The role of chemokine and chemokine receptor genes in genetic susceptibility to HIV infection in South Africa

Desiree C. Petersen

Thesis presented in fulfilment of the requirements for the degree of Master in Health Sciences (Medical Virology) at the University of Stellenbosch

Co-promoter and study supervisor: Dr. Vanessa M. Hayes
Promoter: Prof. Estrelita Janse van Rensburg

March 2002
Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously, in its entirety or in part, submitted it at any University for a degree.

Signature                     Date
Summary

The importance of host genetic factors in susceptibility for human immunodeficiency virus (HIV) infection and disease progression to acquired immunodeficiency syndrome (AIDS) has been shown with the identification of a 32 base pair (bp) deletion in the coding region of the human CC chemokine receptor 5 gene (CCR5). CCR5 serves as a cellular receptor for HIV entry and together with the CD4+ molecule, facilitates the infection of host target cells. The CCR5 deletion mutation, which results in the formation of a truncated receptor, is associated with resistance to HIV-1 infection when found homozygously and slower disease progression to AIDS when found heterozygously. Together with this “protective” genetic variant, two single nucleotide polymorphic (SNP) sites within the coding region of CC chemokine receptor 2 (CCR2-V64I) and 3' untranslated region (3'UTR) of the CXC chemokine, stromal derived factor-1β (SDF1-3'A) genes, have been associated with influencing disease progression to AIDS. To date, screening of genetic variants in the CCR5, CCR2 and SDF1 genes has been largely restricted to Caucasian population groups and it is thus not clear whether certain mutations and/or SNPs are relatively specific to individual population groups or rarely observed in Africans.

In this study, comprehensive mutation detection assays were designed for the entire coding regions of the CCR5 and CCR2 genes and a partial region of the 3'UTR of SDF1, to identify both known and novel “protective” or “causative” mutations and/or SNPs, which may play a role in genetic susceptibility to HIV-1 infection, within a predominantly African population. The CCR5, CCR2 and SDF1 mutation detection assays, based on DGGE (denaturing gradient gel electrophoresis), allowed for the
complete analysis of 10, 5 and 30 individuals per denaturing gel, respectively. The study cohort consisted of approximately 100 HIV seropositive patients and 150 HIV seronegative controls from the diverse ethnic groups of South Africa. Several novel mutations and SNPs with a possible African origin were identified. These novel mutations or collectively occurring novel SNPs may have a significant effect on the normal functioning and expression of chemokines and chemokine receptors and thus influence host susceptibility to HIV-1 infection and/or disease progression to AIDS.
Opsomming

Die belangrikheid van gasheer genetiese faktore by vatbaarheid vir menslike immuniteitsgebrek virus (MIV) infeksie en siekteprogressie na die verworwe immuniteitsgebrek sindroom (VIGS) is aangetoon deur die identifikasie van 'n 32 basis paar (bp) delesie in die koderende streek van die menslike CC chemokien reseptor (CCR5) geen. CCR5 dien as 'n sellulêre reseptor vir MIV toegang en saam met die CD4 molekule, fasiliteer dit die infeksie van gasheer teikenselle. Die CCR5 delesie mutasie, wat lei tot die vorming van 'n afgekapte reseptor, word geassosieer met weerstand teen MIV-1 infeksie wanneer dit homosigoties voorkom en met vertraagde progressie na VIGS wanneer dit heterosigoties voorkom. Saam met hierdie "beskermende" genetiese variant, is daar twee enkel nukleotied polimorfiese (ENP) setels in die koderende streek van die CC chemokien reseptor 2 (CCR2-V64I) en 3' ongetransleerde streek van die CXC chemokien, stromaal afgeleide faktor-1β (SDF1-3'A) gene, wat siekte progressie na VIGS beïnvloed. Tot dusver, is die sifting van genetiese variante in die CCR5, CCR2 en SDF1 gene hoofsaaklik beperk tot die Kaukasiër populasie groepe en dit is dus onduidelik of sekere mutasies en/of ENPs spesifiek tot verskillende populasie groepe behoort, en of dit selde in Swart populasie groepe waargeneem word.

In hierdie studie is omvattende mutasie opsporings toetse ontwerp vir die volledige koderende streke van die CCR5 en CCR2 gene en 'n gedeelte van die 3' ongetransleerde streek van die SDF1 geen om beide bekende en nuwe mutasies en/of ENPs, wat 'n rol mag speel in genetiese vatbaarheid vir MIV-1 infeksie, in 'n oorheersende Swart populasie te identifiseer. Die CCR5, CCR2 en SDF1 mutasie
opsporings toetse, gebaseer op DGGE (denaturerende gradiënt gel elektroferese), is gebruik vir die volledige analise van 10, 5 en 30 individue onderskeidelik. Die studie groep het bestaan uit ongeveer 100 HIV seropositiewe pasiënte en 150 HIV seronegatiewe kontroles, afkomstig vanuit die diverse etniese groepe in Suid-Afrika. Verskeie nuwe mutasies en ENPs met 'n moontlike oorsprong in Swartmense, is geïdentifiseer. Hierdie nuwe mutasies of ENPs wat saam voorkom kan 'n betekenisvolle effek hê op die normale funksionering en uitdrukking van chemokiene en chemokien reseptore en dus 'n invloed hê op die gasheer se vatbaarheid vir MIV-infeksie en/of progressie na VIGS.
Acknowledgements

I wish to extend my sincere thanks to the following people and institutions:

Prof. Estrelita Janse Rensburg, my promoter, for the opportunity to perform research as a student in your department. Your willingness to offer guidance and encouragement is appreciated.

Dr. Vanessa Hayes, my study supervisor and co-promoter, for your guidance, assistance, support and encouragement. Your faith in me as a student has given me the confidence to develop as a scientist.

Dr. Michele Zeier of Tygerberg hospital, Dr. Ashraf Grimwood of Chapel Street Community Health Clinic and Dr. Eftyhia Vardas of the Medical Research Council (MRC) for the collection of HIV seropositive blood samples and clinical information regarding the patients.

Dr. John Gasson of the Western Province Blood Transfusion Service for the co-ordination of the HIV seronegative control blood samples.

Dr. Susan Engelbrecht for your advice and assistance whenever needed.

Annette Laten for performing DNA sequencing and for assisting with the translation of the Afrikaans summary for this thesis.

The staff of the Departments of Medical Virology and Urology, University of Stellenbosch, for your technical assistance, interest and support.
The South African AIDS Vaccine Initiative (SAAVI), the Poliomyelitis Research Foundation, the Medical Research Council (MRC), the Harry Crossley Foundation, Syfrets (GT Searle) and Unistel Medical Laboratories who supported the studies presented in this thesis.

Alec and Ruby Petersen, my parents, for your care, encouragement and support. Your love and guidance has helped me pursue my career and achieve my goals in life. Cheslyn Petersen, my brother, for your support and interest in my studies.
List of abbreviations

