protease cleavage sites, especially in the p2/NC, TFP/p6pol, and p6pol/PR sites. The main limitations of our study relate to the cross-sectional nature of the data sets and the limited availability of well-matched pretreatment controls for use in the B data set. Despite these limitations, the potential impact of our findings on HIV-1 disease progression and response to therapy warrants further investigation, both at the patient level and in vitro using site-directed mutagenesis. The separate monophyletic clustering of B and C cleavage sites suggests that cleavage sites have evolved in a subtype-specific manner. The divergence between ancestral and contemporary sequences in the C data set and the location of an ancestral node distal to the group M MRCA suggest that variation in C cleavage sites began early, prior to the diversification of HIV-1 subtypes. A more detailed investigation of C cleavage sites, both over time and in response to therapy, is in progress. The present study forms the baseline for these ongoing studies.

# Chapter 5

## Molecular Characteristics of Retrospective South African Drug Naive Samples

## 5.1 Introduction

The history of Southern African HIV infection dates back to 1982, when the first South African AIDS cases were reported among the homosexual population (Sher et al, 1989). Until 1987, HIV-1 diagnosis was almost completely limited to the male portion of the population. In 1988, Martin et al reported a HIV-1 prevalence of 20% among homosexual men and 1% among heterosexual men attending a STD clinic in Johannesburg. The same clinic found a1.2% and 0.8% HIV prevalence among black females and males respectively. In 1989, Sher presented data from a number of surveillance studies on the prevalence of HIV infection in South Africa that showed an increase in the number of AIDS cases among the black population, through the heterosexual route of transmission. It was later shown that two HIV-1 epidemics were present in South Africa, a smaller epidemic among the homosexual and bisexual men associated with subtype B and a few subtype D infections, and a subtype C epidemic among the heterosexual population (Williamson et al, 1995; Van Harmelen et al, 1997). By 1992, the reported cases in women roughly equalled those in men. Another study in rural KZN found the prevalence in women to be three times higher in females than in males (Abdool Karim et al, 1992). The prevalence of HIV-1 continued to increase, with 16% of pregnant women attending antenatal clinics testing seropositive in 1997 (Van Harmelen et al, 1999). Van Harmelen noted geographical variation in the extent of the epidemic within South Africa. KZN had the highest prevalence in the country with 26.92%, while the Western Cape had a prevalence of only 6.29% (Van Harmelen et al, 1999). This increased to 36.2% in KZN and 8.7% in the Western Cape by 2002 (Zur Megede et al, 2002). In 2002, Bredell et al reported that while subtype C was still predominant, there was an increase in the number of non-C subtypes detected. Other subtypes were probably introduced independently into different regions of South Africa by immigrants from other African countries (Van Harmelen et al, 1999; Bredell et al, 2002).

Between 1984 and 1990, serum was routinely collected and virus isolated from patients in the Western Cape diagnosed with AIDS and AIDS related conditions (Engelbrecht et al, 1995). To our knowledge, these samples represent the only properly stored archival specimen bank in SA, and thus may serve as an important source of information on relating to the evolutionary history of HIV-1 in South Africa. Accurate surveillance of HIV-1 prevalence and incidence, and systematic monitoring of the genetic diversity, distribution and spread of HIV-1 variants is critical not only to the design of effective

vaccines and intervention strategies, but also to formulate policy and predict the future directions of the global AIDS pandemic. An understanding of the extent of genetic diversity in different geographic regions, and of its changing dynamics over time is also needed to ensure that diagnostic assays are sensitive and appropriate for the population being studied, and that they remain sensitive and specific over time.

As discussed in Chapter 3, recent sequencing in our laboratory has indicated the HIV/AIDS epidemic is still primarily subtype C restricted, while in Cape Town, the epidemic is still diverse, but becoming more subtype C dominated (Bredell et al, 2002; Engelbrecht et al, 2001). To search for underlying genetic differences between these two regional epidemics, we have conducted a collaborative study with colleagues from Tygerberg Hospital, University of Stellenbosch, Cape Town. The study involved comparative sequencing of the RT and PR, as well as *env* in a subset of samples, of representative archival and contemporary sequences from Cape Town, and representative currently circulating and archival HIV-1 strains from Durban and the surrounding areas.

The results of this study provide a baseline for examining the impact of specific amino acid substitutions on the structure, function and biological properties of HIV-1 C viruses and the relationships between genetic variation, resistance and response to therapy.

## 5.2 Methods

### 5.2.1 Samples

Thirty archival samples and nine contemporary samples (2002) were obtained from colleagues at Tygerberg Hospital, Western Cape. Although these samples were randomly chosen from the archival specimen bank to span from the earliest samples in 1984 to 1992, there is a potential sampling bias, as surveillance in the earlier days of the epidemic was not systematic. The samples from the Western Cape were compared with 8 archival samples (1996-1998) collected as part of the national seroprevalence survey, and 46 contemporary (2002) samples from KwaZulu-Natal, a subset of the samples described in Chapter 3.

### 5.2.2 RNA Extraction and Resistance Genotyping

HIV-1 RNA was extracted and sequenced using the Viroseq™ system (Applied Biosystems) according to the manufacturer's protocol and as described in Chapter 3.2.4. Sequences were analysed for resistance as described in Chapter 3.2.5. Signature patterns (VESPA), biologically significant sites (PROSITE, selection pressure (PAML) and the Molecular clock were estimated as described in Chapter 2.1.

## 5.3 Results

HIV-1 subtypes circulating in CT between 1984 and 1992 included subtypes B, C and D, with a predominance of subtype B infections (Table 14). This appears to have changed over time, with subtype C becoming the predominant subtype, although the other subtypes are still present. This is in sharp contrast to KZN, where subtype C appears to have dominated from the onset of the local epidemic (Figure 16). Comparative phylogenetic analysis revealed few genetic changes in biological important sequences, or in signature patterns. However, a number of signature mutations occurred at polymorphic sites, namely the absence of the RT 211K polymorphism found in the subtype C recent isolates (Table 15). This amino acid is under strong positive (diversifying) selection pressure and may offer a fitness advantage to C viruses. When analysed using models that discriminate positive, neutral, and purifying selection pressure at different positions along the length of the protein, the protease and RT were found to be under strong purifying selection with more than 97 % of sites having a $\omega$ value $< 0.5$). Positions under positive selection included 37N and 63P in protease and positions 123D, 174Q, 211K, 245Q and 277R in reverse transcriptase.

The evolutionary rates of the South African HIV-1 C *pol* and *env* genes were 35% and 68% higher than for subtype B, respectively. The mutation rates for *pol* were approximately 0.001 and 0.00189 sites per year for subtypes B and C respectively, while the *env* mutation rate was 0.002 for subtype B and 0.003 for subtype C (Table 16), values that are similar to other previously published rates (Korber et al, 2000). The molecular clock hypothesis was only accepted for the *pol* gene for subtypes B and C.

The relatively high nucleotide mutation rate of the *pol* gene was not reflected in the amino acid sequence. Most mutations were silent (synonymous), a reflection of the non-homogeneous distribution of nucleotide substitutions in the three different codon positions. The mutation rate in the third position was four times higher than in the second position, and 30 times higher than in the first codon position. Substitutions at the third codon position is responsible for many synonymous mutations (Table 17).

Likelihood and genetic distance trees of HIV-1 C polymerase (RT and PR) sequences from South Africa revealed the presence of multiple scattered lineages. The lineages probably arose from different introductions of founder strains. The presence of HIV-1 isolates collected prior to 1992 at internal (basal) branches in three of these lineages suggests that these sequences may represent founder strains that have existed for at least 10 years (see Figure 9 in Chapter 3).

**Table 14. Subtypes Circulating in CT and KZN**

| | |
|---|---|
| Subtypes circulating in CT between 1984 and 1992:<br>subtype B: 15/30 (50%)<br>subtype C: 10/30 (33%)<br>subtype D: 5/30 (17%). | Subtypes circulating in CT between 2000 and 2002:<br>subtype B: 2/9 (22%)<br>subtype C: 6/9 (67%)<br>subtype D: 1/9 (11%) |
| Subtypes circulating in KZN between 1996 and 1998:<br>subtype C: 8/8 (100%) | Subtypes circulating in KZN between 2000 and 2002:<br>subtype C: 45/46 (98%)<br>D/C recombinant: 1/46 (2%) |

**Table 15. Signature mutations in subtype B and C isolates**

| Subtype C isolates | Signature against consensus C | Subtype B isolates | Signature against consensus B |
|---|---|---|---|
| KZN retrospectives | V60I, K211R, R277K | CT retrospective | K277R |
| CT retrospectives | K211R | CT recent | I135V, A272S, K277R, A288S, I293V |
| KZN recent | V60I | | |
| CT recent | G123S | | |

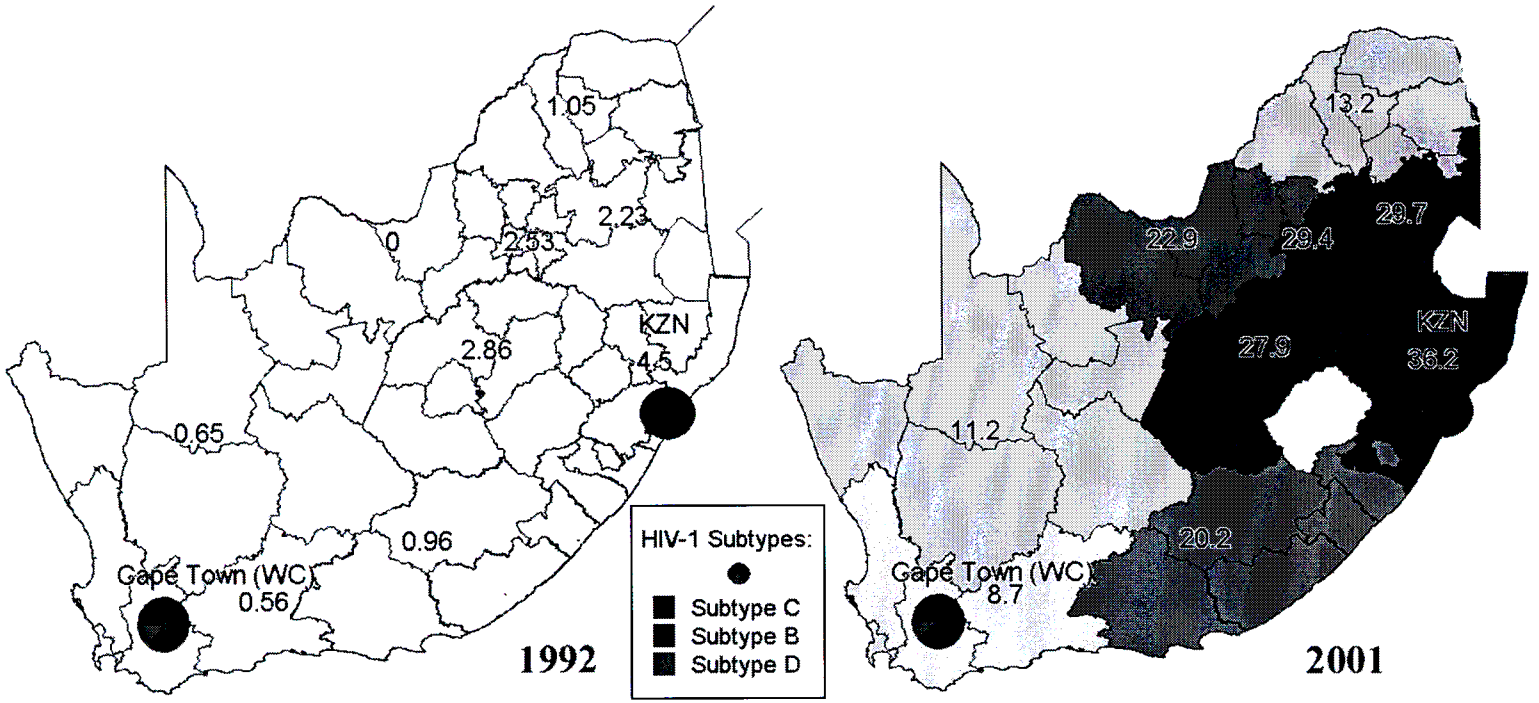



**Figure 16. Changes in HIV-1 subtype distribution and prevalence in CT and KZN from 1992 to 2001. The HIV-1 prevalence is represented by the background colour ranging from white (<1% prevalence) to burgundy (>30% prevalence). Subtype distribution is shown in the pie-charts. Prevalence rates were obtained from the national HIV-1 seroprevalence surveys.**

**Table 16. Estimated optimal likelihood trees and molecular clock likelihood values. The molecular clock is accepted for pol (protease and reverse transcriptase) in subtype B and C, but is rejected for the envelope. Note that the estimated mutation rate was higher in the subtype C dataset for both genes**

| Subtype B Genetic region | Likelihood value | Molecular Clock (TipDate) Likelihood value | Mutation rate (mutations per site per year) ± Stand. Error | Likelihood ratio Test (P value) |
|---|---|---|---|---|
| POL (14) | - 3021.59 | -3032.226 | 0.00114 (±0.000312) | 21.272 (P > 0.01) |
| ENV (16) | -4126.9399 | -4152.625 | 0.00207 (± 0.0012) | 51.372(P < 0.001) |

| Subtype C Genetic region | Likelihood value | Molecular Clock (TipDate) Likelihood value | Mutation rate (mutations per site per year) ± Stand. Error | Likelihood ratio Test (P value) |
|---|---|---|---|---|
| POL (20) | -4270.5180 | -4279.252 | 0.00189 (±0.000248) | 17.22(P >0.05) |
| ENV (23) | -5384.2993 | -5443.709 | 0.00303 (± 0.000877) | 118.819(P < 0.001) |

**Table 17. Rate of mutations at the first, second and third codon position**

| | First | Second | Third |
|---|---|---|---|
| Subtype B | 0.07925 | 0.53754 | 2.38322 |
| Subtype C | 0.16664 | 0.68526 | 2.14810 |

## 5.4 Discussion

South Africa continues to have at least two distinct HIV-1 epidemics characterized by the presence of different genetic subtypes, epidemic timing and HIV-1 prevalence values (Bredell et al, 2002; Zur Megede et al, 2002). The impact of these genetic differences on the transmission dynamics and biological properties of HIV-1 are unknown. Our studies suggest that there are no primary resistance mutations in the protease and RT of drug naive subtype C isolates, as well as no major changes in signature sequences of biologically important sites. However, variation at polymorphic sites may alter the conformation of RT, PR and *env* and lead to indirect changes in replicative capacity, ability to cleave Gag-Pol polyproteins and changes in tropism.

In this study, differences between retrospective and recent sequences from subtype B and C viruses from the Western Cape and KZN were examined in an attempt to identify codon positions that were selected over time and could potentially offer a fitness advantage to C viruses. A few differences were identified, including the RT R211K mutation, that is under strong positive (diversifying) selection pressure in contemporary subtype C isolates but absent in subtype B sequences and Western Cape retrospective subtype C isolates. This mutation, when present with L214F and/or H208Y, has been shown to cause a 2.4- or 8-fold increase in AZT resistance (Sturmer et al, 2003). Another mutation, V60I, was present only in the KZN signature patterns. Both of these mutations do not occur at functional sites and their potential role in viral fitness needs to be further investigated *in vitro*, with site directed mutagenesis and replication assays.

Correlation of subtype distribution with HIV-1 prevalence, suggests that subtype C is outgrowing other subtypes in Cape Town, while remaining predominant in KwaZulu-Natal. This is similar to other regions such as the Congo, Tanzania, Burundi, Kenya, China and Brazil where subtype C is becoming predominant relative to existing subtypes (Soares et al, 2003; Koch et al, 2001; Rodenburg et al, 2001; Brindeiro et al, 1999; Gao et al, 1998; Yu et al, 1998; Janessens and Nkengasong, 1997). Although subtype B has been repeatedly introduced in different parts of the world, it doesn't seem to take off in the same manner as subtype C. Differences in host genetics, socio-economic factors, presence of other pathogens, as wells as differences in tropism and replicative capacity could account for the different outcomes.

Taken together, these findings suggest the diversity and distribution of HIV-1 subtypes in South Africa has changed since the beginning of the local epidemic 22 years ago. Major changes relate to the increasing prevalence of multiple subtypes, and the rapid ongoing

spread of subtype C viruses. Evidence of multiple co-circulating subtypes in the Western Cape emphasizes the need to carefully monitor for the emergence of recombinants. A number of studies have suggested that subtypes A and C may be more fit in vitro, and possibly in vivo, at the population level, but the evidence is far from conclusive (Velazques-Campoy et al, 2001; Velazques-Campoy et al, 2002; Arts et al, 2001; Ball et al, 2003). For example, the study by Ball et al (2003) found that while subtypes B and C displayed equal fitness in Langerhans cells, subtype B was more fit in T-cells and macrophages. They suggest that subtype C was less fit following initial infection and this could lead to slower disease progression (Ball et al, 2003). This slower disease progression could in turn allow for an increased time for transmission (Quinones-Mateu et al, 2000). Continued, ongoing studies such as those described in this thesis, detailing the genetic changes in subtype C in different population groups over time may shed some light on this complex issue. Studies will be most informative if conducted in a geographic region where multiple subtypes co-exist concurrently. This will allow for better comparisons with the least number of variables. The comparison of the Western Cape versus KwaZulu-Natal is an ideal situation.

# Chapter 6

## Emergence and Patterns of Resistance in HIV-1 subtype C Isolates from Treated Patients in KZN, South Africa

## 6 Introduction

The introduction of HAART for the treatment of HIV-1 has greatly reduced the morbidity and mortality associated with this disease (Palella et al, 1998). However, these benefits have mainly been felt in developed countries such as North America and Europe.

Subtype C is one of the most prevalent subtypes worldwide and is especially common in sub-Saharan Africa where the prevalence is estimated to be up to 36% (UNAIDS/WHO: Report, 2000). However, most data on the response of HIV-1 to ARV therapy has been from studies using subtype B. These studies have shown that in previously untreated individuals with drug-susceptible HIV-1 strains, combinations of three or more drugs from two drug classes can lead to prolonged virus suppression and immunological reconstitution (Richman, 1994). However, the drug-induced emergence of resistance mutations can impact on the patient's current and future therapeutic options. The risk of developing resistance is dependent on the level of residual viral replication, the ease of acquisition of a particular mutation, and how this impacts on drug susceptibility and virus fitness (Shafer et al, 2000).

There is a paucity of data relating to non-B subtypes and ARV therapy (subtype C in particular), largely because these viruses are prevalent in resource poor settings, where there is little or no access to ARV therapy (Shafer et al, 1997). Initial reports have shown that non-B subtypes are as susceptible as wild-type subtype B isolates to ARVs and that the drugs are safe and effective when applied to C infections (Pillay et al, 2002b; Frater et al, 2001; Shafer 1997). Drug susceptibility testing, using a recombinant phenotypic assay, has suggested that C viruses may be hypersensitive to Lopinavir (Gonzales et al, 2003).

However, the efficacy of ARV therapy for the treatment of subtype C, in an African context (where opportunistic infections and malnutrition are rife), is unproven. Also, drug resistance may evolve differently in people infected with subtype C (Pillay et al, 2002b). There is increasing evidence that some HIV subtypes may follow different distinct pathways during the development of resistance, although many key resistance mutations occur in both non-B and B subtypes which are exposed to antiretroviral drugs (Palmer et al, 2001; Weidle et al, 2001). Studies from Israel and other African countries have shown that some naturally-occurring polymorphisms in subtype C viruses occur at sites known to cause drug resistance in subtype B (Grossman et al, 2001). Other studies have described natural polymorphisms in the *pol* gene of C viruses that may lead to a more rapid emergence of resistance to NRTIs and NNRTIs, and a reduced *in vitro* responsiveness to PIs (Loveday et al, 2002; Perno et al, 2001). Therefore, there remains the possibility that subtype-specific resistance patterns may emerge in patients failing therapy and that this

may have an impact on the rate or level of drug resistance (Loemba 2002; Kantor et al, 2003). Key PI-associated mutations such as D30N and L90M are significantly less common in subtype C versus subtype B virus following treatment, as are some RT-associated mutations including D67N, K103N and T215Y (Grossman et al, 2001; Cane et al, 2004).

In Chapter 3, we described the naturally-occurring polymorphisms and resistance mutations in a representative subset of pre-treatment adult and paediatric patients from KwaZulu-Natal. In this chapter, we investigate the emergence and pattern of resistance *in vivo* in patients infected with subtype C. At the time, ARV drugs were not freely available in South Africa, except in a few controlled studies and in the private sector. We were fortunate enough to have access to some of these studies that included cohorts from HIV-1 patients co-infected with TB and Karposi Sarcoma (KS) and mothers and babies that participated in a pMTCT study in the Hlabisa district of Northern Natal. Although the numbers are small, these studies provide valuable information on the emergence and pattern of resistance in C-viruses and the interpretation and clinical significance of ARV resistance in the context of a rapidly expanding subtype C epidemic.

## 6.1 Resistance patterns in patients on concomitant TB/HAART and HAART alone

### 6.1.1 Introduction

The use of ARV drugs in South Africa has mostly been limited to the prevention of MTCT (Rollins et al, 2002; Morris et al, 2001) and to patients treated in private-practice. There is a paucity of published reports on the response of South African patients treated with HAART. In addition, there is limited data on the concomitant administration of HAART and TB treatment in subtype C infected patients. TB is also one of the world's major pandemics, and as is the case with HIV, the brunt of the pandemic falls on sub-Saharan Africa (Grange and Zumla, 2002). Despite efforts to control the spread of TB in this area, it continues to be a problem, most likely because of the high HIV prevalence. It has therefore been suggested that strategies to control both TB and HIV simultaneously need to be developed. However, the application of HIV-1 treatment during TB treatment is a controversial issue, as many physicians prefer to delay HAART in patients presenting with TB because of pill burden, drug/drug interactions and toxicity (Mukadi et al, 2001; Dean et al, 2002).

Little is known about evolution of subtype C in the presence of antiretroviral drugs and the impact this will have on response to therapy. Further to that, the implications for second-line and salvage therapy choices are yet to be established. In this section, we present results on the emergence of resistance in patients failing therapy from two separate treatment studies conducted in South Africa. One was conducted in the informal settlement of Khayalitsha, Cape Town, in collaboration with clinicians called the Mèdecin San Frontière (MSF), where 287 treatment naive adults were followed for a median duration of 13.9 months after starting therapy of two NRTIs and one NNRTI. We investigated the first 50 treated patients to reach study end. The other was a pilot study (called the START study) conducted in Durban, KZN on patients dually infected with TB and HIV-1, where concomitant TB/HAART was prescribed.

### 6.1.2 Selection of patients and samples

Plasma samples were collected from 26 HIV-1/TB co-infected patients at baseline and at monthly intervals after commencement of concomitant TB and HIV-1 ARV therapy. Seven patients defaulted and of the 19 remaining patients, sequences were obtained from 16 patients at baseline and at monthly intervals for the two patients failing therapy (START 4 and 11). In addition, we obtained plasma samples from three of five patients failing ARV therapy from the MSF study.

### 6.1.3 Patient treatment history

The patients on TB/HAART received single daily doses of Rifampicin (600 mg), 3TC, ddI and EFV until study end (6 months after starting treatment). The MSF patients were started on a standard regimen of AZT, 3TC and NVP, but two of the patients (MSFCK and MSFED) changed treatment due to a combination of virologic failure, adverse events and/or intolerance. Patient MSFCK stopped therapy at 9 months due to an adverse event, as well as virologic failure. Treatment was recommenced one month later with ddI, d4T and NFV. At 13 months, the NFV had to be changed (to EFV) when the patient developed extra-pulmonary TB. Patient MSFED had two short stops and restarted shortly after commencing treatment, due to AZT intolerance, and had to switch to d4T (before 3 months). All three MSF patients had a clinical history of prior TB (were on cotrimoxizole before and during treatment) and were late stage (stage 3 or 4) AIDS patients. Sequences were obtained from three time points after commencement of therapy for each of the patients.

### 6.1.4 Methods

#### 6.1.4.1 Viral load assays

Plasma was isolated from EDTA blood tubes by spinning at 1000 rpm for 10 minutes. After separation, plasma was stored in 1.0ml aliquots at $-80^0$C until RNA isolation. RNA was extracted using a guanidinium-silica method (Nuclisens Isolation Kit from BioMerieux Inc., formerly Organon Teknika) and an automated extractor (BioMerieux) with no modifications to the protocol. Refer to the 9.0ml lysis buffer protocol in the Nuclisens Isolation Kit package insert. After isolations, HIV-1 RNA levels were measured using the Nuclisens HIV-1 QT Amplification and Detection kit, also with no modifications to the protocol. As 1.0ml of plasma was used for all viral load determination, the lower limit of detection was 40-50 copies per ml; the highest was >5000 000 copies per ml. Reproducibility and accuracy of the viral load assay was assessed every six months through participation in the VQA quality assurance panel.

#### 6.1.4.2 Determination of CD4+ and CD8+ T-cell Subsets

T-cell counts and percentages were determined using a murine monoclonal antibody cocktail of anti-human CD45, CD3, CD4 and CD8 antibodies (TetraOne, Beckman Coulter). Ten microlitres of the antibody cocktail were added to 100 ul of patients' peripheral blood collected in an EDTA container. After ten minutes of incubation, the sample was placed in the TQ-Prep (Beckman Coulter) in order to lyse the red blood cells and fix the leukocytes.

Absolute counts and percentages were determined using a single platform protocol and 100 ul of Flow Count Beads (Beckman Coulter) on the Epics XL-MCL flow cytometer (Beckman Coulter). The CD4+ and CD8+ cell count protocol was designed based on the manufacturer's specifications for TetraOne (Beckman Coulter) analysis. The initial voltages were set using an unstained sample and then a sample with known CD45, CD3, CD4 and CD8 counts and percentages. Colour compensation was conducted to prevent over-estimations in cell counts and percentages by eliminating fluorescent emission overlap from analysis. The reproducibility and accuracy of the T-cell counts were also assessed every three months through participation in the UK NEQUAS quality control panel.

#### 6.1.4.3 Resistance genotyping

Resistance genotyping was performed using the Viroseq kit (Abbott Diagnostics). Briefly, HIV-1 RNA was extracted from plasma using the Sample Prep module from the Viroseq kit, according to the manufacturers protocol. The RNA was reverse transcribed with Moloney murine leukemia virus reverse transcriptase. A 1.8-kb fragment containing the protease (amino acids 1 to 99) and reverse transcriptase (amino acids 1 to 335) regions was then amplified in a 40-cycle PCR with Amplitaq Gold DNA polymerase and AmpErase dUTP/uracil-*N*-glycosidase to minimize the risk of cross-contamination. PCR products were visually quantified by agarose gel electrophoresis. Following purification, the products were sequenced with six of the seven kit primers (primer D was not used) and Big-Dye terminator reagents and run on a 3100 genetic analyzer (Applied Biosystems). Sequences were assembled, translated, and analyzed for the presence of amino acid polymorphisms. A report was generated for each sequence, with mixtures of wild-type and mutant bases being classified as mutant.

#### 6.1.4.4 Sequence analysis

Sequences were analysed as described in Chapter 2. Briefly, sequences were subtyped by phylogenetic analysis using PAUP* and an appropriate model of evolution. Sequences were analysed for positive selection using the codeml program of the PAML software package (Rambaut, 2000). An individual amino acid was considered to be positively selected if the *dn/ds* ratio was significantly greater than 1.0. The most recent common ancestors were reconstructed using the baseml option of the PAML software package and codon models selected by the LRT method. Reconstructed ancestral sequences were saved and used to construct a Maximum-likelihood tree with all sequences available for each patient and other subtype C sequences obtained from GenBank (Rambaut, 2000; Yang,

2000; Yang et al, 2000). The tree was rooted with a homologous region from the HXB2 reference strain. Sequences were screened for the presence of biologically important sites using Prosite, a database of protein families and domains.

#### 6.1.4.5 Prediction of tertiary and secondary structure

RT nucleotide sequences were translated and submitted to Swiss-Model, a site for 3D homology modelling of protein structure (http://www.expasy.org/swissmod/SWISS-MODEL.html), to determine the impact of novel, positively selected sites on the structure of the RT enzyme Models were visualised with Deep View Swiss-pdbViewer, an interface allowing the analysis of several proteins at the same time (Guex and Peitsch, 1997). Secondary structure was predicted using the PhD software program via the Predict protein server (http://www.embl-heidelberg.de/predictprotein/predictprotein.html).

Prediction of secondary structures is a step towards the prediction of the three-dimensional structure of a protein by providing a rough estimate of structural features, and with at least one known homologue available, the PhD method has an expected overall accuracy of 71.4% for proteins (Rost et al, 1994).

### 6.1.5 Results

#### 6.1.5.1 Mutations associated with resistance to RT inhibitors

Patients failing therapy displayed a variety of drug resistance mutations (Figure 17). In START 4, the V106M, Y181C and K219N mutations were detected from one month after starting treatment and were the only amino acids detected at those positions. The L74V, K101E and M184V mutations occurred as quasispecies with the wildtype from two months after starting treatment, but by the fourth month, they were the only amino acids detected at those positions. For START 11, we did not detect any resistance mutations until the second month and these mutations (K65R, V106M and G190A) were the only amino acids detected at those positions. Patient MSFCK displayed the M184V and G190A mutations at the first time point, with the addition of T215F in the second. The third time point showed the addition of the K103N, and the loss of the M184V mutations. This corresponded with a change to d4T, ddI and EFV. Patient MSFED had the D67N, M184V and G190S mutations in the first two time points. The third time point showed the accumulation of the mutations associated with multi nucleoside RT inhibitor (NRTI) resistance (F116Y and Q151M) in addition to the D67N, M184V and G190S mutations. Patient MSFSD did not have any resistance mutations in the first time point, but acquired the K103N and M184V mutations in the second and then losing the M184V again in the

third. Figure 17 shows the pattern of emerging primary drug resistance in these five individual patients in relationship to changes in CD4 T-cell counts and plasma viral load.

### 6.1.5.2 Frequency of polymorphisms before and after treatment

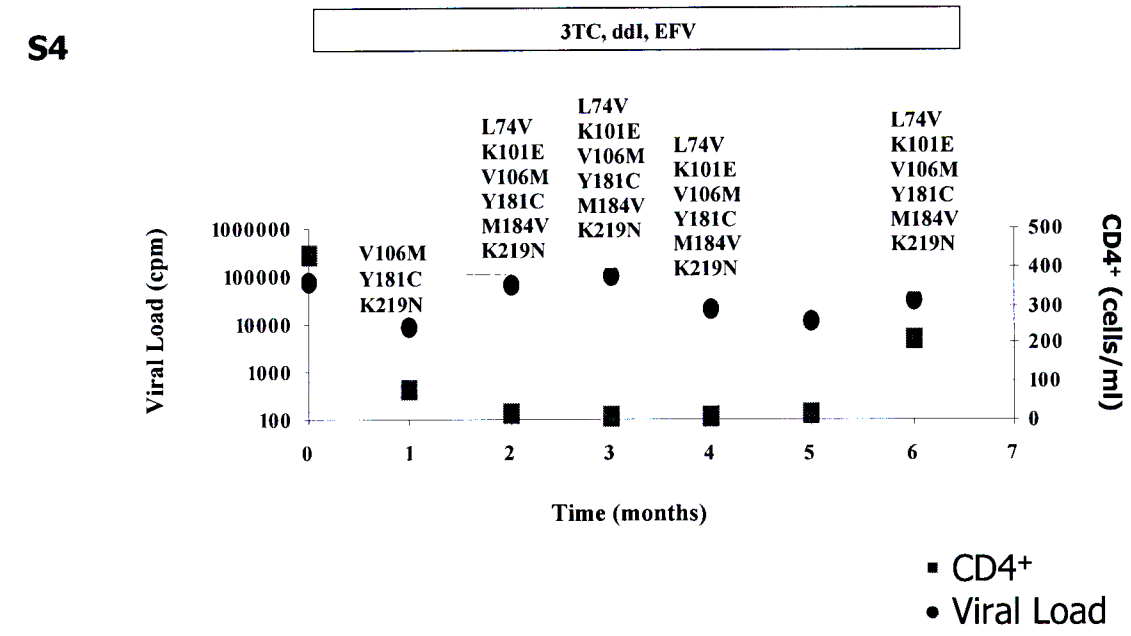
Figure 18a illustrates the frequency of polymorphisms in the RT, before and after treatment for the five patients. Primary resistance mutations were only found after treatment, with the M184V as the highest frequency. Polymorphisms not previously associated with resistance that were detected only after treatment included: K21I, K32N, K49R, V90I, I135M, E203K, K281R and K311Q. Other polymorphisms increased (V36A) or decreased (R123G/S and R277K) in frequency. Although patients were on different drug regimens, this graph highlights the development of polymorphisms in response to ARV treatment.

### 6.1.5.3 Correlation of mutations with alterations in functional properties of RT.

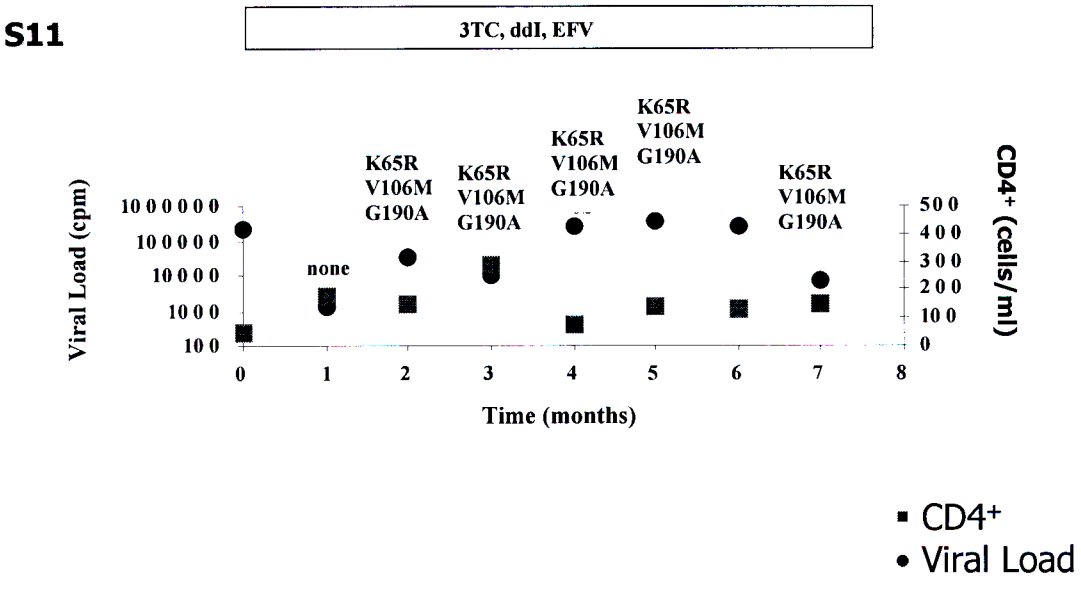
A few differences were observed at sites important for function between drug naïve and those failing treatment. At least three treatment induced genetic changes would be expected to have a significant impact on the functional properties. The first of these was the loss of the cAMP, cGMP dependant PK site at RT 102 to 105, and the gain of a myristoylation site at RT 99 to 104 and a N-glycosylation site at RT 105 to 106 due to the K103N mutation. This is consistent with results from our previous study (Gordon et al, 2003). Another difference was the loss of a MSFCK_M7 site at RT 215 to 218 caused by a T215F mutation in the second and third time points from patient MSFCK. A third difference noted involved the TKP site at position 174-181 in RT, which spans the active site of the RT enzyme. This region was highly conserved in our retrospective subtype B and D sequences (data not shown), but variable in C. Loss of the site is caused either by the loss of a K or R at RT 173 or 174, the loss of a D or E at RT 177, or the loss of a Y at RT 181. There was no consistent pattern seen in those isolates that lost this TKP site.
Lastly, a PKC site at RT 123 to125 was found in 10/16 (62.5%) of the START baselines, which is a great increase compared to our previous study of around 25% (Gordon et al, 2003). This was not found in the MSF isolates, or our retrospective B and D isolates.

### 6.1.5.4 Positively selected sites in the RT

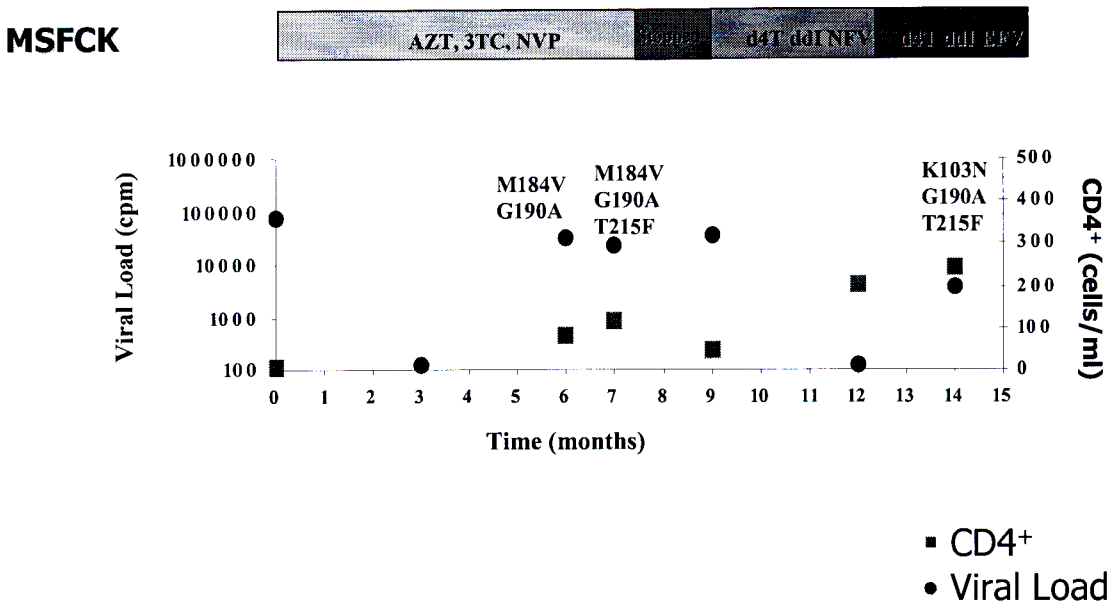
We analysed the selection pressure exerted on the treated isolates in different subgroups: START 4 and 11 samples together; START 4 and START 11 separately; all the MSF


S4
3TC, ddI, EFV
Viral Load (cpm)
1000000
100000
10000
1000
100
V106M
Y181C
K219N
L74V
K101E
V106M
Y181C
M184V
K219N
L74V
K101E
V106M
Y181C
M184V
K219N
L74V
K101E
V106M
Y181C
M184V
K219N
500
400
300
200
100
0
CD4+ (cells/ml)
0
1
2
3
4
5
6
7
Time (months)
CD4+
Viral Load

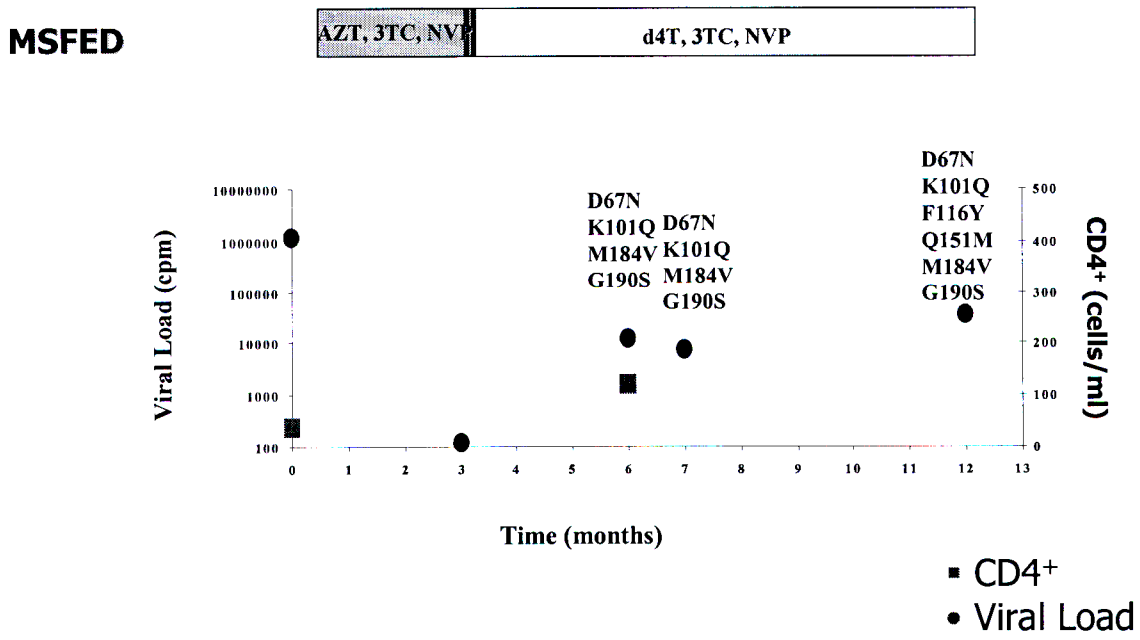
(a)


S11
3TC, ddI, EFV
Viral Load (cpm)
10 0 0 0 0 0
10 0 0 0 0
10 0 0 0
10 0 0
10 0
none
K65R
V106M
G190A
K65R
V106M
G190A
K65R
V106M
G190A
K65R
V106M
G190A
K65R
V106M
G190A
5 0 0
4 0 0
3 0 0
2 0 0
10 0
0
CD4+ (cells/ml)
0
1
2
3
4
5
6
7
8
Time (months)
CD4+
Viral Load

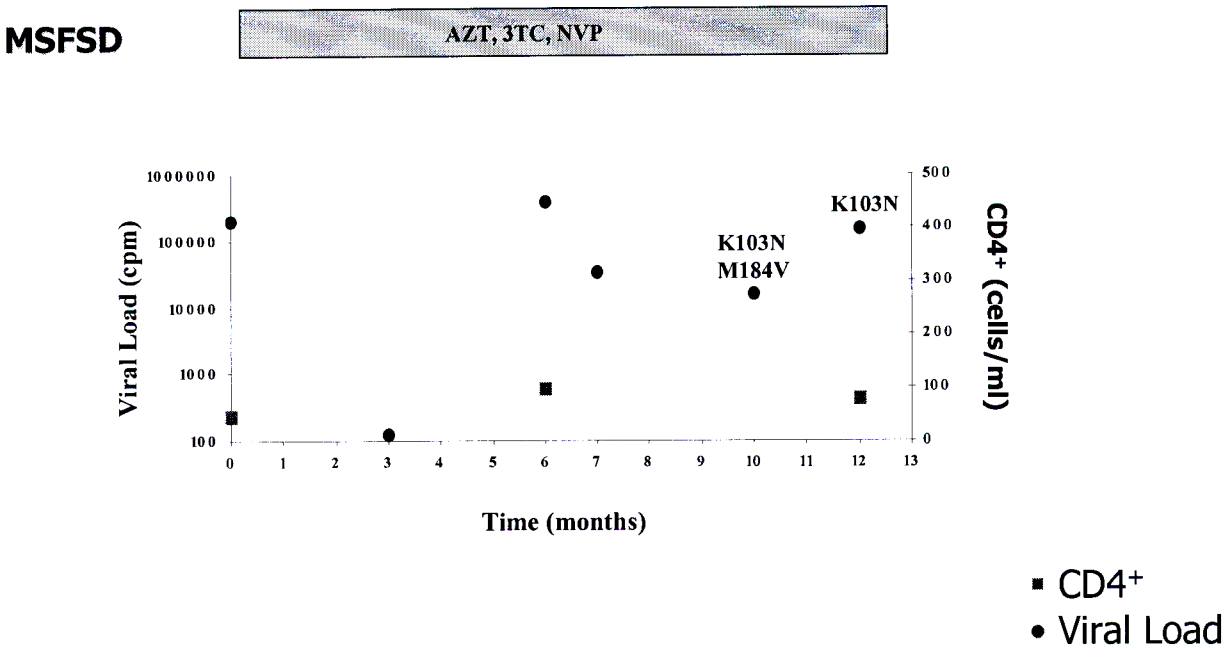
(b)


MSFCK
AZT, 3TC, NVP
d4T, ddI, NFV
Viral Load (cpm)
1000000
100000
10000
1000
100
M184V
G190A
M184V
G190A
T215F
K103N
G190A
T215F
500
400
300
200
100
0
CD4+ (cells/ml)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Time (months)
▪ CD4+
• Viral Load

(c)


MSFED
AZT, 3TC, NVP
d4T, 3TC, NVP
Viral Load (cpm)
10000000
1000000
100000
10000
1000
100
D67N
K101Q
M184V
G190S
D67N
K101Q
M184V
G190S
D67N
K101Q
F116Y
Q151M
M184V
G190S
500
400
300
200
100
0
CD4+ (cells/ml)
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Time (months)
▪ CD4+
• Viral Load

(d)




(e)

**Figure 17. Longitudinal analysis showing the changes in resistance patterns, viral load and CD4+ counts in five patients who developed drug resistance during the first 8 to 15 months of therapy. These patients were: START 4 (a), START11 (b), and MSFCK (c), MSFED (d) and MSFSD (e)**

samples together as well as separately (Table 18). As expected, codons under strong positive selection pressure were at sites known to cause primary (confers resistance on their own) or secondary (confers resistance when occurring with other mutations) drug resistance in subtype B. These were: RT 65, 74, 101, 106, 181, 184, 190, 219 and 334. Interestingly, other codons under positive selection in the START 4 and 11 subgroup included codon 123 which codes for a PKC phosphorylation site in 62.5% of the baselines and in START 11 before and after treatment. Also, codons 173 and 174 were both under positive selection in the START treated samples and form part of a TKP site with codon 181. This TKP site is found in 6/16 (37.5%) of the START baselines (including START 4, but not START 11). This site is lost in START 4 due to the Y181C mutation. The 123, 173 and 174 codons were not under positive selection when START 4 and START 11 were analysed separately.

In START 11, eight mutations caused amino acid change and all were positively selected. Three of the eight (K65R, V106M and G190A) were at sites associated with resistance in sub-B, while mutations at codons 274, 275, 278, 281 and 292 were at sites not previously associated with drug resistance. There were interesting changes observed at these positions, namely, at codon 274, there is a change from V to I in months 2 to 7, but changes at codons 275, 278 and 281 occur at months one and two, but revert to the .

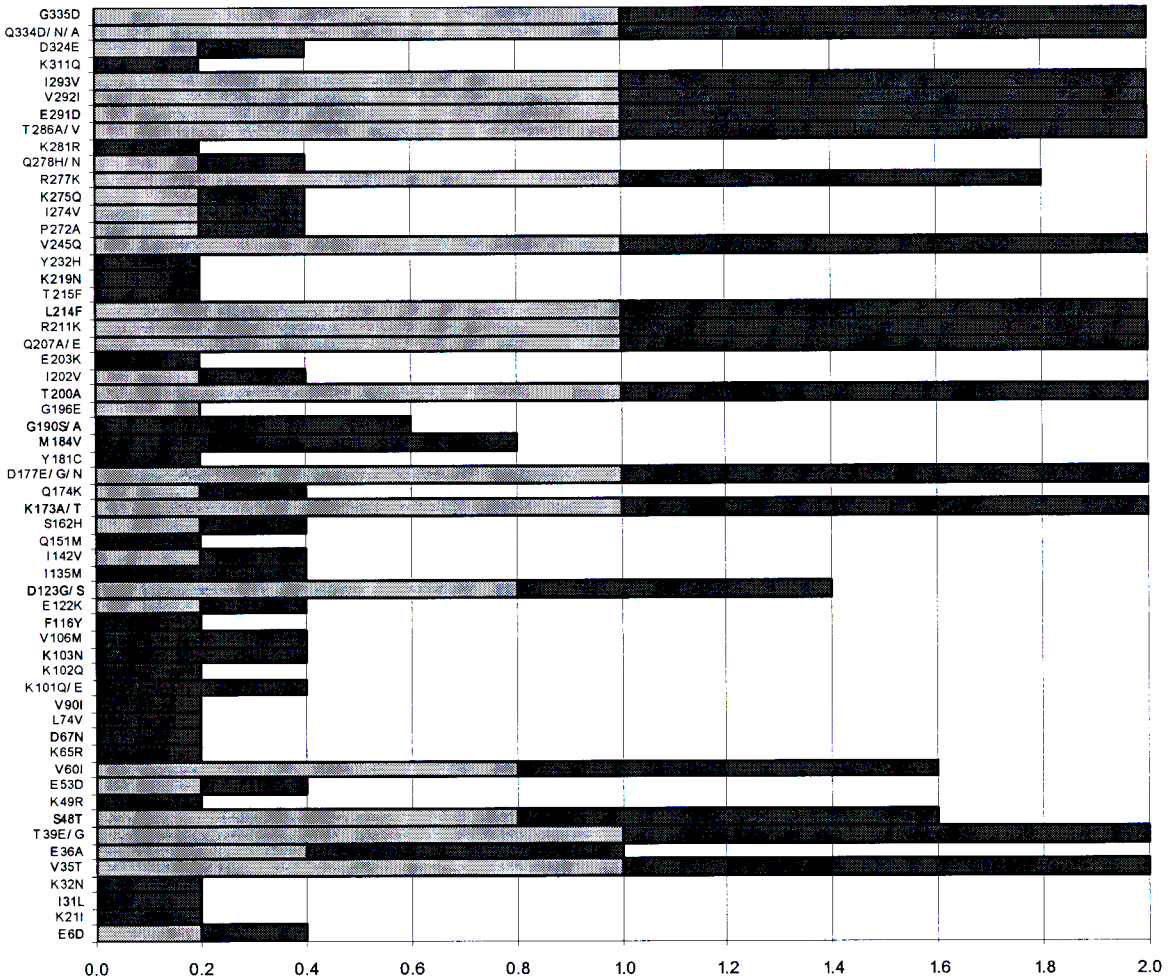



**Figure 18a. Stacked bar graph showing the frequency of RT polymorphisms before (blue) and after (pink) treatment. Values along the x-axis are in increments of 20%. For example, the V35T polymorphism occurred in 100% of the isolates before (blue) treatment, as well as 100% of the isolates after (pink) treatment, while the V60I polymorphism occurred in 80% of the isolates before, as well as after treatment. M184V was the most frequent primary resistance mutation occurring after treatment. The K21I, K32N, K49R, V90I, I135M, E203K, K281R, K311Q mutations were detected after treatment, although at very low frequencies. The frequency of the E36A mutation increased from 40% to 60% after treatment. Changes at other positions (RT123 and RT277) resulted in a decrease in frequency of mutations at these positions.**

baseline sequence from month 3 to month seven. If one takes the ancestral sequences into account, it appears that the mutations at codons 275 and 278 (present in the ancestral sequence) were lost during viral evolution, but regained during therapy (as a result of drug selection pressure) and was lost again, probably because the original sequence had more of a fitness advantage. I292V was only present at month four, which could possibly correlate with a rapid increase in viral load to 240 000 after it had come down to 9200. In START4, 11 mutations resulted in amino acid change. All of those at positions associated with drug resistance were also positively selected, as well as V21I, V202I, E203K and E207A that were positively selected sites, but not previously associated with resistance. V21I and E203K are detected from month two onwards, and could be compensatory mutations in response to the L74V, K101E and M184V drug resistance mutations. In a similar fashion to the mutations seen in START 4, V202I and E207A occur from months one to four and then revert back to the baseline sequence in month six. When the START 4 and 11 baseline samples were removed from their respective subgroups, there were no longer any codons under positive selection, suggesting that the resistance mutations had become fixed very early on.

The MSF samples had fewer codons under positive selection, with eight in the RT. These included codons 103, 151, 184, 190 and 215, also codons known to cause drug resistance to RT inhibitors. Novel codons included RT 135, RT 162 and RT 177. In patient MSFED, seven sites were positively selected, six of these at positions associated with drug resistance. Codon 123 was the only site not previously associated with drug resistance. In patient MSFCK, four of the 13 mutations that resulted in amino acid change were at sites associated with drug resistance (103, 184, 190 and 215) and were positively selected. Of the remaining nine amino acids, five (102, 113, 123, 135 and 196) were also positively selected. Finally, for patient MSFSD, three mutations resulted in amino acid change (K103N, I135M and M184V).

#### 6.1.5.5 Subtyping and Phylogenetic analysis

All START and MSF isolates were classified as subtype C. All isolates from each patient clustered together and were supported by bootstrap values above 90%. As can be seen from the phylogenetic tree (Figure 18b), the branch lengths for the START 4 and 11 time points were very short, indicating a rapid evolutionary rate. This could be interpreted as a rapid evolution to their resistant genotype, followed by very little genetic evolution subsequently. This could also be seen, but to a lesser extent, for the MSF samples.

**Table 18. Positively selected sites in the PR and RT (excluding primary resistance mutations). Samples were analysed together (START baselines and START 4 and 11) and individually.**

| | START baseline | START 4 and 11 | START 4 | START 11 | MSF | ED | CK |
|---|---|---|---|---|---|---|---|
| PR | | 12 | | | | | |
| | | 14 | 14 | 14 | | | |
| | | | 15 | | | | |
| | 19 | 19 | | | | | |
| | 20 | 20 | 20 | | | | |
| | 35 | 35 | | | | | |
| | | | | | 61 | | |
| | 63 | 63 | | | 63 | | |
| RT | | 6 | | | | | |
| | | 21 | 21 | | | | |
| | 36 | | | | | | |
| | 39 | | | | | | |
| | | 48 | | | | | |
| | 60 | | | | | | |
| | | | | | | | 102 |
| | | | | | | | 113 |
| | | 122 | | | | | |
| | 123 | 123 | | | | 123 | 123 |
| | | | | | 135 | | 135 |
| | 142 | 142 | | | | | |
| | 162 | | | | 162 | | |
| | | | | | | | |
| | 173 | 173 | | | | | |
| | 174 | 174 | | | | | |
| | | | | | 177 | | |
| | 178 | | | | | | |
| | 196 | | | | | | 196 |
| | | 202 | 202 | | | | |
| | | 203 | 203 | | | | |
| | 207 | 207 | 207 | | | | |
| | 211 | | | | | | |
| | 214 | | | | | | |
| | | 232 | 232 | | | | |
| | 245 | | | | | | |
| | | | | | | | |
| | 272 | | | | | | |
| | | 274 | | 274 | | | |
| | | 275 | | 275 | | | |
| | 277 | | | | | | |
| | 278 | 278 | | 278 | | | |
| | | 281 | | 281 | | | |
| | 286 | 286 | | | | | |
| | | 292 | | 292 | | | |
| | | 311 | 311 | | | | |
| | 334 | 334 | | 334 | | 334 | |
| | 335 | | | | | | |

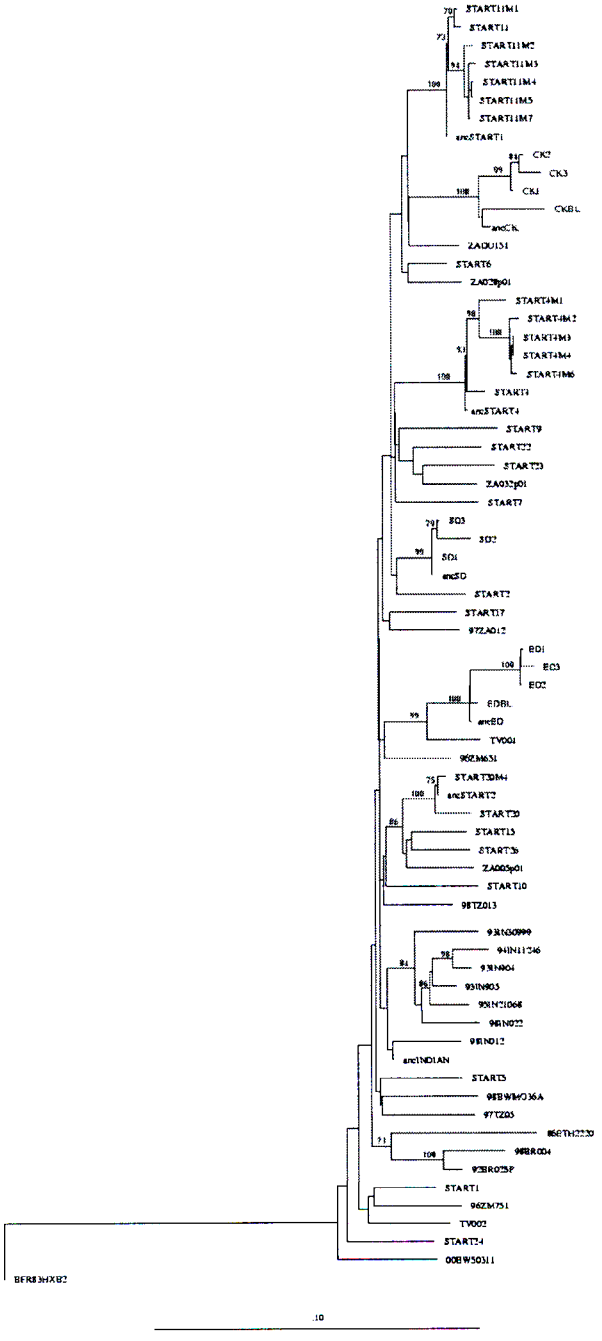



**Figure 18b. Subtype C specific Maximum-likelihood tree of all available START and MSF sequences (including the generate Most Recent Common Ancestor [anc]) and reference subtype C strains from Botswana (BW), Brazil (BR), Ethiopia (ET), India (IN), South Africa (ZA and TV), Tanzania (TZ) and Zimbabwe (ZM). The tree was rooted with HXB2.**

### 6.1.5.7 Changes in predicted tertiary and secondary structure

Primary resistance mutations are either associated with excision of the incorporated ddNTP (eg T215Y), improved discrimination between the NRTI and the natural dNTPs (eg K65R) or changes in conformation of the dNTP-binding pocket (eg K103N) (Parkin et al, 2003; Lennerstrand et al, 2001; Van Laethem et al, 2000). Accessory mutations that compensate for any loss of fitness caused by the drug-induced resistance mutations are usually involved in stabilizing the 3D structure (Courcambeck et al, 2000). 3D homology models generated using Swiss-model (Figure 19) show changes in structure over time at codons that are under strong positive selection, as well as changes at site associated with primary resistance for START 11 (Figure 19a), START 4 (Figure 19b), MSFCK (Figure 19c), MSFED (Figure 19d), MSFSD (Figure 19e). Of note was the K65R mutation, located in the "fingers" domain of RT that caused a marked change in the orientation of that side chain in START 11 from month 2 onwards (Figure 19a). Some positively selected sites in START 4 were at external residues (RT 21, 202, 203, 207 and 311) and probably served to stabilize RT (Figure 19b). Changes at codon 102, which lies in the palm domain of HIV-1 RT and is very close to the NNRTI binding sites, is noted at the same time as the acquisition of the K103N mutation in MSFMSFCK_M14 (Figure 19c). The close proximity of codons 116 and 151 are shown in Figure 19d. Mutations at these positions occurred in MSFED, causing multi NRTI resistance. The position of codon 67 near the opening of the binding groove suggests a possible interaction with the template primer complex (Figure 19d). The I to M mutation at codon 135 in MSFSD appeared to be linked to the appearance, and subsequent disappearance, of the M184V mutation (Figure 19e).

## 6.1.6 Discussion

Antiretroviral therapy can prolong the lives and improve the health of patients with HIV-1 infection. However, the emergence of mutations that cause resistance to ARV drugs can limit the efficacy of ARV treatment regimens. These mutations are often in complex patterns and the interpretation of these genotypic changes and their clinical implication presents a challenge (Shafer et al, 2000; VanDamme, 2001). To compound that, there is a paucity of data available on the efficacy of ARV drugs and the emergence of drug resistance among non-B subtypes, particularly subtype C. Recent reports suggest that subtype C will respond equally as well as subtype B to therapy, but that the impact of signature subtype C sequences on therapy needs to be further investigated (Kantor et al, 2003; Grossman et al, 2001; Loemba et al, 2002).

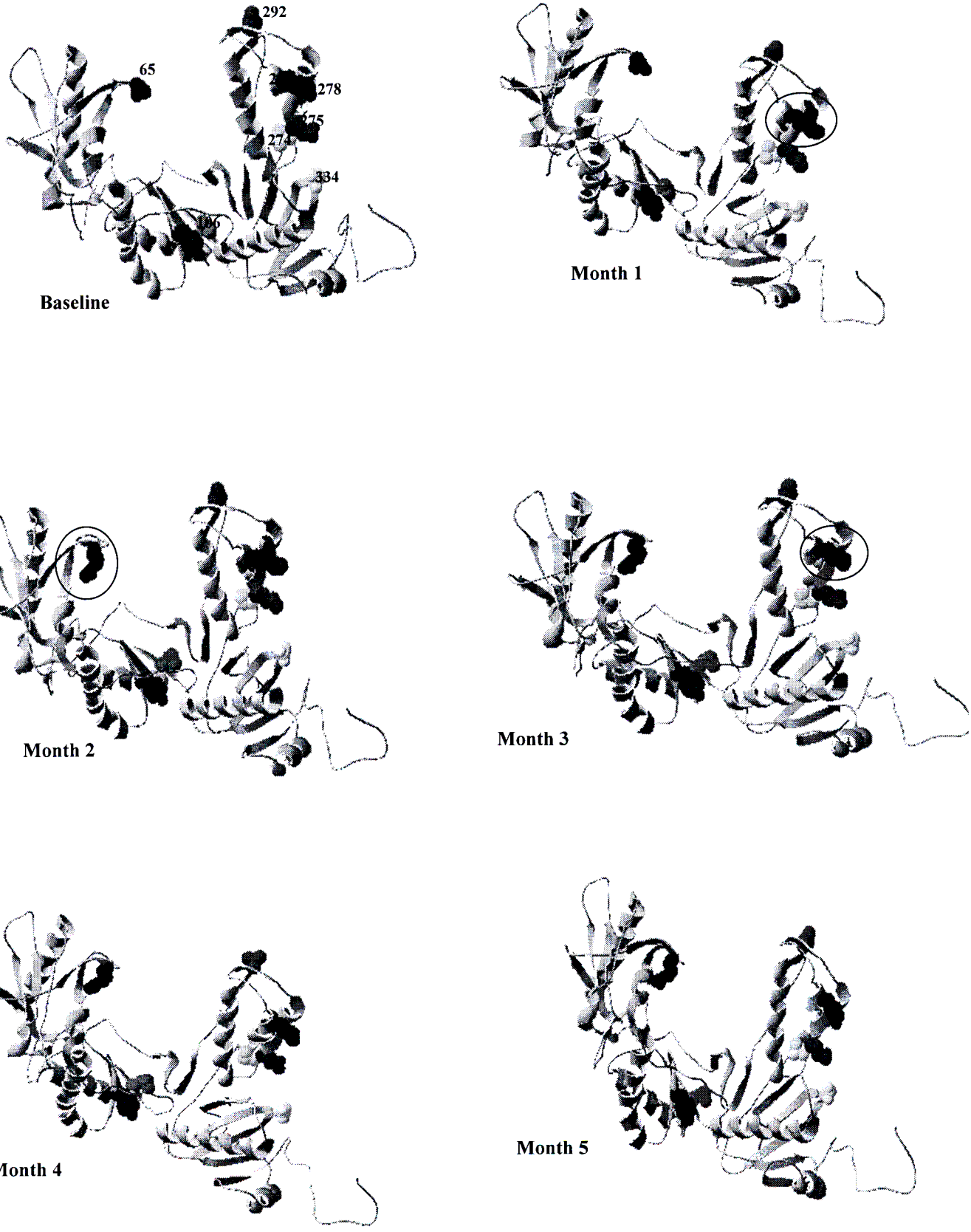



**Figure 19a. The predicted tertiary structures for START11. Sidechains are coloured as follows: RT 65 (blue); 106 (bright green); 181 (red); 190 (orange); 274 (yellow); 275 (pink); 278 (purple); 281 (burgundy); 292 (dark green); 334 (turquoise). The Lysine at RT65 occurs in the "fingers" domain of RT and usually comes into contact with the template-primer complex during normal RT activity. The change to an Arginine at month 2 causes a marked change in the orientation of the side chain. The changes that occur at the positively selected sites are more complex and are located mainly in the "thumb" domain of the RT. At months 1 and 2, it appears that the side chain at RT281 curls upward, with a corresponding downward movement of RT275. The normal orientation is resumed from month 3. The V to I mutation at RT274 from month 2 onwards does not appear to have a dramatic impact on the RT structure.**

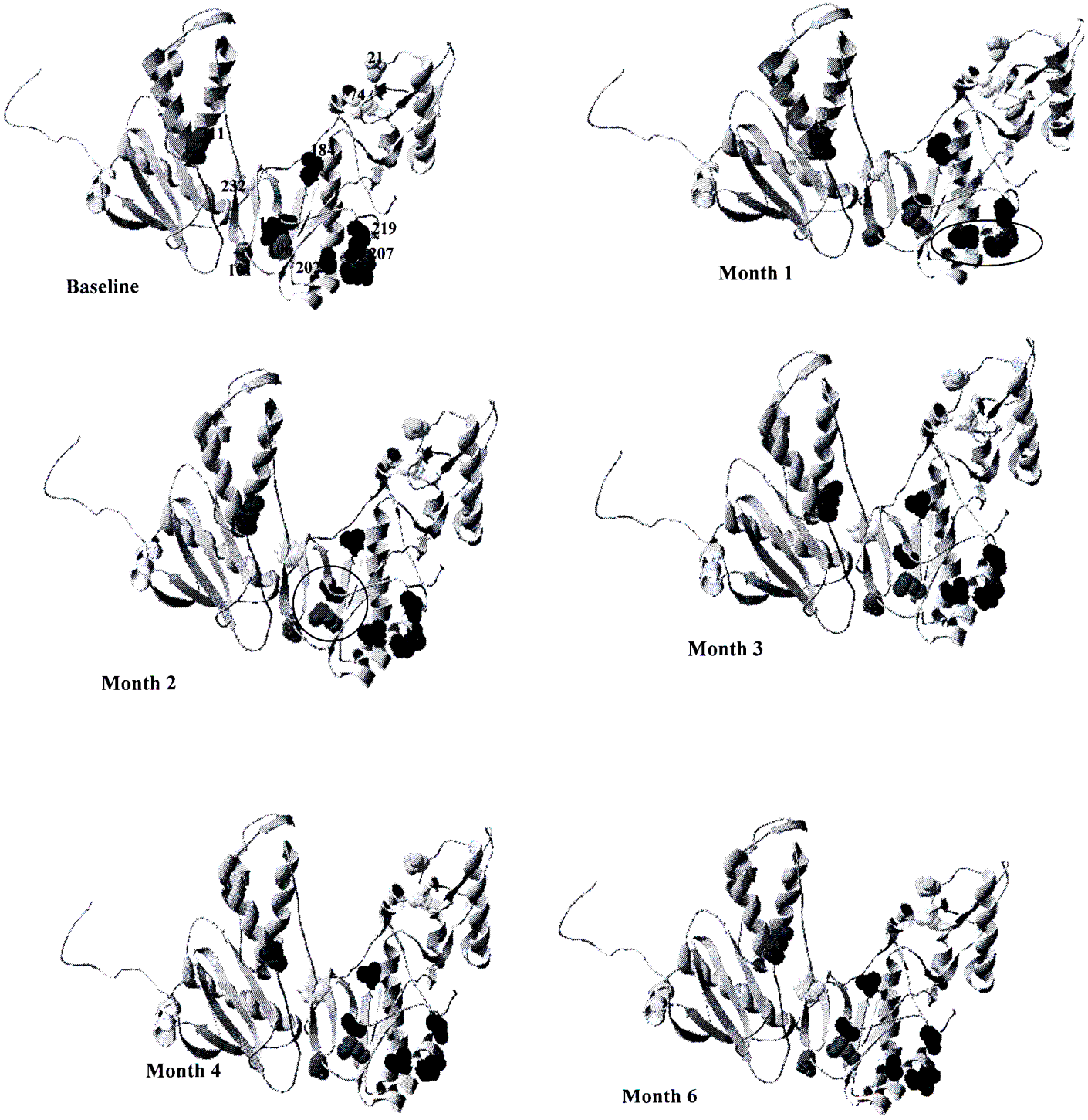
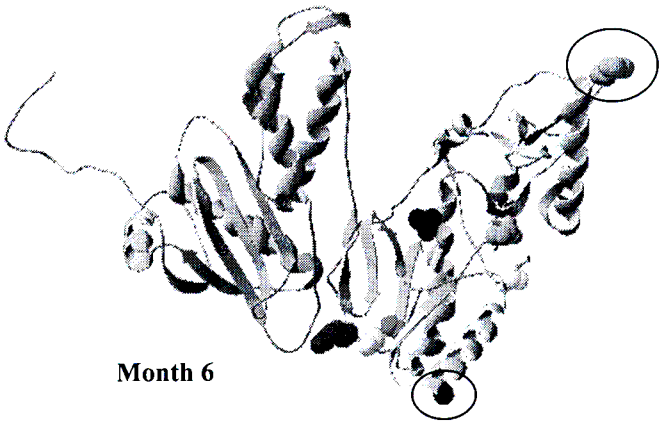
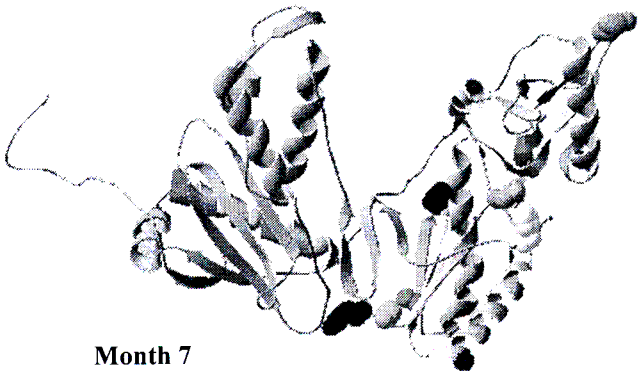



**Figure 19b.The predicted tertiary structures for START4. Sidechains are coloured as followes: RT 21 (dusky pink); 74 (sky blue); 101 (teal); 106 (bright green); 181 (red); 190 (orange); 184 (dark blue); 202 (blue); 203 (green); 207 (brown); 232 (turquoise); 219 (black); 311 (burnt orange). Mutations at RT202, 203 and 207 were positively selected. Although very closely situated to each other, RT203 is more external than the other two residues. Other positively selected residues, RT 311 and RT21 are also situated on the surface of the RT enzyme. Mutations at these locations probably serve to stabilize the enzyme.**

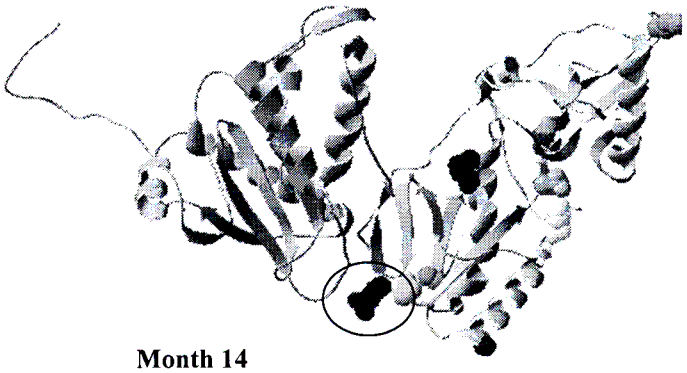

Baseline



Month 6



Month 7



Month 14

**Figure 19c.The predicted tertiary structures for MSFCK. Sidechains are coloured as followes: RT 102 (dark green); 103 (pink); 113 (light blue); 123 (bright blue); 135 (grey); 184 (dark blue); 190 (orange); 196 (bright pink); 215 (cream). Changes in conformation for condons 196, 135 and 123 all occur from MSFCK_M6 onwards. Codon 102 lies in the palm domain of HIV-1 RT and is very close to the NNRTI binding sites, and its change in conformation is noted at the same time as the acquisition of the K103N mutation in MSFCK_M14.**

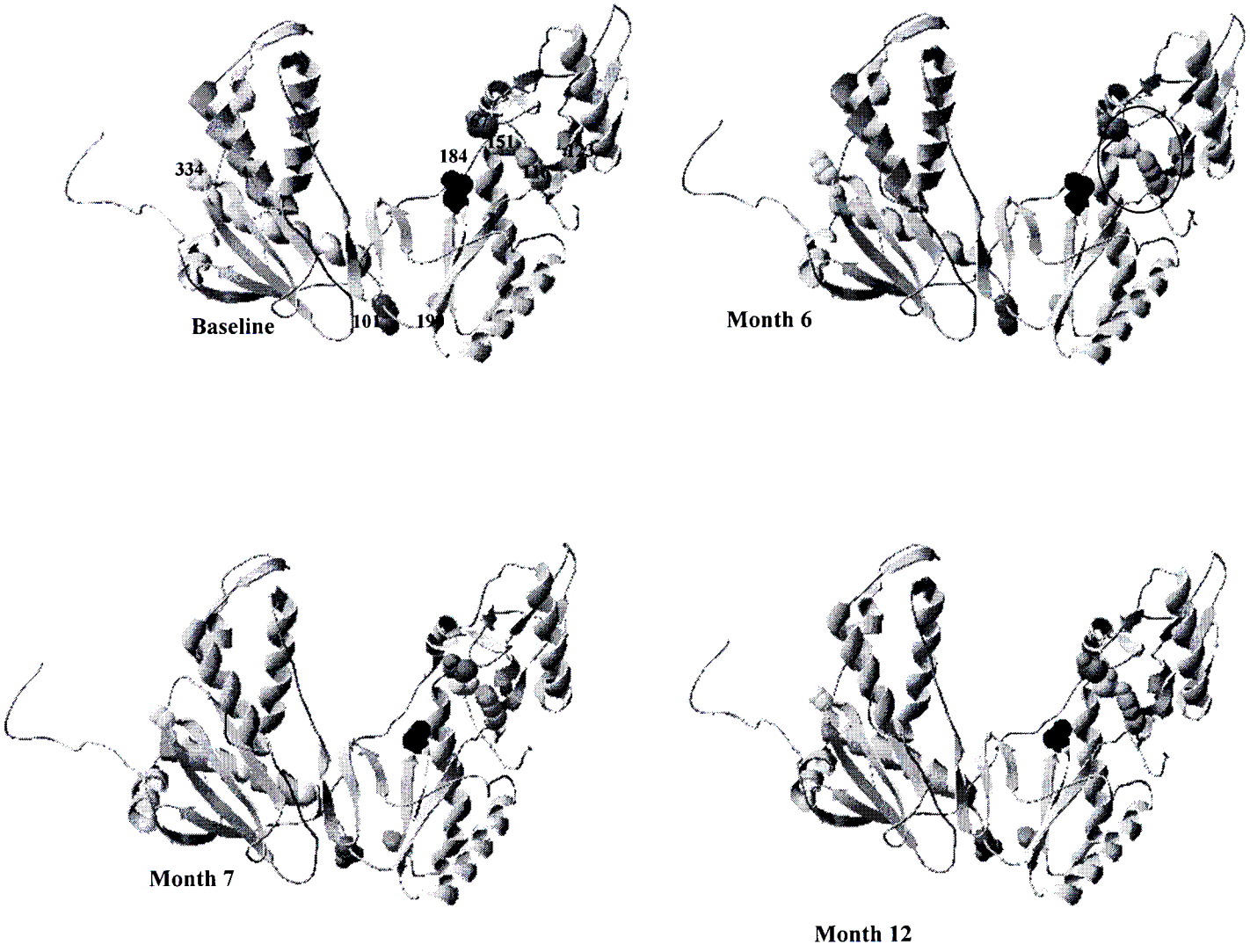
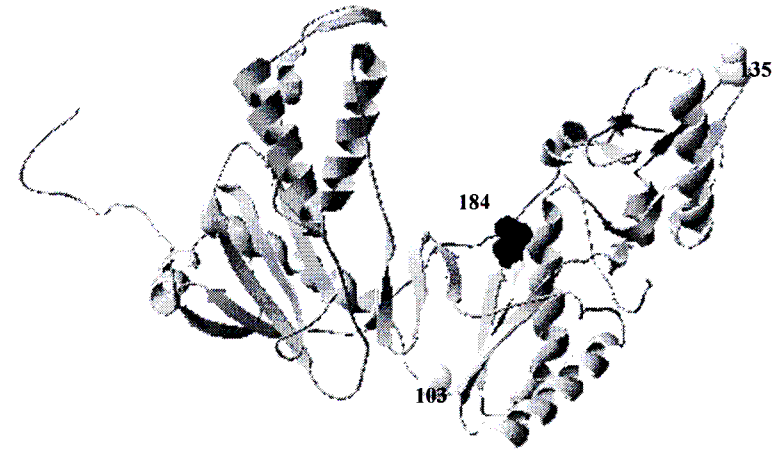
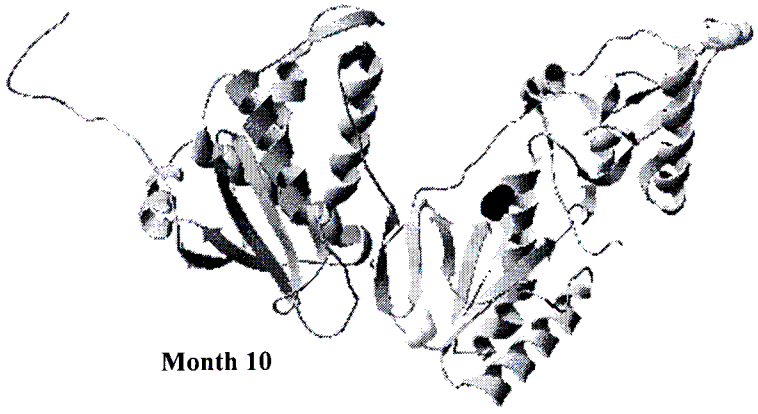


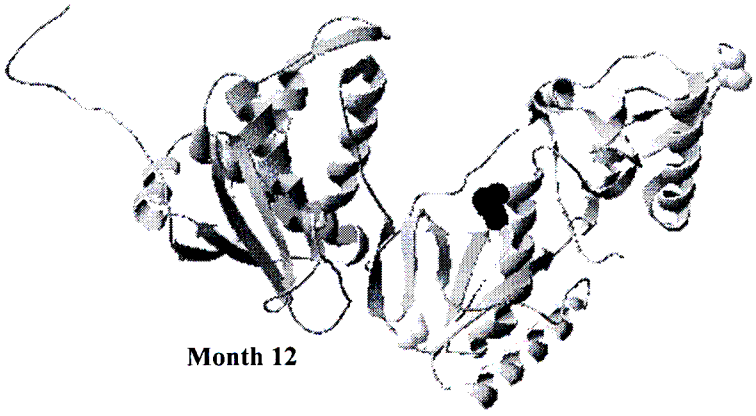

**Figure 19d. The predicted tertiary structures for MSFED. Sidechains are coloured as followes: RT 67 (dusky pink); 101 (teal); 123 (bright blue); 184 (dark blue); 190 (orange); 116 (grey-green); 151 (pink); 334 (lime). Perhaps most important is the close association between codons 116 and 151 in MSFED_M12, which causes multi NRTI resistance. Codon 67 is situated near the opening of the binding groove and is in a good position to interact with template primer complex.**




Month 6

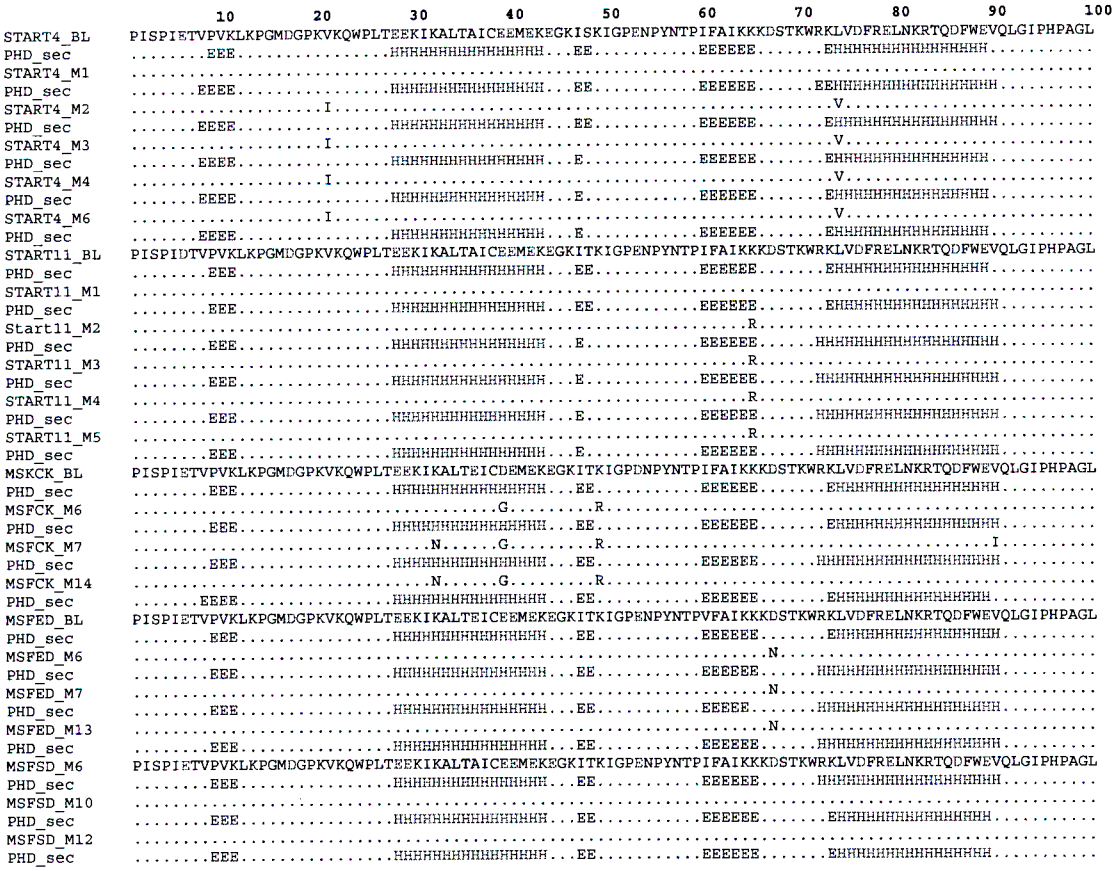


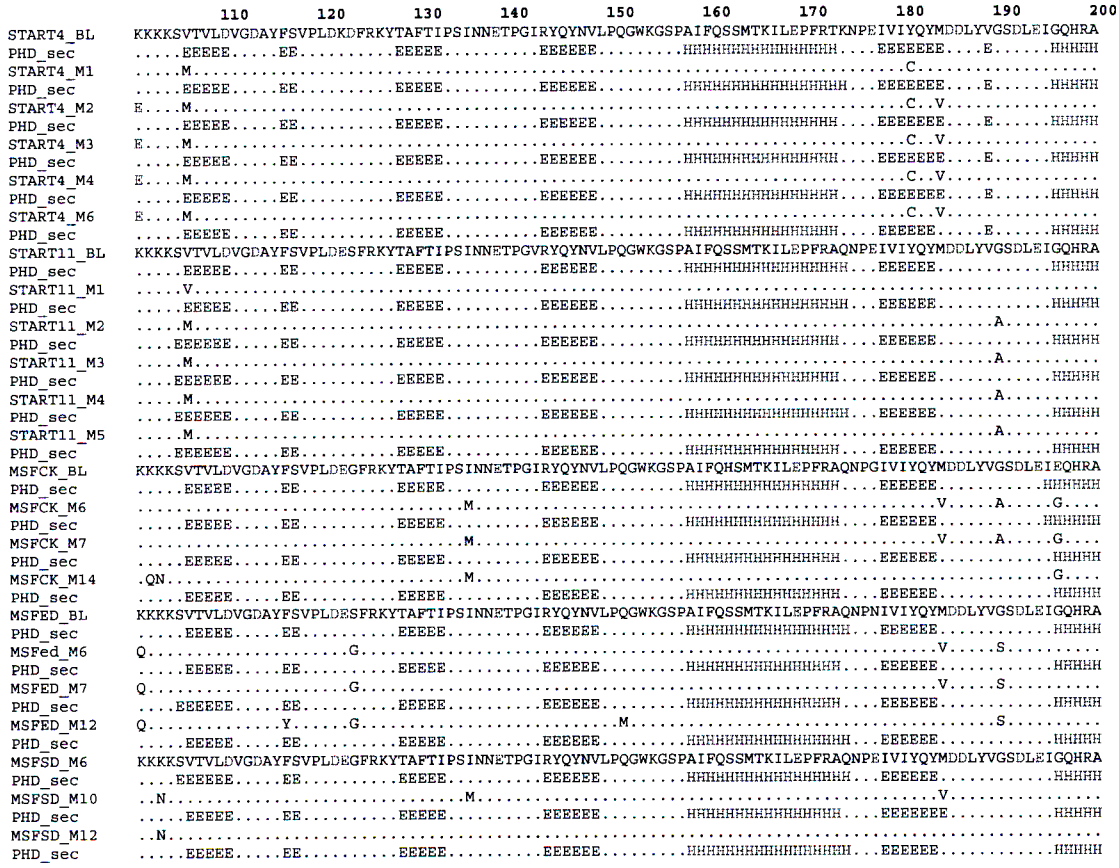
Month 10

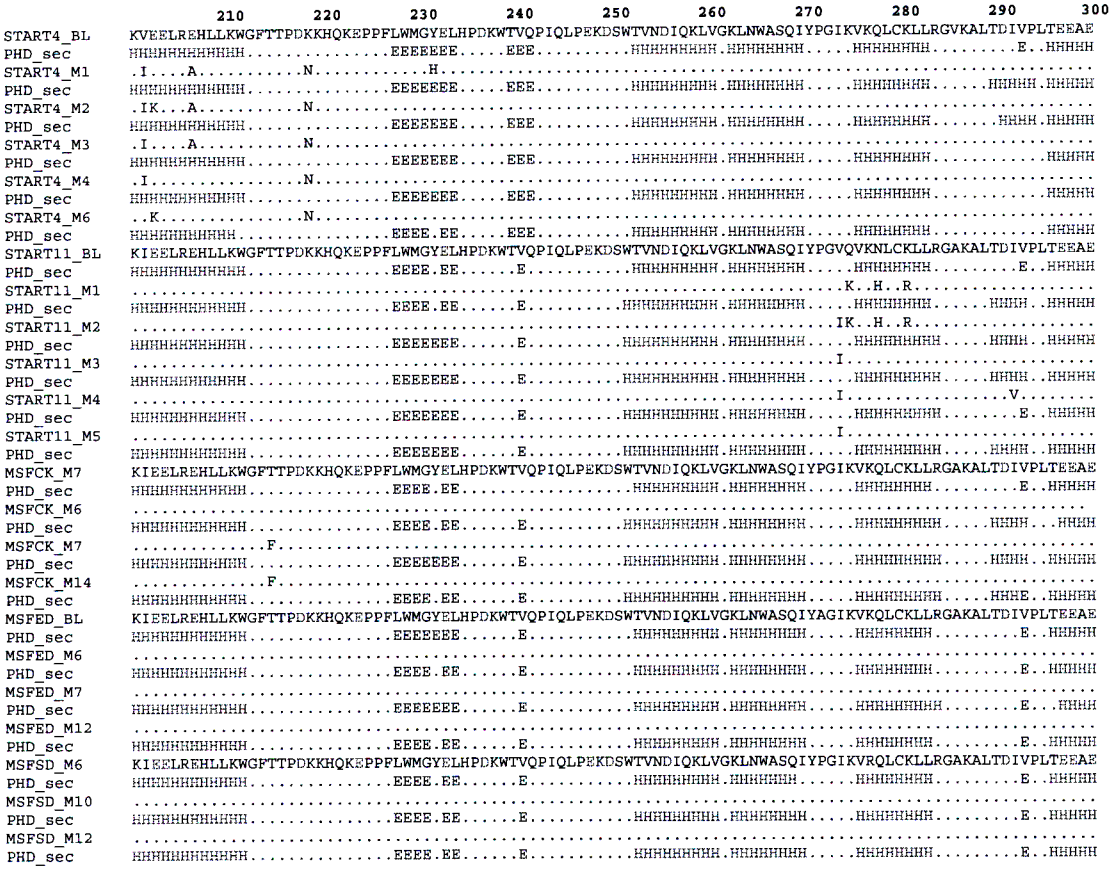


Month 12

**Figure 19e.The predicted tertiary structures for MSFSD. The only change in MSFSD occur at codons 103 (pink), 135 (grey-green) and 184 (dark blue). Codon 135 appears to change from an I to M in association with the M184V mutation and reverts back to the wild type in conjunction with the reversion to wildtype at codon 184.**


10 20 30 40 50 60 70 80 90 100
START4_BL PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
START4_M1 .....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EEHHHHHHHHHHHHHHH.....
START4_M2 .....I.....V.....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHHH.....
START4_M3 .....I.....V.....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....EHHHHHHHHHHHHHH.....
START4_M4 .....I.....V.....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....EHHHHHHHHHHHHHH.....
START4_M6 .....I.....V.....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....EHHHHHHHHHHHHHH.....
START11_BL PISPIDTVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
START11_M1 .....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
Start11_M2 .....R.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....HHHHHHHHHHHHHHHH.....
START11_M3 .....R.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....HHHHHHHHHHHHHHHH.....
START11_M4 .....R.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....HHHHHHHHHHHHHHHH.....
START11_M5 .....R.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...E.....EEEEEE.....HHHHHHHHHHHHHHHH.....
MSKCK_BL PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTEICDEMEKEGKITKIGPDNPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
MSFCK_M6 .....G.....R.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
MSFCK_M7 .....N.....G.....R.....I.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....HHHHHHHHHHHHHHHH.....
MSFCK_M14 .....N.....G.....R.....
PHD_sec .....EEEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
MSFED_BL PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTEICEEMEKEGKITKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
MSFED_M6 .....N.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....HHHHHHHHHHHHHHHH.....
MSFED_M7 .....N.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEE.....HHHHHHHHHHHHHHHHH.....
MSFED_M13 .....N.....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....HHHHHHHHHHHHHHHH.....
MSFSD_M6 PISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....HHHHHHHHHHHHHHH.....
MSFSD_M10 .....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....
MSFSD_M12 .....
PHD_sec .....EEE.....HHHHHHHHHHHHHH...EE.....EEEEEE.....EHHHHHHHHHHHHHH.....