AIDS acquired immunodeficiency syndrome
Ala (A) alanine
Arg (R) arginine
Asn (N) asparagine
Asp (D) aspartic acid
B lymphocytes bone marrow-derived lymphocytes
bp basepair
CCM chemical cleavage of mismatch
CCR CC chemokine receptor
CXCR CXC chemokine receptor
Cys (C) cysteine
DGGE denaturing gradient gel electrophoresis
DHPLC denaturing high-performance liquid chromatography
ECL extracellular loop
EDTA ethylenediaminetetraacetic acid
ELR motif glutamic acid-leucine-arginine motif
env envelope
GC-clamp guanine and cytosine clamp
Gln (Q) glutamine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glu (E)</td>
<td>glutamic acid</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>glycine</td>
</tr>
<tr>
<td>gp</td>
<td>glycoprotein</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine 5'-triphosphate</td>
</tr>
<tr>
<td>HCl</td>
<td>hydrochloric acid</td>
</tr>
<tr>
<td>His (H)</td>
<td>histidine</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>HIV –</td>
<td>HIV seronegative</td>
</tr>
<tr>
<td>HIV +</td>
<td>HIV seropositive</td>
</tr>
<tr>
<td>HLA</td>
<td>human leucocyte antigen</td>
</tr>
<tr>
<td>IL</td>
<td>interleukin</td>
</tr>
<tr>
<td>lle (I)</td>
<td>isoleucine</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodaltons</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>leucine</td>
</tr>
<tr>
<td>LTNP</td>
<td>long-term non-progressor</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>lysine</td>
</tr>
<tr>
<td>MBP</td>
<td>mannose binding protein</td>
</tr>
<tr>
<td>MCP</td>
<td>monocyte chemotactic protein</td>
</tr>
<tr>
<td>Met (M)</td>
<td>methionine</td>
</tr>
<tr>
<td>MIP</td>
<td>macrophage inflammatory protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>M-tropic</td>
<td>macrophage tropic</td>
</tr>
<tr>
<td>NaAc</td>
<td>sodium acetate</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NRAMP1</td>
<td>natural resistance-associated macrophage protein 1</td>
</tr>
<tr>
<td>NSI</td>
<td>non-syncytium inducing</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>proline</td>
</tr>
<tr>
<td>RANTES</td>
<td>regulated on activation normal T cell expressed and secreted</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>SDF1</td>
<td>stromal derived factor-1</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>serine</td>
</tr>
<tr>
<td>SI</td>
<td>syncytium inducing</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSCP</td>
<td>single-stranded conformation polymorphism</td>
</tr>
<tr>
<td>Th</td>
<td>T helper</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>threonine</td>
</tr>
<tr>
<td>T lymphocytes</td>
<td>thymus-derived lymphocytes</td>
</tr>
<tr>
<td>TM</td>
<td>transmembrane</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>tryptophan</td>
</tr>
<tr>
<td>T-tropic</td>
<td>T cell line tropic</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>tyrosine</td>
</tr>
<tr>
<td>UF</td>
<td>urea/formamide</td>
</tr>
<tr>
<td>3' UTR</td>
<td>3' untranslated region</td>
</tr>
<tr>
<td>Valine (V)</td>
<td>valine</td>
</tr>
<tr>
<td>VDR</td>
<td>vitamin D receptor</td>
</tr>
<tr>
<td>V loop</td>
<td>hypervariable loop</td>
</tr>
</tbody>
</table>
# CONTENTS

Summary iii

Opsomming v

Acknowledgements vii

List of abbreviations ix

## Chapter 1: General Introduction 1

1.1. HIV and AIDS 2

1.1.1. HIV infection and AIDS in South Africa 4

1.1.2. Disease progression to AIDS 5

1.2. Chemokine and chemokine receptors 9

1.2.1. Chemokines 11

1.2.2. Chemokine receptors 15

1.2.3. Role of chemokine and chemokine receptors in HIV-1 entry 17

## Chapter 2: Study aim and overview 23

2.1. Aim of the study 24

2.2. Host susceptibility to HIV-1 infection and/or AIDS 25

2.2.1. CCR5 27

2.2.2. CCR2 34

2.2.3. SDF1 37

2.3. Mutation detection 40

2.3.1. Pre-screening methods of mutation detection 41

2.3.1.1. Denaturing gradient gel electrophoresis (DGGE) 41
Chapter 3: CCR5

Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis (DGGE) assay  
(AIDS 15:171-177, 2001)

Chapter 4: CCR2

Novel single nucleotide polymorphisms identified using a comprehensive CCR2-denaturing gradient gel electrophoresis assay  
(Submitted)

Chapter 5: SDF1

Analysis of the SDF1-3' A single nucleotide polymorphism (SNP) within a South African population.

Chapter 6: Discussion

Chapter 7: References

Appendix I: HIV seropositive study cohort

Table 1: Clinical information regarding the HIV seropositive study cohort

Table 2: SNPs analysed individually in the HIV seropositive study cohort

Table 3: SNPs analysed collectively in the HIV seropositive study cohort

Appendix II: Chemical Stock Solutions and Kits

A) Stock solutions used for agarose gel electrophoresis and DGGE

B) Kits used for DNA extraction and DNA sequencing
**Appendix III: Statistical analysis**

1) Allelic frequency  
2) Fisher exact test  
3) Hardy - Weinberg equilibrium (HWE)
Chapter 1

General Introduction
1.1. Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS)

Acquired immunodeficiency syndrome (AIDS) was first diagnosed in 1981, in a few homosexual men [Morbidity and Mortality Weekly Report, 1981a; Morbidity and Mortality Weekly Report, 1981b], but it has spread beyond this risk group and now includes sexually active heterosexual individuals, intravenous drug users, infants, health care workers and haemophiliacs. In 1983 it was clearly demonstrated that the human immunodeficiency virus (HIV) is the infectious agent that causes AIDS [Barre-Sinoussi et al., 1983; Gallo et al., 1983; Gelman et al., 1983]. HIV slowly attacks the immune system, thus destroying the body’s defence against other infections and malignancies that eventually cause AIDS and death.

HIV belongs to the lentivirus family of retroviruses. Lentiviruses are complex retroviruses with several accessory or regulatory genes. Two distinct types of HIV are known, namely HIV-1 and HIV-2. The two HIV types are distinguished on the basis of their genome organisations and phylogenetic relationships with other primate lentiviruses. HIV-1 is divided into three groups M (major), N (novel or non-M-non O) and O (outlier). Group M viruses are the most widespread group of HIV-1 and are further subdivided into 10 subtypes (A – D, F – K). HIV-2 is comprised of 6 diverse subtypes (A - F) [Los Alamos HIV database (hiv-web.lanl.gov)] (see Figure 1). The most common HIV type is HIV-1, with HIV-2 being more prevalent in West Africa [Clavel et al., 1987; Kanki and De Cock, 1994]. It has also been found that HIV-2 results in a much slower disease progression to AIDS than HIV-1 [Marlink et al., 1988; Markowitz, 1993; Marlink et al., 1994].
HIV-1 is divided into three groups M, N and O and Group M is further subdivided into 10 subtypes (A – D, F – K). The most prevalent HIV-1 subtype in South Africa is subtype C. HIV-2 is comprised of six diverse subtypes (A – F).
1.1.1. HIV infection and AIDS in South Africa

Of the estimated 36.1 million worldwide HIV infections, 25.3 million are found in sub-Saharan Africa [UNAIDS, December 2000 (www.unaids.org)]. The AIDS epidemic in South Africa is on the increase and at the beginning of the new century, with an estimated total of 4.7 million HIV infected individuals and thus 11.6% of the total population, South Africa has the largest number of individuals living with HIV/AIDS in the world. Approximately 2.5 million women (15 to 49 years), 2.2 million men (15 to 49 years) and 106,109 children (0 to 14 years) are infected with HIV. Furthermore, the different geographical areas of South Africa show variation in the percentages of HIV infected individuals, with the Kwazulu Natal Province having a prevalence of 36.2% and the Western Cape Province having a prevalence of 8.7%. Studies of HIV seropositive women attending antenatal clinics across South Africa form the basis for provincial and national HIV/AIDS estimates [Department of Health, South Africa, 2001].

South Africa has two independent and distinct HIV-1 epidemics. The first HIV-1 epidemic started in the early 1980s and has remained mainly restricted to homosexual males (subtype B and D), while the second, more predominant HIV-1 epidemic started in the late 1980s and has spread rapidly among heterosexuals (subtype C) [Engelbrecht et al., 1995; Williamson et al., 1995; van Harmelen et al., 1997; Moodley et al., 1998]. Previous studies show that although there is evidence of diverse HIV-1 subtypes in South Africa, HIV-1 subtype C is the most commonly found HIV-1 subtype in all the different provinces of South Africa [Engelbrecht et al., 1995; Engelbrecht et al., 1998; Moodley et al., 1998; Engelbrecht et al., 1999; van Harmelen et al., 1999] (see Figure 1). Presently, subtype C constitutes 56% of
all circulating subtypes of HIV-1 group M viruses in the world [Esparza and Bharmarapravati, 2000] and is the most spread HIV-1 subtype worldwide [UNAIDS, December 2000 (www.unaids.org)]. The main mode of transmission of the HIV-1 subtype C in South Africa is via heterosexual intercourse [van Harmelen et al., 1997] and the key factors which play a role in the spread of the virus is the country’s large migrant workforce, high rates of sexually transmitted diseases, thriving commercial sex worker industry and poverty [Bredell H et al., 1998; Moodley et al., 1998; Cohen, 2000a].

Multiple factors are responsible for influencing an individual’s susceptibility to HIV-1 infection and/or disease progression to AIDS. It has been suggested that the average interval of eight to ten years between HIV-1 infection and disease progression to AIDS is shorter in Africa, when compared to the rest of the world [Grant et al., 1997, Cohen, 2000b]. Besides poor socioeconomic factors (insufficient medical treatment and poverty) and viral factors, it’s possible that host factors (immunological and genetic) may play an important role in susceptibility to HIV-1 infection and/or disease progression to AIDS within certain African ethnic groups of South Africa.

1.1.2. Disease progression to AIDS

The course of HIV infection involves a number of stages and the progression to AIDS does not follow a specific or defined path. A drop in the level of host cells expressing CD4 molecules at their surfaces, such as CD4+ T lymphocytes and CD4+ macrophages, indicates a decrease in immune functioning and progression to a new disease stage. CD4 molecules are surface glycoproteins, which promote immune
responses, such as antibody production. The viral load (amount of virus in the bloodstream) is very high at the point of seroconversion and drops transiently during seroconversion, but only to rise again during the symptomatic phase of HIV infection. Thus measuring the CD4+ cell count and/or the amount of viral load is an indication of how long an individual is infected with HIV (Figure 2) [Reviewed in Daar, 1998; Arnaout et al., 1999; Phair, 1999].

![Diagram showing CD4+ cell count and viral load over time.](Image)

**Figure 2.** The CD4+ cell count and/or the amount of viral load is an indication of the individual's disease progression to AIDS.

HIV disease can be divided into the following stages: asymptomatic infection; early symptomatic infection (CD4+ count of more than 500 cells/mm³ and low viral load); middle symptomatic infection (CD4+ count of 200-500 cells/mm³); late symptomatic infection (CD4+ count of 50-200 cells/mm³ and high viral load); and advanced HIV...
disease (AIDS) [Vlahov et al., 1998; Centers for Disease Control and Prevention (www.cdc.org)]. HIV seropositive patients are classified into groups, depending on their rate of disease progression to AIDS. For the studies presented in this dissertation, the HIV seropositive individuals were classified as follows: slow progressors are individuals who develop AIDS within 10-12 years after infection; normal progressors are individuals who develop AIDS within 5-10 years after infection; and fast progressors are individuals who develop AIDS within 2-5 years after infection. We also find asymptomatic individuals who after 10-12 years have a CD4+ count of more than 500 cells/mm³, a very low viral load and show no symptoms of progressing to AIDS. These HIV seropositive individuals are referred to as long-term non-progressors (LTNPs) [Lifson et al., 1991; Magierowska et al., 1999] (see chapter 2.2). HIV seropositive patients are also grouped into four (I – IV) clinical stages, depending on their specific AIDS-related clinical symptoms [World Health Organisation (www.who.int)].

HIV-1 tropism (preferences for specific host target cells) has been associated with disease progression. Viruses that infect mainly macrophages and to a lesser degree T-lymphocytes are termed macrophage tropic (M-tropic) or non-syncytium inducing (NSI) and are normally present during the early or asymptomatic stages of HIV-1 infection [Roos et al., 1992; Schuitemaker et al., 1992; Connor et al., 1993; Zhu et al., 1993], while viruses that infect mainly T-lymphocytes are referred to as T cell line tropic (T-tropic) or syncytium inducing (SI) and are normally present during the late or symptomatic stages of HIV-1 infection [Tersmette et al., 1988; Tersmette et al., 1989].
Most HIV seropositive patients experience a shift in viral tropism, which results in the conversion of the NSI phenotype to the SI phenotype. This conversion is usually associated with the onset of rapid disease progression to AIDS [Tersmette et al., 1988; Tersmette et al., 1989; Schuitemaker et al., 1992; Connor et al., 1993]. Several studies indicate that the shift in viral tropism, which is largely related to changes in co-receptor usage, is due to mutations in the viral envelope (env) gene (see chapter 1.2.3.) [De Jong et al., 1992; Fouchier et al., 1992; Shioda et al., 1992; Connor et al., 1996; Speck et al., 1997; Connor et al., 1997; Verrier et al., 1999]. A recent study by Treurnicht et al., 2001, suggests that subtype C isolates from South African HIV seropositive individuals, remain phenotypically NSI due to the lack of sequence variation in the viral env gene.

In addition to viral load and viral env genetic variants, other viral factors have been found to influence and thus predict the rate of disease progression to AIDS. These include mutations identified in the gag gene (encodes for the viral capsid and matrix proteins) [Huang et al., 1998], pol gene (encodes for the viral reverse transcriptase) [Merigan et al., 1996] and viral regulatory or accessory genes, namely nef [Deacon et al., 1995; Kirchhoff et al., 1995; Salvi et al., 1998; Rhodes et al., 2000], vif, vpr, vpu, tat [Michael et al., 1995] and rev [Iversen et al., 1995].
1.2. Chemokines and Chemokine Receptors

Chemokines and chemokine receptors play pivotal roles in the genesis, regulation, maintenance and functioning of the host’s immune system. Immunity is the ability of the host to defend itself against infectious agents, foreign cells and abnormal host cells. Two types of immunity exist namely, “innate immunity” (non-specific immunity) and “acquired immunity” (specific/adaptive immunity). “Innate immunity” is present at birth and is the main, first-line defence against invading organisms. It is characterised as being present for life, having no specificity and no memory. “Acquired immunity” is the antithesis of innate immunity, as it is absent at birth, has specificity and memory. Various leukocytes (white blood cells) play vital roles in immunity. These include granulocytes (neutrophils, eosinophils and basophils), monocytes and lymphocytes [Craps and Basie, 1993; Mader, 1996; Peakman and Vergani, 1997; Playfair and Chain, 2001].

Granulocytes are derived from the bone marrow and obtained their name from the large numbers of granules present in their cytoplasm. They constitute approximately 65% of all leukocytes and differentiate into neutrophilis, eosinophils and basophils, which all circulate in the blood. Neutrophils, short-lived phagocytic cells whose granules contain numerous bactericidal substances, are the most common leukocytes of the blood. Eosinophils are leukocytes with large granules that contain highly basic or ‘cationic’ proteins, important in killing large parasites. Basophils are important for inflammatory responses and have large granules, which contain heparin and vasoactive (blood vessel active) amines [Craps and Basle, 1993; Peakman and Vergani, 1997; Playfair and Chain, 2001].
Bone marrow-derived monocytes are the largest nucleated cells of the blood and constitute approximately 5 - 10% of all circulating leukocytes. Monocytes develop into macrophages when they migrate into tissues and become the chief resident phagocyte of the tissues. Several specific forms of macrophages exist, including alveolar macrophages in the lung, Kupffer cells in the liver, mesangial cells in the kidney, microglial cells in the brain and osteoclasts in bone [Craps and Basle, 1993; Peakman and Vergani, 1997; Playfair and Chain, 2001].