110 120 130 140 150 160 170 180 190 200
START4_BL KKKKSVTVLDVGDAYFSVPLDKDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRTKNPEIVIYQYMDDLYVGSDLEIGQHRA
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEEE....E.....HHHHH
START4_M1 .....M.....C.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEEE....E.....HHHHH
START4_M2 E....M.....C..V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEEE....E.....HHHHH
START4_M3 E....M.....C..V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEEE....E.....HHHHH
START4_M4 E....M.....C..V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEEE....E.....HHHHH
START4_M6 E....M.....C..V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....E.....HHHHH
START11_BL KKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSINNETPGVRYQYNVLPQGWKGSPAIFQSSMTKILEPFRAQNPEIVIYQYMDDLYVGSDLEIGQHRA
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH
START11_M1 .....V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH
START11_M2 .....M.....A.....
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
START11_M3 .....M.....A.....
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
START11_M4 .....M.....A.....
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
START11_M5 .....M.....A.....
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
MSFCK_BL KKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQHSMTKILEPFRAQNPGIVIYQYMDDLYVGSDLEIEQHRA
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHHH
MSFCK_M6 .....M.....V.....A.....G....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHHH
MSFCK_M7 .....M.....V.....A.....G....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
MSFCK_M14 .QN.....G....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
MSFED_BL KKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRAQNPNIVIYQYMDDLYVGSDLEIGQHRA
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH
MSFed_M6 Q.....G.....V.....S.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
MSFED_M7 Q.....G.....V.....S.....
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEE.....HHHHH
MSFED_M12 Q.....Y.....G.....M.....S.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH
MSFSD_M6 KKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRAQNPEIVIYQYMDDLYVGSDLEIGQHRA
PHD_sec ....EEEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH
MSFSD_M10 ..N.....M.....V.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH....EEEEEEE.....HHHHH
MSFSD_M12 ..N.....
PHD_sec .....EEEEE.....EE.....EEEEE.....EEEEEE.....HHHHHHHHHHHHHHHH...EEEEEE.....HHHHH



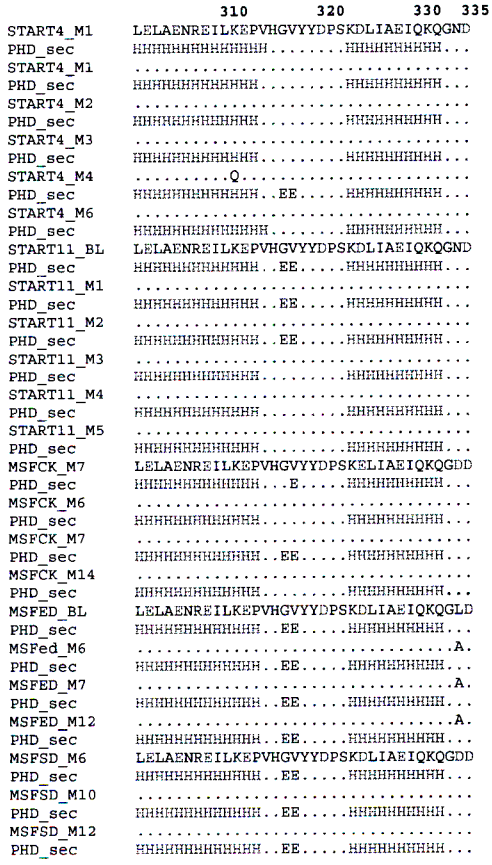


**Figure 20. Comparison of the predicted secondary structures for the 5 patients. Beta sheets are represented by an "E" and helixes are represented by an "H". The time points are given after the sample name and an underscore (eg START11_M1 represents the structure for the month one sample. K65R and D67N appear to cause lengthening of the helix in START11 and MSFED respectively. The V90I mutation in MSFCK_M7 appears to have the same effect. L74V does not appear to change the secondary structure in that region. The secondary structure at RT101-200 appeared to be conserved, with only slight variability at RT189, but this could be an artifact of the software. The length of the helix at RT 290-300 appeared to be very variable, but the pattern was not consistent, nor was it directly related to a mutation in that region. A predicted sheet was conserved at RT315-316 in MSFED and MSFSD, but was variable in the other isolates. Resistance mutations generally did not occur in sheets or helices, except the following mutations: K65R (sheet), L74V (helix), V106M (sheet), Y181C and M184V (sheet).**

This is one of the few reports from South Africa on the response of subtype C to HAART outside of the prevention of mother to child transmission trials, and these preliminary results augur well for the use of RT inhibitors in the treatment of subtype C. Overall the patients responded well to treatment with an accumulative figure of 7/69 patients failing to achieve suppression of viral replication up to 13 months after starting treatment. These data not only increase the information on the response of C viruses to treatment, but also increases the information on the concurrent use of TB and HIV treatment.

Generally, the resistance mutations seen in our isolates were consistent with those seen in subtype B, with the exception of the V106M mutation seen in the two patients on EFV. This mutation has recently been accepted as signature mutation in subtype C patients treated with EFV (Brenner et al, 2003). The V106M (GTG to ATG) mutation is said to confer high-level multi non-nucleoside RT inhibitors (NNRTI) resistance, while the V106A (GTA to GCA) mutation, which occurs more commonly in subtype B isolates, only causes NVP resistance without any cross-resistance to the other NNRTIs. The rapid development of EFV resistance in these patients is consistent with sub optimal treatment that most likely occurred as a result of non-adherence and to a lesser extent, drug/drug interactions of the TB drugs and EFV. Because EFV is one of the drugs of choice in resource poor settings, (it is inexpensive and easily administered as a single daily dose), its use in these settings needs to be closely monitored.

With the exception of START 11, which had already acquired its resistance mutations at the second month and did not alter after that time point, the patients displayed a progressive accumulation of mutations. Multi NRTI resistance was only found in one of the patients (MSFED), who developed mutations at Q151M and F116Y. The Q151M complex confers the ability of RT to discriminate between an analogue and its natural counterpart (Deval et al, 2002). It has been suggested that multi NRTI resistance via

Q151M is more common in non-B subtypes (Kantor et al, 2003). All the MSF samples had developed the M184V mutation, which confers 3TC resistance. This is common, as the M814V is usually the first mutation to appear due to its low genetic barrier. However, this mutation was lost again in 2 of the 3 patients by the third time point. It has been observed that the M184V mutation disappears as rapidly as it appears once the drug selection pressure exerted by the 3TC is removed (Van Laethem, personal communication). This correlates with the change in treatment of MSFCK, but not in MSFSD. Also, the M184V mutation did not appear in START11 at all. Therefore, an alternate explanation could be proposed, since the M184V mutation occurs in the highly conserved YMDD motive of the active site of RT and it is possible that the loss of fitness caused by this mutation is not well tolerated by subtype C viruses. In these 2 patients, it appeared that the K103N mutation emerged as M184V disappeared. However, as we only performed population sequencing and did not clone any of our isolates, we cannot rule out the possibility that the M184V mutation was still present in a minority population. Perhaps it could be said that subtype C viruses do not use the preferred pathways that are seen in the development of subtype B resistance, e.g. the use of the alternate D67N pathway for AZT resistance as seen in patient MSFED.

The drug resistance mutations in the START patients became fixed as early as one month after starting treatment, with very little evolution in subsequent time points. This correlates well with our results on selection pressure which show that these mutations were positively selected for very early on, followed by strong purifying selection at subsequent time points (i.e. a reluctance to change from the resistant genotype).
The amino acids under positive selection in RT included codons that are known to cause drug resistance to RT inhibitors in subtype B, as well as other polymorphisms that have not been associated with drug resistance (e.g. RT 21, 202, 203, 207 and 232 in START 4 and RT 274, 275, 277, 280, and 291 in START 11).

Most of the resistance mutations, such as K65R, K103N, M184V, G190A/S have been reported to cause decreased replication capacity in subtype B isolates (Bloor et al, 2003; White et al, 2002; Huang et al, 2003), while compensatory mutations are usually involved in stabilizing the 3D structure (Wrobel et al, 1998; Courcambeck et al, 2000). It has been suggested that some combinations of mutations give the virus a selective advantage in the presence of various drug combinations (Shafer R et al, 2000; Sturmer et al, 2003; Courcambeck et al, 2000), and therefore screening for positively selected codons could

potentially be useful for identifying novel drug resistance mutations that are peculiar to subtype C, or mutations that compensate for any lack of fitness caused by the resistance mutations. One such mutation is RT 202 which is involved in the stabilization of the 3D structure of RT (Wrobel et al, 1998) and could also be involved in the compensation of any loss of fitness, in a similar way that the RT 208, RT 211 and RT 214 mutations stabilize the 3D structure in response to the 215 resistance mutation (Sturmer et a, 2003). Interestingly, mutations at 3 external residues (K101, K103 and V106) are involved in drug resistance (Wrobel et al, 1998). Most of these external residues are hydrophilic, with the exception of V106, I142 and I195. V106 has a high variability although it is more buried than the other external residues. In a paper by Kantor et al, (2002), RT 207 was reported as a subtype C polymorphism that occurred more commonly in treated than in untreated subtype B isolates and RT 203 was reported to be a treatment related mutation. In the same paper, RT 174 was present at higher rates in treated persons infected with subtype C, suggesting an association with drug resistance in this subtype (Kantor et al, 2002). RT 174, which is under positive selection in the START samples, forms part of a TKP site with codon 181 (a drug resistance site) and is located in the region associated with the template-primer complex interaction (Gonzales et al, 2001). Many of the positions associated with NRTI drug resistance interact with the incoming dNTP or template primer complex during strand synthesis. K65 and Q151 in RT form part of the loop between the β2 and β3 strands in the fingers of HIV-1 and make contact with the incoming dNTP, positioning it in the active site (Courcambeck 2002; Sturmer et al, 2003). Changes in the geometry of the dNTP binding pocket are brought about by the direct interaction of the altered residue with the incoming dNTP (Drosopoulos et al, 1998). Our preliminary results on the structural interactions of novel, positively selected mutations have shown a definite relationship between structural changes at those sites and the development of resistance. These interactions are complex and requires much deeper analysis.

Unfortunately, nearly all the patients developed multi NNRTI resistance, with only one patient (MSFED) still able to use DLV. These patients would have to change to a PI inclusive regimen, which is limited to RTV or Kaletra in those patients on TB/HAART. It would be prudent to make TDF available in South Africa in the future, as it has a higher genetic barrier and would be essential for patients that have limited therapeutic options.
In conclusion, treatment of HIV/AIDS in South Africa is just beginning, and these initial results suggest that the response to ARVs will be the same as seen in Europe and North

America. However, these patients should be closely monitored for viral rebound very early on in treatment (i.e. follow up visits should be scheduled earlier than 6 months). With therapeutic drug monitoring (TDM), one could monitor the drug levels in each patient, and in this way ensure that sufficient levels are reached for optimal treatment. Also, until more is learned about the phenotypic impact of these mutations on subtype C viruses, we cannot rule out the possibility that C viruses can rapidly compensate for the loss of fitness caused by resistance mutations, causing them to become fixed very early after starting treatment.

## 6.2 Characterization of Resistance Patterns in Mother-infant Pairs 6 Weeks After Single Dose NVP

### 6.2.1. Introduction

It has been shown that administration of NVP to HIV-1-infected pregnant women at the onset of delivery, and to the infant during the first 72 hours after birth, decreases transmission of HIV-1 infection from mother-to-infant by nearly 50% (Becker-Pergola et al, 2000; Eshleman et al, 2001a). This regimen is extremely cost-effective (approximately $4 per mother-infant pair), making it the drug of choice to prevent MTCT in developing countries (Saloojee, 2001). However, there are ongoing concerns about the development of drug resistance and its potential impact on subsequent treatment. Resistance to NVP after just a single dose is thought to develop because of the long half-life of the drug in the body. Women receiving a single dose experienced an exposure of 1-2 weeks of active, but diminishing concentrations of NVP, creating selection pressures similar to those experienced during NNRTI monotherapy. Early studies have already demonstrated the rapid emergence of drug-resistant virus during monotherapy with NVP (Havlir et al, 1996; Richman et al, 1994). Mutations present as minor variants within the viral quasispecies are selected for during therapy when other drugs are not present to completely and rapidly suppress replication.

There have been reports stating that viruses carrying the NVP-resistant mutations are less fit than wild-type viruses (Collins et al, 2004). Therefore, these viruses should be quickly overgrown with wild-type virus in the absence of drug selection pressure. A recent study by Kantor et al, (2003b) found that there appeared to be rapid reversion to wild type occurring by eight and 24 weeks after receiving a single dose of NVP. However, there is growing evidence that the K103N mutation can persist as a minor population of the viral quasispecies for many years, which can very quickly re-emerge with the reintroduction of drug selection pressure (Little et al, 2004; Mellors et al, 2004; Palmer et al, 2004). The persistence of such mutations, even at undetectable levels, can influence future therapy options. Current recommendations are that NVP should be made available to women treated with NVP for the prevention of MTCT as a viable prophylactic agent during future pregnancies as well as for future therapy (Wainberg, 2000).

There is little information on whether different HIV-1 subtypes develop resistance to NVP at the same rate and as a result of the same mutations. Recent reports show that while the pattern of resistance is the same in A and D subtypes, the rate of NVP resistance was higher in subtype D isolates (Eshleman et al, 2004a). In a study by Kantor et al (2003b) in

women with subtype C viruses, an alarming 75% of samples had resistance mutations at two weeks after receiving NVP. This report has renewed the debate over the longer-term consequences of NVP resistance in women receiving pMTCT prophylaxis, and could have serious implications for policy in resource-poor settings, where single-dose NVP is administered at delivery. In this study, the pattern of resistance mutations in both the mothers and the infants, 6 weeks after receiving single dose NVP, was investigated. We also used codon-based substitution models and maximum-likelihood methods to better understand the role of pre-treatment polymorphisms on the emergence of NVP resistance, and to assess the biological consequences of genotypic resistance, as defined by changes in the phosphorylation and myristylation domains of the RT enzyme.

### 6.2.2 Methods

#### 6.2.2.1 Sample Information

NVP resistance patterns were examined in 30 mother-infant pairs (including a set of twins) infected with HIV-1 subtype C, who had participated in a pMTCT program. In this program, women received 1 dose of oral NVP intra-partum (200 mg) and the newborn received 1 dose as well (2 mg/kg) within 72 hours of birth. The baseline samples from 11 of the mothers were also analysed.

#### 6.2.2.2 RNA Extraction, RTPCR and Resistance Genotyping

HIV-1 RNA was extracted from 6 week post-partum dried blood spots using the NASBA system (Nuclisens Isolation Kit, Organon Teknika, Boxtel, NL), according to the kit protocol. HIV-1 RNA was amplified and sequenced using the Viroseq™ system (Applied Biosystems) according to the manufacturer's protocol and as described in Chapter 3.2.4. Sequences were analysed for resistance as described in Chapter 3.2.5.

#### 6.2.2.3 Nucleic Acid Sequence Analysis

The codeml program of the PAML software package (Rambaut, 2000) was used to identify prositively selected sites in 12 maternal baseline sequences and 30 mother and infant 6 week samples. All sequences were screened for the presence of biologically important sites using PROSITE, a database of protein families and domains. Mean inter- and intra-patient genetic distances were measured using the Kimura 2-$\alpha$ parameter model with a distance matrix implemented in the MEGA 2.0 package (Kumar et al, 2001).

### 6.2.2.4 Subtyping and Phylogenetic Analysis

Subtyping was performed by phylogenetic tree analysis using PAUP* (Swofford, 1999), as described in Chapter 2.1.3. The relationship between resistant and non-resistant mothers and infants were investigated in using neighbour joining and maximum likelihood methods and the GTR I+G evolutionary model that was selected using MODELTEST 3.0A (Posada and Crandall, 1998). The tree was rooted with a homologous region of the consensus subtype C reference strain obtained from the Los Alamos database (http://hiv.lanl.gov/content/hivdb/SUBTYPE_REF/align.html). Trees were viewed with Treetool and Treeview.

## 6.2.3 Results

### 6.2.3.1 Resistance Associated Mutations

No primary resistance mutations to NNRTI or protease inhibitors were detected in any of the pre-treatment (baseline) sequences. A single mother-infant pair was found to carry the NRTI mutation, G333E, which facilitates dual resistance to AZT and 3TC when present in association with M184V and standard AZT resistance mutations. Six week after the administration of single dose NVP, 11 of 30 (37%) mothers had primary mutations associated with NVP resistance. K103N was the most common mutation found alone (63.6%), or in combination with Y188C (9.0%), Y181C and Y188C (9.0%), or V106M, Y181C, Y188C and G 190A (9.0%). Of the 31 children tested (including a set of twins), 11 (35.5%) had primary resistance six weeks post-NVP. Among infants, Y181C was the most frequent mutation, occurring alone (72.7%) or in combination with K103N (18.2%), or V106M (9.0%) (Table 19).

**Table 19. Resistance mutations seen in the mothers and infants. The study numbers beginning with "m" denotes the mother and those beginning with "c" denotes the infant. "$c_2$" indicates one of the twin pairs.**

| Study No. | Mutations | Study No. | Mutations |
|---|---|---|---|
| mZA129 | V106M, Y181C | cZA129 | - |
| mZA133 | K103N | cZA133 | Y181C |
| mZA134 | K103N | cZA134 | Y181C |
| mZA135 | K103N | cZA135 | V106M, Y181C |
| mZA136 | K103N, V106M, Y181C, Y188C, G190A | cZA136 | K103N, Y181C |
| mZA137 | - | cZA137 | Y181C |
| mZA140 | K103N | cZA140 | K103N, Y181C |
| mZA145 | G333E | cZA145 | G333E |
| mZA146 | K103N, Y181C, Y188C | cZA146 | Y181C |
| mZA151 | K103N | cZA151 | - |
| mZA152 | K103N, Y188C | cZA152 | Y181C |
| mZA154 | - | $c_1$ZA154 | - |
| | | $c_2$ZA154 | Y181C |
| mZA155 | K103N | cZA155 | Y181C |
| mZA157 | K103N | cZA157 | Y181C |

### 6.2.3.2 Positively Selected Sites and Genetic Diversity

Sites that became positively selected in the mothers after single dose NVP treatment included RT207 and RT272. Some sites were only positively selected in the non-resistant samples (mothers and infants) including RT162, RT214 and RT286, while RT277 was only positively selected in the mothers and not in the infants. RT123 and RT245 were the only codons under positive selection in RT in the resistant infants (Table 20).

The nucleotide distances from the mother-child transmission pairs are shown in Table 21. The mean inter-host diversity among maternal samples collected at baseline, and at 6 weeks post NVP, was 5.1% (±0.3) and 5.1% (±0.3), respectively. These distances are comparable to those previously reported for HIV-1-infected adults in KwaZulu-Natal (KZN), South Africa (4.9% ±0.27) (Gordon et al, 2003). The mean inter-host diversity among neonatal samples collected at 6 weeks post-NVP was 5.5% (±0.3). Within host variation between maternal samples collected at baseline and 6 weeks ranged from 0.1% (±0.1) to 1.2% (±0.3) (mean, 0.35% [±0.17]).

The overall mean genetic diversity between the mothers and infants was 5.1% ± 0.003. The highest genetic diversity was seen between the resistant and non-resistant mothers (5.7% ± 0.003) and between the resistant and non-resistant infants (5.6% ± 0.004).

From Figure 21, sequences from matched mother-infant pairs were more closely related to each other than from sequences obtained from unrelated mothers and infants. There was no clustering of samples that harboured resistant mutations. This was also the case in the tree drawn with all 30 mother-infant pairs (Figure 22). A low genetic diversity (represented by short branch lengths) in some of the mother's baseline and respective 6 week sample suggests that the strains circulating in the mother had undergone very little genetic evolution between the baseline and 6 week sampling time. This was usually not the case in the respective infant sample, which generally had longer branch lengths. These observations were supported by the genetic distances estimated using MEGA (Table 21). A possible explanation could be that the transmitted variants were either from another compartment such as the vagina, or minor variants in the mothers. Unfortunately, because there were only 3 sequences in total from each matched pair, not much more can be said about the variants that were transmitted. More time points are needed to obtain better resolution in the tree.

### 6.2.3.3 Analysis of Functional Changes using Prosite

Functional sites in the RT from baseline sequences of 11 mothers, as well as 6 week samples from the mothers and infants were compared. Ten of the resistant mothers

**Table 20. Sites under positive selection in mothers and infants before and after single dose NVP. RT207 and 272 were only positively selected in the mothers after single dose NVP, while RT277 was only positively selected in the mothers and not in the infants. RT123 and 245 were the only codons under positive selection in RT in the resistant infants and RT162, 214 and RT286 were only positively selected in the non-resistant samples (mothers and infants).**

| | Mothers Before NVP (n=11) | Mothers After NVP (6wk) | | Infants After NVP (6wk) | |
|---|---|---|---|---|---|
| | | Nonresistant (n=18) | Resistant (n=12) | Nonresistant (n=19) | Resistant (n=12) |
| PR | | | 19 | | |
| | 20 | 20 | 20 | 20 | |
| | 35 | 35 | 35 | 35 | |
| | | | 60 | 60 | |
| | | | 61 | | |
| | 63 | 63 | 63 | | 63 |
| | | | 64 | | |
| RT | | | 20 | | |
| | | | 35 | | |
| | | | 40 | | |
| | | | | 102 | |
| | 123 | 123 | 123 | 123 | 123 |
| | | 162 | | 162 | |
| | | | 166 | | |
| | 173 | 173 | 173 | 173 | |
| | | | 178 | | |
| | | 207 | 207 | 207 | |
| | 211 | 211 | 211 | 211 | |
| | 214 | 214 | | 214 | |
| | 245 | 245 | 245 | 245 | 245 |
| | | | 248 | | |
| | | 272 | 272 | 272 | |
| | | | 274 | | |
| | 277 | 277 | 277 | | |
| | | | 278 | | |
| | | | 281 | | |
| | 286 | 286 | | 286 | |
| | | | 324 | | |
| | 334 | 334 | 334 | 334 | |
| | | | 335 | 335 | |

**Table 21. Pairwise genetic distance between mother-infant pairs ± the standard error (SE), using the Kimura-2 parameter model as implemented in MEGA.**

| % Diversity ± SE | | | | | | |
|---|---|---|---|---|---|---|
| Intra-host Variation | Distance between linked transmission pairs | | Inter-host Variation | | | |
| Maternal evolution (n=11) | Mom-baseline vs infant 6W (n=42) | Mom-6W vs infant 6W (n=61) | Moms pre-NVP (n=11) | Moms post-NVP (n=30) | Infants post-NVP (n=31) | Cohort Variation (n=87) |
| 0.35%(±0.17) | 0.76%(±0.23) | 0.79%(±0.31) | 5.1%(±0.3) | 5.1%(±0.3) | 5.5%(±0.3) | 5.1%(±0.3) |

n = number of sequences analyzed

replaced a cAMP, cGMP protein kinase phosphorylation (PK) site with a myristoylation and glycosylation site at positions 99-104 and 103-106, respectively (Figure 23). In turn, all resistant infants harboured strains that lacked a tyrosine kinase phosphorylation (TKP) motif (KNPEIVIY) immediately upstream from the polymerase active site, at codons 174 to 181. Viruses containing a Lysine (K) at codon 174 had a TKP site, while viruses containing the 174Q substitution lacked this TKP site. Interestingly, this same phosphorylation motif is also affected by the primary resistance mutation, Y181C, a mutation which replaces the C-terminal tyrosine with a cytosine residue. Prior to NVP, 174K and 174Q substitutions were present at approximately equal frequencies, 55.6% and 44.4%, respectively. Post-NVP, the frequency of 174Q increased to a level of 83.3% among mothers and infants who developed K103N and (or) Y181C resistance mutations. Among women and children who did not develop resistance, the 174Q allele was detected in only 31.6% and 44.4% of cases, respectively.

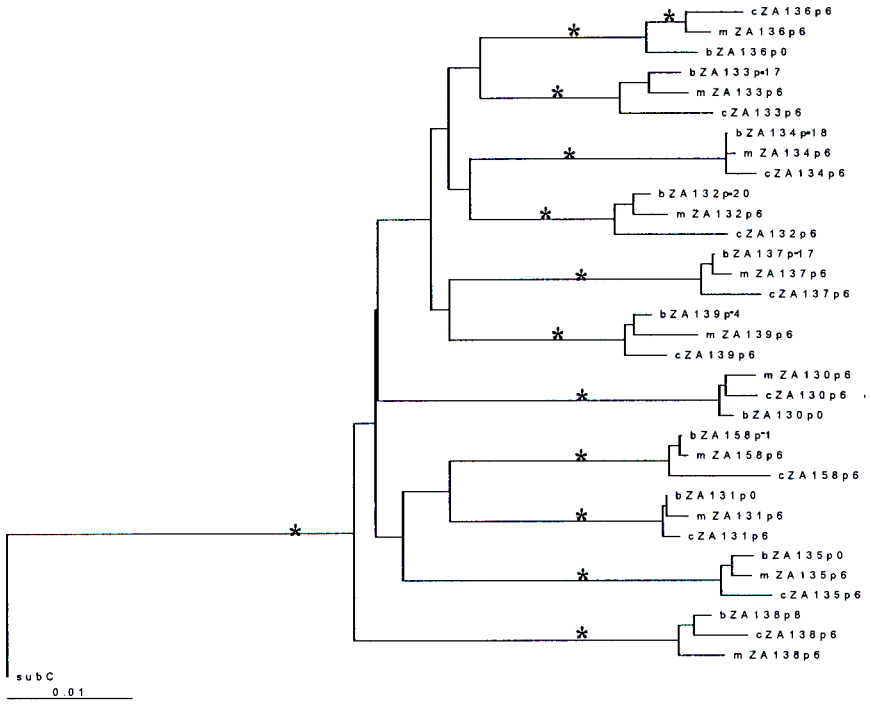



**Figure 21. Neighbour-joining tree of sequences derived from a subset of 11 matched mother-infant pairs. Maternal sequences obtained at baselines and 6 weeks post delivery are denoted by the prefixes "b" and "m" respectively. The infant sequences obtained at 6 weeks, are labelled with the prefix "c". The time from delivery (in weeks) is noted after the letter "p", with negative values representing baseline samples taken before delivery. Bootstrap re-sampling values of >90% are represented by an (*). As expected, sequences from matched mother-infant pairs were more closely related to each other, than from sequences obtained from unrelated mothers and infants.**

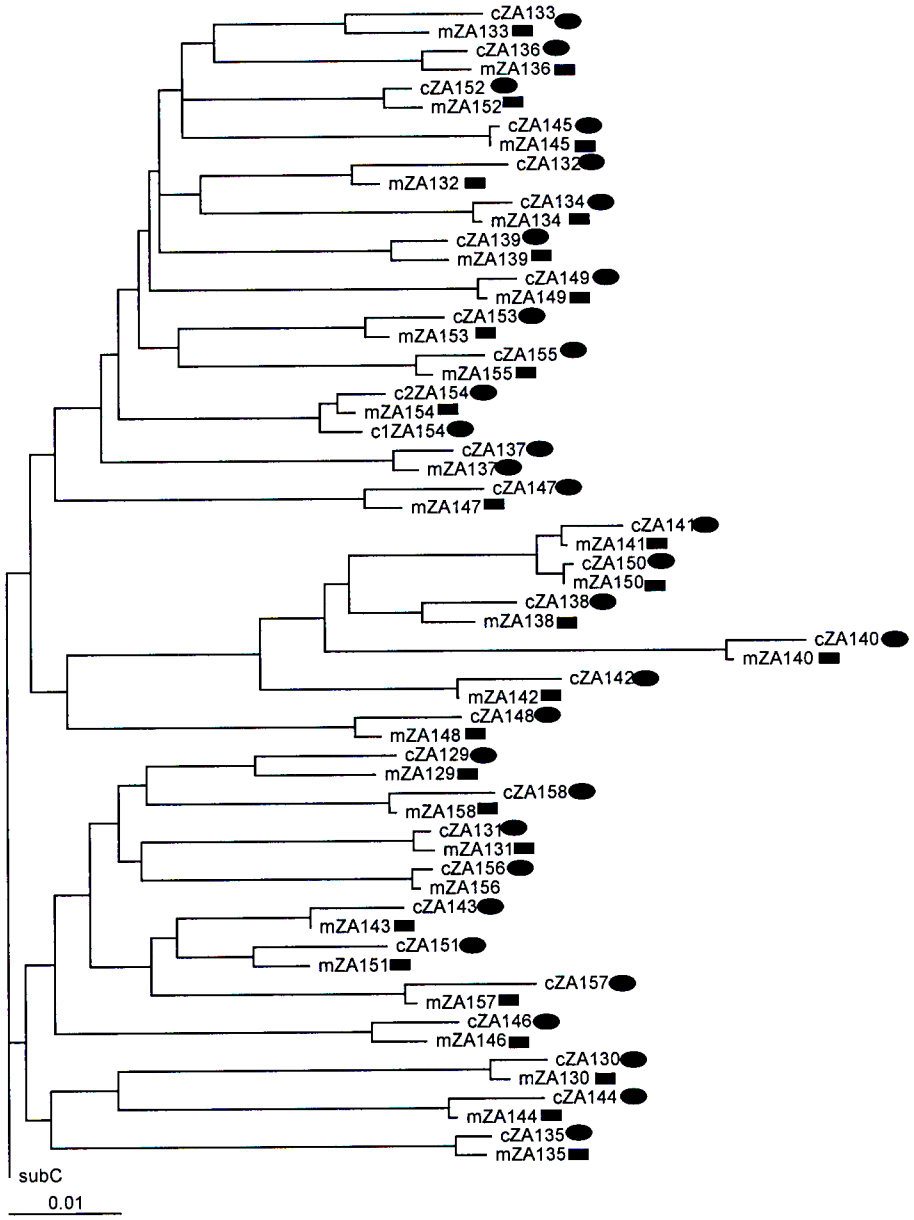



**Figure 22. Maximum-likelihood tree of mother and infant pairs, with a consensus subtype C sequence as the outgroup. Mothers' sequences are denoted by a square and infants' sequences are denoted by a circle. Resistant isolates are coloured in red and non-resistant isolates are coloured in blue. No clustering of resistant isolates was visible from the tree topology.**

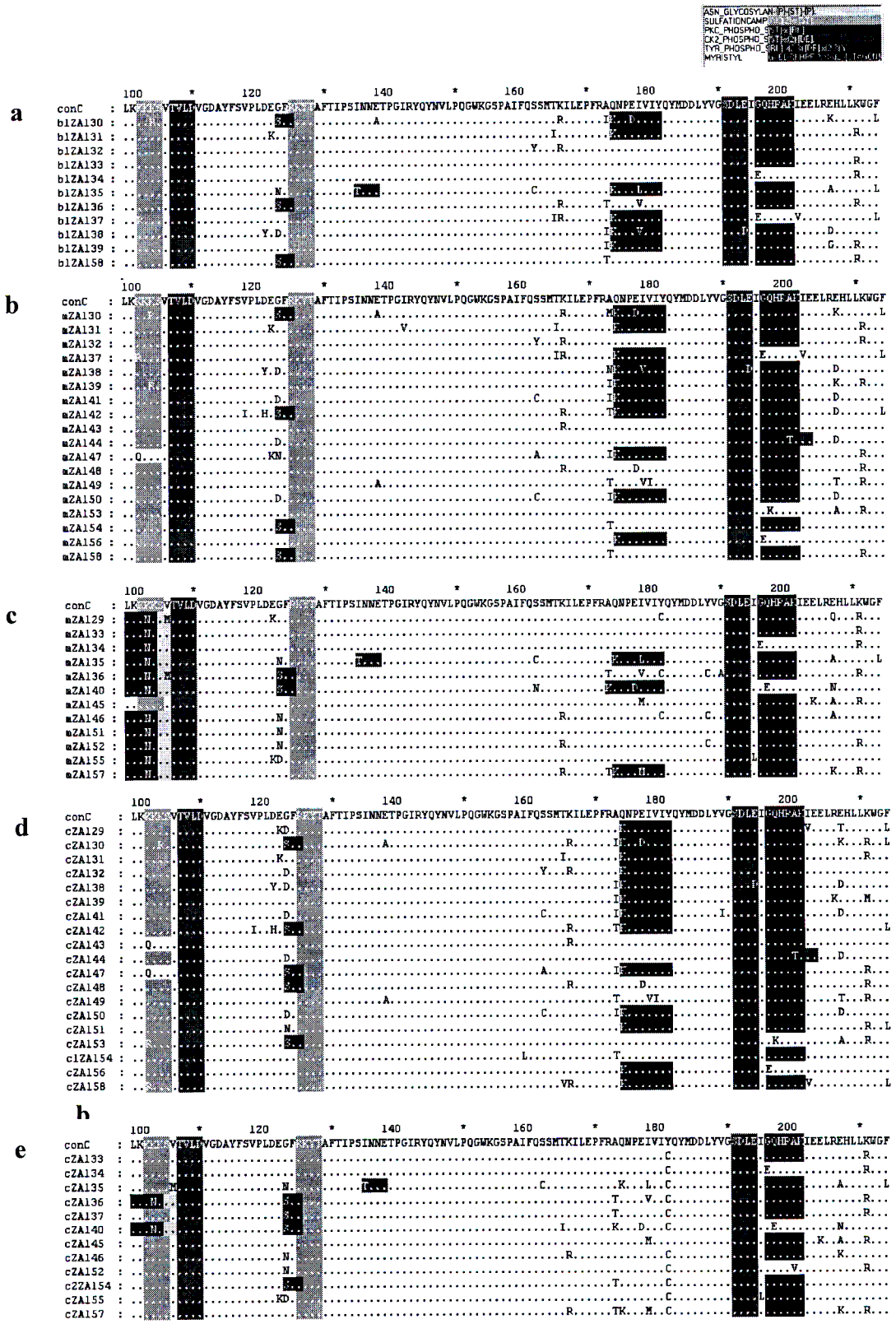



**Figure 23. Comparison of functional sites in selected mothers' baseline samples (a); non-resistant mothers (b) vs resistant mothers (c); and non-resistant infants (d) vs resistant infants (e). The K103N mutation resulted in the loss of a cAMP, cGMP dependant protein kinase phosphorylation site at codons 102 to 105 in RT. This was replaced with a myristoylation site at codons 99 to104 and a glycosylation site at codons 103 to 106. All resistant infants lacked a tyrosine kinase phosphorylation site at codons 174 to 181, near the RT active site.**

### 6.2.4 Discussion

The HIVNET006 and 012 trials account for the majority of the research done on the use of NVP in the prevention of mother to child transmission in developing countries. In these studies, NVP resistance was detected at 6-8 weeks post-partum in approximately 20% of the women who had received NVP as part of their pMTCT program (Musoke et al, 1999; Eshleman et al, 2001b; Eshleman et al, 2002). These results are mainly from subtype A and D viruses. More recent reports from Kantor et al (2003b) have shown an even higher rate of resistance mutations in subtype C infected mothers, with 75% at two weeks and 34% at 6 weeks post partum. This is consistent with the data presented here, where 40% of the mothers and infants tested harboured resistance mutations at six weeks after receiving NVP.

Prior to the introduction of NVP, extensive polymorphism was observed at several amino acid positions associated with maintaining the tertiary structure and facilitating conformational changes in the RT enzyme. Of particular interest for this study, was the identification of four highly variable residues that could accommodate from three to seven different amino acid substitutions at each position. Some of the most common naturally-occurring polymorphisms at these positions, relative to subtype B, were D123G/S, K173A/T/I, Q207E/D and V245E/K, changes which altered the charge of the subtype C RT enzyme. Analysis using codon-based ML substitutions models, indicated that all of these polymorphic sites were under positive selection pressure. There was a strong correlation between the development of resistance and the presence of specific amino acids at positions 173, 207 and 211, in close proximity to the polymerase active site (YMDD), and at several more distal codons, located at positions 123 and 286. Of considerable intrigue, was the finding that variation at position 173 was often linked to changes in the adjacent amino acid at position 174.

In the HIVNET studies, K103N was the most common mutation seen in the mothers and the Y181C mutation was most commonly seen in the infants, which is also consistent with the findings in this chapter. These mutations (K103N and Y181C) occur at positions that introduce or alter putative myristoylation and phosphorylation sites in the RT. Of note, the K103N mutation results in a change from a cAMP, cGMP dependant PK phosphorylation site to a myristoylation and N-glycosylation site. Not much is known about the role of this myristoylation site in RT, although they usually play a role in transportation of Gag and GagPol virus particles to the cell membrane, as well as the increased expression of IL2, a T-cell activator. The most common resistant mutation seen in the infants, Y181C, was linked to the absence of a TKP site in that position. The presence of a Lysine (K) or

Glutamine (Q) at position 174 influenced the presence or absence of the TKP site, respectively. While K174Q substitutions were present at approximately equal frequencies in isolates prior to NVP, the frequency of 174Q increased to a level of 83.3% among mothers and infants who developed K103N and (or) Y181C resistance mutations. It could be possible that the dual combination of 174Q and Y181C, may make it more difficult to re-instate this TKP site following the clearance of NVP. This 174Q/181C combination occurred in 81.8% of children and 36.4% of mothers with primary resistance to NVP. Phosphorylation modulates the activity of many proteins that interact with nucleic acids including DNA and RNA polymerases (Idriss et al, 1999).