Lymphocytes are closely associated with the lymphatic system and constitute the remaining 25 – 35% of leukocytes. They are divided into 2 subgroups, namely B and T lymphocytes, present in the blood at a ratio of approximately 1:5. B lymphocytes (bone marrow-derived) are involved in antigen recognition and as plasma cells in tissues they secrete antibodies into the blood. T lymphocytes (thymus-derived) do not produce antibodies, but are directly involved in distinguishing and attacking cells that bear antigens. A few T lymphocytes are referred to as T helper (Th) cells and play a vital role in regulating immune responses by secreting cytokines (small non-antigen-specific protein molecules). Th cells are further subdivided into Th1 and Th2 cells on the basis of which cytokines they secrete. Th1 cells are important for driving T cell-mediated immunity or delayed hypersensitivity, while Th2 cells are involved in antibody production. Another type of immune cell, derived from the bone marrow, is a dendritic cell. Dendritic cells, found in blood, lymph nodes, bone marrow and tissues, have a specialised function in the activation and priming of lymphocytes. The immune system is thus composed of specialised leucocytes with organised structures and distinct functions [Craps and Basle, 1993; Mader, 1996; Peakman and Vergani, 1997; Playfair and Chain, 2001].
1.2.1. Chemokines

Chemokines, small polypeptides of generally 8-14 kilodaltons (kDa) in size, are a complex superfamily of chemoattractant cytokines secreted during inflammatory responses to recruit and mediate migration of different subsets of leukocytes from the circulation to sites of infection or injury [Oppenheim et al., 1991; Schall, 1991; Miller et al., 1992; Lindley et al., 1993; Baggiolini et al., 1994, Murphy, 1994; Schall et al., 1994]. It has been discovered that the functioning of chemokines is however no longer restricted to cell attraction and migration, as chemokines are multifunctional and have been shown to play a role in regulating various aspects of the host's immune defence system such as, facilitating angiogenesis (the formation and differentiation of blood vessels) [Baggiolini et al., 1997; Rollins, 1997], modulating hematopoiesis (the formation of blood or of blood cells) [Broxomeyer et al., 1989; Rollins, 1997], induction and enhancement of T helper 1 (Th1) and T helper 2 (Th2)-associated cytokine responses [Loetscher et al., 2000], dendritic cell maturation [Sozzani et al., 1998], B [Forster et al., 1994] and T [Vicari et al., 1997] lymphocyte development, and suppressing HIV-1 infection [Cocchi et al., 1995].

All identified chemokines can thus be divided into two functional groups namely, the "inflammatory chemokines" and the "homing (homeostatic) chemokines". The "inflammatory chemokines" are expressed by different types of cells in most tissues and facilitates the migration of leukocytes during an inflammatory response that arises during the invasion of the body by pathogenic organisms. The "homing chemokines" are produced by specific areas of lymphoid tissue and play an important role in the development, maintenance and functioning of the host's immune system [Baggiolini et al., 2000].
Most chemokines are closely related in primary amino acid sequence, secondary and tertiary structures [Oppenheim et al., 1991; Rollins, 1997] Similarity in the genes encoding for these chemokines suggest that they arose by duplication of an ancestral gene [Baggiolini et al., 1994]. Chemokines share the same basic structural features; a relatively short disordered amino-terminus prior to the first cysteine residue is attached to the rest of the molecule by disulphide bonds between the first and third and second and fourth cysteine residues, which stabilises three antiparallel β-pleated sheets that form the core and provide a flat base over which the carboxy-terminal α-helix of 20 – 30 amino acids extends. The chemokines are thus identified by characteristic sequence elements that give rise to certain structural motifs (three β-sheets and a α-helix) [Baggiolini et al 1994; Rollins, 1997].

Presently, more than 40 human chemokines have been identified [Yoshie et al., 2001]. Chemokines are grouped into four subfamilies based on the arrangement of the first two of four highly conserved amino-terminal cysteine residues (see Figure 3):

1. CXC chemokines – They have a single amino acid residue separating the first and second conserved cysteine residues and are also named the α-chemokine family [Schall, 1991].

2. CC chemokines – They have the first two conserved cysteine residues immediately adjacent to each other and are also named the β-chemokine family [Schall, 1991].

3. C chemokines – They lack two of the four conserved cysteine residues as only the second and fourth cysteine residues are preserved and are also named the γ -chemokine family [Kelner et al., 1994].
4. CX3C chemokines – They have three intervening amino acids between the first two conserved cysteine residues and are also named the δ-chemokine family. Unlike all other chemokines, the CX3C chemokines have an extracellular stalk and cytoplasmic domain region, which is separated by a transmembrane (TM) segment [Bazan et al., 1997; Pan et al., 1997].

<table>
<thead>
<tr>
<th>Chemokine subfamily</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CXC chemokines</td>
<td>CXC C C</td>
</tr>
<tr>
<td>2. CC chemokines</td>
<td>CC C C</td>
</tr>
<tr>
<td>3. C chemokines</td>
<td>C C</td>
</tr>
<tr>
<td>4. CX3C chemokines</td>
<td>CXXC C C Extracellular stalk</td>
</tr>
</tbody>
</table>

**Figure 3.** Chemokines are grouped into four subfamilies based on the arrangement of the first two of four highly conserved amino-terminal cysteine residues (Adapted from Ward and Westwick, 1998). Chemokines are not drawn to scale.

Although there are 15 CXC and 24 CC known chemokines, there is only one member each from the C chemokines and the CX3C chemokines that has been identified thus far [Yoshie et al., 2001]. The CXC and CC chemokine subfamilies thus form the two major groups of chemokines. Generally, the CXC chemokines are principally involved in the activation of neutrophils, while the CC chemokines do not activate neutrophils,
but stimulate multiple other cell types, namely monocytes, lymphocytes, basophils and eosinophils [Clark-Lewis et al., 1995; Baggioolini et al., 1997; Rollins, 1997]. Furthermore, most of the genes encoding for the CXC and CC chemokines cluster at human chromosome 4 and 17, respectively [Baggioolini et al., 1994; Rollins, 1997].

The recent identification of novel CXC and CC chemokines however revealed that most of these newly discovered chemokines are highly specific for lymphocytes and also attract dendritic cells [Yoshie et al., 1997]. It was also discovered that their genes are mapped to different chromosomal loci when compared to the “classical”, CXC and CC chemokines gene clusters at chromosomes 4 and 7 [Yoshie et al., 2001]. The CXC chemokines have thus now been divided into 2 groups based on the presence or absence of a glutamic acid-leucine-arginine (ELR) motif preceding the first cysteine. The chemokines with the ELR motif are the “classical” chemokines and attract neutrophils [Schall et al., 1994], while the non-ELR motif chemokines attract activated lymphocytes [Taub et al., 1993, 1995; Liao et al., 1995]. Similarly, while classical” CC chemokines attract monocytes, some of the recently identified novel CC chemokines are specific for lymphocytes and dendritic cells and do not attract monocytes [Yoshie et al., 1997; Zlotnik et al 1999].

In the past, chemokines were named randomly and no specific system was used. Some chemokines have been included with the interleukins, which are cytokines that function as potent intracellular messengers (e.g., interleukin-8, IL-8). Other chemokines were given names describing a function (e.g., monocyte chemotactic protein-1, MCP-1) or named according to the cell type that secretes the chemokine (e.g., stromal derived factor-1, SDF1). A new nomenclature system has been
recently proposed for the chemokine ligands and indicates the chemokine subfamilies (CXC, CC, C and CX3C) followed by L (for ligand) and serial numbers (e.g., IL-8 is CXCL8, MCP-1 is CCL2 and SDF1 is CXCL12) [Zlotnik and Yoshie, 2000]. Due to the relatively recent implication of the new nomenclature system, in this dissertation the old names have been used, followed by the new nomenclature in brackets.

1.2.2. Chemokine receptors

Chemokine receptors are proteins that belong to a large, functionally diverse family of receptors that contain seven hydrophobic domains that are suggested to be transmembrane helices, which span the cell membrane and signal by coupling to heterotrimeric guanosine 5'-triphosphate (GTP)-binding proteins found inside the cell. They are thus designated seven-transmembrane domain, G protein-coupled receptors [Baggiolini et al., 1994; Horuk, 1994, Murphy, 1994]. The function of chemokine receptors involves their binding to a variety of chemokines. The direct interaction between the chemokine receptors and chemokines mediates the multiple cellular effects or biological functions of specific chemokines in the host's immune system [Schall, 1991; Oppenheim et al., 1991; Baggiolini et al., 1994; Horuk, 1994; Murphy, 1994; Baggiolini et al., 1997].

Structural features of the chemokine receptors include an extracellular amino-terminus containing negatively charged residues that may be involved in ligand binding, and a cytoplasmic carboxy-terminus containing a serine and threonine-rich region that may be a target for phosphorylation by G-protein-coupled receptor kinases after ligand binding. The seven transmembrane spanning α-helical domains
of the chemokine receptor contain conserved prolines, while three intracellular and three extracellular loops are composed of hydrophilic amino acids. Disulfide bonds formed by highly conserved cysteine residues are found between the first and second extracellular loop and between the amino-terminus and third extracellular loop. All chemokine receptors also have an amino acid sequence, DRYLAIVHA, in the second intracellular loop domain. The presence of one or two N-linked glycosylation sites is also observed in various chemokine receptors [Horuk, 1994; Murphy, 1994; Strader et al., 1994; Baggilolini et al., 1997; Rollins, 1997].