In the wildtype RT enzyme, 2 beta sheets form a hydrophobic pocket where NVP binds, blocking further polymerization (Zhang et al, 1996). The NVP makes contact with the side chain of the residues of Tyrosine 181 and 188 in the p66 subunit (Spence et al 1995). The Y181C mutation induces conformational change, and consequently decreases the drug binding affinity. Surprisingly, the amino acid at 181 is internal in the RT protein, and therefore would not necessarily play a role in phosphorylation. However, reports of reduced incorporation and high efficiency of phosphorylytic removal of d4T in isolates with the Y181C mutation certainly suggests a role in the biochemical polymerization reaction of RT. Differences in myristoylation and phosphorylation sites have been noted between subtypes (De Oliveira et al, 2004). However, the impact of these changes remains to be established.

One of the key questions is whether resistance was transmitted from mother to infant, or whether the resistance developed independently in the infant. Phylogenetic analysis of 12 mother-infant pairs for which the baseline sequences were available, showed longer branch lengths and greater genetic diversity in most of the infant sequences. Also, resistance was only detected in one of the twin infants and in two mother-infant pairs, resistance was only detected in the infant and not in the mother. This, coupled with the predominance of the Y181C as opposed to the K103N mutation in the infant, supports the independent evolution of resistance in the infant. The studies by Eshleman et al (2004b) and Kantor et al (2003b) report a change in the prevalence of mutations over time, with the Y181C mutations appearing earlier (7 days to 2 weeks) and the K103N mutation later (6-8 weeks). They attribute this difference to a difference in NVP susceptibility and fitness of HIV-1 with these mutations. Therefore, there is a possibility that the burst of viral replication after primary infection in the baby preferentially selects for the Y181C mutation (Krivine et al, 1997). Of course one cannot rule out the possibility that a resistant minor maternal variant could have been transmitted to the infant via the breastmilk.

Transmission by breastfeeding is estimated at 0.5-2% per month (Lee et al, 2003). There have been recent reports of breastmilk shedding of HIV-1 virus in Zimbabwean patients, with up to 65% of samples tested harbouring resistant virus. In these cases, however, the K103N mutation was the most frequent mutation in the breastmilk and one would assume that this mutation would then be more frequent in the infants as well (Lee et al, 2003).

In the studies by Eshleman et al (2001b and 2004b), as well as Kantor et al (2003b), the K103N and Y181C mutations went to undetectable levels in the plasma over time. However, these resistance mutations may still be archived as proviral DNA. Therefore, the critical issue would be whether there are longer-term consequences of briefly selecting out these resistant viruses, such as compromising future treatment efficacy. Some theorise that if NVP was reintroduction at delivery during a second pregnancy, mutants were only likely to reappear after delivery. This coupled with a low rate of resistance, such as that seen in North America (11% to 20%) would be a negligible risk when compared to the benefit of the use of NVP. Unfortunately, it appears that the situation is not the same for subtype C infected mothers, where the rate of resistance mutations is higher. As subtype C is found predominantly in resource-poor settings where NNRTIs would be included in a first-line treatment regimen, the consequences of the development of resistance after a single dose of NVP needs to be carefully considered. These findings, together with data from other studies, reinforce the view that although NVP is effective, affordable and simple to administer, the search for safer regimens to prevent MTCT should be intensified. At the very least, use of NVP for prevention of MTCT needs to be optimised.

## 6.3 Development of ARV Drug Resistance in Plasma and Peripheral Blood Mononuclear Cells (PBMCs) in KS Patients

### 6.3.1 Introduction

Potent combination ARV therapy can reduce plasma HIV levels to below the limit of detection for up to two years or more (Gulick et al, 2000). However, it has been shown that HIV-1 persists in cells and tissues long after HAART has suppressed plasma HIV-1 RNA levels below 50 copies per ml (Wong et al, 1997; Chun et al, 1998). The problem is that ongoing viral replication is often accompanied by the acquisition of resistance mutations (Gunthard et al, 1999; Zhang et al, 1999). Several studies have shown that mutations can develop separately in the different cellular reservoirs (Blankson et al, 2002; Gunthard et al, 2001; Eron et al, 1998). Resistant viruses with mutations different to those seen in the blood have been isolated from the brain, vagina and male genital tract (Solas et al, 2003; Eron et al, 1998). Strain et al (2003) have reported that the emergence of drug resistant variants was delayed in the PBMC DNA, indicating that only a small minority of PBMC are productively infected, despite high levels of viral replication in lymphoid tissue. However, once viral variants that are able to replicate for a certain period enter the latent reservoir, they remain archived for a very long period (Verhofstede et al, 2004). This has serious implications for long-term treatment efficacy.

Kaposi's sarcoma (KS) is usually found in HIV-1 positive patients with late stage disease. KS is a cancer-like disease that usually shows up as spots on the skin (lesions) that look red or purple on white skin, and bluish, brownish or black on dark skin. It is also seen in the linings of the mouth, nose, or eye, but can also spread to the lungs, liver, stomach and intestines, and lymph nodes. Between 1988 and 1996, the incidence of KS in South Africa rose at least threefold and continues to increase as the HIV epidemic grows (Sitas et al, 2001). In many people, HAART can stop the growth or even clear up skin lesions. In addition to HAART, KS treatments include freezing with liquid nitrogen, radiation, surgery and chemotherapy. In this study, the resistance patterns in HIV-1 positive patients presenting with KS and treated with HAART and chemotherapy or HAART alone are investigated, with particular attention to the difference in viral populations present in the plasma and the PBMCs.

### 6.3.2 Methods

#### 6.3.2.1 Sample information

Venous samples were obtained at enrolment (baseline), day 1, 3, 7, 14, 28, month 3 and month 6, from HIV-1 positive patients with Karposi Sarcoma attending the dermatology

clinic at King Edward VIII Hospital, Durban, South Africa. In total, 20 patients were enrolled in an ARV program to treat KS patients and were given AZT, 3TC and EFV, either in combination with chemotherapy or HAART alone. Buffy coat and cell free plasma were separated by centrifugation. PBMCs were isolated from the buffy coat by Ficoll-Histopaque separation according to the manufacturers protocol. PBMCs were counted and stored as dry pellets at approximately $1X10^6$ cells per aliquot. Plasma and PBMCs were stored at $-80^{o}C$ until used. Resistance genotyping was performed prospectively on samples showing viral rebound. Earlier time-points from those patients were sequenced retrospectively.

#### 6.3.2.2 RNA Extraction, RTPCR and Resistance Genotyping

Samples were tested for viral load and CD4+ counts as described in Chapter 3.2.2. HIV-1 RNA was extracted, amplified by RTPCR and sequenced using the Viroseq™ system (Applied Biosystems) according to the manufacturer's protocol and as described in Chapter 3.2.4. Sequences were analysed for resistance as described in Chapter 3.2.8.

#### 6.3.2.3 Subtyping and Phylogenetic Analysis

Subtyping was performed by phylogenetic tree analysis using PAUP* Swofford, 1999), as described in Chapter 2.1.3. Briefly, sequences were compared to subtype reference strains in the Los Alamos subtype database (http://hiv-web.lanl.gov/content/hiv-db/SUBTYPE_REF/align.html). Following degapping with the degapped option in PAUP*, phylogenetic trees were generated on a Linux computer with the F84 model of substitution and the neighbor-joining method (version 4.0b2a) of PAUP* (Swofford, 1999). Trees were rooted with a homologous region of HIV-1 group O (OCM_MP5180). The subtype C and B sequences were examined separately in trees rooted with a subtype C and B reference strain, respectively. Trees were viewed with Treetool and Treeview.

### 6.3.3 Results

#### 6.3.3.1 Comparison of Resistance Mutations in Plasma and PBMCs

Four patients showed signs of increase in plasma viral load and were genotyped, two were from the HAART/Chemotherapy arm (KAR05 and KAR11), and two were on HAART alone (KAR01 and KAR09). The plasma and PBMC viral load for the four patients were compared and these results are shown in Figure 24. KAR01 and KAR05 were tested for resistance at month 6. KAR01 developed the V108I, Y181C and M184V mutations and KAR05 developed the K103N and Y188C mutations. KAR09 and KAR11 were tested at

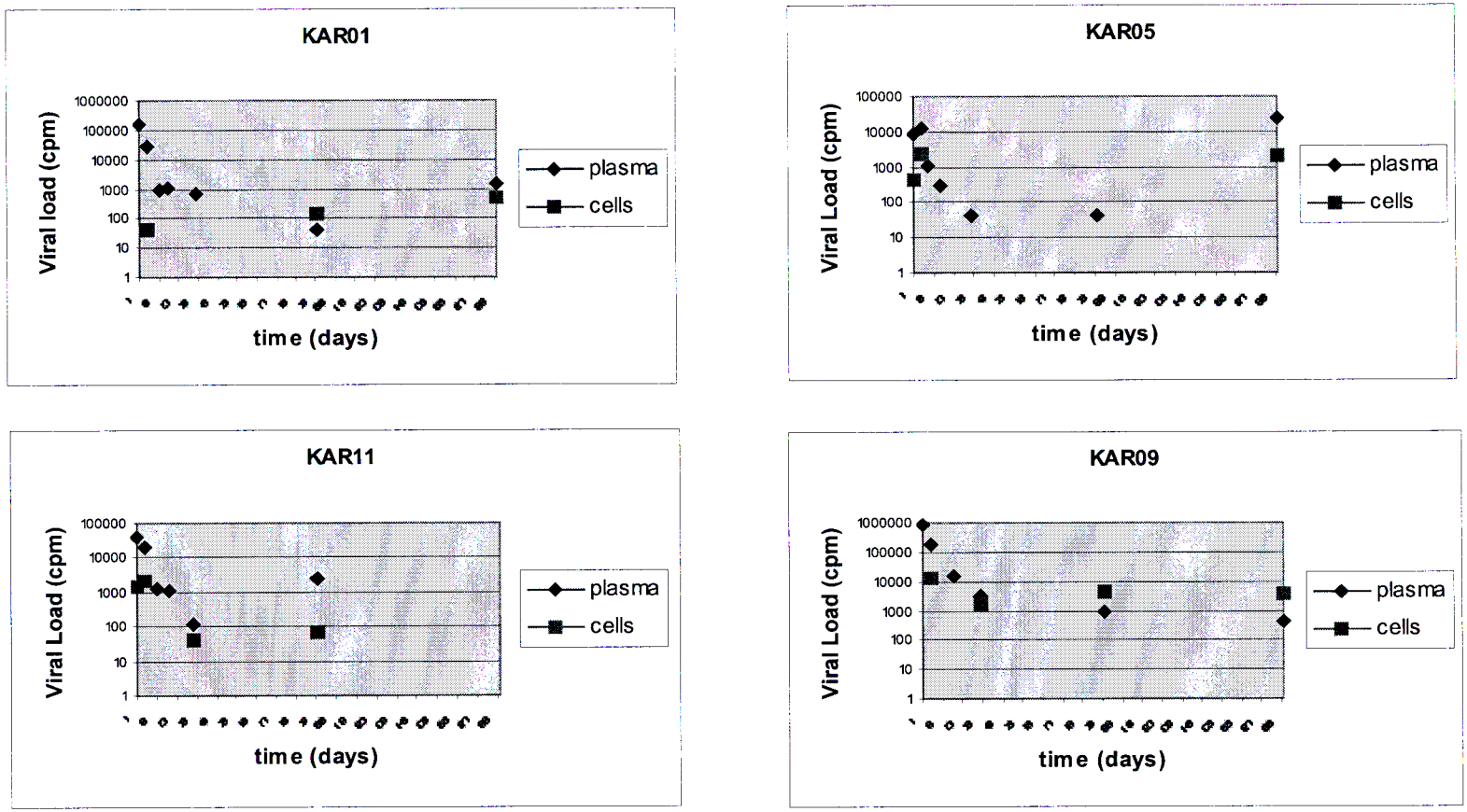



**Figure 24. KAR01's plasma viral load was still at detectable levels at 30 days after starting treatment, and only went to below detection (<40 cpm) at 3 months. In contrast, the PBMC viral load gradually increased to just below 1000cpm at 6 months. The plasma viral load for KAR05 decreased more rapidly and was already <40 cpm at 30 days. Unfortunately, the PBMC viral load quickly increased after starting treatment and was still above 1000cpm at 6 months. For KAR11, the plasma viral load decreased after starting treatment, but never went to below detectable levels up to three months after starting treatment. The PBMC viral load displayed an initial increase, followed by a decrease to <40cpm at 30 days. By three months, the PBMC viral load had increased marginally to 62 cpm. In KAR09, neither the plasma nor PBMC viral load went to below 40 cpm by three months.**

month 3, with KAR09 showing the V106I and Y181C mutations. No drug resistance mutations were detected in KAR11. Sequences from the plasma and PBMC samples obtained at 6 months were compared for both KAR01 and KAR05. No resistance mutations were found in the PBMC sample for KAR01. For KAR05, the PBMC sequence developed the V106M mutation in addition to the mutations found in the plasma.

#### 6.3.3.2 Evolution of Resistance

Analysis of the baseline plasma samples from these patients showed that KAR09 had already harboured the resistance mutations before starting the treatment. The other patients did not display any resistance mutations at baseline.

Although no resistance mutations were detected in the PBMCs of KAR01, there were differences identified between plasma and PBMC sequences at codons 32, 53, and 169 in the RT (Figure 25). The V32I mutation was seen in the month 6 plasma sample and a V32K mutation in the PBMC sample from the same time point. An E53D mutation was only seen in the month 6 plasma sample and a D169E mutation was only found in the PBMC sample. The E219D and K300E mutations were in both the plasma and PBMC month-6 samples.

For KAR05, an additional plasma sample was sequenced from Day28. No resistance mutations were detected at this time point. Polymorphisms at RT 36, 72, 128, 158 and 277 were only fount in the Day28 sample. Differences at sites not related to resistance were also seen between the plasma and PBMC samples, namely at codons 207, 281 and 292. The Q207V and K281R mutations were seen only in the month 6 plasma sample and the I292L mutation was only seen in the PBMC sample. Overall, there did not appear to be a pattern of an accumulation of mutations in either of the resistant isolates.

We also examined additional plasma and PBMCs sequences at baseline and Day3 for KAR11, the patient that did not appear to develop resistance. In this patient, 16 polymorphisms were seen in the RT region, all of which were in the month 3 plasma sample. The amino acid sequence was very conserved in the other time points.

Finally, additional sequences were obtained at Day3 and Day28 for KAR09. Polymorphisms were seen at 5 sites in the RT region, 58, 215, 245, 326 and 334. The T215A mutation was seen in the Day3 plasma sample, but was not seen again at later time points. Also, the E334D mutation was seen in the Day3 PBMC sample and was not seen at later time points. V245E was seen in the Day28 plasma sample, but was also not seen at month 3. T58P and I326L were only seen in the month 3 sample. Again, there did not appear to be an accumulation of mutations over time.

**KAR01**

```
KAR01D4pla 1   PISPIETVPVKLKPGMDGPKVKQWPLTEEKIVALTEICKEMEKEGKITKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
KAR01M6pla 1   ............................I.....................D.....................................................
KAR01M6pel 1   ............................K.....................................................................

KAR01D4pla 101 KKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILDPFRAKNPDIVIYQYMDDLYVGSDLEIGQHRA
KAR01M6pla 101 .......I.....................................................................C..V.................
KAR01M6pel 101 ..........................................................................E..........................

KAR01D4pla 201 KIEELREHLLRWGLTTPEKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYSGIKVRQLCKLLRGAKALTDIVTLTEEAK
KAR01M6pla 201 ................D.....................................................................................E
KAR01M6pel 201 ................D.....................................................................................E

KAR01D4pla 301 LELAETREIL
KAR01M6pla 301 ..........
KAR01M6pel 301 ..........
```

**KAR05**

```
KAR05D0pla   1   PISPIETVPVQLKPGMDGPKVKQWPLTEEKIKALTEICEEMEKEGKITKIGPENPYNTPIFAIKKKDGTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
KAR05D28pla  1   ................................A.....................................G.....................
KAR05M6pla   1   ..........................................................................................
KAR05M6pel   1   ..........................................................................................

KAR05D0pla   101 KKKKSVTVLDVGDAYFSVPLYEGFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRAQNPDIVIYQYMDDLYVGSDLEIGQHRA
KAR05D28pla  101 ........................A.....................T.....................................
KAR05M6pla   101 ..N.....................................................................................C.....
KAR05M6pel   101 ..N..M..................................................................................C.....

KAR05D0pla   201 KIEELRQHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLIRGVKALTDIVPLTEEAE
KAR05D28pla  201 ..........................................................................K...............
KAR05M6pla   201 ......V...................................................................R...............
KAR05M6pel   201 ..........................................................................L.........

KAR05D0pla   301 LELAENREILKEPVHGVYYDPSKDLIAEIQKQGDD
KAR05D28pla  301 .............................
KAR05M6pla   301 .............................
KAR05M6pel   301 ..................
```

**KAR11**

```
KAR11D0pla   1   PISPIETIPVKLKPGMDGPKVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
KAR11D0pel   1   ..........................................................................................
KAR11D3pla   1   ..........................................................................................
KAR11D3pel   1   ..........................................................................................
KAR011M3pla  1   .......V..Q........................D.................I.....................................

KAR11D0pla   101 KKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTRILEPFRAQNPEIVIYQYMDDLYVGSDLEIGQHRA
KAR11D0pel   101 ..........................................................................................
KAR11D3pla   101 ..........................................................................................
KAR11D3pel   101 ..........................................................................................
KAR011M3pla  101 ...................Y.................V....................K.......K.....................

KAR11D0pla   201 KIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWSVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEEAE
KAR11D0pel   201 ..........................................................................................
KAR11D3pla   201 ..........................................................................................
KAR11D3pel   201 ..........................................................................................
KAR011M3pla  201 .........K..............................H.....N.T.....................K........V.............

KAR11D0pla   301 LELAENREILKEPVHGVYYDPSKDLIAEIQKQGNE
KAR11D0pel   301 .............................
KAR11D3pla   301 .............................
KAR11D3pel   301 .............................
KAR011M3pla  301 ............................QG
```

**KAR09**

```
KAR9D0pla    1   PISPIETVPVKLKPGMDGPKVKQWPLTEEKINALIEICTEMEKEGKISKIGPENPYNTPVFAIKKKDGTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGL
KAR09D3pla   1   ..........................................................................................
KAR09D3pel   1   ..........................................................................................
KAR09D28pla  1   ..........................................................................................
KAR09M3pla   1   ..........................................................P.......................................

KAR9D0pla    101 KKKKSITVLDVGDAYFSVPLDKEFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPEIVICQYMDDLYVGSDLEIGQHRT
KAR09D3pla   101 .....:.....................................................................:.....................
KAR09D3pel   101 .....:.....................................................................:.....................
KAR09D28pla  101 .....:.....................................................................:.....................
KAR09M3pla   101 .....:.....................................................................:.....................

KAR9D0pla    201 KIEELRQHLLQWGFTTPDKKYQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLKGTKALTEVVPLTEEAE
KAR09D3pla   201 ..............A.....................................................................................
KAR09D3pel   201 ..........................................................................................
KAR09D28pla  201 ........................................E.................................................
KAR09M3pla   201 ..........................................................................................

KAR9D0pla    301 LELAENREILKEPVHGVYYDPSKDLIAELQKQGEG
KAR09D3pla   301 .............................
KAR09D3pel   301 ............................D.
KAR09D28pla  301 .............................
KAR09M3pla   301 ........................L.........
```

**Figure 25. Comparison of RT sequences in the plasma and PBMCs for all patients. Resistance mutations are shaded in yellow.**

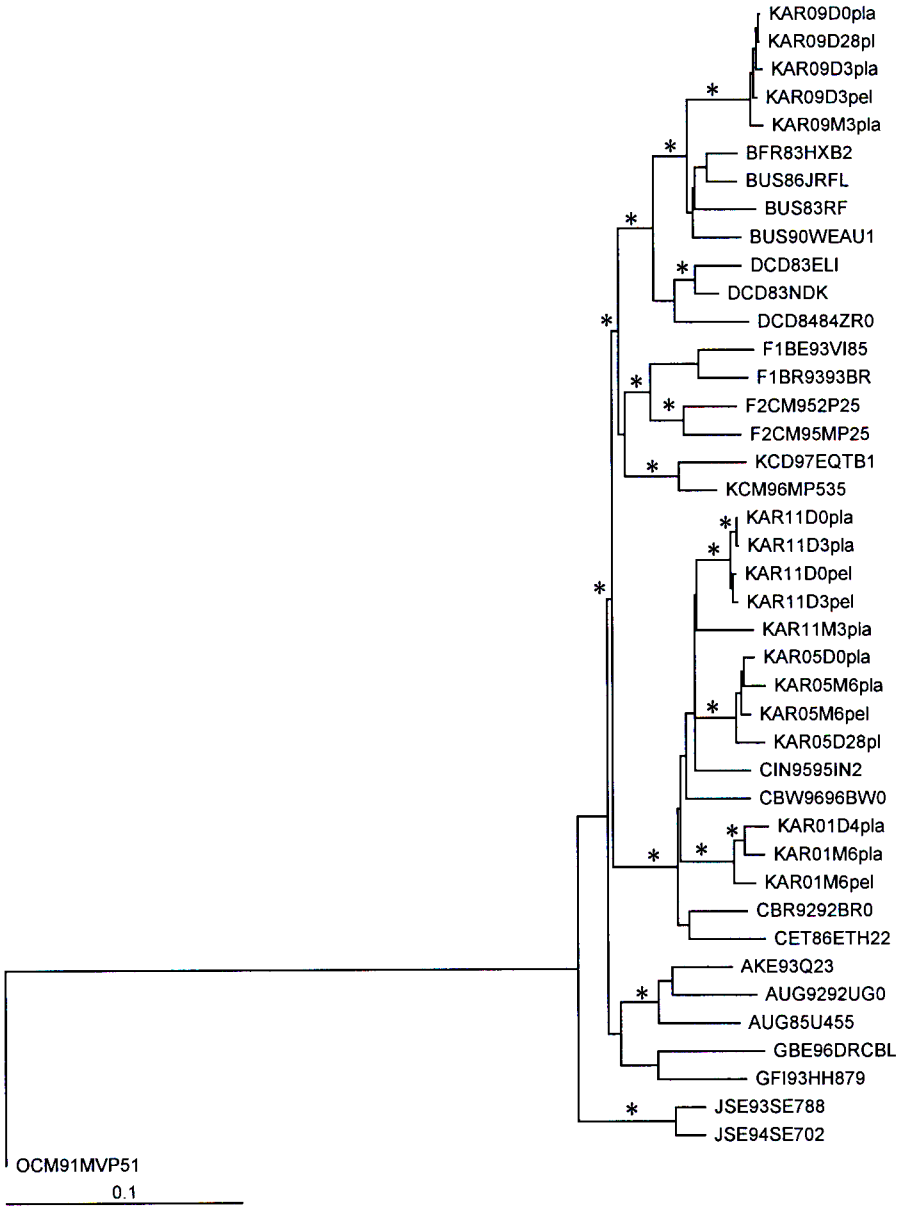



**Figure 26. Neighbour-joining tree of the resistant isolates and subtype reference strains. Bootstrap values >90% are represented by a "*". KAR09 isolates clearly clustered with the subtype B reference strains, while KAR01, KAR05 and KAR11 fell within the subtype C cluster. As expected, sequences from each isolate clustered together.**

### 6.3.3.3 Subtyping and Phylogenetic Analysis

Three of the four patients were infected with subtype C strains (KAR01, KAR05 and KAR11) and one with subtype B (KAR09) (Figure 26). Figure 27 shows the relationship between the patient isolates (plasma and PBMC) obtained from different time points. KAR09 showed very little genetic evolution between the different time points, characterised by very short branch lengths. The PBMC sequence was dispersed within the sequences obtained from the different plasma time points. KAR01 and KAR05 also did not show any clustering of plasma or PBMC sequences. Interestingly, there did appear to be some separation of plasma and PBMC sequences of the earlier isolates from KAR11.

### 6.3.4 Discussion

During HIV infection, virus may persist as latent reservoirs, where virus replication is restricted, but there is also the potential to rekindle productive viral infection when treatment is interrupted (Neuman et al, 1999; Davey et al, 1999). It may also persist at sites with low drug penetration, where there is ongoing viral replication. This replication may then lead to drug resistance in those cells or tissues (Gunthard et al, 2001). In this study, the development of resistance in plasma and PBMCs was investigated in patients failing therapy, using sequential samples with detectable viral load. Resistant virus was detected in three patients, two were classified as subtype C and the third as subtype B. Resistance was not related to the presence or absence of chemotherapy in the treatment regimen. In the two patients infected with subtype C, resistance was detected in the PBMCs of only one of the two, with the pattern of resistance different in the plasma and the PBMCs. Interestingly, the PBMCs had a greater number of resistance-associated mutations. However, the isolates were not cloned, and some mutations may have been missed by population sequencing. Future work will include cloning all isolates (plasma and PBMC), as well as tissue biopsies, so that the evolution of resistance in these patients can be more fully elucidated.

Although PBMCs have the potential to harbour a small latent reservoir of infectious virus under HAART, the baseline plasma sample from KAR05 did not appear to have resistance, so it is unlikely that the PBMCs were initially infected with resistant virus (Zhang et al, 1999). It has been shown that the PBMCs can be dynamically replenished during ongoing viral replication (Gunthard et al, 2001). Therefore, it is more likely that the PBMCs were seeded with resistant virus from another reservoir or compartment, potentially the macrophages or KS lesion. KS is known to stimulate the HIV-1 trans-activating regulatory gene (*tat*), which in turn activates transcription of the HIV-1 provirus

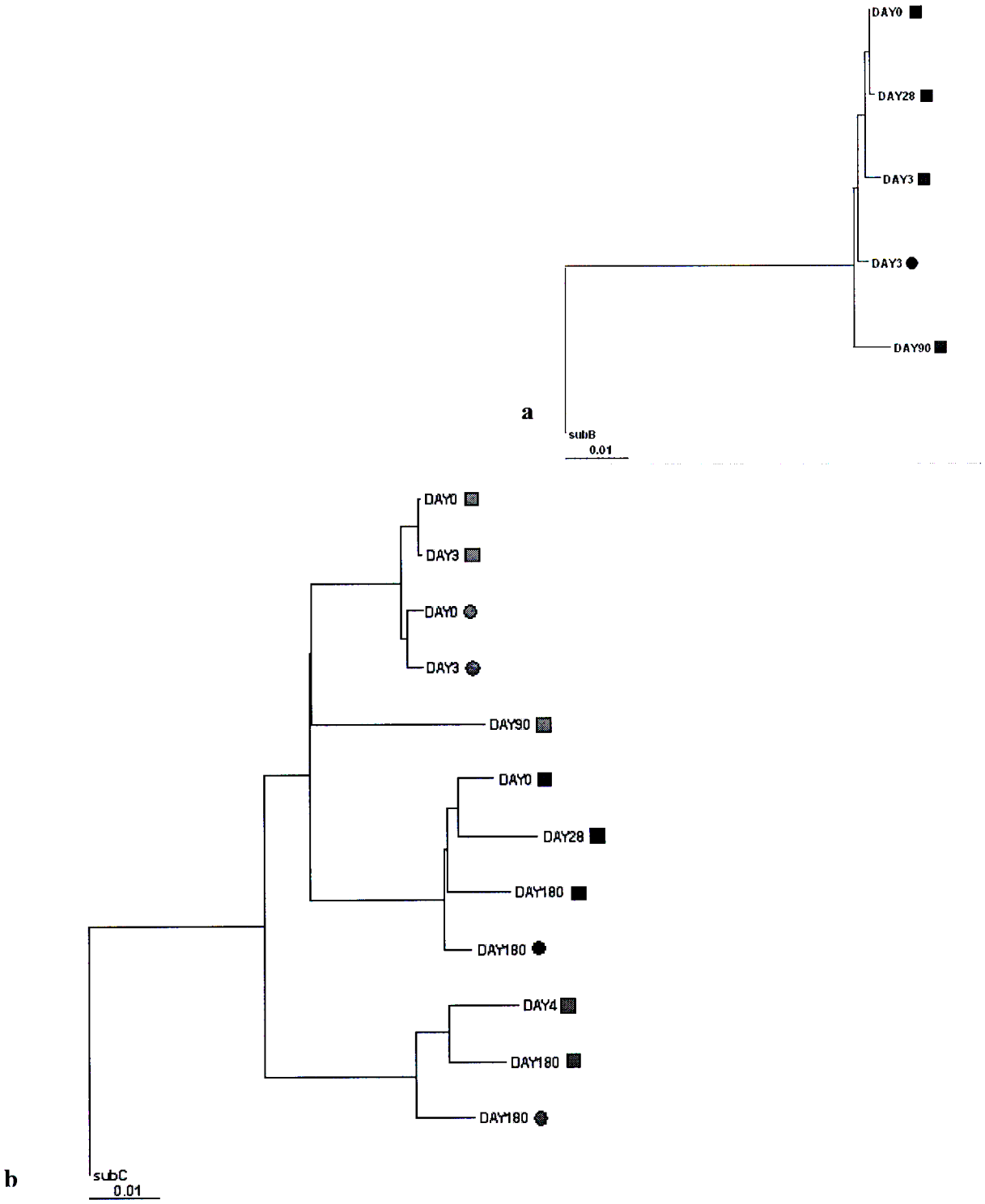



**Figure 27(a) An HXB2 rooted neighbour-joining tree using sequences from all the KAR09 (subtype B) available time points (represented by purple squares and circles). There was no clustering of PBMCs or plasma sequences, with very little genetic diversity between the sequences obtained for this patient. (b) Neighbour-joining tree using sequences from all the time points available for KAR01 (blue), KAR05 (red) and KAR11 (green) (subtype C isolates). The tree was rooted with a subtype C consensus sequence. Coloured squares represent plasma sequences and coloured circles represent PBMC sequences.**

(Vogel et al, 1998). The PBMCs do not usually contribute to the majority of rebounding virus in the plasma and an alternate reservoir could possibly be contributing to the resistance seen in the plasma, particularly since one of the resistant patients did not have detectable resistance mutations in the PBMCs. Unfortunately there were no other sequences available from earlier time points for the PBMCs or from the KS lesions for comparison. The difference in the evolution of resistance in the two subtype C infected patients could also suggest that host-specific factors are also contributing to the evolution of the virus *in vivo*, as has been shown by Anastassopoulou et al (2003).

Interestingly, no resistance was found in the plasma or PBMCs of KAR11. This increase in viremia could have been related to a burst of residual viral replication in that patient (Gunthard et al, 1999; Havlir et al, 2000). Of course, it is possible that a minor resistant population exists in this patient, which was missed by population-based sequencing. The patient infected with subtype B had detectable NNRTI resistance mutations at baseline. This highlights the importance of performing routine resistance testing at baseline. Also, this also highlights the problem that resistance mutations can develop in the PBMCs and once they are archived, they remain a concern for the efficacy of future therapy in those patients.

## 6.4 Surveillance of ARV Drug Resistance in a HIV Clinic in KZN, South Africa

### 6.4.1 Introduction

Antiretroviral therapy can prolong the lives and improve the health of patients with HIV-1 infection. However, resource-poor countries have been slow to provide access to ARV drugs. This is in the process of changing as the availability of generic ARV drugs increases. The planned treatment of large numbers of HIV-1-infected persons in South Africa has underscored the urgent need to develop simple, standardized once-daily drug dosing regimens that are inexpensive, specific, and appropriate for use in Africa. Unfortunately, the emergence of mutations that cause resistance to ARV drugs can limit the efficacy of treatment regimens. Therefore it is of paramount importance that effective and affordable tools are developed for monitoring treatment responses, not only in individual patients, but also at the population level. Although South Africa has only recently implemented its national ARV program, there are patients who have already been treated in the private sector, prior to the establishment of the official National Guidelines for ARV therapy. As a result, some patients may have received sub-optimal regimens. Equally as important, these patients may have received little or no formal adherence training. In light of this, the screening for drug resistance is important to ensure that each patient receives optimal therapy and avoids the use of ineffective drugs. This section focuses on the surveillance of drug resistance in a single HIV clinic. It provides information on the prevalence and pattern of resistance mutations in a treatment experienced subtype C population.

### 6.4.2 Study Population

Samples were obtained from HIV-1 positive adults and children attending the Sinikithemba clinic at McCords Hospital. McCords is a state subsidised, independent, not for profit, hospital in Durban, KZN. The first 100 patients on treatment presenting at the clinic were included in the analysis. These consisted of 96 adults (46% females and 54% males) and 4 children (75% males and 25% females); see Table 22. Some patients had already been treated at other centres, including the private sector, before coming to the Sinikithemba HIV Clinic. Most of these patients had been on treatment for a longer period and this was reflected in their treatment histories that included mono and dual therapy. Patients started on treatment at the Sinikithemba Clinic were all given HAART. In some cases, patients were changed from ddI, d4T and EFV to CBV and EFV because CBV was cheaper and more easily administered. A list of the various treatment regimens for the entire cohort are shown in Table 23.

**Table 22. Characteristics of and laboratory results for patients in the study**

| | | All Patients (n=100) | Adults (n=96) | Children (n=4) |
|---|---|---|---|---|
| # males | | 53/98* (54%) | 50/94 (54%) | 3/4 (75%) |
| # females | | 45/98* (46%) | 44/94 (46%) | 1/4 (25%) |
| # heterosexual | | 82/83* (99%) | 82/83 (99%) | N/A |
| #homosexual | | 1/83* (1%) | 1/83 (1%) | N/A |
| Ethnicity | Black | 90/100 (90%) | 86/100 (86%) | 4 (4%) |
| | Indian | 4/100 (4%) | 4/100 (4%) | 0 (0%) |
| | White | 4/100 (4%) | 4/100 (4%) | 0 (0%) |
| | Coloured | 2/100 (2%) | 2/100 (2%) | 0 (0%) |
| Average age (years) | | 33.5 | 35.3 | 7 |
| Average Viral Load | | 26396 | 26423 | N/A |
| Average CD4 count | | 194 | 183 | 515 |
| Average stage of Disease (WHO) | Stage1 | 2/85* (2%) | 2/83 (2%) | 0/2 (0%) |
| | Stage2 | 7/85* (8%) | 7/83 (8%) | 0/2 (0%) |
| | Stage3 | 59/85* (69%) | 58/83 (70%) | 1/2 (50%) |
| | Stage4 | 17/85* (20%) | 16/83 (19%) | 1/2 (50%) |
| Patients currently on treatment | | 86/100 (86%) | 82/100 (82%) | 4/100 (4%) |
| Patient not currently on treatment | | 14/100 (14%) | 14/100 (14%) | 0/100 (0%) |
| Average time on treatment (months) | | 7.31 | 6.92 | 17 |

*** Total number of patients for which the information was available**

### 6.4.3 RNA Extraction and RTPCR and Resistance Genotyping

Samples were tested for viral load and CD4+ counts as described in Chapter 3.2.2. HIV-1 RNA was extracted, amplified by RTPCR and sequenced using the Viroseq™ system (Applied Biosystems) according to the manufacturer's protocol and as described in Chapter 3.2.4. Sequences were analysed for resistance as described in Chapter 3.2.8.

### 6.4.4 Subtyping and Phylogenetic Analysis

Subtyping was performed by phylogenetic tree analysis using PAUP* Swofford, 1999), as described in Chapter 2.1.3 and Chapter 3.2.5. Sequences were analysed for recombination using SimPlot as described in Chapter 2.1.4.

### 6.4.4 Results

A total of 23 isolates with detectable viral load were sequenced. Of these, 21 isolates (91%) were classified as subtype C. The remaining two isolates were classified as an A/G recombinant (Figure 28) and subtype A (Figure 29). The breakpoint in the A/G recombinant was midpoint in the PR, with all of RT being classified as subtype A (Figure 28). Eighteen (18%) of the patients that were enrolled displayed drug resistance mutations. Their resistance profiles and treatment histories are shown in Table 24. Of particular importance was the finding that all but one of the resistant patients that had been on a NNRTI inclusive regimen had lost the NNRTI class of drugs, with >90% of the resistant isolates harbouring viruses resistant to NNRTIs (Figure 30a) and >70% of those harbouring multi-NNRTI resistant viruses (Figure 30b). The most common NNRTI drug resistance mutations that were seen were the K103N/S (33%) and V106M (27.7%) mutations. The V106M mutation occurred in 4 of 6 patients currently receiving EFV, (MC027, MC035, MC089 and MC177). This mutation also occurred in an isolate from a patient receiving a NVP-containing regimen. Of the 5 patients currently receiving NVP mutations at codon 103 were most common (K103N=2 and K103S=1). Other common NVP related drug resistance mutations seen were the Y188L/C (27.7%), G190A (22.2%), A98G (11.1%), V179D (11.1%), K101E/Q (11.1%), V106A (5.5%), Y181C (5.5%), P225H (5.5%) and F227L (5.5%) mutations. The F227L mutation occurred in conjunction with the V106A mutation, which is said to augment NVP resistance (Shafer, 2003). The most common NRTI mutation was the M184V mutation (38.8%), which occurred in 88% of the resistant isolates that were currently on 3TC. Other NRTI mutations were the Thymidine analogue mutations (TAMS) [D67N/G (27.75%), K70R/E (16.6%), L215F/Y,

## Table 23. Treatment data of patients in the study.

| Patients currently receiving dual-NRTI regimens | Months |
|---|---|
| No of Patients | 5 |
| Average duration of treatment | 15 |
| Different dual-NRTI regimens | Average duration of regimen |
| ddI, d4T (2) | 9 |
| 3TC, AZT (2) | 23 |
| CBV (1) | 12 |
| | |
| Patients currently receiving HAART, first line | |
| No of patients | 47 |
| Different HAART regimens: | Average duration of regimen |
| CBV, EFV (19) | 4 |
| ddI, d4T, EFV (17) | 4 |
| d4T, 3TC, EFV (3) | 3 |
| ddI, d4T, NVP (5) | 8 |
| CBV, NVP (2) | 9 |
| RTV, CBV (1) | 2 |
| | |
| Patients currently receiving HAART, not first line | |
| No of patients | 33 |
| No of patients with previous mono-NRTI regimens | 2 |
| Different mono-NRTI regimens: | Average duration of treatments |
| RTV (1) | 1 |
| d4T (1) | 1 |
| No of patients with previous dual-NRTI regimens | 11 |
| Different dual-NRTI regimens: | Average duration of treatments |
| ddI, d4T (6) | 8 |
| d4T, 3TC (1) | 1 |
| 3TC ,AZT (2) | 10 |
| SQV, RTV (1) | 1 |
| CBV (3) | 2 |
| | |
| No of patients with previous HAART regimens | 22 |
| Different HAART regimens | Average duration of treatments |
| ddI, d4T, EFV (12) | 6 |
| ddI, d4T, NVP (8) | 5 |
| ddI, AZT, NVP (1) | 18 |
| ABC, ddI, d4T (1) | 8 |
| | |
| Current HAART regimens: | Average duration of treatments |
| CBV, EFV (19) | 3 |
| ddI, d4T, EFV (3) | 6 |
| d4T, 3TC, EFV (3) | 3 |
| ddI, d4T, NVP (1) | 6 |
| CBV, NVP (3) | 4 |
| ddI, AZT, EFV (1) | 2 |
| LPV/r, ddI, d4T (1) | 7 |
| IDV, RTV, d4T, 3TC (1) | 3 |
| | |
| No of patients changed treatment > once | 5 |
| | |
| Patients not currently on treatment | 15 |
| No of patients with previous mono-NRTI regimens | 3 |
| No of patients with previous dual-NRTI regimens | 1 |
| No of patients with previous HAART regimens | 11 |
| Number of different HAART regimens | Average duration of treatments |
| 3TC, AZT, NVP (1) | - |
| d4T, 3TC, NVP (1) | 13 |
| ddI, d4T, EFV (6) | 9 |
| ddI, d4T, NVP (2) | 4 |
| CBV, EFV (1) | 1 |
| | |
| No of patients currently receiving the following drugs: | |
| 3TC | 10 |
| AZT | 3 |
| ddI | 29 |
| d4T | 36 |
| CBV | 44 |
| NVP | 11 |
| EFV | 66 |
| RTV | 2 |
| IDV | 1 |
| SQV | 1 |
| LPV/r | 1 |

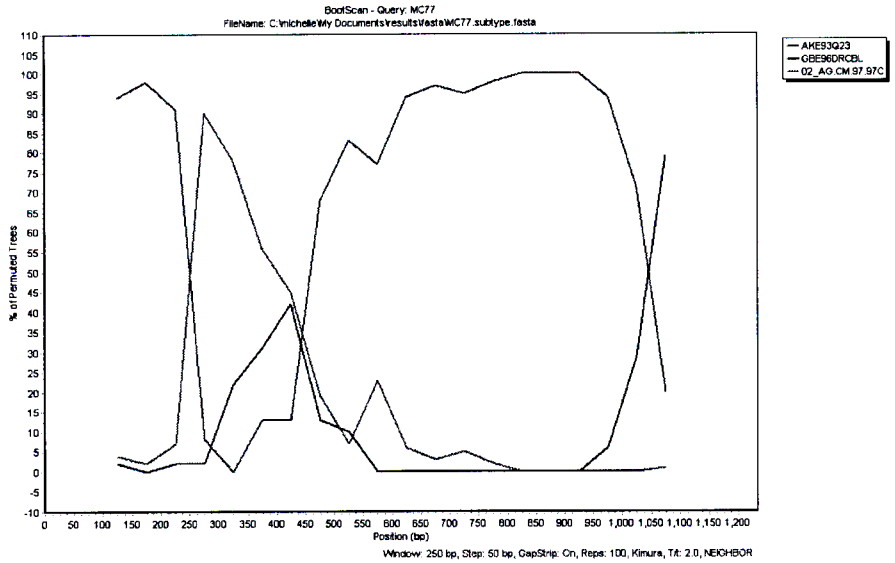



**Figure 28. Bootscan result for MC77. The analysis was performed with a sliding window of 250 bp and a step of 50 bp. The breakpoints were estimated to be just before the end of PR and aproximately150 bp in RT.**

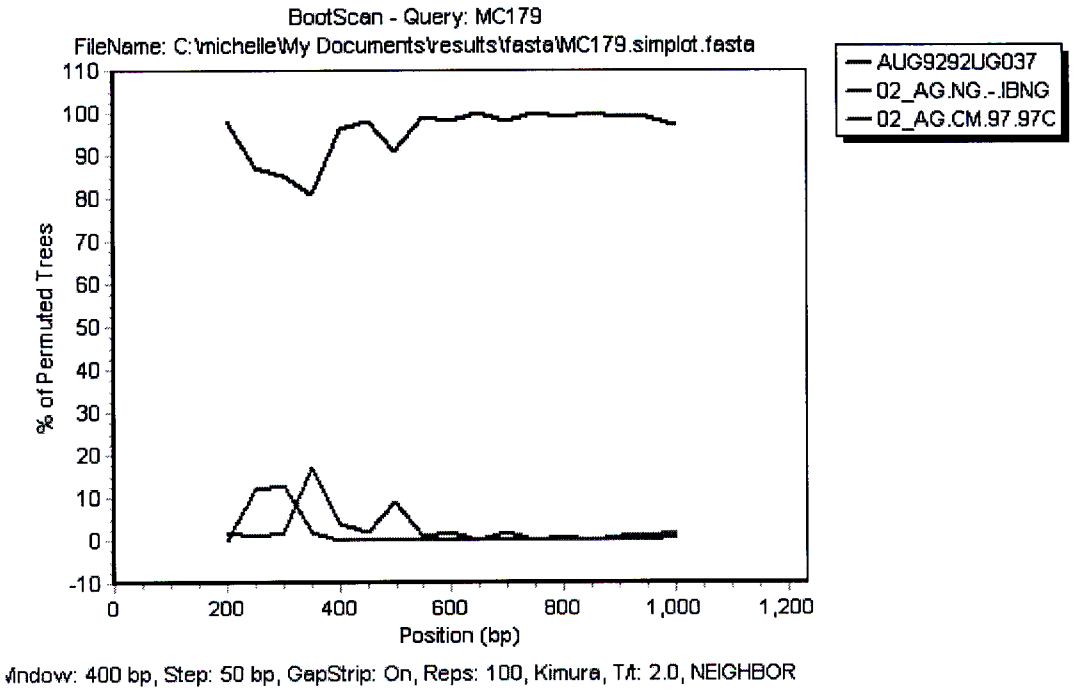



**Figure 29. Bootscan result for MC179. The analysis was performed with a sliding window of 400bp and a step of 50 bp.**

T69N/A (11.1%), L210W/S (11.1%), K219Q/E (11.1%), G333E (5.5%) and V75I/L (5.5%)]. While these mutations develop as a result of the selection pressure exerted by the drugs in the patient's regimen, like the NNRTIs, they may also cause cross-resistance to other drugs in that class (Figure 31), and thereby limiting the patient's future treatment options.

Some mutations that were seen have rarely been reported in the literature. These include the T69I and K70E mutations, which have rarely been found in treated subtype B isolates and have not previously been reported for subtype C. The K103S has only recently been characterized as conferring resistance to all the NNRTIs and at the time of writing this manuscript, there were no reported subtype C isolates with this mutation. Five patients were currently receiving dual therapy. Sequences were obtained for four of the patients, however only 2 of the isolates showed drug resistance mutations (MC109 and MC242). These patients were both receiving AZT and 3TC, both of whom were children. Common to both isolates was the M184V mutation, which causes resistance to 3TC. The other mutations were a variety of TAMs, which are associated with resistance to AZT. In addition, patient MC242 developed the A98G mutation, which is associated with reduced susceptibility to the NNRTIs. Seventy-five of the patients were WHO stage 3 or 4, with 60 of the patients having experienced recent co-infections. Almost half of these infections were mycobacterial. Comparison of the frequency of polymorphisms in the RT between treated patients with and without resistance mutations are shown in Figure32. While polymorphisms at RT positions 35, 173, 200, 207, 245, 293 occurred at equal frequencies in resistant and non-resistant isolates, the R211K mutation was found more frequently in resistant isolates. Polymorphisms at codons 135, 138 and 324 occurred only in resistant isolates (>40%).

### 6.4.5 Discussion

HAART is just becoming available in South Africa, with just a few ARV clinics operational. The high cost and complexity of treatment and monitoring has been a major hurdle in the quest to make treatment available for all HIV-1 infected South Africans. The availability of affordable generic drugs is now making the access to ARVs more attainable. However, coupled with this increased access to ARVs is the increased possibility for the emergence and/or transmission of ARV resistant viruses. Resistance is the result of sub-optimal or inconsistent regimens often due to poor counselling or interrupted drug supply (Richman, 1994). Inadequate infrastructure to ensure high quality patient monitoring may further exacerbate the problem. Avoiding resistance is even more

**Table 24. Genotypic profiles of viral isolates for each patient**

| Sample ID | Stage | Subtype | Drugs used in therapy | Viral load | Time on treatment | Co-infections | PR resistance mutations | RT resistance mutations |
|---|---|---|---|---|---|---|---|---|
| MC017 | 3 | C | CBV, ddI, d4T, NVP | 1 400 | 26 | peripheral neuropathy | M36I, V82I, I93L | D67N, K101Q, M184V, G190A |
| MC027 | 3 | C | ddI, d4T, EFV | 4 400 | 12 | chronic diarrhea; vomiting | M36I, I93L | V106M |
| MC028 | - | C | ddI, d4T, NVP | 910 | 5 | peripheral neuropathy | M36I, L63V, I93L | K70E, V106A, V118I, G190A, F227L |
| MC035 | 3 | C | ddI, d4T, EFV | 11 000 | 17 | - | K20R, M36I, L63P, I93L | K65R, T69I, K103N, V106M |
| MC037 | B | C | ddI, 3TC, NVP | 8 700 | 16 | pneumonia; mild oral thrush | K20R, M36I, D60E, L63P, I93L | K101E, G190A |
| MC043 | 3 | C | DdI | 12 000 | 0.5 | TB | M36I, D60E, I93L | A98G, K103N, V179D, Y188L |
| MC077 | 4 | A/G | CBV, ddI, d4T, EFV | 1 500 | 22 * | Osophageal candida | L10I, M36I, L63T | V179D |
| MC181 | | C | d4T, 3TC, EFV | 440 000 | | - | L10I, M36I | M184V, Y188L, P225H |
| MC106 | 2 | C | ddI, d4T, NVP | 4 300 | 10 | - | M36I, L63V, I93L | Y181C |
| MC109 | - | C | 3TC, AZT | 24 000 | 9 | pneumonia; oral candida, occ diahroea | M36I, I93L | D67N, K70R, M184V, K219E |
| MC89 | 3-4 | C | ddI, d4T, EFV, NVP | 5 300 | 6 | Taxoplasmosis | M36I, D60E, L63P, I93L | V75I, K103N, V106M, Y188C |
| MC177 | 3 | C | ddI, d4T, EFV | 3 000 | 7 | - | M36I, L63Q, I93L | D67G, T69A, V106M, Y188L |
| MC179 | 4 | A | RTV, CBV | 12 000 | 2 | - | K20R, M36I, L63V, I93L | G190A |
| MC197 | 4 | C | d4T, 3TC, EFV, CBV, NVP | 84 000 | 1 | PTB | M36I, L63A, I93L | T69N, K103N |
| MC217 | 3 | C | CBV, NVP | 4 800 | 16 | - | M36I, D60E, L63P, I93L | D67N, K70R, M184V, Y188L, T215F, K219E |
| MC210 | - | C | 3TC, AZT, NVP | 310000 | 3 | - | L10I, M36V, L63V, I93L | K103N, M184V |
| MC242 | 4 | C | 3TC, AZT | 4500 | 36 | - | M36I, L63P, I93L | M41L, E44D, A98G, M184V, L210W, T215Y |
| MC244 | 3 | C | 3TC, AZT, NVP | 15000 | 36 | - | M36I, I93L | M41L, E44A, D67N, K103S, V106M, M184V, L210S, T215Y, G333E |
| MC010 | 3 | C | ddI, d4T, EFV | 890 | 3 | herpes simplex, oral candida; otitis media | M36I, L63P, I93L | none |
| MC056 | 3 | C | ddI, d4T | 8 200 | 8 | on TB Rx | K20R, M36I, I93L | none |
| MC071 | 4 | C | ddI, d4T, NVP | 90 000 | 6 | neuropathy; myelopathy; bacterial Infection | K20R, M36I, I93L | none |
| MC087 | 3 | C | ddI, d4T | 1300 | 9 | TB | M36I, I93L | none |
| MC092 | 3 | C | ddI, d4T, CBV, EFV | 69000 | 3 | peripheral neuropathy | L10M, M36I, L63I, I93L | none |

* Not currently on treatment

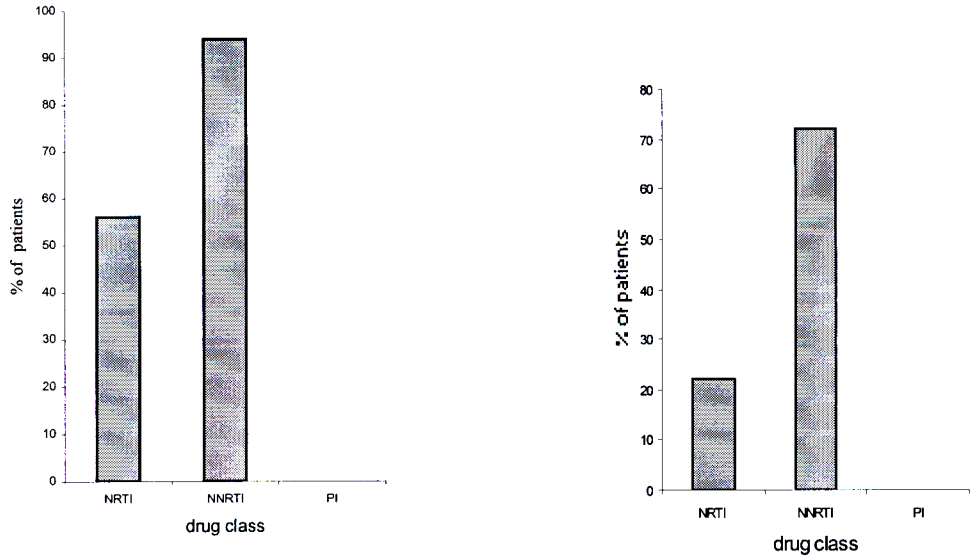



**Figure 30. a) Frequency of resistance to the NRTIs, NNRTIs and PIs in the resistant isolates. b) Frequency of multi- NRTI and NNRTI resistance in the isolates. From both figures, a high rate of NNRTI resistance is evident.**

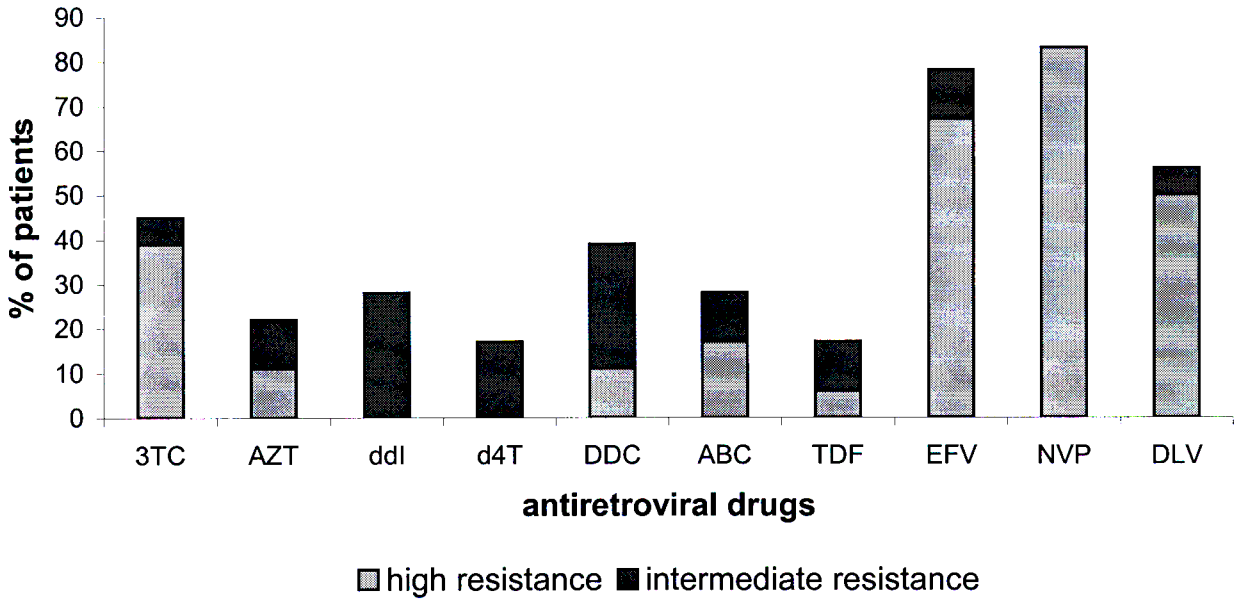



**Figure 31. Levels of cross-resistance to the RT inhibitors. Most resistant isolates showed a high level of resistance to all the NNRTI with intermediate to high-level resistance also seen in the NRTIs, albeit to a lesser extent.**

important in countries where first line and second line treatment options are limited. Thus, surveillance coupled with intense program monitoring is needed to maintain program effectiveness. This study was conducted to monitor the prevalence of resistance in a local HIV clinic as well as to increases the data on the response of subtype C viruses to therapy. Interestingly, one of the resistant isolates were classified an A/G recombinant by phylogenetic analysis. This is the first report of an A/G recombinant isolated from KwaZulu Natal. Other non-C isolates have previously been reported in Cape Town with one C/D recombinant reported for KwaZulu-Natal (Gordon et al, 2003).

From those included in the study, resistance mutations were detected in 18 (18%). This figure is better than the estimated prevalence of 36% resistant virus among Ugandan patients and 40-57% resistance among patients from Cote d'Ivoire on HAART (Vergne et al, 2003). The high prevalence of resistance in these countries is thought to be a consequence of the inappropriate use of mono and dual therapy with NRTIs. This could also explain some of the resistance seen in the patients in this study. Although treatment in this setting was under controlled conditions, some of the patients had come from the private sector prior to the establishment of the National Guidelines for ARV treatment and where there was not the same level of control. Consequently, some of the patients were treated with sub-optimal regimens.

Most of the resistance mutations that were seen are known to cause resistance to the NNRTIs. This is probably due to the low genetic barrier to resistance in this class of drugs. Of particular interest is the V106M mutation seen in two patients, which is a putative subtype C signature mutation in EFV experienced patients (Brenner et al, 2003). From the previous study (Chapter 6.1), I have found that the V106M mutation emerges as early as one month after starting treatment. The resistance profile of MC028 was very interesting because, although the patient had stopped treatment, there were a host of mutations present, some causing primary resistance to the NNRTIs, while others such as K70E and F227L are not known to cause resistance to any current RT inhibitors. It is possible that the K70E mutation is an intermediate mutation before the development of the K70R variant (a known TAM). Most of the polymorphisms found in the RT that are not associated with resistance in the context of subtype B, were equally prevalent in isolates with and without resistance. Two polymorphisms were peculiar to isolates without resistance: mutations at codon 122 occurred in 80% of patients while mutations at codon 334 occurred in 60% without resistance. There were also a few polymorphisms that only occurred in resistant isolates: mutations at codons 135 occurred in 56% of resistant isolates, while mutations at codons 138 and 324 occurred in 44% of resistant isolates; and

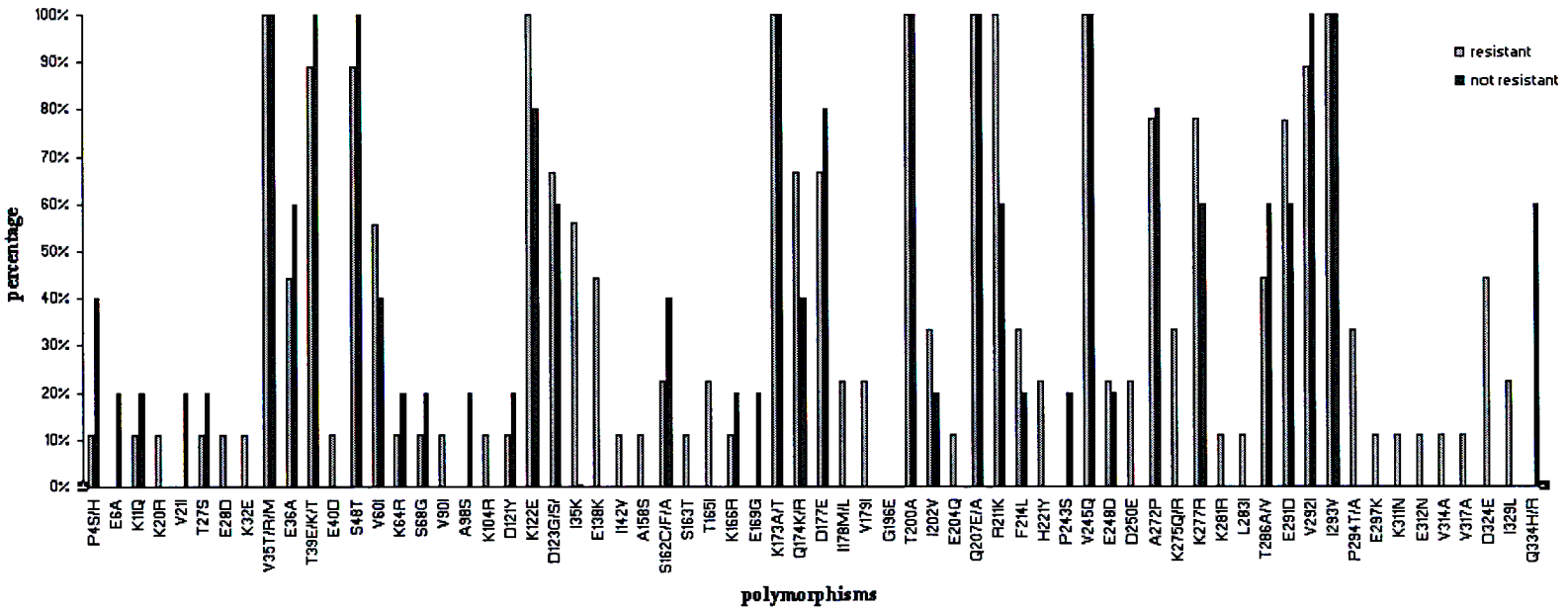



**Figure 32. Comparison of polymorphisms in the RT between treated patients with and without resistance mutations. Polymorphisms occurred at equal frequencies in resistant and non-resistant isolates at RT positions 35, 173, 200, 207, 245, 293, while the R211K mutation was found more frequently in resistant isolates. Mutations at codons 135, 138 and 324 were found in >40% and only in resistant isolates.**

mutations at codons K275 and P294 occurred in 33% of resistant isolates. Similarly, polymorphisms in the PR were common to resistant and wild-type isolates, with no particular mutations occurring more frequently in either group.

Most of the patients included in the study were in the late stages of the disease (stage 3 and 4), and with an average CD4 count of 194. This is not surprising, as most people in South Africa do not know their HIV status until very late in the infection, when they begin to show symptoms of AIDS. As a result, most of these patients (60%) had experienced recent co-infections which further complicated patient management, particularly TB, which limits the drugs that can be administered to these patients as well as co-infections such as oesophageal candidiasis which have been associated with a more rapid progression to death (Morgan et al, 2002) There have also been reports of a more rapid progression to AIDS if HAART is initiated after CD4 counts fall below 200 cells/ml, than if initiated at counts between 200 to 350 (Chan et al, 2002). However, there is still some benefit in treating late stage AIDS patients, as adults and children with advanced AIDS who rarely survive more than two years, are increasing their survival time as well as quality of life. (Morgan et al, 2002).

A major problem for most HIV-1 positive patients contemplating starting ARV treatment is cost of the ARVs and monitoring (CD4s and viral loads). Those that can afford the drugs, may not be able to afford the monitoring. Sometimes the patients start treatment, but later find that they can no longer afford the ARVs due to a change in financial circumstances, and stop treatment until they can purchase them at a later date. Twenty five (25%) of the patients in the study had stopped taking their drugs at some point in their treatment. Eight of these developed resistance.

Some researchers feel that, particularly in resource poor setting where the choice of drugs for salvage therapy would be even more limited, the choice of first ARV regimen is of greatest importance, and therefore should be well tolerated, have a relatively low pill burden and ease of administration, and have a high genetic barrier to resistance in the event of less than perfect adherence (Palella et al, 2002; Bartlett et al, 2001). Based on ease of administration and relative tolerability, the preferred "backbone" NRTI combination is AZT and 3TC (now available as CBV), with d4T and 3TC or AZT and ddI as alternatives (Staszewsk et al, 2001). Many researchers have found that EFV with 2

NRTIs is as effective as 2 NRTIs with a PI such as NFV and INV (Bartlett et al, 2001; Albrechts et al, 2001).

Finally, although a high level of multi-NNRTI resistance was seen in the patients, all the resistant patients were still susceptible to the PI class of drugs. As South Africa is limited in the drug options that are available for salvage therapy, first line regimens need to be chosen wisely, and it is equally important to have effective second line regimens in place, given that some level of resistance is inevitable. Resistance genotyping will aid clinicians who are currently working blindly, ignorant of the resistance status of patients who are clinically failing treatment. This is particularly true where the patient cannot produce a complete clinical history. It is hoped that these data will assist in patient treatment management, and facilitate public health policy and planning.

# Chapter 7

## *In Vitro* Antiretroviral Drug Susceptibility of Subtype C Strains

## 7.1 Phenotyping of HIV-1 isolates from South Africa

### 7.1.1 Introduction

HIV replication is markedly inhibited by highly active antiretroviral therapy (HAART). However, the development of drug resistance is an important factor limiting the effectiveness of ARV treatment. Mutations associated with resistance to RTIs and PIs have been extensively characterized (Hirsch et al, 2000; Schinazi et al, 2000). Guidelines by expert panels have recommended the routine use of drug resistance testing in the treatment of HIV-1 infected patients (US Dept Health; Hirsh et al, 2000). Current methods of HIV resistance testing include *in vitro* phenotyping (drug-susceptibility assays) and genotypic assays that detect mutations known to confer drug resistance. Genotypic tests are more commonly used in clinical settings because of their wider availability, lower cost and quicker turnaround, although phenotypic tests are a more direct measurement of HIV-1 susceptibility. However, there are limitations to both of these tests. They are both unable to detect minor drug resistant populations and there is limited knowledge about the clinical significance of certain combinations of mutations and of certain levels of phenotypic drug resistance. Most HIV drug susceptibility studies have involved subtype B strains, although non-B subtypes cause most of the infections worldwide. In a study conducted by Vergne et al, (2000) on naturally occurring polymorphisms in non-B subtypes, they found the prevalence of major mutations associated with resistance to NRTIs, NNRTIs and PIs very low. However they found many accessory mutations related to resistance to NRTIs and PIs. A significant number of amino acid polymorphisms exist among subtype A, B and C sequences (Brennan et al, 1997; Becker-Pergola et al, 2000). The high prevalence of the L10I, K20R, M36I, and L89M mutations in non-subtype B strains is largely known (Pieniazek et al, 2000; Tanuri et al, 1999). Velazquez-Campoy (2001) have shown that subtypes A and C proteases have a higher biochemical fitness than subtype B proteases against the currently available PIs. However, this does not necessarily reflect a lower *in vivo* efficacy, because other factors may also be important in antiviral potency (Klabe et al, 1998).

The recombinant virus assay (RVA). is an innovative method for generating viruses for drug susceptibility testing (Boucher et al, 1996; Maschera et al, 1995; Kellam et al, 1994). RVA involves the production of viable virus *in vitro* by homologous recombination of RT-PCR products from plasma virus with an infectious PR or RT deleted subtype B cloned HIV-1 provirus. The resulting recombinant viruses derive all their biological properties from the molecular clone, except for RT and PR, which are encoded by the genes from the patient isolates. The drug susceptibility of the viruses is measured using a MTT cell-

killing assay (Pauwels et al, 1988). Living cells convert MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) into a blue product (formazan), and therefore the amount of formazan reflects the number of cells protected by the drug against killing by the virus. The production of formazan can be quantified spectrophotometrically, and this coupled with the possibility of performing the test with microtiter plates, facilitates testing of large numbers of viral isolates. The biochemical and phenotypic roles of the natural polymorphisms found in the PRs of C viruses are poorly understood. In this chapter, we characterize the drug susceptibility of subtype C isolates from South Africa to the currently available PR and RT inhibitors. The phenotyping was done in collaboration with colleagues from Brazil.

## 7.2 Methods

### 7.2.1 Samples

10 patient samples were phenotyped (Table25). The majority (8/10) were drug naive, while two were known to have had prior treatment with d4T, 3TC and NVP. Seven of the isolates were classified as subtype C. For comparison, the D/C recombinant described in chapter 3 was included, as well as a retrospective subtype B isolate from Cape Town. One of the treated isolates was also classified as subtype B.

### 7.2.2 Preparation of PCR products for transfection

Viroseq RT PCR products were lyophilized and transported to the Molecular Virology Laboratory in the Genetics Department at the University of Rio de Janeiro, Brazil. The products were reconstituted with 40ul sterile filtered water and the protease and RT regions were amplified separately using the primers described by Maschera et al, 1995 (protease) and Boucher et al, 1996 (RT). Briefly, primers 5'GGG AAG ATC TGG CCT TCC TAC AAG GG 3' (forward) and 5' GGC AAA TAC TGG AGT ATT GTA TGG 3' (reverse), 200uM dNTPs, 1.5mM $Mg^{2+}$, 250 nM primer and 0.75U AmpliTaq Gold (Applied Biosystems) were used to amplify the protease regions under the following conditions: 35 cycles of 94°C for 30 seconds, 55°C for 55 seconds and 72°C for 1 minute, after an initial denaturation of 94°C for 12 minutes. The following primers were used to amplify the RT region with the same conditions: RT 5' GGA CAT AAA GCT ATA GGT ACA G 3' (forward) 5'CTG CCA GTT CTA GCT CTG CTT C3' (reverse). The vectors used were: pGEMT3ΔPR, which carries defective HIV HXB2 genomic cDNA that lacks the gene for PR (ΔPR) (Maschera et al, 1995) and ΔRT which lacks the first 290 bp of RT.

**Table 25. List of samples and their resistance mutations**

| Seq ID | Subtype | Ethnicity* | Drug Exposure | NASBA Viral Load (cpm) | PR Mutations | RT Mutations | Year of Isolation |
|---|---|---|---|---|---|---|---|
| ZA010p01 | C | Black | naive | 24 000 | M36I, I93L | A98G | 2001 |
| ZA012p01 | C | White | naive | 220 000 | M36I, L63P, I93L | V118I | 2001 |
| ZA024p01 | C | Indian | naive | 250 000 | L63I, V77I, I93L | K103N | 2001 |
| ZA023p01 | C | Indian | naive | 280 000 | L63T, V77I, I93L | K103N, G190A | 2001 |
| ZA021p01 | D/C | Indian | naive | 180 000 | M36L | - | 2001 |
| ZA028p01 | C | Black | naive | 35 000 | M36I, L63P, I93L | - | 2001 |
| ZA032p01 | C | Black | naive | 62 000 | M36I, I93L | - | 2001 |
| R555p91 | B | - | naive | - | L63P | M41L, L210W, T215Y | 1991 |
| PT01p02 | B | Indian | 3TC, d4T, NVP | 36 856 | L63P, V77I, I93L | D67, K70R, M184V, G333E, K101E, G190A | 2002 |
| PT02p02 | C | Indian | 3TC, d4T, NVP | 13 375 | K20R, M36I | K65R, M184V, K103N, Y181C | 2002 |

***South Africans of different descents**

The vectors were prepared by overnight digestion with either BsteII (ΔPR) or SmaI (ΔRT). Each PCR product was co-precipitated with the respective vector (i.e. PR products with the protease vector). Briefly, 1 ug of the PCR product and 1 ug of the vector was precipitated with 3 times the volume of ethanol and 10% of a 3M sodium acetate solution and centrifuged at 12000 rpm for 20 minutes. The supernatant was aspirated and the pellet washed with 70% ethanol. The supernatant was aspirated again and the pellet was dried on a heating block for 5 minutes.

### 7.2.3 Cell Electroporation

Two days before performing the transfection, cell cultures were prepared with 2.5 X $10^5$ MT2 cells/ml. At Day0, cells were counted and spun at 2000 rpm for 8 minutes and resupended at 2 X $10^7$ cells/ml in RPMI with 20% FBS. Cells were left on ice for 10 minutes. Each PCR product/vector precipitate was resuspended with 25 ul sterile water. 250 ul of cells and each resuspended PCR product/vector were added to separate cuvettes and electroporated with a BioRad Gene Pulser (900 μF capacitance; high ohms; 250 V). 125 ul RPMI was added to the cuvette to gently wash the cells away from the sides of the cuvette and clumps were gently resuspended. The electroporated cells were transferred to a flask with 5 ml RPMI and 25 ul fresh cells (2 X $10^7$cells/ml) were added to the flasks. Cell cultures were incubated at 37°C and monitored daily for syncytium formation. Fresh medium was added to maintain an optimal cell concentration of approximately 0.2 X $10^6$ cells/ml. Viral supernatants were harvested by centrifugation (1000 x g for 10 minutes) when the culture contained 100% full blown syncytia, which indicated that nearly all the cells in the culture were involved in giant cell formation. Viral stocks were stored at –70°C for subsequent titration and sensitivity testing. Homologous recombination led to the generation of a chimeric virus containing PR or RT sequences derived from patient viruses. A chimeric virus containing the NL4-3 PR (HXB2/NL4-3-PR) was also generated as a subtype B control virus to obtain the reference $IC_{50}$ during the phenotypic assay.

### 7.2.4 Determining the virus $TCID_{50}$

The virus culture to be tested was thawed and serially diluted (1 in 10) up to $10^{-4}$.
A 0.2 X $10^6$ concentration of cells was prepared and infected with the different virus dilutions by spinning at 1200 x g for 2 hrs in a process called "spinoculation".

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS |
| PBS | neat | neat | $10^{-1}$ | $10^{-1}$ | $10^{-2}$ | $10^{-2}$ | $10^{-3}$ | $10^{-3}$ | $10^{4}$ | $10^{-4}$ | PBS |
| PBS | | | | | | | | | | | PBS |
| PBS | | | | | | | | | | | PBS |
| PBS | | | | | | | | | | | PBS |
| PBS | | | | | | | | | | | PBS |
| PBS | | | | | | | | | | | PBS |
| PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS |

**Figure 33. Template of the plate layout for determining the virus $TCID_{50}$.**

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | | | | | | |
| 3 | | | | | | |
| 4 | **Dilutions** | **Number of Infected Wells** | **Number of Uninfected Wells** | **Total Number Infected** | **Total Number Uninfected** | **% Total Infected** |
| 5 | 1.00E-05 | 12 | =B10-B5 | =SUM(B5:B8) | =SUM(C5:C5) | =D5/(D5+E5) |
| 6 | 1.00E-06 | 8 | =B10-B6 | =SUM(B6:B8) | =SUM(C5:C6) | =D6/(D6+E6) |
| 7 | 1.00E-07 | 1 | =B10-B7 | =SUM(B7:B8) | =SUM(C5:C7) | =D7/(D7+E7) |
| 8 | 1.00E-08 | 0 | =B10-B8 | =SUM(B8:B8) | =SUM(C5:C8) | =D8/(D8+E8) |
| 9 | | | | | | |
| 10 | **Num Wells** | 12 | | | | |
| 11 | **mls/well** | 0.01 | | | | |
| 12 | | | | | | |
| 13 | **Prop.Dist** | =(H10-.05)/(H10-I10) | | | | |
| 14 | **Log TCID** | =J10-B13 | | | | |
| 15 | **TCID50** | =10^B14 | | | | |
| 16 | **1/TCID50** | =1/B15 | | | | |
| 17 | **TCID50/ml** | **=B16/B11** | | | | |
| 18 | | | | | | |
| 19 | **pfu/ml** | **=B17*0.69** | | | | |

| | G | H | I | J |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | **Above 50%** | **% Above 50%** | **% Below 50%** | **Log Dilution Above 50%** |
| 5 | =IF(F5>0.5,"TRUE","FALSE") | =IF(AND(G5="TRUE",G6="FALSE"),F5,0) | =IF(AND(G4="TRUE",G5="FALSE"),F5,0) | =IF(H5>0,LOG(A5),0) |
| 6 | =IF(F6>0.5,"TRUE","FALSE") | =IF(AND(G6="TRUE",G7="FALSE"),F6,0) | =IF(AND(G5="TRUE",G6="FALSE"),F6,0) | =IF(H6>0,LOG(A6),0) |
| 7 | =IF(F7>0.5,"TRUE","FALSE") | =IF(AND(G7="TRUE",G8="FALSE"),F7,0) | =IF(AND(G6="TRUE",G7="FALSE"),F7,0) | =IF(H7>0,LOG(A7),0) |
| 8 | =IF(F8>0.5,"TRUE","FALSE") | =IF(AND(G8="TRUE",G9="FALSE"),F8,0) | =IF(AND(G7="TRUE",G8="FALSE"),F8,0) | =IF(H8>0,LOG(A8),0) |
| 9 | | | | |
| 10 | | =SUM(H5:H8) | =SUM(I5:I8) | =SUM(J5:J8) |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |

**Figure 34. Template of the Excel spreadsheet for the calculation of the virus $TCID_{50}$.**

After the spinoculation, the cells were resuspended in 3 ml of RPMI and added to the plate in duplicate as shown in Figure 33. Plates were incubated in a $CO_2$ incubator at 37°C and developed on the 6th or 7th day by adding 40 ul MTT (7mg/ml) to each well and incubating at 37°C for 30 min to 2 hrs. One hundred microlitres of lysis solution was added to each well (0.4% HCl + 2% Trition X 100 up to 100 ml with isopropanol) and mixed by pipetting until all crystals had dissolved. Plates were read at 490 and 640 nm. The $TCID_{50}$ was determined as shown in Figure 34. All statistical treatments for calculation of the $IC_{50}$s for the isolates were performed by using the Analyze-it program (version 1.62) for Microsoft Excel statistics package and Sigmaplot software.

### 7.2.5 Performing the Phenotyping

Cells were infected with the virus to be phenotyped (using the virus' TC $ID_{50}$), in conjunction with a wildtype reference strain (pNL4-3) and a "mock infection" of cells without virus (Figure 35). The reference strain was included on each plate. The required volumes of filtered drugs were added and serially diluted (in duplicate), leaving the last row without the drug (just virus and cells) and an equal volume of cells was added to each of the wells. Plates were incubated in a $CO_2$ incubator at 37°C and developed on the 6th or 7th day in the same way as described above.

### 7.2.6 Calculation of drug susceptibility

Susceptibilities of chimeric virus to Amprenavir, Indinavir, Lopinavir, Ritonavir, Saquinavir, and Nelfinavir were determined by 3-(4,5-dimethylthiazol-2yl)-2,5-diphenyltetrazolium bromide (MTT)-based cell viability assays, each of which was performed in duplicate, as described previously (Hertogs et al, 1998). The percentage of viable cells obtained by the MTT assay after seven days of infection was plotted semi-logarithmically against the concentrations of the drugs tested. A Hill's three-parameter nonlinear regression was performed to obtain the sigmoid curve of viable cells and the $IC_{50}$ for each virus tested. The fold-change of resistance was calculated by comparing the isolate's $IC_{50}$ with the $IC_{50}$ for the drug susceptibility wild-type reference strain (Figure 36).

## 7.3 Results

Attempts to obtain a recombinant clone for the PR region of ZA012 and ZA024, as

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS |
| PBS | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | pNL4-3 | PBS |
| PBS | AZT | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | Drug free | PBS |
| PBS | AZT | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | Drug free | PBS |
| PBS | 3TC | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | Drug free | PBS |
| PBS | 3TC | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | 1 in 2 | Drug free | PBS |
| PBS | M | O | C | K | | I | N | F | E | C | PBS |
| PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS | PBS |

**Figure 35. Template of the plate layout for determining the virus drug susceptibility. The drugs were serially diluted in duplicate, with 2 drugs per plate. Each plate contained the reference strain pNL4-3.**

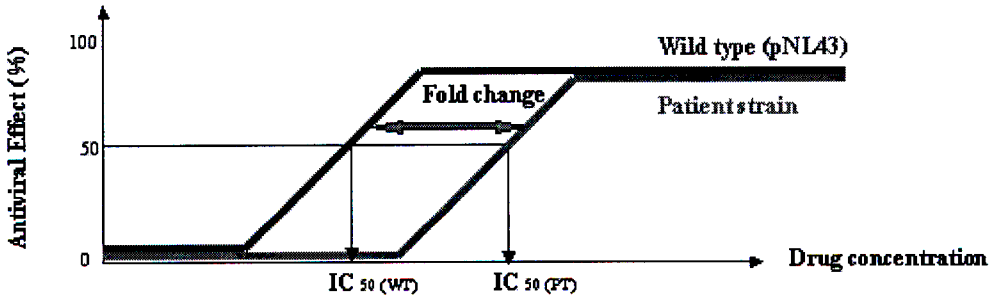



**Figure 36. Calculation of fold change. The fold change was calculated by dividing the patient strain $IC_{50}$ by the reference strain $IC_{50}$.**

well as the RT region for ZA010, were unsuccessful. As a result, the response of these isolates to the drugs targeting these regions could not be determined. Phenotypic resistance to one or more drugs was detected in samples ZA024, R555, PT01 and PT02 (Table 26). ZA024, a drug naive isolate harbouring the K103N mutation, showed a greater than 100 fold resistance to EFV, and a 16-fold resistance to NVP. The two treatment experienced isolates, PT01 and PT02, both showed resistance to the NNRTI drugs tested. In addition, PT02 had a >100-fold resistance to 3TC, while PT01 only showed a 20-fold resistance to 3TC when compared to pNL4.3, even though both patients had a similar clinical history. ZA023 did not show phenotypic resistance, possibly because the K103N mutation was present as a minority strain and was not detected by the phenotyping assay. Most of the isolates showed hypersusceptibility to Lopinavir, with the $IC_{50}$s for those isolates being up to 16.2 times lower than that for the reference strain HXB2/NL4-3-PR. There were no other observable differences in drug susceptibility to the commercially available PIs tested.

### 7.4 Discussion

HIV-1 sequences vary widely around the world, but mutations that potentially contribute to resistance to drug targeting the PR and RT of HIV-1 occur in both B and non-B subtypes. In spite of this, very little information is available on the drug susceptibility of non-B subtypes, particularly subtype C. Although HIV-1 subtype C does not appear to have any primary mutations in its signature PR and RT sequence, it does has several accessory mutations that are associated with resistance in the context of subtype B (Grossman et al, 2000; Vergne et al, 2000). Since accessory/compensatory mutations are often associated with the restoration of viral fitness in the presence of crippling primary mutations (Hirsch et al, 2000; Schinazi et al, 2000) the impact of the mutations seen in subtype C on the long-term efficacy of ARVs is not known.

In this chapter, samples from drug naive subtype C, B and D/C isolates and 2 drug experienced isolates (one subtype C and one subtype B), were phenotyped using the recombinant virus assay described by Hertogs et al, (1998). The drug naive subtype C isolates showed a similar response to the RT inhibitors as the wild type strain. However, the drug experienced subtype C isolate, which harboured the M184V

**Table 26. Drug susceptibility of the isolates to the available PR and RT inhibitors. The $IC_{50}$ of the reference NL43 strain is shown in red (below).**

| Samples | AZT | 3TC | ddI | ddC | d4T | ABV | NVP | EFV | IDV | RTV | NFV | SQV | APV | LPV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (4.0)[a] | (4.5) | (3.5) | (3.5) | (3.0) | (3.0) | (8.0) | (6.0) | (3.0) | (3.5) | (4.0) | (2.5) | (2.5) | |
| ZA012 | 0,5[b] | 0.29 | 0.5 | 0.2 | 0.7 | 0.8 | 0.4 | 0.8 | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. |
| ZA024 | 0.7 | 0.34 | 0.6 | 0.7 | 0.8 | 0.6 | 16 | **>100** | N.D. | N.D. | N.D. | N.D. | N.D. | N.D. |
| ZA023 | 0.6 | 0.45 | 0.8 | 0.4 | 0.8 | 0.4 | 0.2 | 0.5 | 1 | 0.5 | 0.5 | 0.8 | 0.9 | 0.05 |
| ZA021 | 0.5 | 1 | 0.6 | 1 | 0.6 | 0.8 | 0.7 | 3 | 0.5 | 0.7 | 0.8 | 0.6 | 0.8 | 0.17 |
| ZA028 | 0.6 | 1.5 | 1.5 | 2 | 1.2 | 1.3 | 3.2 | 1 | 1 | 0.7 | 0.7 | 0.6 | 0.68 | 0.1 |
| ZA032 | 1.5 | 3.5 | 1.7 | 1.2 | 0.7 | 0.9 | 0.6 | 0.5 | 0.3 | 0.5 | 0.7 | 0.5 | 0.3 | 0.18 |
| R555 | **>100** | **10** | 1.7 | 0.9 | 3.2 | 2 | 0.7 | 0.5 | 1.1 | 1.7 | 2 | 1 | 1.1 | 0.2 |
| PT01 | 2.5 | **20** | 1 | 1 | 0.5 | 0.9 | **>100** | **>100** | 1 | 0.9 | 2.2 | 1 | 0.9 | 0.18 |
| PT02 | 0.6 | **>100** | 1.1 | 1.2 | 0.7 | 1.2 | **>100** | **>100** | 1.3 | 1.2 | 0.5 | 0.8 | 0.8 | 0.28 |
| ZA010 | N.D. | N.D. | N.D. | N.D. | N.D. | ND | N.D. | N.D. | 1 | 0.5 | 0.3 | 0.4 | 1 | 0.05 |
| **NL43$IC_{50}$(nM)** | **10n M** | **200n M** | **10n M** | **53n M** | **66n M** | **5µM** | **100nM** | **2.5n M** | **12n M** | **47n M** | **62n M** | **12n M** | **100n M** | **32n M** |

**a- Biological cut-off generated by Harrigan et al. (2001) using similar phenotyping technology. The values are expressed in fold resistance compared to NL4-3. This paper has not determined the cut-off for Lopinavir.**

**b- Values are expressed in fold resistance compared to NL4-3. The values in bold represents samples with resistance level above cut-off.**

mutation, showed a greater fold-resistance when compared to the wild-type strain than the subtype B isolate with the same resistance mutation. However the significance of this is unknown, as the M184V mutation usually confers high-level (>100 fold) resistance to 3TC (Havlir et al, 2000; Descamps et al, 2000). As resistance mutations often occur in complex patterns, it is likely that the effect of the other mutations present in each isolate is influencing the response to 3TC.

The impact of the accessory mutations in PR on drug susceptibility was minimal, even with up to 3 accessory mutations in a single isolate. Unfortunately, the vector used in this assay did not incorporate the 3' end of *gag*, which included sites that have been linked to PI resistance, which could influence the susceptibility of the virus.

The assay is also limited in that a subtype B backbone was used as the vector for the recombination, and whether the responses will be the same with a subtype C backbone is yet to be determined. The assay did not detect resistance to the K103N mutation in ZA023, probably because the mutation was present as a minor population in that sample, which is another of the limitations of the assay.