Chemokine receptors are classified according to the specific chemokines that they bind and the chemokine nomenclature thus indicates the ligand subfamilies (CXC, CC, C, CX3C) followed by R (for receptor) and then a serial number based on the chronological order in which the chemokine receptor was identified (e.g., CXCR4) [Zlotnik and Yoshie, 2000]. Presently, there are six CXC chemokine receptors, 11 CC chemokine receptors, one C chemokine receptor and one CX3C chemokine receptor that have been cloned and characterised [Yoshie et al., 2001]. Similarly to the chemokine genes, chemokine receptor genes were mainly mapped at specific loci on chromosomes 2 and 3 [Murphy, 1994; Baggilolini et al., 1997; Rollins, 1997], but some of the recently identified chemokine receptors map to completely different chromosomal loci [Yoshie et al., 2001].

Various studies regarding structure-function relations have been performed to understand the ligand requirements for binding and the functional activation of specific cell surface receptors [Horuk, 1994]. It has been suggested that chemokines have two sites of interaction with their chemokine receptors, namely the amino-
terminal region and the core structure (antiparallel β-pleated sheets) that follows the second cysteine residue. The initial interaction thus takes place between the core region of the chemokine and the extracellular domain of the chemokine receptor, which then facilitates further binding of the amino-terminal of the chemokine for receptor activation [Clark-Lewis et al., 1995].

It has been observed that "inflammatory chemokines" and their chemokine receptors have highly promiscuous ligand-receptor relationships and thus a single chemokine receptor can exhibit overlapping ligand specificities by recognising more than one chemokine, or conversely, a single chemokine can bind more than one chemokine receptor [Murphy, 1994; Baggionini et al., 1997; Rollins, 1997]. "Homing chemokines" however are selective for their chemokine receptors and thus have more specific ligand-receptor relationships [Yoshie et al., 1997; Zlotnik et al., 1999; Zlotnik and Yoshie, 2000].

1.2.3. Role of chemokine and chemokine receptors in HIV-1 entry

The first indication that chemokines play a role in HIV-1 infection came from a study done by Cocchi et al., 1995, who showed that the CC or β-chemokines, RANTES (regulated on activation normal T cell expressed and secreted) (CCL5), MIP-1α (macrophage inflammatory protein alpha) (CCL3) and MIP-1β (macrophage inflammatory protein beta) (CCL4), secreted by CD8+ cells (cytotoxic T cell lymphocytes), could act as potent inhibitors by preventing infection by M-tropic or NSI HIV-1 viruses. Later studies showed SDF1 (CXCL12) act as an inhibitor of T-cell tropic (T-tropic) or syncytium inducing (SI) HIV-1 viruses and also possibly influence viral replication [Bleul et al., 1996a; Oberlin et al., 1996].
The discovery that specific chemokine receptors, expressed in various cells, serve as HIV-1 co-receptors (see Table 1) [Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Doranz et al., 1996; Dragic et al., 1996; Feng et al., 1996], further emphasised the importance of understanding the precise way in which chemokines influence HIV-1 pathogenesis. Chemokines could suppress or even prevent HIV-1 infection by either direct competition with the virus for binding to the chemokine receptors (see Figure 4) or by downregulating the biosynthesis of the chemokine receptors [Bleul et al., 1996a; Oberlin et al., 1996; Samson et al., 1996a; Combadiere et al., 1996; Raport et al., 1996; Amara et al., 1997]. Elevated levels of chemokines have been observed in exposed yet uninfected individuals and has been associated with delaying the disease progression to AIDS in HIV seropositive patients [Paxton et al., 1996; Paxton et al., 1998; Ullum et al., 1998].

After the identification of the CD4 molecule as the primary receptor for HIV-1 [Dalgeish et al., 1984; Klatzman et al., 1984], it became evident that an additional co-receptor was required for infection. Subsequently, several studies identified this additional co-receptor as a chemokine receptor [Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Doranz et al., 1996; Dragic et al., 1996; Feng et al., 1996]. There are multiple sites on chemokine receptors, especially in their amino terminus and extracellular loops, which can interact with the virus. Contact sites vary depending on the specific virus infecting the target cell and specific chemokine receptor utilised for entry. An example is the amino terminus of CCR5 (CC chemokine receptor 5), including aspartic acids at position 2 and 11 and a glutamic acid at position 18, which is required for M-tropic HIV-1 env and membrane fusion [Dragic et al., 1998] (see chapter 2.2.1).
Entry of HIV-1 into target cells is a multistep process, which requires the virally encoded env protein complex to interact with the CD4 molecule and subsequently, the specific chemokine receptor on the target cell surface. The HIV-1 env protein complex is initially synthesised as a precursor, glycoprotein (gp) 160, which is then extensively glycosylated and proteolytically cleaved by a cellular convertase into a surface subunit, gp120 and a transmembrane subunit, gp41 [Earl et al., 1991; Willey et al., 1988].

Structural analysis have shown that gp120 contains five hypervariable loop structures (VI – V5) [Starcich et al., 1986] and several studies show that the V3 loop of gp120 plays a critical role in determining fusion specificity [Chesebro et al., 1991; Hwang et al., 1991; Hwang et al., 1992]. HIV-1 viruses selective for macrophages therefore have a different sequence in the third hypervariable (V3) loop of gp120 than those HIV-1 viruses selective for lymphocytes. Single amino acid changes in the V3 loop have been shown to alter chemokine receptor utilisation [De Jong et al., 1992; Fouchier et al., 1992; Shoida et al., 1992; Speck et al., 1997; Verrier et al., 1999]. The first amino terminus of gp41 contains a hydrophobic, glycine-rich “fusion” peptide that is necessary for the fusion of the cellular and viral membranes [Moore et al., 1993; Sattentau et al., 1995; Lapham et al., 1996].

HIV-1 infection is thus initiated by high affinity binding of the viral env gp120 to a CD4 molecule on the surface of the target cell [Dalgleish et al., 1984; Klatzmann et al., 1984; Maddon et al., 1986, Lasky et al., 1987]. This results in the formation of a CD4-gp120 complex, which induces a conformational change in the viral envelope to enable gp120 to bind to the chemokine receptor [Trkola et al., 1996; Wu et al., 1996;
Speck et al., 1997, Kwong et al., 1998; Rizzuto et al., 1998]. This further results in the exposure of the hydrophobic amino-terminal “fusion” peptide of viral env gp41 for eventual virus-host cell fusion [Moore et al., 1993; Sattentau et al., 1995; Lapham et al., 1996] (see Figure 4).

![Diagram](image)

**Figure 4.** HIV entry is initiated by the interaction of the virion envelope glycoproteins, gp120 and gp41, with two cellular host receptors, of which one is a CD4 molecule and the other, a chemokine receptor, whose natural ligands are specific chemokines.

Two principle co-receptors have been described, namely CCR5 for HIV-1 M-tropic or NSI viruses [Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Dragic et al., 1996] and CXCR4 for HIV-1 T-tropic or SI viruses [Feng et al., 1996]. There is thus a current model for co-receptor usage by different viruses. M-tropic or NSI viruses (termed R5 viruses) infect macrophages, monocytes and T lymphocytes by utilising the CD4 molecules and CCR5 or “less efficient” co-receptors, such as CCR2, CCR3
as co-receptors for entry [Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Dragic et al., 1996]. T-tropic or SI viruses (termed X4 viruses) infect T-lymphocytes by utilising CD4 molecules and CXCR4 as co-receptors for entry [Feng et al., 1996]. The T-tropic or SI viruses may also show dual tropism (termed R5X4 viruses) and can thus utilise both CXCR4 and CCR5 as co-receptors and other additional chemokine receptors, such as CCR2 and CCR3 for entry [Choe et al., 1996; Doranz et al., 1996; Simmons et al., 1996; Berger et al., 1998] (see Table 1 and Figure 5).

**Table 1.** The main chemokine receptors involved in HIV entry, their biological ligands, cell expression and utilisation by specific HIV strains.