The retrospective subtype B isolate was resistant to both AZT and 3TC *in vitro*. Although this patient was said to be drug naive, it is very possible that the patient had received treatment. At that time (around 1991), the most likely treatment would have been dual therapy with AZT and 3TC. The public health importance of this finding is that drug resistant HIV-1 was present in South Africa more than a decade ago. Although our studies have shown minimal resistance in the general drug naive population in KwaZulu-Natal, this result suggests that the population in the Cape had greater access to treatment in the past than patients in KwaZulu-Natal and could consequently result in the transmission of resistant strains. This emphasises the urgent need for screening for resistance in drug naive patients in this region.

Interestingly, the subtype C isolates showed a hyper-susceptibility (a lower $IC_{50}$) to Lopinavir which was not seen in the wild-type strain. Further analysis by Gonzales et al (2003) using site-directed mutagenesis showed that this hyper-susceptibility was linked to the presence of the I93L mutation found in most subtype C isolates. Although other studies have also reported no difference in the IC50s for subtype B and subtype C strains when the other PIs were tested, they did not included Lopinavir among the antiretroviral compounds tested in their phenotyping assays (Harrigan et al, 2001; Velazquez-Campoy et al, 2001). It has been reported that hyper-susceptibility can influence the clinical outcome in patients on Amprenavir treatment, whose

isolates carry the N88S mutation (Velazquez-Campoy et al, 2001). Gonzales et al (2003) suggest that the *in vitro* phenotypic hyper-susceptibility of subtype C isolates with the I93L mutation could result in different clinical outcomes, but caution that these findings need to be further tested in controlled clinical trials comparing the virological responses of patients infected with subtype B HIV isolates and those infected with subtype C HIV isolates receiving Lopinavir.

# Chapter 8

## General Discussion

Since subtype C began its spread across southern Africa in the late 1980s, major outbreaks have now occurred in every country in this region. Zimbabwe, Botswana and South Africa have been most severely hit by the outbreak of HIV-1 subtype C infections. Despite this, few studies have examined the genetic diversity and molecular phylogeny of subtype C viruses in South Africa, and particularly in KwaZulu-Natal where the HIV-1 prevalence is highest (Rollins et al, 2002). One of the main objectives of this thesis was therefore to define the spectrum of naturally occurring resistance-associated polymorphisms and mutations in RT and protease of subtype C viruses in KwaZulu-Natal, as South Africa begins its national ARV roll-out.

With the exception of two primary resistance mutations, K103N and G190A, which occurred in a single husband-wife pair, no other major resistance-associated mutations to the available PIs and RTIs were found in a cohort of 72 drug naive isolates from KwaZulu-Natal. However, similar to other studies from Zimbabwe and Botswana, signature subtype C polymorphisms were found in these isolates, particularly in the PR region, that have been linked to increased catalytic activity in subtypes A and C (Kantor et al, 2002; Velazquez-Campoy et al, 2001). A few of these isolates were phenotyped and were found to be as susceptible as wild-type subtype B to the available PIs and RTIs (similar to a study by Harrigan et al, 2001). Interestingly, the isolates were also found to be hypersusceptible to LPV, possibly linked to the I93L signature subtype C mutation in the PR (Gonzalez et al, 2003). To better understand the mechanisms of resistance and viral pathogenesis, sites under positive selection were identified using the methods proposed by Yang (2000). Most of the codons in the PR and RT genes (>95%) were under strong purifying (negative) selection pressure, particularly at sites related to viral structure and function. The remaining 5% of amino acids were under strong diversifying (positive) selection pressure. Seven of these amino acids (PR 12S and 19I; RT 36A, 39E, 123G, 211K, and 245Q) were present in both the KwaZulu-Natal and subtype C consensus sequence but not in the consensus sequences of subtypes A, B, and D, suggesting that these signature residues may offer a subtype-specific fitness advantage to C viruses. Soares et al (2003a) have shown an association between ARV treatment and the D/G123S mutation in subtype C isolates from Brazil. The R211K signature mutation in the RT was absent from retrospective subtype C isolates from the Western Cape and KwaZulu-Natal. R211K has been associated with resistance in the RT in subtype B isolates suggesting that the evolution from an Arginine to a Lysine at that position could confer over time a fitness advantage in the presence of drug selection pressure. The V60I mutation, which occurs as a signature in our isolates (although also prevalent in other subtype C viruses), was not under positive selection, and did not occur at a site related to function. Whether this

mutation plays a role in viral fitness has yet to be determined. Naturally occurring polymorphisms also resulted in significant variation in the number and type of phosphorylation sites in both the *pol* and *env* genes. Of note was the loss of a glycosylation site at the beginning of the V3 loop in *env* which was associated with the presence of a serine PKC phosphorylation site at position 11-13 in the V3. This suggests a potential linkage between deglycosylation and phosphorylation in the V3 loop of C viruses related to tropism in C viruses.

Some of these isolates were phenotyped using the RVA, in collaboration with our colleagues in Brazil, and showed similar responses to the RT inhibitors as the wild type subtype B strain. Interestingly, the subtype C isolates showed a hyper-susceptibility (a lower $IC_{50}$) to Lopinavir which was not seen in the wild-type strain. Further analysis by Gonzales et al (2003) using site-directed mutagenesis showed that this hyper-susceptibility was linked to the presence of the I93L mutation found in most subtype C isolates. This could influence the clinical outcome of patients infected with subtype C and treated with Lopinavir, but requires further investigation. Also, a subtype B backbone was used as the vector for the recombination, and whether the responses will be the same with a subtype C backbone is yet to be determined.

Phylogenetic analysis of the drug naive isolates showed that the epidemic in KwaZulu Natal is characterized by multiple circulating HIV-1 subtype C sublineages in both the Indian and Black communities, characteristic of multiple introductions of subtype C into South Africa from across its many borders. The presence of HIV-1 isolates collected prior to 1992 at internal branches of the phylogenetic tree suggests that these sequences may represent founder strains that have existed in KwaZulu-Natal for at least 10 years. The increased prevalence of multiple subtypes, including subtype C, in the Western Cape highlights the change in diversity and distribution of HIV-1 subtypes in South Africa since the beginning of the epidemic. Archival material available from the Western Cape may serve as an important source of information on the evolutionary history of HIV-1 in South Africa and is being investigated further by colleagues from the Western Cape.

The effect of signature subtype C polymorphisms in protease on the 12 protease cleavage sites showed that seven sites (those with structural or enzymatic functions) are highly conserved in all subtypes and that they are under strong purifying selection pressure (De Oliveira et al, 2003; Ericson-Viitanen et al, 1989; Kaplan et al, 1994; Krausslich et al, 1989; Pettit et al, 1994). The remaining five cleavage sites exhibited extensive variability across all subtypes, with C viruses being significantly more variable than subtype B at all sites except TFP/p6*pol*, which is responsible for protease activation. p2/NC was the most variable cleavage site in subtype C (42.4%). This site controls the rate of Gag and Gag-Pol processing. The variation in cleavage sites between subtypes suggests that there may be

important differences in the way that B and C viruses regulate polyprocessing and virion assembly (De Oliveira et al., 2004; Petit et al, 2003; Petit et al, 2002). These differences could have a major impact on the pathogenesis of HIV-1 subtypes and on response to therapy.

A second objective was to determine the impact of naturally-occurring polymorphisms in subtype C viruses on response to ARV therapy in an African setting (i.e. using PI-exclusive regimens). Overall the patients responded well to treatment with only 10-18% of patients on HAART-based regimens developing drug resistance. Generally, the resistance mutations seen in our isolates were consistent with those seen in treatment experienced subtype B isolates. Of note was the high level of resistance to the entire class of NNRTIs seen in all the cohorts. This could be reflective of the predominant use of NNRTI-based regimens, as well as the low genetic barrier in this class of drugs. Other possible explanations for the high rate of NNRTI resistance were non-adherence in patients due to stigma or lack of funds, as well as the use of sub-optimal regimens in patients treated in the private sector prior to release of the national guidelines for ARV therapy. Comparison of resistance mutations in the PBMCs and plasma of patients presenting with KS showed that resistance could evolve separately at both sites, but this was not consistent, and was probably influenced by host genetic and immunologic factors. Resistance was high (40%) in matched subtype C infected mother and infant pairs, 6 weeks after receiving a single dose of NVP to prevent mother-to-child-transmission. The pattern of resistance differed, with K103N most common in the mothers and Y181C most common in the infants, consistent with reports from Eshleman et al (2003b) and Kantor et al (2003a). Of note were the changes in functional properties caused by these mutations, by the introduction or alteration of putative myristoylation and phosphorylation sites in the RT. The exact role of these alterations need to be further investigated.

Screening for positively selected sites could be useful for identifying novel drug resistance-associated or compensatory mutations. Novel mutations that occurred in one or more patients after therapy and were positively selected included: RT 21, 102, 123, 135, 162, 196, 202, 203, 207, 232, 245, 272, 274, 275, 278, 281, 292, 311 and 334. Some of these have recently been reported as treatment associated mutations, namely RT 135, 202 and 207, and could be acting as compensatory mutations (Kantor et al, 2002).

Resistance mutations are often accompanied by structural changes that either change the conformation of the RT binding pocket or stabilize the 3D structure of the enzyme (Sturmer et al, 2003; Gonzales et al, 2001; Wrobel et al, 1998). Our results on the structural interactions of novel, positively selected mutations have shown a definite relationship between structural changes at those sites and the development of resistance. However, these interactions are complex and require more in-depth analysis.

These data not only increase the information on the response of C viruses to treatment, but also increases the information on HIV treatment in the context of co-infections with other pathogens. TB and other opportunistic infections are a problem in resource-poor settings and the concomitant use of HAART and TB treatment could have great public health benefits. However, the rapid development of resistance to EFV with the V106M mutation in TB patients who fail therapy suggests that the use of this drug needs to be closely monitored. Therapeutic drug monitoring is one of the methods that can be used to ensure that sufficient drug levels are achieved in each patient. The high percentage of multi-NNRTI resistance in the surveillance study has serious implications for the treatment management of patients in resource poor settings, particularly because of the use of NNRTI-base drug regimens in these regions.

Recent reports on the prevalence of resistance mutations in mothers and infants after just a single dose of NVP have increased reluctance on the part of the government to use it in its pMTCT program. Although the K103N and Y181C mutations associated with NVP resistance have been reported to disappear from the plasma over time (Eshleman et al, 2001; Eshleman et al, 2003; Kantor et al, 2003), there is the possibility, however, that they are archived as proviral DNA and could conceivably compromise future treatment efficacy. Given the high prevalence of HIV-1 infection in South African pregnant women, the potential risk of developing multi-NNRTI resistance in these patients cannot be taken lightly (Rollins et al, 2002). Therefore, while NVP is cheap and effective in preventing MTCT, efforts should increase to find a safer alternative. Mothers going onto ARV therapy after receiving NVP in the past should be closely monitored for resistance. Perhaps these patients should go straight on to a PI inclusive regimen, which is also inherently more potent. Kaletra (LPV/r) is currently being used as part of the second-line drug regimen in patients failing therapy, and although subtype C isolates have shown hypersensitivity to LPV *in vitro,* its efficacy in salvage therapy with a compromised regimen has yet to be determined (Gonzalez et al, 2003). The rapid development of high level NNRTI resistance could compromise the effectiveness of the national drug program unless potent salvage therapies are prescribed, including more potent NRTIs such as TDF.

In South Africa, resistance testing is not provided as "standard-of-care" for patient treatment management as it is in North America and Europe. Although expensive, resistance testing provides extensive insight into the presence of drug resistant variants in the population of viruses within an individual and prevents the use of ineffective drugs in a drug regimen (Van Laethem et al 1999; Shafer et al 2000). Prospective controlled studies

have shown that patients and physicians who have access to genotypic resistance data respond better than those who do not have access to these assays (Durant et al, 1999; Baxter et al, 2000; Cohen et al, 2000). As a result, expert panels have recommended the use of resistance testing in the treatment of HIV-1 infected patients (US Dept Health; Hirsch et al, 2003). Phenotypic resistance tests are a more direct measurement of HIV-1 susceptibility, as the effect of multiple mutations can be taken into account, although it is much more expensive and time-consuming. Also, RVA phenotypic tests use a subtype B backbone as the vector for the recombination, which may not give a true reflection of susceptibility of non-B subtypes to the currently available ARVs. Genotypic tests are more commonly used in clinical settings because of their wider availability, lower cost and quicker turnaround. The Viroseq resistance genotyping kit has been successfully used in both research and diagnostic settings. It has proven to be reliable for use with non-B subtypes and is a rapid convenient method for determining resistance genotypes (Eshleman et al, 2004). Unfortunately, the presence complex patterns of resistance mutations in an isolate may make the genotype difficult to interpret. A limitation of both the genotypic and phenotypic tests is that they are unable to detect minor drug resistant populations. This is why it is not recommended to perform resistance testing on patients who are not currently taking their ARVs.

In the future, the impact on of positively selected phosphorylation sites on viral replicative capacity and the role of variable subtype C cleavage sites on response to treatment with PIs needs to be investigated, both at the patient level, and in vitro using site-directed mutagenesis. The role of novel, positively selected resistance mutations in the development of resistance, whether conferring resistance on their own, or in combination with mutations at other sites, needs further investigation.

**Conclusion**

In conclusion, these initial results suggest that local circulating subtype C strains will be as susceptible to ARV therapy as subtype B isolates from Europe and North America. However, the long-term efficacy of HAART on subtype C has yet to be determined. Changes in phosphorylation sites and the development of novel, positively selected mutations under drug selection pressure, suggests that subtype C can rapidly compensate for any loss of fitness caused by drug resistance mutations.

## References


1. Abdool Karim Q, Abdool Karim SS, Singh B, et al. Seroprevalence of HIV infection in rural South Africa. *AIDS* 1992; **6:**1535-1539.

2. Abebe A, Demissie D, Goudsmit J, et al. HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS* 1999; **13:**1305–1311.

3. Abebe A, Lukashov VV, Pollakis G, et al. Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. *AIDS* 2001; **15:**1555-1561.

4. Akaike, H. A new look at statistical model identification. *IEEE Trans Automatic control.* 1997; **19:**716-723.

5. Albrecht H. The value of resistance testing. *AIDS Clin.Care* 2002; **14:**46.

6. Albrecht MA, Bosch RJ, Hammer SM, et al. Nelfinavir, efavirenz, or both after the failure of nucleoside treatment of HIV infection. *N Engl J Med* 2001; **345:**398-407.

7. Altschul S, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Res* 1997; **25:**3389-3402.

8. Anastassopoulou C, Paraskevis D, Sypsa VA, et al. Genetic Evolution of Human Immunodeficiency Virus Type 1 in Two Spouses Responding Successfully to Highly Active Antiretroviral Therapy. *AIDS Res and Hum Retroviruses* 2003; **19:**65–71.

9. Anisimova M, Bielawski JP, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* 2001; **18:**1585–1592.

10. Archer RH, Dykes C, Gerondelis P, et al. Mutants of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase resistant to nonnucleoside reverse transcriptase inhibitors demonstrate altered rates of RNase H cleavage that correlate with HIV-1 replication fitness in cell culture. *J Virol* 2000; **74:**8390-8401.

11. Arion D, Sluis-Cremer N, Parniak MA. Mechanism by which phosphonoformic acid resistance mutations restore 3'-azido-3'-deoxythymidine (AZT) sensitivity to AZT-resistant HIV-1 reverse transcriptase. *J Biol Chem* 2000; **275:**9251-9255.

12. Arion D, Kaushik N, McCormick S, et al. Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochem* 1998; **37:**15908-15917.

13. Arion D, Borkow G, Gu Z, et al. The K65R mutation confers increased DNA polymerase processivity to HIV-1 reverse transcriptase. *J Biol Chem* 1996; **271:**19860-19864.

14. Arts EJ, Ball SC, Qunones-Mateu ME, et al. Competition assays clearly indicates that subtype C HIV-1 isolates are less fit than isolates of other subtypes. *1st IAS Conference on HIV Pathogenesis and Treatment*, Buenos Aires, Argentina, 2001.

15. Auerbuck D, Shapiro MJ, Engelhard D, et al. Resistance mutations in children infected with Non-Subtype B following HAART [abstract TuPeB4599]. *XIV International AIDS Conference*, Barcelona, Spain, 2002.

16. Bacheler L, Jeffrey S, Hanna G, et al. Genotypic correlates of phenotypic resistance to efavirenz in virus isolates from patients failing nonnucleoside reverse transcriptase inhibitor therapy. *J Virol* 2001; **75:**4999-5008.

17. Bacheler LT, Anton ED, Kudish P, et al. Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. *Antimicrob.Agents Chemother* 2000; **44:**2475-2484.

18. Back NK, Berkhout B. Limiting deoxynucleoside triphosphate concentrations emphasize the processivity defect of lamivudine-resistant variants of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* 1997; **41:**2484-2491.

19. Back NK, Nijhuis M, Keulen W, et al. Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme. *Embo J* 1996; **15:**4040-4049.

20. Baldwin ET, Bhat TN, Gulnik S, et al. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* 1995; **3:**581-590.

21. Ball SC, Abraha A, Collins KR, et al. Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J Virol* 2003; **77:**1021-1038.

22. Balzarini J, Pelemans H, Esnouf R, et al. A novel mutation (F227L) arises in the reverse transcriptase of human immunodeficiency virus type 1 on dose-escalating treatment of HIV type 1-infected cell cultures with the nonnucleoside reverse transcriptase inhibitor thiocarboxanilide UC-781. *AIDS Res Hum Retroviruses* 1998; **14:**255-260.

23. Barre-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983; **220:**868-871.

24. Bartlett JA, DeMasi R, Quinn J, et al. Overview of the effectiveness of triple combination therapy in antiretroviral-naive HIV-1 infected adults. *AIDS* 2001; **15:**1369-1377.

25. Baskar PV, Ray SC, Rao R, et al. Presence in India of HIV type 1 similar to North American strains. *AIDS Res Hum Retroviruses* 1994; **10:**1039-1041.

26. Becker-Pergola G, Mellquist JL, Guay L, et al. Identification of diverse HIV type 1 subtypes and dual HIV type 1 infection in pregnant Ugandan women. *AIDS Res Hum Retroviruses* 2000; **16:**1099–1104.

27. Berkowitz RD, Alexander S, Bare C, et al. CCR5- and CXCR4-utilizing strains of human immunodeficiency virus type 1 exhibit differential tropism and pathogenesis in vivo. *J Virol* 1998; **72:**10108-10117.

28. Berkowitz RD, van't Wout AB, Kootstra NA, et al. R5 strains of human immunodeficiency virus type 1 from rapid progressors lacking X4 strains do not possess X4-type pathogenicity in human thymus. *J Virol* 1999; **73:**7817-7822.

29. Beyrer C, Razak MH, Lisam K et al. Overland heroin trafficking routes and HIV-1 spread in south and south-east Asia. *AIDS* 2000; **14:**75-83.

30. Billich, S., M. T. Knoop, J. Hansen, P. Strop, J. Sedlacek, R. Mertz, and K. Moelling. 1988. Synthetic peptides as substrates and inhibitors of human immunodeficiency virus-1 protease. *J Biol Chem*; **263:**17905–17908.

31. Bjorndal A, Sonnerborg A, Tscherning C, et al. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses* 1999; **15:**647–653.

32. Bjorndal A, Deng H, Jansson M, et al. Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J Virol* 1997; **71:**7478–7487.

33. Blankson J, Persaud D, Siliciano RF. The challenge of viral reservoirs in HIV infection. *Annu Rev Med* 2002; **53:**557-593.

34. Bleiber G, Munoz M, Ciuffi A, et al. Individual contributions of mutant protease and reverse transcriptase to viral infectivity, replication, and protein maturation of antiretroviral drug-resistant human immunodeficiency virus type 1. *J Virol* 2001;**75:**3291-300.

35. Bloor S, Kempf K, Hertogs K, Alcorn T, Larder B. Patterns of HIV drug resistance in routine clinical practice: survey of almost 12000 samples from the USA in 1999. *Antivir Ther* 2000; **5:**169.

36. Boden D, et al. HIV-1 drug resistance in newly infected individuals. *J Am Med Assoc* **282:**1135-1141.

37. Bollyky PL, Wester W, Peter T, et al. Genotypic variation in HIV-1 reverse transcriptase and protease in subtype C samples from Botswana in treated and untreated patients [abstract TuPeC4825]. *XIV International AIDS Conference*, Barcelona, Spain, 2002.

38. Bongertz V, Bou-Habib DC, Brigido LF, et al. HIV-1 diversity in Brazil: genetic, biologic, and immunologic characterization of HIV-1 strains in three potential HIV vaccine evaluation sites. Brazilian Network for HIV Isolation and Characterization. *J Acquir Immune Defic Syndr* 2000; **23:**184-193.

39. Bossi P, Legrand O, Faussat AM, et al. P-glycoprotein in blood CD4 cells of HIV-1-infected patients treated with protease inhibitors. *HIV Med* 2003; **4:**67-71.

40. Bossi P, Mouroux M, Yvon A, et al. Polymorphism of the human immunodeficiency virus type 1 (HIV-1) protease gene and response of HIV-1-infected patients to a protease inhibitor. *J Clin Microbiol* 1999; **37:**2910-2912.

41. Boucher CA, Keulen W, van Bommel T, et al. Human Immunodeficiency virus type 1 drug susceptibility determination by using recombinant viruses generated from patient sera tested in a cell-killing assay. *Antimicrob Agents Chemother* 1996; **40:**2404-2409.

42. Boucher CA, Cammack N, Schipper P, et al. High-level resistance to (-) enantiomeric 2'-deoxy-3'-thiacytidine in vitro is due to one amino acid substitution in the catalytic site of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* 1993; **37:**2231-2234.

43. Boyer PL, Sarafianos SG, Arnold E, Hughes SH. Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. *J Virol* 2001; **75:**4832-4842

44. Boyer PL, Hughes SH. Analysis of mutations at position 184 in reverse transcriptase of human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 1995; **39:**1624-1628.

45. Bredell H, Hunt G, Casteling A, et al. HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa. *AIDS Res Hum Retroviruses* 2002; **18:**681-683.

46. Bredell H, Williamson C, Sonnenberg P, et al. Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Res Hum Retroviruses* 1998; **14:**677-684.

47. Brennan CA, Lund JK, Golden A, et al. Serologic and phylogenetic characterization of HIV-1 subtypes in Uganda. *AIDS* 1997; **11:**1823-1832.

48. Brenner B, Turner D, Oliveira M, et al. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *AIDS* 2003; **17:**F1-5.

49. Briant L, Wade CM, Puel J, et al. Analysis of envelope sequence variants suggests multiple mechanisms of mother-to child transmission of human immunodeficiency virus type 1. *J Virol* 1995; **69:**3778–3788.

50. Brindeiro RM, Diaz RS, Sabino EC, et al. Brazilian Network for HIV Drug Resistance Surveillance (HIV-BResNet): a survey of chronically infected individuals. *AIDS* 2003; **17:**1063-1069.

51. Brindeiro PA, Brindeiro RM, Mortensen C, et al. Testing genotypic and phenotypic resistance in human immunodeficiency virus type 1 isolates of clade B and other clades from children failing antiretroviral therapy. *J Clin Microbiol* 2002; **40:**4512-4519.

52. Brindeiro R, Vanderborght B, Caride E, et al. Sequence diversity of the reverse transcriptase of human immunodeficiency virus type 1 from untreated Brazilian individuals. *Antimicrob Agents Chemother* 1999; **43:**1674-1680.

53. Brown AJ, Precious HM, Whitcomb J, et al. Reduced susceptibility of HIV-1 to protease inhibitors from patients with primary HIV infection by three distinct routes [abstract 424]; *8th Conference on Retroviruses and Opportunistic Infections*, Chicago, 2001.

54. Brun-Vezinet F, Ingrand D, Deforges L, et al. HIV-1 sensitivity to zidovudine: a consensus culture technique validated by genotypic analysis of the reverse transcriptase. *J Virol Methods* 1992; **37:**177-188.

55. Byrnes VW, Emini EA, Schleif WA, et al. Susceptibilities of human immunodeficiency virus type 1 enzyme and viral variants expressing multiple resistance-engendering amino acid substitutions to reserve transcriptase inhibitors. *Antimicrob Agents Chemother* 1994; **38:**1404-1407.

56. Byrnes VW, Sardana VV, Schleif WA, et al. Comprehensive mutant enzyme and viral variant assessment of human immunodeficiency virus type 1 reverse transcriptase resistance to nonnucleoside inhibitors. *Antimicrob Agents Chemother* 1993; **37:**1576-1579.

57. Caliendo AM, Savara A, An D, et al. Effects of zidovudine-selected human immunodeficiency virus type 1 reverse transcriptase amino acid substitutions on processive DNA synthesis and viral replication. *J Virol* 1996; **70:**2146-2153.

58. Camacho R, Deforche K, Valadas ME, et al. Nelfinavir resistance in HIV-1 subtype B and G infected patients: evidence for different pathways and novel mutations associated with failure of nelfinavir based regimens. *Antivir Ther* 2004; **9:**S114.

59. Cane PA, Osman H, Smit E. Effect of HIV-1 subtype on development of NNRTI resistance mutations in patients failing first line therapy *Antivir Ther* 2004; **9:**S115.

60. Caride E, Brindeiro R, Hertogs K, et al. Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. *Virol* 2000; **275:**107-115.

61. Casado JL, Hertogs K, Ruiz L, et al. Non-nucleoside reverse transcriptase inhibitor resistance among patients failing a nevirapine plus protease inhibitor-containing regimen. *AIDS* 2000; **14:**F1-7.

62. Cassol S, Gill MJ, Pilon R, et al. Quantification of HIV-1 RNA from dried plasma spots collected on filter paper. *J Clin Microbiol* 1997; **35:**2795-2801.

63. Cecilia D, KewalRamani VN, O'Leary J, et al. Neutralization profiles of primary human immunodeficiency virus type 1 isolates in the context of coreceptor usage. *J Virol* 1998; **72:**6988-6996.

64. Cecilia D, Kulkarni SS, Tripathy SP, et al. Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. *Virol* 2000; **271:**253–258.

65. Chakrabarti S, Panda S, Chatterjee A, et al. HIV-1 subtypes in injecting drug users & their non-injecting wives in Manipur, India. *Indian J Med Res* 2000; **111:**189-194.

66. Chan DC, Chutkowski CT, Kim PS. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc Natl Acad Sci U S A* 1998; **95:**15613-15617.

67. Chen Z, Li Y, Schock HB, et al. Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. *J Biol Chem* 1995; **270:**21433-21436.

68. Chow YK, Hirsch MS, Kaplan Jcet al. HIV-1 error revealed. *Nature* 1993; **364:**679.

69. Chun, TW, Engel D, Berrey MM, Shea T, Corey L, Fauci AS. Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection. *Proc Natl Acad Sci USA* 1998; **95:**8869–8873.

70. Cilliers T, Nhlapo J, Coetzer M, et al. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. *J Virol* 2003; **77:**4449-4456.

71. Cilliers T, Patience T, Pillat C, et al. Sensitivity of HIV type 1 subtype C isolates to the entry inhibitor T-20. *AIDS Res Hum Retroviruses* 2004; **205:**477-82.

72. Clavel F, Race E, Mammano F. HIV drug resistance and viral fitness. *Adv Pharmacol* 2000; **49:**41–66.

73. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 1995; **267:**483-489.

74. Cohen DA, Scriber R. An STD/HIV prevention intervention framework. *AIDS Patient care STDS* 2000; **14:**37-45.

75. Collins J, Thompson MG, Paintsil E, et al. Competitive fitness of nevirapine resistant human immunodeficiency virus type 1 mutants. *J Virol* 2004; **78:**603-611.

76. Colonno RJ, Friborg J, Rose RE, et al. Identification of amino acid substitutions correlated with reduced atazanavir susceptibility in patients treated with atazanavir containing regimens. *Antivir Ther* 2002; **7:**S6.

77. Colonno RJ, Hertogs K, Larder B, Limoli K, Heilek-Snyder G, Parkin N. BMS-232632 sensitivity of a panel of HIV-1 clinical isolates resistant to one or more approved protease inhibitors. *Antivir Ther* 2000; **5:**7.

78. Condra JH, Holder DJ, Schleif WA, et al. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. *J Virol* 1996; **70:**8270-8276.

79. Condra JH, Petropoulos CJ, Ziermann R, Schleif WA, Shivaprakash M, Emini EA. Drug resistance and predicted virologic responses to human immunodeficiency virus type 1 protease inhibitor therapy. *J Infect.Dis* 2000; **182:**758-765.

80. Condra JH, Schleif WA, Blahy OM, et al. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature* 1995; **374:**569-571.

81. Conway B, Wainberg MA, Hall D, et al. Development of drug resistance in patients receiving combinations of zidovudine, didanosine and nevirapine. *AIDS* 2001; **15:**1269-1274.

82. Cornelissen M, van den Burg R, Zorgdrager Fet al. pol gene diversity of five human immunodeficiency virus type 1 subtypes: evidence for naturally occurring mutations that contribute to drug resistance, limited recombination patterns, and common ancestry for subtypes B and D. *J Virol* 1997; **71:**6348-6358.

83. Cote HC, Brumme ZL, Harrigan PR. Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir. *J Virol* 2001; **75:**589-594.

84. Couto-Fernandez JC, Morgado MG, Bongertz V, et al. HIV-1 subtyping in Salvador, Bahia, Brazil: a city with African sociodemographic characteristics. *J Acquir Immune Defic Syndr* 1999; **22:**288-293.

85. Craig C, Race E, Sheldon J, et al. HIV protease genotype and viral sensitivity to HIV protease inhibitors following saquinavir therapy. *AIDS* 1998; **12:**1611-1618.

86. Croteau G, Doyon L, Thibeault D, et al. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol* 1997; **71:**1089-1096.

87. Courcambeck J. Molecular modelling appracah to explain the susceptibility of PMPA for drug resistant HIV-1 reverse transcriptase K65R, Q151M and M184V. *Antivir Ther* 2000; **7:**S26.

88. Csillag C. HIV-1 subtype C in Brazil. *Lancet*1994; **344:**1354.

89. de Jong JJ, Goudsmit J, Lukashov VV, et al. Insertion of two amino acids combined with changes in reverse transcriptase containing tyrosine-215 of HIV-1 resistant to multiple nucleoside analogs. *AIDS* 1999; **13:**75-80.

90. De Oliveira T, Engelbrecht S, Janse van Rensburg E, et al. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol* 2003; **77:**9422-9430.

91. De Oliveira T, Salemi M, Gordon M, et al. Mapping Sites of Positive Selection and Amino Acid Diversification in the HIV Genome: An Alternative Approach to Vaccine Design? *Genetics* 2004; **167:**1047-1058.

92. De Oliveira T, Bishop K, Danaviah S, et al. Changing dynamics of HIV-1 subtype diversification in Africa. *8th HIV Dynamics and Evolution Meeting* Paris, 2001.

93. De Oliveira T, Miller R, Tarin M, et al. An integrated Genetic Data Environment (GDE)-Based LINUX interface for the Analysis of HIV-1 and other microbial sequences. *Bioinformatics* 2003; **19:**153-154.

94. de Ronde A, van Dooren M, van Der Hoek L, et al. Establishment of new transmissible and drug-sensitive human immunodeficiency virus type 1 wild types due to transmission of nucleoside analogue-resistant virus. *J Virol* 2001; **75:**595-602.

95. Dean GL, Edwards SG, Ives NJ, et al. Treatment of tuberculosis in HIV-infected persons in the era of highly active antiretroviral therapy. *AIDS* 2002; **16:**75-83.

96. Deeks SG. International perspectives on antiretroviral resistance. Nonnucleoside reverse transcriptase inhibitor resistance. *J Acquir Immune Defic Syndr* 2001; **26:**S25-33.

97. Demeter LM, Shafer RW, Meehan PM, et al. Delavirdine susceptibilities and associated reverse transcriptase mutations in human immunodeficiency virus type 1 isolates from patients in a phase I/II trial of delavirdine monotherapy (ACTG 260). *Antimicrob Agents Chemother* 2000; **44:**794-797.

98. Department of Health/Directorate Health systems Research: 7th National HIV survey of women attending antenatal clinics of the public health services in South Africa. October/November 2000. Pretoria: Directorate Health Systems Research, Department of Health, 2001.

99. Descamps D, Flandre P, Calvez V, Peytavin G, Meiffredy V, Collin G, Delaugerre C, Robert-Delmas S, Bazin B, Aboulker JP, Pialoux G, Raffi F, Brun-Vezinet F. Mechanisms of virologic failure in

previously untreated HIV-infected patients from a trial of induction-maintenance therapy. Trilege (Agence Nationale de Recherches sur le SIDA 072) Study Team. *JAMA* 2000; **283:**205-211.

100. Descamps D, Calvez V, Collin G, et al. Line probe assay for detection of human immunodeficiency virus type 1 mutations conferring resistance to nucleoside inhibitors of reverse transcriptase: comparison with sequence analysis. *J Clin Microbiol* 1998; **36:**2143-2145.

101. Deval J, Selmi B, Boretto J, et al. The molecular mechanism of multidrug resistance by the Q151M human immunodeficiency virus type 1 reverse transcriptase and its suppression using alpha-boranophosphate nucleotide analogues. *J Biol Chem* 2002; **277:**42097-42104

102. Devereux HL, Emery VC, Johnson MA, et al. Replicative fitness in vivo of HIV-1 variants with multiple drug resistance-associated mutations. *J Med Virol* 2001; **65:**218-224.

103. Dias Tavares M, Brandao Varella R, Baia Fereira S, et al. Large Scale use of antiretroviral drugs is not associated with a high prevalence of genotypic resistance mutations in treatment-naive HIV-1 infected individuals in Rio de Janeiro, Brazil [abstract 778]. *2nd IAS Conference on HIV Pathogenesis and Treatment*, Paris, 2003.

104. Dickover RE, Garratty EM, Plaeger S, et al. Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. *J Virol* 2001; **75:**2194–2203.

105. Domingo E, Escarmis C, Menendez-Arias L, Holland J. Viral quasispecies and fitness variations. In: Domingo E, Webster R, Holland J, editors. Origin and evolution of viruses. San Diego: Academic Press. 1999 pp:141–161.

106. Domingo E, and Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 1997; **51:**151–178.

107. Downing R, Pieniazek D, Hu DJ, et al. Genetic characterization and phylogenetic analysis of HIV-1 subtype C from Uganda. *AIDS Res Hum Retroviruses* 2000; **16:**815-819.

108. Doyon L, Croteau G, Thibeault D, et al. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol* 1996; **70:**3763-3769.

109. Dronda F, Casado JL, Moreno S, et al. Phenotypic cross-resistance to nelfinavir: the role of prior antiretroviral therapy and the number of mutations in the protease gene. *AIDS Res Hum Retroviruses* 2001; **17:**211-215.

110. Drosopoulos WC, Prasad VR. Increased misincorporation fidelity observed for nucleoside analog resistance mutations M184V and E89G in human immunodeficiency virus type 1 reverse transcriptase does not correlate with the overall error rate measured in vitro. *J Virol* 1998; **72:**4224-4230.

111. Duliege AM, Amos CI, Felton S, et al. Birth order, delivery route, and concordance in the transmission of human immunodeficiency virus type 1 from mothers to twins. International Registry of HIV-Exposed Twins. *Pediatr* 1995; **126:**625-632.

112. Dumans AT, Soares MA, Machado ES, et al. Synonymous genetic polymorphisms within Brazilian human immunodeficiency virus Type 1 subtypes may influence mutational routes to drug resistance. *J Infect Dis* 2004; **189:**1232-1238.

113. Dumans AT, Soares MA, Pieniazek D, et al. Prevalence of protease and reverse transcriptase drug resistance mutations over time in drug-naïve human immunodeficiency virus type 1-positive individuals in Rio de Janeiro, Brazil. *Antimicrob Agents Chmother* 2002; **46:**3075-3079.

114. Dunn, D. S., B. Hurtel, C. Beyer, et al. 1997. Protection of SIV mac-infeced macaque monkeys against superinfection by a simian immunodeficiency virus expressing envelope glycoproteins of HIV type 1. *AIDS Res Hum Retroviruses* **13:**913-922.

115. Eastman PS, Boyer E, Mole Let al. Nonisotopic hybridization assay for determination of relative amounts of genotypic human immunodeficiency virus type 1 zidovudine resistance. *J Clin Microbiol*1995; **33:**2777-2780.

116. Eastman PS, Mittler J, Kelso R, et al. Genotypic changes in human immunodeficiency virus type 1 associated with loss of suppression of plasma viral RNA levels in subjects treated with ritonavir (Norvir) monotherapy. *J Virol* 1998; **72:**5154-5164.

117. Eckert DM, Malashkevich VN, Hong LH, et al. Inhibiting HIV-1 entry: discovery of Dpeptide inhibitors that target the gp41 coiled-coil pocket. *Cell* 1999; **99:**103-115.

118. Emini EA, Graham DJ, Gotlib L, et al. HIV and multidrug resistance. *Nature* 1993; **364:**679.

119. Engelbrecht S, de Villiers T, Sampson CC, et al. Genetic analysis of the complete gag and env genes of HIV type 1 subtype C primary isolates from South Africa. *AIDS Res Hum Retroviruses* 2001; **17:**1533-1547.

120. Engelbrecht S, Laten JD, Smith TL, et al. Identification of env subtypes in fourteen HIV type 1 isolates from South Africa. *AIDS Res Hum Retroviruses* 1995; **11:**1269-271.

121. Erickson JW, Gulnik SV, Markowitz M. Protease inhibitors: resistance, cross-resistance, fitness and the choice of initial and salvage therapies. *AIDS* 1999; **13:**S189-204.

122. Ericson-Viitanen S, Manfredi J, Viitanen P, et al. Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res Hum Retrovir* 1989; **5:**577–591.

123. Eron J, Vernazza PL, Johnston DM, et al, Resistance of HIV-1 to antitroviral agents in blood and seminal plasm; implications for transmission. *AIDS* 1998; **12:**F181-189.

124. Eshleman SH, Guay LA, Mwatha A, et al. Characterization of nevirapine resistance mutations in women with subtype A vs. D HIV-1 6-8 weeks after single-dose nevirapine (HIVNET 012). *J Acquir Immune Defic Syndr* 2004a; **35:**126-30.

125. Eshleman SH, Wang J, Guay LA, Cunningham SP, Mwatha A, Brown ER, Musole P, Mmiro F, Jackson JB. Distinct patterns of selection and fading of K103N and Y181C are seen in women with subtype A vs D HIV-1 after single dose nevirapine: HIVNET 012. *Antivir.Ther* 2004b; **9:**S59.

126. Eshleman SH, Jackson JB. Nevirapine resistance after single dose prophylaxis. *AIDS Rev* 2002; **4:**59-63.

127. Eshleman SH, Becker-Pergola G, Deseyve M, et al. Impact of human immunodeficiency virus type 1 (hiv-1) subtype on women receiving single-dose nevirapine prophylaxis to prevent hiv-1 vertical transmission (hiv network for prevention trials 012 study). *J Infect Dis* 2001a; **184:**914-917.

128. Eshleman SH, Mracna M, Guay LA, et al. Selection and fading of resistance mutations in women and infants receiving nevirapine to prevent HIV-1 vertical transmission (HIVNET 012). *AIDS* 2001b; **15:**1951-1957.

129. Esnouf RM, Ren J, Hopkins AL, et al. Unique features in the structure of the complex between HIV-1 reverse transcriptase and the bis(heteroaryl)piperazine (BHAP) U-90152 explain resistance mutations for this nonnucleoside inhibitor. *Proc Natl Acad Sci USA* 1997; **94:**3984-3989.

130. Esnouf R, Ren J, Ross C, et al. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nat Struct Biol* 1995; **2:**303-308.

131. Esparza J, and Bhamarapravati N. Accelerating the development and future availability of HIV-1 vaccines: Why, when, where, and how? *Lancet* 2000; **355:**2061-2066.

132. Essex M. Human immunodeficiency viruses in the developing world. *Adv Virus Res* 199; **53:**71-88.

133. Esu-Williams E, Mulanga-Kabeya C, Takena H, et al. Seroprevalence of HIV-1, HIV-2, and HIV-1 group O in Nigeria: evidence for a growing increase of HIV infection. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; **16:**204-10.

134. EuroGuidelines Group for HIV Resistance. Clinical and laboratory guidelines for the use of HIV-1 drug resistance testing as part of treatment management: recommendations for the European setting. *AIDS* 2001; **15:**309-320.

135. Falloon J, Piscitelli S, Vogel S, et al. Combination therapy with amprenavir, abacavir, and efavirenz in human immunodeficiency virus (HIV)-infected patients failing a protease-inhibitor regimen: pharmacokinetic drug interactions and antiviral activity. *Clin Infect Dis* 2000; **30:**313-318.

136. Felsenstein J. PHYLIP – Phylogeny Inference Package (Version 3.2), *Cladistics* 1989; **5:**164-166.

137. Fitch WM. Rate of change of concomitantly variable codons. *J Mol Evol* 1971; **1:**84-96.

138. Fitzgibbon JE, Howell RM, Haberzettl CA, et al. Human immunodeficiency virus type 1 pol gene mutations which cause decreased susceptibility to 2',3'-dideoxycytidine. *Antimicrob Agents Chemother* 1992; **36:**153-7.

139. Fontanet AL, Messele T, Dejene A, et al. Age- and sex-specific HIV-1 prevalence in the urban community setting of Addis Ababa, Ethiopia. *AIDS* 1998; **12:**315-322.

140. Frankel F, Turner D, Brenner B et al. Impaired rescue of chain-terminated DNA synthesis associated with the L74V mutation in HIV-1 reverse transcriptase. *Antiviral Ther* 2004; **9:**S33.

141. Frater AJ, Beardall A, Ariyoshi K, et al. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. *AIDS* 2001; **15:**1493-1502.

142. Frost SD, Nijhuis M, Schuurman R, et al. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *J Virol* 2000; **74:**6262-6268.

143. Fujiwara M, Kodama EN, Okamoto M, et al. Characterization of human immunodeficiency virus type 1 strains resistant to the non-nucleoside reverse transcriptase inhibitor RD4-2217. *Antivir Chem Chemother* 1999; **10:**315-320.

144. Fujiwara T, Sato A, el-Farrash M, et al. S-1153 inhibits replication of known drug-resistant strains of human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 1998; **42:**1340-1345.

145. Gadkari DA, Moore D, Sheppard HW, et al. Transmission of genetically diverse strains of HIV-1 in Pune, India. *Indian J Med Res* 1998; **107:**1-9.