<table>
<thead>
<tr>
<th>Receptor</th>
<th>Ligands</th>
<th>Expression</th>
<th>Viruses</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5</td>
<td>RANTES (CCL5), MIP-1α (CCL3), MIP-1β (CCL4)</td>
<td>Monocytes, microglia, Th1 cells, memory T cells, dendritic cells</td>
<td>R5, R5X4, HIV-2</td>
<td>Samson et al., 1996a; Deng et al., 1996; Dragic et al., 1996; Doranz et al., 1996</td>
</tr>
<tr>
<td>CXCR4</td>
<td>SDF1 (CXCL12)</td>
<td>Neutrophils, monocytes, microglia, Th1 and Th2 cells, naive T cells, B cells, dendritic cells</td>
<td>R5, X4, R5X4, HIV-2</td>
<td>Bleul et al., 1996a; Oberlin et al., 1996; Feng et al., 1996</td>
</tr>
<tr>
<td>CCR2</td>
<td>MCP-1 (CCL2), MCP-2 (CCL8), MCP-3 (CCL7), MCP-4 (CCL13), MCP-5 (CCL12)</td>
<td>Basophils, Monocytes, Th1 and Th2 cells, dendritic cells</td>
<td>R5X4, HIV-2</td>
<td>Charo et al., 1994; Doranz et al., 1996</td>
</tr>
<tr>
<td>CCR3</td>
<td>Eotaxin (CCL11), MCP-2 (CCL8), MCP-3 (CCL7), MCP-4 (CCL13), RANTES (CCL5)</td>
<td>Eosinophils, basophils monocytes, microglia, Th2 cells, dendritic cells</td>
<td>R5X4, HIV-2</td>
<td>Ponath et al., 1996; Daugherty et al., 1996; Choe et al., 1996; Doranz et al., 1996</td>
</tr>
<tr>
<td>CCR8</td>
<td>I-309 (CCL1)</td>
<td>Monocytes, B cells, Th2 cells, dendritic cells</td>
<td>R5X4, HIV-2</td>
<td>Tiffany et al., 1997, Rucker et al., 1997</td>
</tr>
</tbody>
</table>
Early HIV disease

Other chemokine receptors
e.g. CCR2, CCR3

CCR5

Target cell

NSI virus

SDF1 is the natural ligand
for CXCR4

CCR5/CXCR4

Target cell

SI virus

Late HIV disease

Figure 5. Schematic illustration of the current model for co-receptor usage by different viruses.
Chapter 2

Study aim and overview
2.1. Study aim

The aim of the study involved firstly, the designing of comprehensive mutation detection assays, based on denaturing gradient gel electrophoresis (DGGE) (see chapter 2.3), for the entire coding regions of the CCR5 and CCR2 genes, as well as the 3' untranslated region (3' UTR) of the SDF1 gene, for genetic analysis. The reason for this specific approach was due to most previous studies having been restricted to the genes mentioned above, but mainly focusing on Caucasian populations. The second part of the aim was thus to identify and determine the distribution of previously reported and novel mutations of the CCR5, CCR2 and SDF1 genes, which may play a role in host genetic susceptibility to HIV-1 pathogenesis within a South African population. Our South African population with its diverse ethnic groups (described in chapter 3) is ideal for the identification of African-related protective or causative novel mutations, which can be used in association studies. Mutations that occur at an allelic frequency of 0.01 or greater, within a specific population group, are referred to as single nucleotide polymorphisms (SNPs). It is likely that HIV-1 pathogenesis has a multifactorial nature and we therefore cannot exclude the possibility that SNPs in a large number of so-called "weaker genes" could collectively determine an individual's "risk profile" for susceptibility to HIV-1 infection and/or disease progression to AIDS.
2.2. Host susceptibility to HIV-1 infection and/or progression AIDS

Although most people are susceptible to HIV-1 infection, there are a few rare individuals who remain uninfected despite high-risk exposure to the virus. These “exposed yet uninfected” persons have been identified and include commercial sex workers [Rowland Jones et al., 1995; Fowke et al., 1996; Plummer et al., 1999], discordant couples who have unprotected sex [Clerici et al., 1992; Langlade-Demoyen et al., 1994; Beretta et al., 1996; Mazzoli et al., 1997; Bernard et al., 1999, Goh et al., 1999; Stanford et al., 1999], babies of HIV seropositive mothers [Rowland-Jones et al., 1993; De Maria et al., 1994], intravenous drug users [Barcellini et al., 1995], health care workers who have had needlestick injuries with HIV-infected blood [Clerici et al., 1994; Pinto et al., 1995; Bernard et al., 1999] and haemophiliacs who have received HIV contaminated blood products [O’Brien and Dean, 1997a; Zagury et al., 1998, Salkowitz et al., 2001].

Another observation is that the disease course and clinical outcome of HIV-1 infection varies widely among individuals, even those infected from a common source [Liu et al., 1997]. The median time from HIV-1 infection to the development of AIDS is generally eight to ten years, but approximately 5-10% of HIV seropositive individuals are defined as being long-term nonprogressors because they remain asymptomatic in the absence of treatment, have normal CD4 cell counts and low or undetectable viral loads for ten years or longer [Lifson et al., 1991; Sheppard et al., 1993; Cao et al., 1995; Munoz et al., 1995; Pantaleo et al., 1995; Rinaldo et al., 1995; Haynes et al., 1996; Learmont et al., 1999; Magierowska et al., 1999]. Conversely, approximately 10% of all HIV seropositive individuals (as observed in some people of African-ethnic origin) display rapid disease progression and develop
AIDS within five years after HIV-1 infection [Phair et al., 1992; Haynes et al., 1996; Grant et al., 1997; Cohen, 2000b].

Numerous studies have indicated that susceptibility for HIV-1 infection and the rates of HIV-1 disease progression are determined by certain parameters, which include viral characteristics and host immunological and genetic factors. The role of host factors in susceptibility to HIV-1 infection and/or disease progression to AIDS has been advanced by the discovery of HIV-1 suppressive chemokines [Cocchi et al., 1995]. HIV-1 suppressive chemokines have been identified as the natural ligands for the HIV entry coreceptors, namely specific chemokine receptors, and can thus modulate the efficiency of HIV-1 infection (discussed in chapter 1) [Bleul et al., 1996a; Oberlin et al., 1996; Samson et al., 1996a; Combadiere et al., 1996; Raport et al., 1996; Amara et al., 1997].

Elevated levels of the β-chemokines, RANTES (CCL5), MIP-1α (CCL3) and MIP-1β (CCL4) and low expression of CCR5 have been associated with relative resistance to M-tropic HIV-1 infection of CD4 lymphocytes from persons who remain uninfected, despite multiple high-risk sexual exposures [Paxton et al., 1996; Paxton et al., 1998]. An overproduction of MIP1β (CCL4) has also been associated with slower disease progression to AIDS [Ullum et al., 1998]. The importance of host genetic factors has thus been underscored by the identification of mutations in genes encoding specific chemokines (eg. SDF1 (CXCL12) and RANTES (CCL5)) and chemokine receptors (egs. CCR5 and CCR2), which are associated with influencing host susceptibility to HIV-1 infection and/or disease progression to AIDS (see chapters 3 to 5).
There are other determinants influencing host genetic susceptibility for HIV-1 infection and/or disease progression to AIDS. Of this host genetic factors include certain human leucocyte antigen (HLA) allelic distributions of the human major histocompatibility complex [Carrington et al., 1999a; Hendel et al., 1999] and genetic variants identified in the mannose binding protein (MBP) [Garred et al., 1997], vitamin D receptor (VDR) [Hill et al., 1998], interleukin-4 (IL-4) [Nakayama et al., 2000], interleukin 1 (IL-1) [Witkin et al., 2001], haptoglobin [Delanghe et al., 1998] and natural resistance-associated macrophage protein 1 (NRAMP1) [Marquet et al. 1999] genes.

2.2.1. CC chemokine receptor 5 (CCR5)

CCR5 is a seven transmembrane G-coupled protein consisting of 352 amino acids and belongs to the CC or β-chemokine receptor family. The natural ligands for CCR5 are the β-chemokines, RANTES (CCL5), MIP-1α (CCL3) and MIP-1β (CCL4), [Samson et al., 1996a]. The second extracellular loop of CCR5 has been shown to be the principal determinant for ligand selectivity [Samson et al., 1997], while residues in the CCR5 amino terminus [Blanpain et al., 1999a; Zhou et al., 2000] and disulfide bonds that stabilise the extracellular loops of CCR5 [Blanpain et al., 1999b] are required for high-affinity ligand binding.

CCR5 is considered the principle co-receptor for M-tropic or NSI strains of HIV-1 and together with the CD4 molecule, it facilitates HIV-1 entry during the asymptomatic phase of infection [Alkhatib et al., 1996; Choe et al., 1996; Deng et al., 1996; Dragic et al., 1996]. Various chemokine receptors have been found to be involved in HIV-1 infection in vitro, but only CCR5 has been clearly shown to have a role in sexual
transmission of the virus [Berger et al., 1999; Loetscher et al., 2000]. The expression and functioning of the CCR5 protein at the cell surface may thus be an indication of an individual’s susceptibility to HIV-1 infection and/or disease progression to AIDS.

Initially, determinants of CCR5 co-receptor function were identified using human and murine CCR5 and/or CCR2B chimeras, but the results from several studies were difficult to interpret collectively due to differences in chimeric constructs, HIV-1 isolates and the detection systems that were used to test the role of CCR5 extracellular domains in co-receptor function [Atchison et al., 1996; Rucker et al., 1996; Alkhatib et al., 1997; Bieniasz et al., 1997; Doranz et al., 1997; Picard et al., 1997; Wang et al., 1999]. All the studies did however prove that multiple domains of CCR5 are involved in co-receptor activity and that it’s amino terminus plays a vital role in mediating membrane fusion and facilitating entry of various HIV-1 isolates.

Site-directed mutagenesis studies have provided a more informative understanding of the determinants of CCR5 co-receptor function. Negatively charged and tyrosine residues in the amino terminus of CCR5 (Asp-2, Tyr-2, Tyr-10, Asp-11, Tyr-14, Tyr-15 and Glu-18) have been shown to be important in gp120-CCR5 binding and virus entry [Doranz et al., 1997; Dragic et al., 1998; Farzan et al., 1998; Rabut et al., 1998; Blanpain et al., 1999a]. A few other residues of the CCR5 amino terminus that have also been found to be involved in co-receptor function are Ser-6, Ser-7, Ile-9, Asn 13, Gln-21 and Lys-22 [Farzan et al., 1998; Rabut et al., 1998; Blanpain et al., 1999a]. Residues in the extracellular loops (ECLs) of CCR5 that were found to influence co-receptor function include, Gln-93 in ECL 1 [Kuhmann et al., 1997]; Gly-163 in the transmembrane helix 4/ ECL 2 junction [Siciliano et al., 1999]; Tyr-184, Ser-185 and
Arg-197 in ECL 2 [Doranz et al., 1997; Ross et al., 1998]; and Asp-276 and Gln-280 in ECL 3 [Doranz et al., 1997; Farzan et al., 1998]. The residues 184 and 185, as well as residues 197 and 276, have to be substituted simultaneously in order to disrupt fusion and entry of HIV-1 [Doranz et al., 1997; Ross et al., 1998].

The human CCR5 gene was cloned from a human genomic DNA library based on its similarity to a murine CC chemokine receptor clone [Samson et al., 1996a]. CCR5 is located at band position p21 of chromosome 3 [Liu et al., 1996] and comprises four exons and two introns. Exons 2 and 3 are not interrupted by an intron and exon 4 contains the entire CCR5 open reading frame (ORF), which is 1056 base pairs (bp) in length (see Figure 1). CCR5 also shows dual promoter usage, with the presence of a weak promoter upstream of exon 1 and a strong downstream promoter, which includes the intronic region between exon 1 and 3 [Mummidi et al., 1997].

**Figure 1.** CCR5 is organised into four exons and two introns and exons 2 and 3 are not interrupted by an intron. Exon 4 contains the entire CCR5 ORF. Black blocks represent the non-coding regions and the grey block represents the coding region. Exons and introns are not drawn to scale.
Previously identified mutations in the CCR5 gene, include a 32bp deletion (CCR5Δ32), which results in C-terminal residues being translated out of frame and the formation of a truncated protein that is not expressed at the cell surface, so HIV-1 is unable to bind and infect target cells [Liu et al., 1996; Samson et al., 1996b]. The CCR5Δ32 comprises nucleotides 794 to 825 of the cDNA sequence and results in a reading frameshift after amino acid 174, inclusion of 7 novel amino acids and truncation at codon 182 [Liu et al., 1996]. The truncated CCR5 protein thus lacks the last three of seven putative transmembrane regions as well as regions involved in G-protein coupling and signal transduction [Samson et al., 1996b].

Individuals homozygous for CCR5Δ32 seem to have a highly protective effect against HIV-1 infection [Liu et al., 1996; Samson et al., 1996b; Dean et al., 1996; Huang et al., 1996; Zimmerman et al., 1997], although a few exceptions to the rule have been reported. These exceptions include individuals who are infected with viruses that utilise CXCR4 as a co-receptor [Biti et al., 1997; Balotta et al., 1997 O'Brien et al., 1997b; Theodorou et al., 1997]. Heterozygosity for CCR5Δ32 has been shown to offer partial protection and is associated with delaying the progression to AIDS by 2-4 years in HIV seropositive individuals [Liu et al., 1996; Samson et al., 1996b; Dean et al., 1996; Huang et al., 1996; Zimmerman et al., 1997].

CCR5Δ32 is however largely confined to the Caucasian population (allelic frequency = 0.092) and rarely observed in Africans [Samson et al., 1996b; Huang et al., 1996; Martinson et al., 1997]. Haplotype studies indicate that CCR5Δ32 originated approximately 700 years ago (range 275 – 1875 years) at a single point in Northeastern Europe [Libert et al., 1998; Stephens et al., 1998]. It has been
suggested that the selective pressure caused by the bubonic plague could account for the high prevalence of $CCR5\Delta 32$ in the Caucasian population [Stephens et al., 1998]. It has also been hypothesised that the Vikings disseminated $CCR5\Delta 32$ in Europe during the eighth to the tenth century [Lucotte and Mercier, 1998; Lucotte, 2001]. A recent study by Sullivan et al., 2001 shows that HIV is providing a selective pressure for $CCR5\Delta 32$.

Other genetic variants, which may also influence the functioning and expression of the CCR5 protein at the cell surface and thus affect HIV-1 infection and/or disease progression to AIDS, have been identified in both the CCR5 promoter [Kostrikis et al., 1998; Martin et al., 1998; McDermott et al., 1998; Mummidi et al., 1998] and coding regions [Dean et al., 1996; Ansari-Lari et al., 1997; Carrington et al., 1997; Quillent et al., 1998; Carrington et al., 1999b; Petersen et al., 2001], in a number of different population groups.

A recent study by our group, using an African-based population, resulted in the identification six previously reported (including $CCR5\Delta 32$) and seven novel mutations in the coding region of CCR5 (see Figure 2) [Petersen et al., 2001] (see chapter 3). Considering the structural features of CCR5 and the CCR5 mutations we identified, we can suggest that besides $CCR5\Delta 32$, two novel CCR5 mutations could influence the expression and functioning of CCR5. These include a non-conservative mutation at codon 2 (D2V), which lies in the CCR5 amino terminus and a nonsense mutation at codon 225 (R225X), which results a premature stop codon and the formation of a truncated CCR5 protein.
Figure 2. Schematic illustration of the human CCR5 protein. A recent study by our group resulted in the identification of previously reported (indicated by the blue shaded amino acid residues) and novel (indicated by the red shaded amino acid residues) CCR5 mutations within an African-based population group [Petersen et al., 2001] (see chapter 3). The starting site of CCR5Δ32 is indicated at amino acid residue 185.
Limited functional studies have been done for a few of the reported CCR5 mutations [Howard et al., 1999; Blanpain et al., 2000]. Howard et al., 1999, investigated six CCR5 mutations located between amino acid residues 1 and 100. Three of these variants, I12T, C20S and A29S, are expressed on the cell surface, but do not allow ligand binding. The I12T and C20S variants were also found to lack co-receptor function, while the A29S supported HIV-1 infection. The remaining three variants, I42F, L55Q and A73V, were all found to have co-receptor function, but their ligand binding affinities for specific chemokines were altered.