146. Gallant JE, Deresinski S. Tenofovir disoproxil fumarate. *Clin Infect Dis* 2003; **37:**944-950.

147. Gao HQ, Boyer PL, Arnold E, et al. Effects of mutations in the polymerase domain on the polymerase, RNase H and strand transfer activities of human immunodeficiency virus type 1 reverse transcriptase. *J Mol Biol* 1998; **277:**559-572.

148. Gao Q, Gu Z, Hiscott J, et al. Generation of drug-resistant variants of human immunodeficiency virus type 1 by in vitro passage in increasing concentrations of 2',3'-dideoxycytidine and 2',3'-dideoxy-3'-thiacytidine. *Antimicrob Agents Chemother* 1993; **37:**130-133.

149. Garcia F, Plana M, Ortiz GM, et al. Virological and immunological consequences of structure interruptions in chronic HIV-1 infection. *AIDS* 2002; **15:**F29-40.

150. Goldman N, Yang Z. A codon-based model of nucleotide substitution for proteincoding DNA sequences. Molecular Biology and Evolution 1994; **11:**725-736.

151. Gonzales MJ, Machekano RN, Shafer RW. Human immunodeficiency virus type 1 reversetranscriptase and protease subtypes: classification, amino acid mutation patterns, and prevalence in a northern California clinic-based population. *J Infect Dis* 2001; **184:**998-1006.

152. Gonzalez LM, Brindeiro RM, Aguiar RS, et al. Impact of nelfinavir resistance mutations on in vitro phenotype, fitness, and replication capacity of human immunodeficiency virus type 1 with subtype B and C proteases. *Antimicrob Agents Chemother* 2004; **48:**3552-3555.

153. Gonzalez LM, Brindeiro RM, Tarin M, et al. In vitro hypersusceptibility of human immunodeficiency virus type 1 subtype C protease to lopinavir. *Antimicrob Agents Chemother* 2003; **47:**2817-2822.

154. Gonzales MJ, Wu TD, Taylor J, et al. Extended spectrum of HIV-1 reverse transcriptase mutations in patients receiving multiple nucleoside analog inhibitors. *AIDS* 2003; **17:**791-799.

155. Gordon M, De Oliveira T, Bishop K, et al. Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: Implications for vaccine and antiretroviral control strategies. *J Virol* 2003; **77:**2587-2599.

156. Gotte M, Arion D, Parniak MA, et al. The M184V mutation in the reverse transcriptase of human immunodeficiency virus type 1 impairs rescue of chain-terminated DNA synthesis. *J Virol* 2000; **74:**3579-3585.

157. Gotte M, Wainberg MA. Biochemical mechanisms involved in overcoming HIV resistance to nucleoside inhibitors of reverse transcriptase. *Drug Resist Updat* 2000; **3:**30-38.

158. Goudsmit J, de Ronde A, de Rooij E, et al. Broad spectrum of in vivo fitness of human immunodeficiency virus type 1 subpopulations differing at reverse transcriptase codons 41 and 215. *J Virol* 1997; **71:**4479-4484.

159. Goudsmit J, de Ronde A, Ho DD, et al. Human immunodeficiency virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. *J Virol* 1996; **70:**5662-5664.

160. Grange JM, Zumla A. The global emergency of tuberculosis: what is the cause? *J R Soc Health* 2002; **122:**78-81.

161. Grez M, Dietrich U, Balfe P, et al. Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J Virol* 1994; **68:**2161-2168.

162. Grossman Z, Istomin V, Averbuch D, et al. Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C. *AIDS* 2004; **18:**909-915.

163. Grossman Z, Paxinos E, D Auerbuch, et al. D30N is not the preferred resistance pathway in subtype C patients treated with nelfinavir. *Antiviral Ther* 2002a; **7:**S30.

164. Grossman Z, Vardinon N, Burke M, et al. Use of nucleoside analogues and development of resistance in subtype B and subtype C infected patients [abstract TuPeB4602]. *XIV International AIDS Conference*, Barcelona, Spain, 2002b.

165. Grossman Z, Vardinon N, Chemtob D, et al. Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. *AIDS* 2001; **15:**1453-1460.

166. Gu Z, Gao Q, Fang H, et al. Identification of a mutation at codon 65 in the IKKK motif of reverse transcriptase that encodes human immunodeficiency virus resistance to 2',3'-dideoxycytidine and 2',3'-dideoxy-3'-thiacytidine. *Antimicrob Agents Chemother* 1994; **38:**275-281.

167. Gu Z, Gao Q, Li X, et al. Novel mutation in the human immunodeficiency virus type 1 reverse transcriptase gene that encodes cross-resistance to 2',3'-dideoxyinosine and 2',3'-dideoxycytidine. *J Virol* 1992; **66:**7128-7135.

168. Guay LA, Musoke P, Fleming T, et al. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet* 1999; **354:**795-802

169. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997; **18:**2714-2723.

170. Gulick RM, Mellors JW, Havlir D, et al. 3-year suppression of HIV viremia with indinavir, zidovudine, and lamivudine. *Ann Intern Med* 2000; **133:**35-39.

171. Gulnik SV, Suvorov LI, Liu B, et al. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry* 1995; **34:**9282-9287.

172. Gunthard D, Havlir D, Fiscus, et al, Residual human immunodeficiency virus (HIV) type 1 RNA and DNA in lymph nodes and HIV RNA in genetal secretions and in cerebrospinal fluid after supppression of viremia for 2 years. *J Infect Dis* 2001; **183:**1318-1327.

173. Gunthard HF, Frost SD, Leigh-Brown AJ, et al.: Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiviral therapy. *J Virol* 1999; **73:**9404–9412.

174. Gurtler L. Recent aspects of HIV and HIV-related diseases. *Infection* 1997; **25:**71-3.

175. Gurtler L, Hauser PH, Eberle J, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J Virol* 1994; **68:**1581-1585.

176. Hahn BH, Shaw GM, De Cock KM, Sharp PM. *AIDS* as a zoonosis: scientific and public health implications. *Science* 2000; **287:**607-614.

177. Hammer SM, Vaida F, Bennett KK, et al. AIDS Clinical Trials Group 398 Study Team. Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial. *JAMA* 2002; **288:**169-180.

178. Handema R, Terunuma H, Kasolo F, et al. Emergence of new HIV-1 subtypes other than Subtype C among antenatal women in Lusaka, Zambia. *AIDS Res Hum Retroviruses* 2001; **17:**759-763.

179. Hanna GJ, Johnson VA, Kuritzkes DR, et al. Patterns of resistance mutations selected by treatment of human immunodeficiency virus type 1 infection with zidovudine, didanosine, and nevirapine. *J Infect Dis* 2000; **181:**904-911.

180. Hanna SL, Yang C, Owen S, et al. Variability of critical epitopes within HIV-1 heptad repeat domains for selected entry inhibitors in HIV-infected populations worldwide [corrected]. *AIDS* 2002; **16:**1603-1608.

181. Harouse JM, Gettie A, Tan RC, et al. Distinct pathogenic sequela in rhesus macaques infected with CCR5 or CXCR4 utilizing SHIVs. *Science* 1999; **284:**816-819.

182. Harrigan PR, Alexander CS. Selection of drug-resistant HIV. *Trends Microbiol* 1999; **7:**120-123.

183. Harrigan PR, Wynhoven B, Montaner J, et al. Mutations at reverse transcriptase codon 103**:** phenotypic resistance to non-nucleoside reverse transcriptase inhibitor and clinical correlates. *Antivir Ther* 2003; **8:**120.

184. Harrigan PR, Stone C, Griffin P, et al. Resistance profile of the human immunodeficiency virus type 1 reverse transcriptase inhibitor abacavir (1592U89) after monotherapy and combination therapy. CNA2001 Investigative Group. *J Infect Dis* 2000; **181:**912-920.

185. Harrigan PR, Kinghorn I, Bloor S, et al. Significance of amino acid variation at human immunodeficiency virus type 1 reverse transcriptase residue 210 for zidovudine susceptibility. *J Virol* 1996; **70:**5930-5934.

186. Hartigan JA. Clustering. *Annu Rev Biophys Bioeng* 1973; **2:**81-101.

187. Haubrich RH, Kemper CA, Hellmann NS, et al. The clinical relevance of non-nucleoside reverse transcriptase inhibitor hypersusceptibility: a prospective cohort analysis. *AIDS* 2002; **16:**F33-40.

188. Havlir DV, Eastman S, Gamst A, et al. Nevirapine-resistant human immunodeficiency virus: kinetics of replication and estimated prevalence in untreated patients. *J Virol* 1996; **70:**7894-7899.

189. Havlir DV, Hellmann NS, Petropoulos CJ, et al. Drug susceptibility in HIV infection after viral rebound in patients receiving indinavir-containing regimens. *JAMA* 2000; **283:**229-234.

190. Havlir DV, Richman DD. Viral dynamics of HIV: implications for drug development and therapeutic strategies. *Ann Int Med* 1996; **124:**984–994.

191. Hensle K. HIV in Southern Africa. *Harv AIDS Rev* 1998;10-3.

192. Herring BL, Ge YC, Wang B, et al. Segregation of human immunodeficiency virus type 1 subtypes by risk factor in Australia. *J Clin Microbiol* 2003; **4:**4600-4604.

193. Hertogs K, Bloor S, De Vroey V, et al. A novel human immunodeficiency virus type 1 reverse transcriptase mutational pattern confers phenotypic lamivudine resistance in the absence of mutation 184V. *Antimicrob Agents Chemother* 2000; **44:**568-573.

194. Hertogs K, Bloor S, Kemp SD, et al. Phenotypic and genotypic analysis of clinical HIV-1 isolates reveals extensive protease inhibitor cross-resistance: a survey of over 6000 samples. *AIDS* 2000; **14:**1203-1210.

195. Hertogs K, De Bethune MP, Miller V, et al. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. *Antimicrob Agents Chemother* 1998; **42:**269-276.

196. Hirsch MS, Brun-Vezinet F, Clotet B, et al. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type **1:** 2003 recommendations of an International *AIDS* Society-USA Panel. *Clin Infect Dis* 2003; **37:**113-128.

197. Hirsch MS, Brun-Vezinet F, D'Aquila RT, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society-USA Panel 2000; *JAMA* **283:**2417-2426.

198. Hirsch MS, Conway B, D'Aquila RT, et al. Antiretroviral drug resistance testing in adults with HIV infection: implications for clinical management. International *AIDS* Society--USA Panel. *JAMA* 1998; **279:**1984-1991.

199. Hoelscher M, Kim B, Maboko L, et al. High proportion of unrelated HIV-1 intersubtype recombinants in the Mgeya region of southwest Tanzania. *AIDS* 2001; **15:**1461-1470.

200. Hoffman NG, Schiffer CA, Swanstrom R. Covariation of amino acid positions in HIV-1 protease. *Virol* 2003; **314:**536-548.

201. Hoffman TL, Doms RW. HIV-1 envelope determinants for cell tropism and chemokine receptor use. *Mol Membr Biol* 1999; **16:**57-65.

202. Holland JJ, de la Torre JC, Clarke D et al. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J Virol* 1991; 6**5:**2960–2967.

203. Holm-Hansen C, Ayehunie S, Johansson B, et al. HIV-1 proviral DNA sequences of env gp41 PCR amplificates from Tanzania. *Apmis* 1996; **104:**459-464.

204. Hooker DJ, Tachedjian G, Solomon AE, et al. An in vivo mutation from leucine to tryptophan at position 210 in human immunodeficiency virus type 1 reverse transcriptase contributes to high-level resistance to 3'-azido-3'-deoxythymidine. *J Virol* 1996; **70:**8010-8018.

205. Hsiou Y, Ding J, Das K, et al. The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. *J Mol Biol* 2001; **309:**437-445.

206. Hu DJ, Buve A, Baggs J, et al. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS* 1999; **13:**873-881.

207. Huang H, Chopra R, Verdine GL, et al. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* 1998; **282:**1669-1675.

208. Huang W, Gamarnik A, Limoli K, et al. Amino acid substitutions at position 190 of human immunodeficiency virus type 1 reverse transcriptase increase susceptibility to delavirdine and impair virus replication. *J Virol* 2003; **77:**1512-1523.

209. Huang W, Limoli K, Sartoris MM, et al. Complex interactions involving multiple amino acid substitutions alter NNRTI susceptibility. *Antivir Ther* 1999; **4:**50-51.

210. Hulo N, Sigrist CJA, Le SauxV et al, Recent improvements to the PROSITE database. *Nucleic Acids Res* 2004; **1:**134-137.

211. Hurst M, Faulds D. Lopinavir. *Drugs* 2000; **60:**1371-1379.

212. Hwang SS, Boyle TJ, Lyerly HK, et al. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* 1991; **253:**71-74.

213. Ikuta K, Suzuki S, Horikoshi H, et al. Positive and negative aspects of the human immunodeficiency virus protease: development of inhibitors versus its role in *AIDS* pathogenesis. *Microbiol Mol BiolRev* 2000; **64:**725–745.

214. Iversen AK, Shafer RW, Wehrly K, et al. Multidrug-resistant human immunodeficiency virus type 1 strains resulting from combination antiretroviral therapy. *J Virol* 1996; **70:**1086-1090.

215. Jacks T, Power MD, Masiarz FR, et al. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 1998; **33:**280–283.

216. Jackson JB, Becker-Pergola G, Guay LA, et al. Identification of the K103N resistance mutation in Ugandan women receiving nevirapine to prevent HIV-1 vertical transmission. *AIDS* 2000; **14:**F111-115.

217. Jacobsen H, Yasargil K, Winslow DL, et al. Characterization of human immunodeficiency virus type 1 mutants with decreased sensitivity to proteinase inhibitor Ro 31-8959. *Virol* 1995; 206**:**527-534.

218. Jameel S, Zafrullah M, Ahmad M, et al. A genetic analysis of HIV-1 from Punjab, India reveals the presence of multiple variants. *AIDS* 1995; **9:**685-690.

219. Janssens W, Buve A, Nkengasong JN: The puzzle of HIV-1 subtypes in Africa. *AIDS* 1997; **11:**705–712.

220. Japour AJ, Mayers DL, Johnson VA, et al. Standardized peripheral blood mononuclear cell culture assay for determination of drug susceptibilities of clinical human immunodeficiency virus type 1 isolates. The RV-43 Study Group, the *AIDS* Clinical Trials Group *Virology* Committee Resistance Working Group. *Antimicrob Agents Chemother* 1993; **37:**1095-1101.

221. Jeeninga RE, Hoogenkamp M, Armand-Ugon M, et al. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. *J Virol* 2000; **74:**3740-3751.

222. Ji J, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying a hypervariable region of the HIV-1 env gene. *Virol* 1994; **199:**323-330.

223. Johansson B, Sherefa K, Sonnerborg A. Multiple enhancer motifs in HIV type 1 strains from Ethiopia. *AIDS Res Hum Retroviruses* 1995; **11:**761-764.

224. Johnston ER, Zijenah LS, Mutetwa S, et al. High frequency of syncytium-inducing and cxcr4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral treatment. *J Virol* 2003; **77:**7682-7688.

225. Joly V, Moroni M, Concia E, et al. Delavirdine in combination with zidovudine in treatment of human immunodeficiency virus type 1-infected patients: evaluation of efficacy and emergence of viral resistance in a randomized, comparative phase III trial. The M/3331/0013B Study Group. *Antimicrob Agents Chemother* 2000; **44:**3155-3157.

226. Kanki PJ, Hamel DJ, Sankale JL, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis* 1999; 179**:**68-73.

227. Kantor R, Katzenstein D. Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS Rev* 2003a; **5:**25-35.

228. Kantor R, Lee E, Johnston E et al. Rapid flux in non-nucleoside reverse transcriptase inhibitor resistance mutations among subtype C HIV-infected women after single dose nevirapine [abstract 78]. *XIIth annual HIV Drug Resistance Workshop*, Cabo San Lucas, Mexico, 2003b.

229. Kantor R, Katzenstein D, Cane PA, et al. Classical subtype B resistance mutations and novel candidate non-subtype B resistance mutations in treated persons with non-B HIV-1 isolates [abstract TuPeB4614]. *XIV International AIDS Conference,* Barcelona, Spain, 2002a.

230. Kantor R, Zijenah LS, Shafer RW, et al. HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy. *AIDS Res Hum Retroviruses* 2002b; **18:**1407-1413.

231. Kapiga SH, Bang H, Spiegelman D, et al. Correlates of plasma HIV-1 RNA viral load among HIV-1-seropositive women in Dar es Salaam, Tanzania. *J Acquir Immune Defic Syndr* 2002; **30:**316-323.

232. Kaplan AH, Krogstad P, Kempf DJ, et al. Human immunodeficiency virus type 1 virions composed of unprocessed Gag and Gag-Pol precursors are capable of reverse transcribing viral genomic RNA. *Antimicrob Agents Chemother* 1994; **38:** 2929-2933.

233. Kaplan JC, Hirsch MS. Therapy other than reverse transcriptase inhibitors for HIV infection. *Clin Lab Med* 1994; **14:**367-391.

234. Karacostas V, Wolffe EJ, Nagashima K, et al. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* 1993; **193:**661-671.

235. Katlama C, Clotet B, Plettenberg A, et al. The role of abacavir (ABC, 1592) in antiretroviral therapy-experienced patients: results from a randomized, double-blind, trial. CNA3002 European Study Team. *AIDS* 2000; **14:**781-789.

236. Kato K, Sato H, Takebe Y. Role of naturally occurring basic amino acid substitution in the human immunodeficiency virus type 1 subtype E envelope V3 loop on viral coreceptor usage and cell tropism. *J Virol* 1999; **73:**5520-5526.

237. Katzenstein DA, Bosch RJ, Hellmann N, et al. Phenotypic susceptibility and virological outcome in nucleoside-experienced patients receiving three or four antiretroviral drugs. *AIDS* 2003; **17:**821-830.

238. Katzenstein DA. Editorial comment: resistance testing-an integral part of HIV management. *AIDS Read* 2003; **13:**36.

239. Katzenstein D. HIVresistanceWeb; www.hivresistanceweb.com/protected/po/dk-00apr-vert.shtml. May 2000.

240. Kazatchkine MD, Van PN, Costagliola D, et al. Didanosine dosed once daily is equivalent to twice daily dosing of patients on double or triple combination anti-retroviral therapy. A1454-147 Team. *J Acquir Immune Defic Syndr*2000; **15:** 24418-24424.

241. Kellam P, Larder BA. Recombinant virus assay: a rapid, phenotypic assay for assessment of drug susceptibility of human immunodeficiency virus type 1 isolates. *Antimicrob Agents Chemother* 1994; **38:**23-30.

242. Kellam P, Boucher CA, Larder BA. Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proc Natl Acad Sci USA* 1992; **89:**1934-1938.

243. Kempf DJ, Isaacson JD, King MS, et al. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. *J Virol* 2001; **75:**7462-7469.

244. Keulen W, Nijhuis M, Schuurman R, et al. Reverse transcriptase fidelity and HIV-1 variation. *Science* 1997; **275:**229.

245. Kilby JM, Lalezari JP, Eron JJ, et al, The safety, plasma pharmacokinetics and antiviral activity of subcutaneous enfuvirtide (T-20), a peptide inhibitor of gp41-mediated virus fusion, in HIV-infected adults. *AIDS Res Hum Retroviruses* 2002; **18:**685-693.

246. Kilby JM, Hopkins S, Venetta TM, et al. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat Med* 1998; **4:**1302-1307.

247. King RW, Winslow DL, Garber S, et al. Identification of a clinical isolate of HIV-1 with an isoleucine at position 82 of the protease which retains susceptibility to protease inhibitors. *Antiviral Res* 1995; **28:**13-24.

248. Kleim JP, Bender R, Kirsch R, et al. Mutational analysis of residue 190 of human immunodeficiency virus type 1 reverse transcriptase. *Virology* 1994; **200:**696-701.

249. Koch N, Ndihokubwayo JB, Yahi N, et al. Genetic analysis of hiv type 1 strains in bujumbura (burundi): predominance of subtype c variant. *AIDS Res Hum Retroviruses* 2001; **17:**269-273.

250. Kohlstaedt LA, Wang J, Friedman JM, et al. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 1992; **256:**1783-1790.

251. Korber, B. HIV Signature and Sequence Variation Analysis. In: Computational Analysis of HIV Molecular Sequences. Chapter 4, pp 55-72. AG Rodrigo and GH Learn (eds). Dordrecht, Netherlands, Kluwer Academic Publishers, 2001.

252. Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science* 2000; **288:**1789-1796

253. Korber B, Foley B, Kuiken Bet al. Numerbing positions in HIV relative to HXB2CG. In: Human Retroviruses and *AIDS* (Korber B, Foley B, Hahn B, McCutchan F, Mellors J, Sodrosk J, eds.). Los Alamos National Laboratory, Los Alamos, N.Mex, 1998.

254. Korber B, Myers G. Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Res Hum Retroviruses* 1992; **8:**1549-1560.

255. Koulinska IN, Ndung'u T, Mkakagile D, et al. A new human immunodeficiency virus type 1 circulating recombinant from Tanzania. *AIDS Res Hum Retroviruses* 2001; **17:**423-431.

256. Kozal MJ, Shafer RW, Winters MA, et al. HIV-1 syncytium-inducing phenotype, virus burden, codon 215 reverse transcriptase mutation and CD4 cell decline in zidovudine-treated patients. *J Acquir Immune Defic Syndr* 1994; **7:**832-838.

257. Kozal MJ, Shafer RW, Winters MA, et al. A mutation in human immunodeficiency virus reverse transcriptase and decline in CD4 lymphocyte numbers in longterm zidovudine recipients. *J InfectDis* 1993; **167:**526-532.

258. Krausslich H, Ingraham FH, Skoog M, et al. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. *Proc Natl Acad Sci* 1989; **86:**807-811.

259. Kreiss JK, Koech D, Plummer FA, et al. AIDS virus infection in Nairobi prostitutes. Spread of the epidemic to East Africa. *N Engl J Med* 1986; **314:**414–418.

260. Krivine A, Le Bourdelles S, Firtion G, et al. Viral kinetics in HIV-1 perinatal infection. *Lancet* 1997; **350:**493.

261. Idriss H, Kawa S, Damuni Z, et al. HIV-1 reverse transcriptase is phosphorylated in vitro and in a cellular system. *Int J Biochem Cell Biol* 1999; **31:**1443-1452.

262. Kuiken C, Korber B, Shafer RW. HIV sequence databases. *AIDS Rev* 2003; **5:**52-61.

263. Kuiken C, Foley B, Hahn B, et al. HIV sequence compendium. Los Alamos National Laboratory, Los Alamos, N.Mex, 2002.

264. Kuiken C, Foley B, Hahn B, et al. Human Retroviruses and AIDS. A compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.Mex, 1999.

265. Kumar S, Tamura K, Jakobsen IB, et al. MEGA2: Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA, 2001.

266. Kuritzkes DR. Preventing and managing antiretroviral drug resistance. *AIDS Patient Care STDS* 2004; **18:**259-273.

267. Kuritzkes DR, Sevin A, Young B, et al. Effect of zidovudine resistance mutations on virologic response to treatment with zidovudine-lamivudine-ritonavir: genotypic analysis of human immunodeficiency virus type 1 isolates from AIDS clinical trials group protocol 315.ACTG Protocol 315 Team. *J Infect Dis* 2000; **181:**491-497.

268. Lacey SF, Larder BA. Mutagenic study of codons 74 and 215 of the human immunodeficiency virus type 1 reverse transcriptase, which are significant in nucleoside analog resistance. *J Virol* 1994; **68:**3421-3424.

269. Lalezari JP, Henry K, O'Hearn M, e al. Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America. *N Engl J Med* 2003; 34**8:**2175-2185.

270. Lalvani A, Shastri JS. HIV epidemic in India: opportunity to learn from the past. *Lancet* 1996; **347:**1349-1350.

271. Lanier E, Irlbeck D, Ross L, et al. Prediction of NRTI options by linking RT genotype and phenotype breakpoints [abstract 586]. *10th Conference on Retroviruses and Opportunistic Infections*, Boston, MA, 2003.

272. Lanier ER, Ait-Khaled M, Scott J, et al. Antiviral efficacy of abacavir in antiretroviral therapy-experienced adults harbouring HIV-1 with specific patterns of resistance to nucleoside reverse transcriptase inhibitors. *Antivir Ther* 2004; **9:**37-45.

273. Lanier R, Scott J, Steel H, Hetherington B, et al. Multivariate analysis of predictors of response to abacavir: comparison of prior antiretroviral therapy, baseline HIV RNA, CD4 count and viral resistance. *Antivir Ther* 1999; **4:**S56.

274. Larder BA, Bloor S, Kemp SD, et al. A family of insertion mutations between codons 67 and 70 of human immunodeficiency virus type 1 reverse transcriptase confer multinucleoside analog resistance. *Antimicrob Agents Chemother* 1999; **43:**1961-1967.

275. Larder BA, Stammers DK. Closing in on HIV drug resistance. *Nat Struct Biol* 1999; **6:**103-106.

276. Larder BA, Kemp SD, Harrigan PR. Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. *Science* 1995; **269:**696-699.

277. Larder BA, Kellam P, Kemp SD. Convergent combination therapy can select viable multidrug-resistant HIV-1 in vitro. *Nature* 1993; **365:**451-453.

278. Larder BA, Chesebro B, Richman DD. Susceptibilities of zidovudine-susceptible and -resistant human immunodeficiency virus isolates to antiviral agents determined by using a quantitative plaque reduction assay. *Antimicrob Agents Chemother* 1990; **34:** 436-441.

279. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 1989; **246:**1155-1158.

280. Laurent C, Diakhate N, Gueye NF, et al. The Senegalese government's highly active antiretroviral therapy initiative: an 18-month follow-up study. *AIDS* 2002; **16:**1363-1370.

281. Lawrence J, Schapiro J, Winters M, et al. Clinical resistance patterns and responses to two sequential protease inhibitor regimens in saquinavir and reverse transcriptase inhibitor-experienced persons. *J Infect Dis* 1999; **179:**1356-1364.

282. Lazaro JB, Boretto J, Selmi B, et al. Phosphorylation of AZT-resistant human immunodeficiency virus type 1 reverse transcriptase by casein kinase II in vitro: effects on inhibitor sensitivity. *Biochem Biophys Res Commun* 2000; **275:**26-32.

283. Ledergerber B, Egger M, Opravil M, et al. Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. *Lancet* 1999; **353:**863-868.

284. Lee E, Kantor R, Johnston E. Breast milk shedding of drug-resistant subtype C HIV-1 and among women receiving single-dose nevirapine [abstract 96], *10th Conference on Retroviruses and Opportunistic Infections*, Boston, MA, 2003.

285. Leigh Brown A, Frost SD, Mathews WC, et al. Transmission fitness of drug-resistant human immunodeficiency virus and the prevalence of resistance in the antiretroviral-treated population. *J Infect Dis* 2003; **187:**683.

286. Leigh Brown AJ, Precious HM, Whitcomb JM, et al. Reduced susceptibility of human immunodeficiency virus type 1 (HIV-1) from patients with primary HIV infection to nonnucleoside reverse transcriptase inhibitors is associated with variation at novel amino acid sites. *J Virol* 2000; **74:**10269-273.

287. Lennerstrand J, Stammers DK, Larder BA. Biochemical mechanism of human immunodeficiency virus type 1 reverse transcriptase resistance to stavudine. *Antimicrob Agents Chemother* 2001; **45:**2144-2146.

288. Li Y, Rey-Cuille MA, Hu SL. N-linked glycosylation in the V3 region of HIV type 1 surface antigen modulates coreceptor usage in viral infection. *AIDS Res Hum Retroviruses* 2001; **17:**1473-1479.

289. Liitsola K, Tashkinova I, Laukkanen T, et al. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. *AIDS* 1998; **12:**1907-1919.

290. Little SJ, Koelsch KK, Ignacio CC, et al. Persistence of transmitted drug-resistant virus among subjects with primary HIV infection deferring antiretroviral therapy. [abstract 36LB]. *11th Conference on Retroviruses and Opportunistic Infections,* San Francisco, 2004.

291. Little SJ, Holte S, Routy JP, et al. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; **347:**385-394.

292. Little SJ. Is transmitted drug resistance in HIV on the rise? (editorial) *BR MED J* 2001; **322:**1074.

293. Little SJ. Transmission and prevalence of HIV resistance among treatment-naïve subjects. *Antivir Ther* 2000; **5:**33–40.

294. Loeb DD, Hutchison CA 3rd, Edgell MH, et al. Mutational analysis of human immunodeficiency virus type 1 protease suggests functional homology with aspartic proteinases. *J Virol* 1989; **63:**111-121.

295. Loemba H, Brenner B, Parniak MA, et al. Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against nonnucleoside inhibitors of RT. *Antimicrob Agents Chemother* 2002; **46:**2087-2094.

296. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; **73:**152-160.

297. Losman B, Biller M, Olofsson S, et al. The N-linked glycan of the V3 region of HIV-1 gp120 and CXCR4-dependent multiplication of a human immunodeficiency virus type 1 lymphocyte-tropic variant. *FEBS Lett* 1991; **454:**47-52.

298. Louwagie J, Janssens W, Mascola J, et al. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol* 1995; **69:**263-271.

299. Loveday C, van Hooff F, Johnson M. Inferior virologic responses to highly active antiretroviral therapy in patients with HIV-1 subtype C infection: a case controlled study. *Antiviral Ther* 2002; **7:**S145.

300. Loveday C, et al. High prevalence of multiple drug resistance mutations in a UK HIV/AIDS patient population. *AIDS* 1999; **13:**627-628.

301. Lukashov VV, Huismans R, Jebbink MF, et al. Selection by AZT and rapid replacement in the absence of drugs of HIV type 1 resistant to multiple nucleoside analogs. *AIDS Res Hum Retroviruses* 2001; **17:**807-818.

302. Lukashov VV, Cornelissen MT, Goudsmit J, et al. Simultaneous introduction of distinct HIV-1 subtypes into different risk groups in Russia, Byelorussia and Lithuania. *AIDS* 1995; **9:**435-439.

303. Maeda Y, Venzon DJ, Mitsuya H. Altered drug sensitivity, fitness, and evolution of human immunodeficiency virus type 1 with pol gene mutations conferring multi-dideoxynucleoside resistance. *J Infect Dis* 1998; **177:**1207-1213.

304. Maguire M, Gartland M, Moore S, et al. Absence of zidovudine resistance in antiretroviral-naive patients following zidovudine/lamivudine/protease inhibitor combination therapy: virological evaluation of the AVANTI 2 and AVANTI 3 studies. *AIDS* 2000; **14:**1195-1201.

305. Mahalingam B, Louis JM, Hung J, et al. Structural implications of drug-resistant mutants of HIV-1 protease: high-resolution crystal structures of the mutant protease/substrate analogue complexes. Proteins 2001; **43:**455-464.

306. Maitra A, Singh B, Banu S, et al. Subtypes of HIV type 1 circulating in India: Partial envelope sequences. *AIDS Res Hum Retroviruses* 1999; **15:**941–944.

307. Mammano F, Trouplin V, Zennou V, et al. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J Virol* 2000; **74:**8524-8531.

308. Mammano F, Petit C, Clavel F. Resistance-associated loss of viral fitness in human immunodeficiency virus type **1:** phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* 1998; **72:**7632-7637.

309. Mandal D, Jana S, Bhattacharya SK, et al. HIV type 1 subtypes circulating in eastern and northeastern regions of India. *AIDS Res Hum Retroviruses* 2002; **18:**1219-1227.

310. Mandal D, Jana S, Panda S, et al. Distribution of HIV-1 subtypes in female sex workers of Calcutta, India. *Indian J Med Res* 2000; **112:**165-172.

311. Markland W, Rao BG, Parsons JD, Black J, Zuchowski L, Tisdale M, Tung R. Structural and kinetic analyses of the protease from an amprenavir-resistant human immunodeficiency virus type 1 mutant rendered resistant to saquinavir and resensitized to amprenavir. *J Virol* 2000; **74:** 7636-7641.

312. Markowitz M, Conant M, Hurley A, et al. A preliminary evaluation of nelfinavir mesylate, an inhibitor of human immunodeficiency virus (HIV)-1 protease, to treat HIV infection. *J Infect Dis*1998; **177:**1533-1540.

313. Martin DJ, Schoub BD, Padayachee GN, et al. One year surveillance of HIV-1 infection in Johannesburg, South Africa. *Trans R Soc Trop Med Hyg*. 1990; **84:**728-730.

314. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol* 1999; **73:**3744-3752.

315. Martinez-Picado J, Sutton L, De Pasquale MP, et al. Human immunodeficiency virus type 1 cloning vectors for antiretroviral resistance testing. *J Clin Microbiol* 1999; **37:**2943-2951.

316. Maschera B, Furfine E, Blair ED. Analysis of resistance to human immunodeficiency virus type 1 protease inhibitors by using matched bacterial expression and proviral infection vectors. *J Virol* 1995; **69:**5431-5436.

317. Masquelier B, Race E, Tamalet C, et al. Genotypic and phenotypic resistance patterns of human immunodeficiency virus type 1 variants with insertions or deletions in the reverse transcriptase (RT): multicenter study of patients treated with RT inhibitors. *Antimicrob Agents Chemother* 2001; **45:**1836-1842.

318. Mastro TD, Kunanusont C, Dondero TJ, et al. Why do HIV-1 subtypes segregate among persons with different risk behaviors in South Africa and Thailand? *AIDS* 1997; **11:**113-116.

319. Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. *J Mol Biol* 2002; **317:**523-540.

320. McCormack GP, Glynn JR, Crampin AC, et al. Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi. *J Virol* 2002; **76:**12890-12899.

321. McCormick-Davis C, Dalton SB, Singh DK, et al. Comparison of Vpu sequences from diverse geographical isolates of HIV type 1 identifies the presence of highly variable domains, additional invariant amino acids, and a signature sequence motif common to subtype C isolates. *AIDS Res Hum Retroviruses* 2000; **16:**1089-1095.

322. Mellors J, Palmer S, Nissley D, et al: Low-frequency NNRTI-resistant variants contribute to failure of efavirenz-containing regimens. [abstract 39]. *11th Conference on Retroviruses and Opportunistic Infections*, San Francisco, 2004.

323. Mervis RJ, Ahmad N, Lillehoj EP, et al, The gag gene products of human immunodeficiency virus type **1:** alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *J Virol* 1988; **62:**3993-4002.

324. Meyer PR, Matsuura SE, Mian AM, et al. A mechanism of AZT resistance: an increase in nucleotide-dependent primer unblocking by mutant HIV-1 reverse transcriptase. *Mol Cell* 1999; **4:**35-43.

325. Meyer PR, Matsuura SE, Schinazi RF, et al. Differential removal of thymidine nucleotide analogues from blocked DNA chains by human immunodeficiency virus reverse transcriptase in the presence of

physiological concentrations of 2'-deoxynucleoside triphosphates. *Antimicrob Agents Chemother* 2000; **44:**3465-3472.

326. Miller MD, McColl DJ, White KL, et al. Genotypic and phenotypic characterization of patient-derived HIV-1 isolates containing the K65R mutation in reverse transcriptase [abstract 904]. *43rd InterScience Conference on Antimicrobial Agents and Chemotherapy*, Chicago, IL, 2003.

327. Miller V, Stark T, Loeliger AE, et al. The impact of the M184V substitution in HIV-1 reverse transcriptase on treatment response. *HIV Med* 2002; **3:**135-145.

328. Miller MD, Margot NA, Lamy PD, et al. Adefovir and tenofovir susceptibilities of HIV-1 after 24 to 48 weeks of adefovir dipivoxil therapy: genotypic and phenotypic analyses of study GS-96-408. *J Acquir Immune Defic Syndr* 2001; **27:**450-458.

329. Miller V, Larder BA. Mutational patterns in the HIV genome and cross-resistance following nucleoside and nucleotide analogue drug exposure. *Antivir Ther* 2001; **6:**25-44.

330. Miller V, Ait-Khaled M, Stone C, et al. HIV-1 reverse transcriptase (RT) genotype and susceptibility to RT inhibitors during abacavir monotherapy and combination therapy. *AIDS* 2000; **14:**163-171.

331. Mirochnick M, Clarke DF, Dorenbaum A. Nevirapine: pharmacokinetic considerations in children and pregnant women. *Clin Pharmacokinet* 2000; **39:**281-293.

332. Mo H, Parkin N, Stewart K, et al. I84A and I84C mutations in protease confer high-level resistance to protease inhibitors and impair replication capacity. *Antivir Ther* 2003; **8:**56.

333. Mofenson LM. Mother-child HIV-1 transmission: Timing and determinants. *Obstet Gynecol Clin North Am* 1997; **24:**759-784.

334. Molina JM, Marcelin AG, Pavie J, et al. Didanosine (ddI) in treatment experienced HIV-infeted patients: results from a randomized double-blind study (AI454- 176 Jaguar) [abstract H-447]. *43rd InterScience Conference on Antimicrobial Agents and Chemotherapy*, Chicago, IL, 2003.

335. Molla A, Brun S, Garren K, et al. Patterns of resistance to lopinavir in protease inhibitor-experienced patients following viral rebound during lopinavir/ritonavir therapy. *Antivir Ther* 2001; **6:**49.

336. Molla A, Korneyeva M, Gao Q, et al. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med* 1996; **2:**760-766.

337. Montano MA, Nixon CP, Ndung'u T, et al. Elevated tumor necrosis factor-alpha activation of human immunodeficiency virus type 1 subtype C in Southern Africa is associated with an NF-kappaB enhancer gain-of-function. *J Infect Dis* 2000; **181:**76-81.

338. Montano MA, Novitsky VA, Blackard JT, et al. Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. *J Virol* 1997; **71:**8657-8665.

339. Moodley D, Moodley J, Coovadia H, et al. South African Intrapartum Nevirapine Trial (SAINT) Investigators A multicenter randomized controlled trial of nevirapine versus a combination of zidovudine and lamivudine to reduce intrapartum and early postpartum mother-to-child transmission of human immunodeficiency virus type 1. *J Infect Dis* 2003; **187:**725-735.

340. Morison L, Buve A, Zekeng L, et al. HIV-1 subtypes and the HIV epidemics in four cities in sub-Saharan Africa. *AIDS* 2001; **15:**S109–S116.

341. Morris L, Pillay C, Chezzi C, et al. Low frequency of the V106M mutation among HIV-1 subtype C-infected pregnant women exposed to nevirapine. *AIDS* 2003; **17:**1698-1700.

342. Morris L, Cilliers T, Bredell H, et al. CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis. *AIDS Res Human Retroviruses* 2001; **17:**697–701.

343. Morris L, Pillay C, Gray G, et al. HIV-1 drug resistance and mother-to-child transmission. *SAMJ* 2001; **56:**614-616.

344. Mukadi YD, Maher D, Harries A. Tuberculosis case fatality rates in high HIV prevalence populations in sub-Saharan Africa. *AIDS* 2001; **15:**143-152.

345. Muller B, Patschinsky T, Krausslich HG. The late-domaincontaining protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *J Virol* 2002; **76:**1015–1024.

346. Murray MC, Embree EJ, Ramdahin SG, Anzala AO, Njenga S, Plummer FA. Effect of human immunodeficiency virus (HIV) type 1 viral genotype on mother-to-child transmission of HIV-1. *J Infect Dis* 2000; **181:**746–749.

347. Musoke P, Guay LA, Bagenda D, et al. A phase III study of the safety and pharmacokinetics of nevirapine in HIV-1-infected pregnant Ugandan women and their neonates (HIVNET 006). *AIDS* 1999; **13:**479-486.

348. Myers G, Korber B, Berzofsky JA, et al. Human retroviruses and AIDS. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, N.Mex, 1997.

349. Naeger LK, Margot NA, Miller MD. Tenofovir (PMPA) is less susceptible to pyrophosphorolysis and nucleotide-dependent chain-terminator removal than zidovudine or stavudine. *Nucleosides Nucleotides Nucleic Acids* 2001; **20:**635-639.

350. Nakayama EE, Shioda T, Tatsumi M, et al. Importance of the N-glycan in the V3 loop of HIV-1 envelope protein for CXCR-4 but not CCR-5-dependent fusion. *FEBS Lett*1998; **426:**367-372.

351. Navaratne L, Tong W, Christie I, et al. Is L90M mutation selected more frequently in non-B subtype HIV1 infected patients failing on Nelfinavir-containing HAART? [abstract TuPeB4592]. *XIV International AIDS Conference,* Barcelona, Spain, 2002.

352. Neild PJ, Gazzard BG. HIV-1 infection in China. *Lancet* 1997; **350:**963.

353. Neilson JR, John GC, Carr JK, et al. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. *J Virol* 1999; **73:**4393-4403.

354. New Data on the Prevention of Mother-to-Child Transmission of HIV and their Policy Implications. Conclusions and Recommendations. WHO Technical Consultation on Behalf of UNFPAUNICEF/WHO/UN*AIDS* Inter-Agency Task Team on Mother-to-Child Transmission of HIV, Geneva, 2001.

355. Ng H. AIDS in Africa: a regional overview. *Harv AIDS Rev* 2000, 2-5.

356. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 1998; **148:**929-936.

357. Nijhuis M, Deeks S, Boucher C. Implications of antiretroviral resistance on viral fitness. *Curr Opin Infect Dis* 2001; **14:**23-28.

358. Nijhuis M, Schuurman R, de Jong D, et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* 1999; **13:**2349-2359.

359. Novitsky V, Flores-Villanueva PO, Chigwedere P, et al. Identification of most frequent HLA class I antigen specificities in Botswana: relevance for HIV vaccine design. Hum Immunol.2001; **62:**146-156.

360. Novitsky VA, Montano MA, McLane MF, et al. Molecular cloning and phylogenetic analysis of human immunodeficiency virus type I subtype C: a set of 23 full-length clones from Botswana. *J Virol* 1999; **73:**4427-4432.

361. Nowak P, Karlsson AC, Naver L, Bohlin AB, Piasek A, Sonnerborg A. The selection and evolution of viral quasispecies in HIV-1 infected children. *HIV Med* 2002; **3:**1-11.

362. Nowak MA, Bonhoeffer S, Shaw GM, et al. Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations. *J Theor Biol* 1997; **184:**203-217.

363. Obel AO, Sharif SK, McLigeyo SO, et al. Acquired immunodeficiency syndrome in an African. *East Afr Med J* 1984; **61:**724-726.

364. Olsen DB, Stahlhut MW, Rutkowski CA, et al. Non-active site changes elicit broad-based cross-resistance of the HIV-1 protease to inhibitors. *J Biol Chem* 1999; **274:**23699-23701.

365. Op de Coul E, Coutinho RA, van der Schoot A, et al. The impact of immigration on env HIV-1 subtype distribution among heterosexuals in the Netherlands: influx of subtype B and non-B strains. *AIDS* 2001; **15:**2277-2286.

366. Oude EB, Back NK, Berkhout B. Increased polymerase fidelity of the 3TC-resistant variants of HIV-1 reverse transcriptase. *Nucleic Acids Res* 1997; **25:**3212-3217.

367. Palella FJ, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N Engl J Med* 1998; **338:**853–860.

368. Palmer S, Boltz V, Maldarelli F, et al: Emergence and long-term persistence of NNRTI-resistant variants in patients starting and stopping NNRTI-containing regimens. [abstract 37]. *11th Conference on Retroviruses and Opportunistic Infections*, San Francisco, 2004.

369. Palmer S, Margot N, Gilbert H, et al. Tenofovir, adefovir, and zidovudine susceptibilities of primary human immunodeficiency virus type 1 isolates with non-B subtypes or nucleoside resistance. *AIDS Res Hum Retroviruses* 2001; **17:**1167-1173.

370. Palmer S, Shafer RW, Merigan TC. Highly drug-resistant HIV-1 clinical isolates are cross-resistant to many antiretroviral compounds in current clinical development. *AIDS* 1999; **13:**661-667.

371. Para MF, Glidden DV, Coombs RW, et al. Baseline human immunodeficiency virus type 1 phenotype, genotype, and RNA response after switching from long-term hard-capsule saquinavir to indinavir or soft-gel-capsule saquinavir in AIDS clinical trials group protocol 333. *J Infect Dis* 2000; **182:**733.

372. Parkin NT, Hellmann NS, Whitcomb JM, Kiss L, Chappey C, Petropoulos CJ. Natural variation of drug susceptibility in wild-type human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 2004; **48:**437-443.

373. Parkin NT, Chappey C, Petropoulos CJ. Improving lopinavir genotype algorithm through phenotype correlations: novel mutation patterns and amprenavir cross-resistance. *AIDS* 2003a; **17:**955-961.

374. Parkin N, Chappey C, Petropoulos C, et al. HIV-1 reverse transcriptase mutations that suppress zidovudine resistance also increase in vitro susceptibility to tenofovir, but not stavudine. *Antivir Ther* 2003b; **8:**34.

375. Parkin NT, Chappey C, Maranta M, et al. Genotypic and phenotypic analysis of a large database of patient samples reveals distinct patterns of cross-resistance. *Antivir Ther* 2001; **6:**49.

376. Parkin NT, Lie YS, Hellmann N, et al. Phenotypic changes in drug susceptibility associated with failure of human immunodeficiency virus type 1 (HIV-1) triple combination therapy. *J Infect Dis* 1999; **180:**865-870.

377. Partaledis JA, Yamaguchi K, Tisdale M, et al. In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol* 1995; **69:**5228-5235.

378. Pasquier C, Cayrou C, Blancher A, et al. Molecular evidence for mother-to-child transmission of multiple variants by analysis of RNA and DNA sequences of human immunodeficiency virus type 1. *J Virol* 1998; **72:**8493–8501.

379. Patick AK, Duran M, Cao Y, et al. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother* 1998; **42:**2637-2644.

380. Patick AK, Mo H, Markowitz M, et al. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob Agents Chemother* 1996; **40:**292-297.

381. Pauwels R, Balzarini J, Baba M, et al. Rapid and automated tetrazolium based colorimetric assay for the detection of anit-HIV compounds. *J Virol Methods* 1988; **20:**309-321.

382. Peeters M, Vincent R, Perret JL, et al. Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-**1:** syncytium-inducing variants seem rare among subtype C HIV-1 viruses. *J Acquir Immune Defic Syndr Hum Retrovirol* 1999; **20:**115–121.

383. Perez-Alvarez L, Thomson MM, Villahermosa ML, et al. HIV-1 subtype G and BG recombinant viruses in Spanish natives: evidence of characteristic mutations in reverse transcriptase and protease. *AIDS* 2001; **15:**1907-1910.

384. Perno CF, Cozzi-Lepri A, Balotta C, et al. Impact of mutations conferring reduced susceptibility to lamivudine on the response to antiretroviral therapy. *Antivir Ther* 2001; **6:**195-198.

385. Petra study team Efficacy of three short-course regimens of zidovudine and lamivudine in preventing early and late transmission of HIV-1 from mother to child in Tanzania, South Africa, and Uganda (Petra study): a randomised, double-blind, placebo-controlled trial. *Lancet* 2002; **359:**1178-1186.

386. Petropoulos C, Chappey C, Parkin NT. High-level resistance to HIV-1 non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the absence of known resistance mutations [H-451]. *43rd InterScience Conference on Antimicrobial Agents and Chemotherapy.* Chicago, IL, 2003

387. Petropoulos CJ, Parkin NT, Limoli KL, et al. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother* 2000; **44:**920-928.

388. Pettit SC, Gulnik S, Everitt L, et al. The dimmer interfaces of protease and extra-protease domains influence the activation of protease and the specificity of GagPol cleavage. *J Virol* 2003; **77:**366–374.

389. Pettit SC, Henderson GJ, Schiffer CA, et al. Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit of enhance the rate of cleavage by viral protease. *J Virol* 2002; **76:**10226-10233.

390. Pettit SC, Moody MD, Wehbie RS, *et al.* The p2 domain of human immuno-deficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J Virol* 1994; **68:**8017-8027.

391. Pieniazek D, Rayfield M, Hu DJ, et al. Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group. *AIDS* 2000;**14:**1489-1495.

392. Pillay C, Gray G, G Stevens, et al. Emergence of resistance mutations in children treated with didanosine plus stavudine after treatment to prevent mother-to-child transmission. *Antiviral Ther* 2002a; **7:**S61.

393. Pillay D et al. The impact of transmitted resistance on time to CD4< 350 cells/ml. *Antivir Ther* 2002b; **7:** 147.

394. Ping LH, Nelson JA, Hoffman IF, et al. Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J Virol* 1999; **73:**6271-6281.

395. Polzer S, Dittmar MT, Schmitz H, et al. Loss of N-linked glycans in the V3-loop region of gp120 is correlated to an enhanced infectivity of HIV-1. *Glycobiology* 2001; **11:**11-19.

396. Ponnighaus JM, Fine PE, Bliss L. The Karonga Prevention Trial: a leprosy and tuberculosis vaccine trial in northern Malawi. I. Methods of the vaccination phase. *Lepr Rev* 1993; **64:**338–356.

397. Ponnighaus JM, Fine PE, Bliss L, et al. The Lepra Evaluation Project (LEP), an epidemiological study of leprosy in northern Malawi. *Methods Lepr Rev* 1987; **58:**359–375.

398. Posada DK, Crandall A. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 1998; **14:**817-818.

399. Prado JG, Wrin T, Beauchaine J, et al. Amprenavir-resistant HIV-1 exhibits lopinavir cross-resistance and reduced replication capacity. *AIDS* 2002; **16:**1009-1017.

400. Pybus OG, Charleston MA, Gupta S, et al. The epidemic behaviour of the hepatitis C virus. *Science* 2001; **292:**2323-2325.

401. Qari SH, Pieniazek D, Heneine W. Resistance-related polymorphisms in HIV-1 non-B subtype protease influence the resistance pathway and amplify resistance to protease inhibitors. *Antivir Ther* 2004; **9:**S45.

402. Quinn JB, Borroto-Esoda K, Hinkle J, et al. Overview of the genotypic findings form emtricitabine-treated HIV+ patients [abstract H-908]. *43rd InterScience Conference on Antimicrobial Agents and Chemotherap*, Chicago, IL, 2003.

403. Quinones-Mateu ME, Arts EJ. Fitness of drug resistant HIV-**1:** methodology and clinical implications. *Drug Resist Updat* 2002; **5:**224-233.

404. Quinones-Mateu ME, Ball SC, Marozsan J, et al. A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. *J Virol* 2000; **74:**9222-9233.

405. Raffi F, Reliquet V, Ferre V, et al. The VIRGO study: nevirapine, didanosine and stavudine combination therapy in antiretroviral-naïve HIV-1-infected adults. *Antivir Ther* 2000; **5:**267-272.

406. Rambaut A, Posada D, Crandall K, et al. The causes and consequences of HIV evolution. *Nature Genetics* 2004; **5:**52-61.

407. Rambaut A. Estimating the rate of molecular evolution: Incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 2000; **16:**395-399.

408. Ravaux I, et al. Persistence in HIV-1 protease of resistance mutations in absence of drug selective pressure three years after sexual transmission of a multiclass drug resistant variant. *Antivir Ther* 2003; **8:** S415.

409. Rayfield MA, Downing RG, Baggs J, et al. A molecular epidemiologic survey of HIV in Uganda. HIV Variant Working Group. *AIDS* 1998; **12:**521-527.

410. Reinis M, Vandasova J, Stankova M, et al. Human immunodeficiency virus 1 strains resistant to nucleoside inhibitors of reverse transcriptase in isolates from the Czech Republic as monitored by line probe assay and nucleotide sequencing. *Acta Virol* 2001; **45:**279-286.

411. Renjifo B, Chaplin B, Mwakagile M, et al. Epidemic expansion of HIV type 1 subtype C and recombinant genotypes in Tanzania. *AIDS Res Hum Retroviruses*. 1998; **14:**635-638.

412. Rhee SY, Gonzales MJ, Kantor R, et al. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003; **31:**298-303.

413. Richardson BA, Hughes JP. Modeling breastmilk infectivity in HIV-1 infected mothers. *Biometrics* 2003; **59:**179-185.

414. Richman D. Principles of HIV resistance testing and overview of assay performance characteristics. *Antivir Ther* 2000; **5:** 27-31.

415. Richman DD. Resistance, drug failure, and disease progression. *AIDS Res Hum Retroviruses* 1994; **10:**901-905.

416. Richman DD, Havlir D, Corbeil J, et al. Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. *J Virol* 1994; **68:**1660-1666.

417. Richman DD. Antiretroviral drug resistance. *AIDS* 1991; **5:**S189-94.

418. Ridky, T. W., C. E. Cameron, J. Cameron, J. Leis, T. Copeland, A. Wlodawer, I. T. Weber, and R. W. Harrison. 1996. Human immunodeficiency virus, type 1 protease substrate specificity is limited by interactions between substrate amino acids in adjacent enzyme subsites. *J Biol Chem* **271:**4709–4717.

419. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science* 2000; **288:**55-56.

420. Robertson DL, Sharp PM, McCutchan FE, et al. Recombination in HIV-1. Nature 1995; **374:**124-126.

421. Rodenburg CM, Li Y, Trask SA, et al. The UNAIDS and NIAID Networks for HIV Isolation and Characterization. Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. *AIDS Res Hum Retroviruses* 2001; **17:**161-168.

422. Rollins NC, Dedicoat M, Danaviah S, et al. Prevalence, incidence, and mother-to-child transmission of HIV-1 in rural South Africa. *Lancet* 2002; **360:**389.

423. Rose RE, Gong YF, Greytok JA, et al. Human immunodeficiency virus type 1 viral background plays a major role in development of resistance to protease inhibitors. *Proc Natl Acad Sci USA* 1996; **93:**1648-1653.

424. Ross L, Johnson M, DeMasi R, et al. Viral genetic heterogeneity in HIV-1-infected individuals is associated with increasing use of HAART and higher viremia. *AIDS* 2000; **14:**813-819.

425. Rost B, Sander C, Schneider R. PHD-an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci* 1994; **10:**53-60.

426. Rousseau CM, Nduati RW, Richardson BA, et al. Longitudinal analysis of human immunodeficiency virus type 1 RNA in breast milk and of its relationship to infant infection and maternal disease. *J Infect Dis* 2003;**187:**741-747.

427. Rousseau MN, Vergne L, Montes B, et al. Patterns of resistance mutations to antiretroviral drugs in extensively treated HIV-1-infected patients with failure of highly active antiretroviral therapy. *J Acquir Immune Defic Syndr* 2001; **26:**36-43.

428. Sabino EC, Shpaer EG, Morgado MG, et al. Identification of human immunodeficiency virus type 1 envelope genes recombinant between subtypes B and F in two epidemiologically linked individuals from Brazil. *J Virol* 1994; **68:**6340-6346.

429. Salminen, M, Carr J, Burke D, et al. Identification of breakpoints in intergenotypic recombinants of HIV-1 by Bootscanning. *AIDS Res Hum Retroviruses* 1995; **11:**1423-1425.

430. Saloojee H, Violari A. Regular review: HIV infection in children. *BR MED J* 2001; **323:**670-674.

431. Sarafianos SG, Clark ADJ, Tuske S, et al. Trapping HIV-1 reverse transcriptase before and after translocation on DNA. *J Biol Chem* 2003; **278:**16280-16288.

432. Sarafianos SG, Das K, Ding J, Boyer PL, Hughes SH, Arnold E. Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *ChemBiol* 1999a; **6:**R137-46.

433. Sarafianos SG, Das K, Clark ADJ, et al. Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc Natl Acad Sci USA* 1999b; **96:**10027-10032.

434. Saraswathy TS, Ng KP, Sinniah M. Human immunodeficiency virus type 1 subtypes among Malaysian intravenous drug users. *Southeast Asian J Trop Med Public Health* 2000; **31:**283-286.

435. Sarkar S, Das N, Panda S, et al. Rapid spread of HIV among injecting drug users in north-eastern states of India. *Bull Narc* 1993; **45:**91-105.

436. Schapiro JM, Winters MA, Stewart F, et al. The effect of highdose saquinavir on viral load and CD4+ T-cell counts in HIV-infected patients. *Ann Intern Med* 1996; **124:**1039-1050.

437. Schinazi RF, Lloyd RMJ, Nguyen MH, et al. Characterization of human immunodeficiency viruses resistant to oxathiolane-cytosine nucleosides. *Antimicrob Agents Chemother* 1993; **37:**875-881.

438. Schinazi RF, Larder BA, Mellors JW. Mutations in retroviral genes associated with drug resistance: 2000-2001 update. *Int Antivir News* 2000; **5:**65-91.

439. Schooley RT, Clumeck N, Haubrich R, et al. A dose-ranging study to evaluate the antiretroviral activity and safety of amprenavir alone and in combination with abacavir in HIV-infected adults with limited antiretroviral experience. *Antivir Ther* 2001; **6:**89-96.

440. Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003; **31:** 3381-3385.

441. Scott WR, Schiffer CA. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure Fold Des* 2000; **8:**1259-1265.

442. Selmi B, Deval J, Boretto J, et al. Nucleotide analogue binding, catalysis and primer unblocking in the mechanisms of HIV-1 reverse transcriptase-mediated resistance to nucleoside analogues. *Antivir Ther* 2003; **8:**143-154.

443. Servais J, Plesseria JM, Lambert C, et al. Genotypic correlates of resistance to HIV-1 protease inhibitors on longitudinal data: the role of secondary mutations. *Antivir Ther*2001; **6:**239-248.

444. Seth P, Sharma UK. Recovery of human immunodeficiency virus from asymptomatic prostitutes from Tamil Nadu. *Indian J Med Res* 1991; **93:**277-279.

445. Sevin AD, DeGruttola V, Nijhuis M, et al. Methods for investigation of the relationship between drug-susceptibility phenotype and human immunodeficiency virus type 1 genotype with applications to *AIDS* clinical trials group 333. *J Infect Dis* 2000; **182:**59-67.

446. Shafer R. Genotypic Testing for HIV-1 Drug Resistance. 2003. viewed July 2003, <http://hivdb.stanford.edu/>

447. Shafer RW, Kantor R Gonzales J. The genetic basis of HIV-1 resistance to reverse transcriptase and protease inhibitors. *AIDS Rev* 2000; **2:**211-228.

448. Shafer RW, Hsu P, Patick AK, et al. Identification of biased amino acid substitution patterns in human immunodeficiency virus type 1 isolates from patients treated with protease inhibitors. *J Virol* 1999; **73:**6197-6202.

449. Shafer RW, Winters MA, Palmer S, Merigan TC. Multiple concurrent reverse transcriptase and protease mutations and multidrug resistance of HIV-1 isolates from heavily treated patients. *Ann Intern Med* 1998; **128:**906-911.

450. Shafer RW, Eisen JA, Merigan TC, et al. Sequence and drug susceptibility of subtype C reverse transcriptase from human immunodeficiency virus type 1 seroconverters in Zimbabwe. *J Virol* 1997; **71:**5441-5448.

451. Shafer RW, Winters MA, Iversen AK, Merigan TC. Genotypic and phenotypic changes during culture of a multinucleoside-resistant human immunodeficiency virus type 1 strain in the presence and absence of additional reverse transcriptase inhibitors. *Antimicrob Agents Chemother* 1996; **40:**2887-2890.

452. Shafer RW, Kozal MJ, Winters MA, et al. Combination therapy with zidovudine and didanosine selects for drug-resistant human immunodeficiency virus type 1 strains with unique patterns of pol gene mutations. *J Infect Dis* 1994; **169:**722-729.

453. Shah FS, Curr KA, Hamburgh ME, et al. Differential influence of nucleoside analog-resistance mutations K65R and L74V on the overall mutation rate and error specificity of human immunodeficiency virus type 1 reverse transcriptase. *J Biol Chem* 2000; **275:**27037-27044.

454. Sham HL, Kempf DJ, Molla A, et al. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob Agents Chemother* 1998; **42:**3218-3224.

455. Shankarappa R, Chatterjee R, Learn GH, et al. Human immunodeficiency virus type 1 env sequences from Calcutta in Estearn India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. *J Virol* 2001; **75:**10479-10487.

456. Shao XW, Malmsten A, Lennerstrand J, et al. Use of HIV-1 reverse transcriptase recovered from human plasma for phenotypic drug susceptibility testing. *AIDS* 2003; **17:**1463-1471.

457. Shao W, Everitt L, Manchester M, et al. Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants. *Proc Natl Acad Sci USA* 1997; **94:**2243-2248.

458. Sharma B, Kaushik N, Upadhyay A, et al. A positively charged side chain at position 154 on the beta8-alphaE loop of HIV-1 RT is required for stable ternary complex formation. *Nucleic Acids Res* 2003; **31:**5167-5174.

459. Sharma PL, Crumpacker CS. Decreased processivity of human immunodeficiency virus type 1 reverse transcriptase (RT) containing didanosine-selected mutation Leu74Val: a comparative analysis of RT variants Leu74Val and lamivudine-selected Met184Val. *J Virol* 1999; **73:**8448-8456.

460. Sharma PL, Crumpacker CS. Attenuated replication of human immunodeficiency virus type 1 with a didanosine-selected reverse transcriptase mutation. *J Virol* 1997; **71:**8846-8851.

461. Sharp PM, Robertson DL, Gao F, et al. Origins and diversity of human immunodeficiency viruses. *AIDS* 1994; **8:**S27-42.

462. Shehu-Xhilaga M, Kraesslich HG, Pettit S, et al. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol* 2001; **75:**9156–9164.

463. Sher A, Gazzinelli RT, Jankovic D, et al. Cytokines as determinants of disease and disease interactions. *Braz J Med Biol Res* 1998; **31:**85-87.

464. Sher R HIV infection in South Africa, 1982-1988-a review. *S Afr Med J* 1989; **76:**314-318.

465. Sherefa K, Sallberg M, Sonnerborg A: Evidence of no change in V3 loop antibody recognition pattern in HIV type 1 infected Ethiopians between 1988 and 1993. *AIDS Res Hum Retroviruses* 1994; **10:**1551–1556.

466. Shibata R, Siemon C, Czajak SC, et al. Live, attenuated simian immunodeficiency virus vaccines elicit potent resistance against a challenge with a human immunodeficiency virus type 1 chimeric virus. *J Virol* 1997; **71:**8141-8148.

467. Shirasaka T, Kavlick MF, Ueno T, et al. Emergence of human immunodeficiency virus type 1 variants with resistance to multiple dideoxynucleosides in patients receiving therapy with dideoxynucleosides. *Proc Natl Acad Sci USA* 1995; **92:**2398-2402.

468. Shulman NS, Machekano RA, Shafer RW, et al. Genotypic correlates of a virologic response to stavudine after zidovudine monotherapy. *J Acquir Immune Defic Syndr* 2001; **27:**377-380.

469. Siepel AC, Halpern AL, Macken C, et al. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res Hum Retroviruses* 1995; **11:**1413-1416.

470. Simmonds P, Zhang LQ, McOmish F, et al. Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. *J Virol* 1991; **65:**6266–6276.

471. Simon F, Mauclere P, Roques P, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat Med* 1998; **4:**1032-1037.

472. Sista PR, Melby T, Greenberg M, et al. Characterization of baseline and treatment-emergent resistance to T- 20 (enfuvirtide) observed in phase II clinical trials: substitutions in gp41 amino acids 36-45 and enfuvirtide susceptibility of virus isolates. *Antivir Ther* 2002; **7:**S16-S17.

473. Sitas F, Newton R. Kaposi's sarcoma in South Africa. *J Natl Cancer Inst Monogr* 2001; (28):1-4.

474. Smith SW, Overbeek CR, Woese W, et al. The Genetic Data Environment: An expandable GUI for multiple sequence analysis. *Comput Appli Biosci* 1994; **10:**671-675.

475. Soares MA, De Oliveira T, Brindeiro RM, et al. A specific subtype C of human immunodeficiency virus type 1 circulates in Brazil. *AIDS* 2003a; **17:**11-21.

476. Soares EA, Santos RP, Pellegrini JA, et al. Epidemiologic and molecular characterization of human immunodeficiency virus type 1 in southern Brazil. *J Acquir Immune Defic Syndr* 2003b; **34:**520-526.

477. Soto-Ramirez LE, Renjifo B, Mclane MF, et al. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. *Science* 1996; **271:**1291-1293.

478. Spence RA, Kati WM, Anderson KS, et al. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* 1995; **267:**988-993.

479. Spira S, Wainberg MA, Loemba H, et al. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrob Chemother* 2003; **51:**229-240.

480. Srikanth P, John TJ, Jeyakumari H, et al. Epidemiological features of acquired immunodeficiency syndrome in southern India. *Indian J Med Res* 1997; **105:**191-197.

481. St Clair MH, Martin JL, Tudor-Williams G, et al. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science* 1991; **253:**1557-1559.

482. Stammers T. Do condoms actively prevent non-HIV STIs? *Br J Gen Pract* 2001; **51:**229.

483. Staszewski S, Keiser P, Montaner J, et al. Abacavir-lamivudine-zidovudine vs indinavir-lamivudine-zidovudine in antiretroviral-naive HIV-infected adults: A randomized equivalence trial. *JAMA* 2001; **285:**1155-1163.

484. Strain MC, Gunthard HF, Havlir DV, et al. Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: Intrinsic stability predicts lifelong persistence *Proc Natl Acad Sci USA* 2003; **100:**4819–4824.

485. Sturmer M, Staszewski S, Doerr H-W, et al. Correlation of Phenotypic Zidovudine Resistance with Mutational Patterns in the Reverse Transcriptase of Human Immunodeficiency Virus Type **1:** Interpretation of Established Mutations and Characterization of New Polymorphisms at Codons 208, 211, and 214. *Antimicrob Agents Chemother* 2003; **47:**54–61.

486. Swanstrom R, Wills JW. Retroviral gene expression. II. Synthesis, processing, and assembly of viral proteins, p. 263–334. In J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses.Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y, 1997

487. Swofford D. PAUP 4.0: Phylogenetic analysis using parsimony (and other methods), 4.0b2a. Sunderland, Mass: Sinauer Associates, Inc, 1999.

488. Tanuri A, Caridea E, Dantas MC, et al. Prevalence of mutations related to HIV-1 antiretroviral resistance in Brazilian patients failing HAART. *J Clin Virol* 2002; **25:**39-46.

489. Tanuri A, Jesus dC, Brindeiro R, et al. Construction of a selectable nef-defective live-attenuated human immunodeficiency virus expressing Escherichia coli gpt gene. *Virology* 2000; **268:**79-86.

490. Tanuri A, Vicente AC, Otsuki K, et al. Genetic variation and susceptibilities to protease inhibitors among subtype B and F isolates in Brazil. *Antimicrob Agents Chemother* 1999; **43:**253-258.

491. Tatt ID, Barlow KL, Nicool A, et al. The public health significance of HIV-1 subtypes. *AIDS* 2001; **15:**S59-S71.

492. Tessmer U, Krausslich HG. Cleavage of human immunodeficiency virus type 1 proteinase from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J Virol* 1998; **72:**3459-3463.

493. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22:**4673-4680.

494. Tien PC, Chiu T, Latif A, et al. Primary subtype C HIV-1 infection in Harare, Zimbabwe. *J Acquir Immune Defic Syndr Hum Retrovirol* 1999; **20:**147-153.

495. Tisdale M, Alnadaf T, Cousens D. Combination of mutations in human immunodeficiency virus type 1 reverse transcriptase required for resistance to the carbocyclic nucleoside 1592U89. *Antimicrob Agents Chemother* 1997; **41:**1094-1098.

496. Tisdale M, Myers RE, Maschera B, et al. Cross-resistance analysis of human immunodeficiency virus type 1 variants individually selected for resistance to five different protease inhibitors. *Antimicrob Agents Chemother* 1995; **39:**1704-1710.

497. Tisdale M, Kemp SD, Parry NR, et al. Rapid in vitro selection of human immunodeficiency virus type 1 resistant to 3'-thiacytidine inhibitors due to a mutation in the YMDD region of reverse transcriptase. *Proc Natl Acad Sci USA* 1993; **90:**5653-5656.

498. Tobin NH, Frenkel LM. Human immunodeficiency virus drug susceptibility and resistance testing. *Pediatr Infect Dis J* 2002; **21:**681-683.

499. Tomasselli AG, Heinrickson RL. Specificity of retroviral proteases: an analysis of viral and nonviral protein substrates. *Methods Enzymol* 1994; **241:**279–301.

500. Tong W, Lu CD, Sharma SK, et al. Nucleotide-induced stable complex formation by HIV-1 reverse transcriptase. *Biochem* 1997; **36:**5749-5757.

501. Torti C, Pozniak A, Nelson M, et al. Distribution of K103N and/or Y181C HIV-1 mutations by exposure to zidovudine and non-nucleoside reverse transcriptase inhibitors. *J Antimicrob Chemother* 2001; **48:**113-116.

502. Tozser J, Bagossi P, Boross P, et al. Effect of serine and tyrosine phosphorylation on retroviral proteinase substrates. *Eur J Biochem* 1999; **265:**423-429.

503. Trask SA, Derdeyn CA, Fideli U, et al. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. *J Virol* 2002; **76:**397-405.

504. Tscherning C, Alaeus A, Fredriksson R, et al. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* 1998; **241:**181-188.

505. Tsega E, Mengesha B, Nordenfelt E, et al. Serological survey of human immunodeficiency virus infection in Ethiopia. *Ethiop Med J* 1988; **26:**179-184.

506. Tsuchie H, Saraswathy TS, Sinniah M, et al. HIV-1 variants in South and South-East Asia. *Int J STD AIDS* 1995; **6:**117–120.

507. Turner D, Brenner B, Moisi D, et al. Nucleotide and amino acid polymorphisms at drug resistance sites in non-B subtype HIV-1 variants. *Antivir Ther* 2004; **9:**S99.

508. UK Collaborative Group on Monitoring the Transmission of HIV Drug Resistance. Analysis of prevalence of HIV-1 drug resistance in primary infections in the United Kingdom. *Br Med J* 2001; **322:**1087-1088.

509. UNAIDS/WHO Report on the global AIDS epidemic-update 2002. UNAIDS, Geneva, 2002. <http://www.unaids.org/epidemicupdatedec01/report /index_html>

510. UNAIDS 2001. Regional HIV/AIDS statistics and features, end of 2001. UNAIDS, Geneva, 2001.

511. UNAIDS/WHO working group on global HIV/AIDS and STD surveillance: Report on the global HIV/AIDS epidemic, December 2000. UNAIDS, Geneva, 2000.

512. UNAIDS/World Health Organization 1998. HIV/AIDS: regional statistics and features, December 1997. In Global HIV/AIDS and STD surveillance 1998.

513. US Department of Health and Human Services Panel on Clinical Practices for Treatment of HIV Infection A. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents (The living document, November 10, 2003), 2003 <http://www.AIDSinfo.nih.gov/guidelines/>

514. Van Harmelen J, Williamson C, Kim B, et al. Characterization of full-length HIV-1 type 1 subtype C sequences from South Africa. *AIDS Res Hum Retroviruses* 2001; **17:**1527-1531.

515. Van Laethem K, Schmit JC, Pelemans H, et al. Presence of 2',5'-Bis-O-(tert-butyldimethylsilyl)-3'-spiro-5"-(4"-amino-1",2"-oxath iole-2",2"-dioxide) (TSAO)-resistant virus strains in TSAO-inexperienced HIV patients. *AIDS Res Hum Retroviruses* 2000; **16:**825-833.

516. Van Harmelen JH, Van der Ryst E, Loubser AS, et al. A predominantly HIV-1 subtype C-restricted epidemic in South African urban populations. *AIDS Res Hum Retroviruses* 1999; **15:**395-398.

517. van Harmelen J, Wood R, Lambrick M, et al. An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 1997; **11:**81-87.

518. Van Laethem K, Witvrouw M, Pannecouque C, et al. Mutations in the non-nucleoside binding-pocket interfere with the multi-nucleoside resistance phenotype. *AIDS* 2001; **15:**553-561.

519. van Rensburg EJ, Smith TL, Zeier M, et al. Change in co-receptor usage of current South African HIV-1 subtype C primary isolates. *AIDS* 2002; **16:**2479-2480.

520. Vandamme AM, Houyez F, Banhegyi D, et al. Laboratory guidelines for the practical use of HIV drug resistance tests in patient follow-up. *Antivir Ther* 2001; **6:**21-39.

521. Velazquez-Campoy A, Todd MJ, Vega S, et al. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. *Proc Natl Acad Sci USA* 2001; **98:**6062-6067.

522. Velazquez-Campoy A, Vega S, Fleming E, et al. Protease inhibition in African subtypes of HIV-1. *AIDS Rev* 2003; **5:**165-171.

523. Velazquez-Campoy A, Vega S, Freire E. Amplification of the effects of drug resistance mutations by background polymorphisms in HIV-1 protease from African subtypes. *Biochemistry* 2002; **41:**8613-8619.

524. Verhofstede C, Demecheleer E, De Cabooter N, et al. Diversity of the Human Immunodeficiency Virus Type 1 (HIV-1) env Sequence after Vertical Transmission in Mother-Child Pairs Infected with HIV-1 Subtype A. *J Virol* 2003; **77:**3050-3057.

525. Verhofstede C, Noe A, Demecheleer E, et al. Drug-Resistant Variants That Evolve During Nonsuppressive Therapy Persist in HIV-1-Infected Peripheral Blood Mononuclear Cells After Long-Term Highly Active Antiretroviral Therapy. *J Acquir Immune Defic Syndr* 2004; **35:**473-483.

526. Vidal C, Arnedo M, Garcia F, et al. Genotypic and phenotypic resistance patterns in early-stage HIV-1-infected patients failing initial therapy with stavudine, didanosine and nevirapine. *Antivir Ther* 2002; **7:**283-287.

527. Wainberg MA. HIV resistance to nevirapine and other non-nucleoside reverse transcriptase inhibitors. *J Acquir Immune Defic Syndr* 2003; **34:**S2-7.

528. Wainberg MA, White AJ. Current insights into reverse transcriptase inhibitor-associated resistance. *Antivir Ther* 2001; **6:**11-19.

529. Wainberg M. HIVresistanceWeb: Is Resistance to nevirapine a serious consideration in studies of the prevention of perinatal transmission www.hivresistanceweb.com/protected/po/dk-00apr-vert.shtml. May 2000.

530. Wainberg MA, Miller MD, Quan Y, et al. In vitro selection and characterization of HIV-1 with reduced susceptibility to PMPA. *Antiviral Ther* 1999; **4:**87-94.

531. Wainberg MA, Hsu M, Gu Z, et al. Effectiveness of 3TC in HIV clinical trials may be due in part to the M184V substitution in 3TC-resistant HIV-1 reverse transcriptase. *AIDS* 1996; **10:**S3-10.

532. Wang D, Larder B. Enhanced prediction of lopinavir resistance from genotype by use of artificial neural networks. *J Infect Dis* 2003; **188:**653-660.

533. Wang D, Larder B, Revell A, Harrigan R, Montaner J. 2003. A neural network model using clinical cohort data accurately predicts virological response and identifies regimens with increased probability of success in treatment failures. *Antivir Ther* 2003; **8:**112.

534. Wei X, Decker JM, Liu H, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob Agents Chemother* 2002; **46:**1896-1905.

535. Wei X, Ghosh SK, Taylor ME, et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 1995; **373:**117-122.

536. Weidle PJ, Kityo CM, Mugyenyi P, et al. Resistance to antiretroviral therapy among patients in Uganda. *J Acquir Immune Defic Syndr* 2001; **26:**495-500.

537. Wensing AM, Boucher CA. Worldwide transmission of drug-resistant HIV. *AIDS Rev* 2003; **5:**140-155.

538. Whitcomb JM, Parkin NT, Chappey C, et al. Broad nucleoside reversetranscriptase inhibitor cross-resistance in human immunodeficiency virus type 1 clinical isolates. *J Infect Dis* 2003; **188:**992-1000.

539. Whitcomb JM, Huang W, Limoli K, et al. Hypersusceptibility to non-nucleoside reverse transcriptase inhibitors in HIV-**1:** clinical, phenotypic and genotypic correlates. *AIDS* 2002; **16:**F41-47.

540. White KL, Margot NA, Wrin T, et al. Molecular mechanisms of resistance to human immunodeficiency virus type 1 with reverse transcriptase mutations K65R and K65R+M184V and their effects on enzyme function and viral replication capacity. *Antimicrob Agents Chemother* 2002; **46:**3437-3446.

541. Wilbe K, Salminen M, Laukkanen T, et al. Characterization of novel recombinant HIV-1 genomes using the branching index. *Virol* 2003; **316:**116-125.

542. Williamson C, Engelbrecht S, Lambrick M, et al. HIV-1 subtypes in different risk groups in South Africa. *Lancet* 1995; **346:**782.

543. Wilson JW. Update on antiretroviral drug resistance testing: combining laboratory technology with patient care. *AIDS Read* 2003; **13:**25-30.

544. Winslow DL, Garber S, Reid C, et al. Selection conditions affect the evolution of specific mutations in the reverse transcriptase gene associated with resistance to DMP 266. *AIDS* 1996; **10:**1205-1209.

545. Winters MA, Merigan TC. Variants other than aspartic acid at codon 69 of the human immunodeficiency virus type 1 reverse transcriptase gene affect susceptibility to nuleoside analogs. *Antimicrob Agents Chemother* 2001; **45:**2276-2279.

546. Winters MA, Schapiro JM, Lawrence J, et al. Human immunodeficiency virus type 1 protease genotypes and in vitro protease inhibitor susceptibilities of isolates from individuals who were switched to other protease inhibitors after long-term saquinavir treatment. *J Virol* 1998a; **72:**5303-5306.

547. Winters MA, Coolley KL, Girard YA, et al. A 6-basepair insert in the reverse transcriptase gene of human immunodeficiency virus type 1 confers resistance to multiple nucleoside inhibitors. *J Clin Invest* 1998b; **102:**1769-1775.

548. Winters MA, Shafer RW, Jellinger RA, et al. Human immunodeficiency virus type 1 reverse transcriptase genotype and drug susceptibility changes in infected individuals receiving dideoxyinosine monotherapy for 1 to 2 years. *Antimicrob Agents Chemother* 1997; **41:**757-762.

549. Wolinsky SM, Wike CM, Korber BT, et al. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* 1992; **255:**1134–1137.

550. Womble DD. GCG: The Wisconsin Package of sequence analysis programs. *Methods Mol Biol* 2000; **132:**3-22.

551. Wombel D, Liu X, Bambra C, et al. Genetic polymorphism of envelope V3 region of HIV type 1 subtypes A, C, and D essential for efficient Gag polyprotein processing and viral infectivity. *J Virol* 1996; **72:**3459–3463.

552. Wong JK, Hezareh M, Gunthard HF, et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* 1997; **278:**1291-1295.

553. Wu TD, Schiffer CA, Gonzales MJ, et al. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J Virol* 2003; **77:**4836-4847.

554. Xu L, Hue S, Taylor S, et al. Minimal variation in T-20 binding domain of different HIV-1 subtypes from antiretroviral-naive and -experienced patients. *AIDS* 2002; **16:**1684-1686.

555. Yahi N, Tamalet C, Tourres C, et al. Mutation L210W of HIV-1 reverse transcriptase in patients receiving combination therapy. Incidence, association with other mutations, and effects on the structure of mutated reverse transcriptase. *J Biomed Sci* 2000; **7:**507-513.

556. Yahi N, Tamalet C, Tourres C, et al. Mutation patterns of the reverse transcriptase and protease genes in human immunodeficiency virus type 1-infected patients undergoing combination therapy: survey of 787 sequences. *J Clin Microbiol* 1999; **37:**4099-4106.

557. Yang C, Li M, Shi YP, et al. Genetic diversity and high proportion of intersubtype recombinants among HIV type 1-infected pregnant women in Kisumu, western Kenya. *AIDS Res Hum Retroviruses* 2004; **20:**565-574.

558. Yang C, Li M, Newman RD, et al. Genetic diversity of HIV-1 in western Kenya: subtypes specific differences in mother-to-child transmission *AIDS* 2003; **17:**1667–1674.

559. Yang R, Xia X, Kusagawa S, et al. On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan province of China. *AIDS*. 2002; **16:**1401-1407.

560. Yang, Z. Phylogenetic Analysis by Maximum Likelihood (PAML), version 3.0. University College London, London, England, 2000.

561. Yang Z, Nielsen R, Goldman N, et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000; **155:**431-449.

562. Yang Z. A space-time process model for the evolution of DNA sequences. *Genetics* 1995; **139:**993-1005.

563. Yerly S, Vora S, Rizzardi P, et al. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. *AIDS* 2001; **15:**2287-2292.

564. Young SD, Britcher SF, Tran LO, et al. L-743, 726 (DMP-266): a novel, highly potent nonnucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* 1995; **39:**2602-2605.

565. Yu XF, Chen J, Shao Y, et al. Two subtypes of HIV-1 among injection-drug users in southern China. *Lancet* 1998; **351:**1250.

566. Zennou V, Mammano F, Paulous S, et al. Loss of viral fitness associated with multiple Gag and Gag-Pol processing defects in human immunodeficiency virus type 1 variants selected for resistance to protease inhibitors in vivo. *J Virol* 1998; **72:**3300-3306.

567. Zewdie DW: *AIDS* Control Program and Research in Ethiopia [abstract 75]. *24th Ethiopian Annual Medical Conference*, Addis Ababa, 1988.

568. Zhang D, Caliendo AM, Eron JJ, et al. Resistance to 2',3'-dideoxycytidine conferred by a mutation in codon 65 of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* 1994; **38:**282-287.

569. Zhang H, Orti G, Du Q, et al. Phylogenetic and phenotypic analysis of HIV type 1 env gp120 in cases of subtype C mother-to-child transmission. *AIDS Res Hum Retroviruses* 2002; **18:**1415-1423.

570. Zhang H, Dornadula G, Wu Y, et al. Kinetic analysis of intravirion reverse transcription in the blood plasma of human immunodeficiency virus type 1 infected individuals: direct assessment of resistance to reverse transcriptase inhibitors in vivo. *J Virol* 1996; **70:**628-634.

571. Zhang L, Ramratnam B, Tenner-Racz K, et al. Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med* 1999; **340:** 1605–1613.

572. Zhang LQ, MacKenzie P, Cleland A, et al. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* 1993; **67:**3345–3356.

573. Zhang YM, Imamichi H, Imamichi T, et al. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol.* 1997; **71:**6662-6670.

574. Zhu TF et al. An African HIV-1 sequence from 1959 and implication for the origin of the epidemic. *Nature* 1998; **391:**594-597.

575. Ziermann R, Limoli K, Das K, et al. A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J Virol* 2000; **74:**4414-4419.

576. Zolopa AR, Shafer RW, Warford A, et al. HIV-1 genotypic resistance patterns predict response to saquinavir-ritonavir therapy in patients in whom previous protease inhibitor therapy had failed. *Ann Intern Med* 1999; **131:**813-821.

577. Zur Megede J, Engelbrecht S, De Oliveira T, et al. Novel evolutionary analyses of full-length HIV-1 subtype C molecular clones from Cape Town, South Africa. *AIDS Res Hum Retroviruses* 2002; **18:**1327-1332.

578. Zybarth G, Carter C. Domains upstream of the protease (PR) in human immunodeficiency virus type I Gag-Pol influence PR autoprocessing. *J Virol* 1995; **69:**3878–3884.

## Appendix A

The control files are described below.

- Seqfile, outfile and treefile specifies the names of the sequence data file, main result file, and the tre structure file respectively.
- Noisy controls how much output you want on the screen
- Verbose controls how much output is in the result file
- Runmode = 0 means evaluation of the tree topologies specified in the trees structure file. Runmode = 1 or 2 means heirustic tree search by the star-decomposition algorithm etc.
- Model specifies the model of nucleotide substitution
- Mgene is used for combined analysis of data from multiple genes or the three codon positions.
- Fix_kappa specifies whether k in K80, F84 or HKY85 is given at a fixed value or is to be estimated by iteration from the data.
- Fix_alpha and alpha work in a similar way, where alpha refers to the shape parameter α of the gamma distribution for variable substitution rates across sites (Yang, 1994)
- Fix_rho and rho work in a similar way and concern independence or correlation of rates at adjacent sites, where rho is the correlation parameter of the auto-discrete gamma model (Yang, 1995).
- NparK specifies nonparametric models for variable and Markov-dependent rates across sites: npark=1 or 2 means several categories of independent rates for sites, while npark=3 or 4 means the rates are Markov-dependent at adjacent sites.
- Clock specifies models concerning rate constancy among lineages.
- Nhomo is for baseml only and concerns the frequency parameters in the F81, F84, HKY85, TN93 or REV models.
- GetSE indicates whether we wnat estimates of the standard errors of estimated parameters.
- RateAncestor=1 works with runmode=0. Rates are calculated for sites along the sequence (output in rates file) and performs marginal ancestral reconstruction (output in rst file)
- Small_Diff is a small value used in the difference approxiamtion of derivatives.
- Cleandata = 1 means sites involving ambiguity characters or alignment gaps are removed from all sequences. This leads to faster calculation.
- Method: this variable controls the iteration algorithm for estimating branch lengths under a model of no clock. method = 0 implements the old algorithm in PAML, which updates all parameters including branch lengths simultanwously.Method = 1 specifies an algorithm which updates branch lengths one by one. Method=1 does not work under the clock models.
- Ndata: specifies the number of separate data sets in the file.