Another study investigated the functional consequences of 16 CCR5 mutations described in various population groups. Ten of these variants (I12L, I42F, R60S, A73V, S215L, R223Q, ΔK228, G301V, A335V and Y339F were found to result in functional responses similar to that of the normal CCR5 protein. The remaining six variants (C20S, A29S, L55Q, C101X, C178R and FS299) were found to alter either the expression, ligand binding affinities or co-receptor functioning of CCR5. The C101X and FS299 variants resulted in poor expression of CCR5 at the cell surface and thus an alteration in responses to chemokines. The C20S, C178R and A29S variants altered chemokine ligand binding affinities, while the L55Q variant resulted in the inability to mediate co-receptor function [Blanpain et al., 2000].

The functional analysis differs for some of the CCR5 mutations (C20S, A29S, I42F, L55Q and A73) investigated in both of the studies mentioned above. Further functional studies are thus required for CCR5 mutations, so that their possible effects on CCR5 expression, ligand binding and co-receptor functioning can be fully elucidated and confirmed.
2.2.2. CC chemokine receptor 2 (CCR2)

CCR2, classified as a CC or β-chemokine receptor, is utilised as an additional co-receptor by a minority of dual-tropic HIV-1 strains, however at a much lower efficiency than CCR5 and CXCR4 [Doranz et al., 1996]. The amino terminus of CCR2 is vital for co-receptor function [Rucker et al., 1996; Frade et al., 1997]. The natural ligands for CCR2 include the β-chemokines, monocyte chemoattractant protein (MCP)-1 (CCL2), 2 (CCL8), 3 (CCL7), 4 (CCL13) and 5 (CCL12). MCP-1 (CCL2) and MCP-3 (CCL7) have been shown to inhibit HIV-1 replication of both M-tropic and T-tropic viruses [Reviewed in Kalinkovich et al., 1999; Lee and Montaner, 1999]. The CCR2 amino terminus is an important determinant for chemokine selectivity and high-affinity ligand binding [Monteclaro and Charo, 1996; Monteclaro and Charo, 1997]. The first extracellular loop of CCR2 also contains a high-affinity ligand-binding site and is essential for receptor activation [Han et al., 1999].

The cloning of the CCR2 gene resulted in the identification of it’s 2 isoforms, CCR2A and CCR2B [Charo et al., 1994], which are the result of alternative splicing of a single gene. CCR2A (374 amino acids) and CCR2B (360 amino acids) only differ in their carboxyl tails and are both functional seven transmembrane G-coupled proteins. The predominant isoform, CCR2B, is expressed at the cell surface and in the cytoplasm, while CCR2A is mainly found in the cytoplasm due to cytoplasmic retention signals in it’s carboxyl tail [Wong et al., 1997].

The CCR2 gene, located at band position p21 of chromosome 3 and 17 kilobases (kb) from the CCR5 gene [Daugherty and Springer, 1997], spans about 7kb and is organised into 3 exons and 2 introns. Exon 2 and part of exon 3 contain the CCR2A
coding region, while the entire coding region for \textit{CCR2B} is found within exon 2 [Wong \textit{et al.}, 1997] (see Figure 3). The 5' untranslated region of the \textit{CCR2} gene plays an essential role in transcriptional activation and tissue specific expression of \textit{CCR2} [Yamamoto \textit{et al.}, 1999].

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure3.pdf}
\caption{\textit{CCR2} has two gene transcripts. The coding region for \textit{CCR2A} is found in exon 2 and part of exon 3, while exon 2 contains the entire coding region for \textit{CCR2B}. Black blocks represent the non-coding regions and the grey blocks represent the coding regions. Exons and introns are not drawn to scale.}
\end{figure}

A common genetic variant has been identified in the first transmembrane domain of the \textit{CCR2} gene. It results in a conservative amino acid change from valine to isoleucine at codon 64 (\textit{CCR2V64I}) [Smith \textit{et al.}, 1997a]. Although the mechanism by which the \textit{CCR2V64I} SNP influences HIV pathogenesis still needs to be elucidated [Lee \textit{et al.}, 1998; Mariani \textit{et al.}, 1999], it was found to delay the onset of AIDS in seroconverter patients by 2-4 years in both the heterozygous and
homozygous state, but does not offer any resistance to HIV-1 infection [Smith et al., 1997a; Smith et al., 1997b; Anzala et al., 1998; Kostrikis et al., 1998; Rizzardi et al., 1998]. Some studies, however, do not confirm the effect of the CCR2V64I SNP on disease progression [Michael et al., 1997; Eugen-Olsen et al., 1998; Ioannidis et al., 1998].

Allelic frequencies for the CCR2V64I SNP range from 0.1 to 0.25 in different population groups [Smith et al., 1997a; Michael et al., 1997; Williamson et al., 2000] and it's effect is more apparent in Africans when compared to Caucasians [Mummidi et al., 1998]. Initially, it was suggested that the CCR2V64I SNP is in linkage disequilibrium with the CCR5Δ32 mutation, but it was shown that the CCR2V64I SNP occurs invariably on a chromosome with the wildtype CCR5 allele [Kostrikis et al., 1998]. The CCR2V64I SNP has however been found to be in linkage disequilibrium with the 59653-T promoter variant of CCR5 [Kostrikis et al., 1998; Martinson et al., 2000], but the functional significance of this finding is unknown.

A second CCR2 SNP has been previously reported and involves a silent mutation (AAC – AAT) occurring at codon 260 (CCR2N260) [Genbank database (www.ncbi.nlm.nih.gov)]. Previously, our group found the T allele to occur at a much higher allelic frequency than the commonly reported C allele within the South African population groups [Petersen et al., 2001]. Recently, novel CCR2 mutations and SNPs have been identified and are discussed in chapter 4. Further investigations regarding the potential effect of CCR2 gene variants, could ultimately result in a better understanding of the precise role of CCR2 in HIV-1 entry.
2.2.3. Stromal derived factor-1 (SDF1) (CXCL12)

SDF1 (CXCL12), a member of the CXC or α-chemokine family, is the natural ligand for the HIV-1 co-receptor CXC receptor 4 (CXCR4) and suppresses infection of T-tropic or SI viruses [Bleul et al., 1996a; Oberlin et al., 1996] by down-regulating CXCR4 surface expression and thus interfering with HIV-1 fusion and entry [Amara et al., 1997; Signoret et al., 1997]. The SDF1 amino terminal residues 1 to 8 forms an important receptor-binding site, but only residues 1 and 2 are required for receptor activation. Residues 12 to 17, located in the loop region, also form a vital receptor-binding site and it has been proposed as an important initial docking site of SDF1 with CXCR4 [Crump et al., 1997]. The carboxy terminus of SDF1 has no function in receptor binding and activation, but does enhance the biological activity of the SDF1 amino terminus [Luo et al., 1999]. SDF1 is an extremely efficacious chemoattractant for a variety of cells, including T-lymphocytes and monocytes [Bleul et al., 1996b] and also plays a fundamental role in development [Nagasawa et al., 1996].

The SDF1 gene, approximately 10 kb in length, was cloned from mouse bone marrow stromal cells and encodes 2 isoforms, namely SDF1α (89 amino acids) and SDF1β (93 amino acids). The SDF1β protein has four additional amino acid residues in the carboxy terminus [Tashiro et al., 1993; Nagasawa et al., 1994]. The first 21 amino acid residues of the SDF1α and SDF1β proteins form an amino acid-cleaved signal peptide and a SDF1α form processed at the carboxy terminus to generate a 67 residue protein has been purified [Bleul et al., 1996b]. The SDF1α and SDF1β isoforms are encoded by three and four exons, respectively and are the result of alternative splicing of a single SDF1 gene. Although most of the CXC chemokines
are located on the long arm of chromosome 4, the \textit{SDF1} gene is an exception and it's location is at band q11 of chromosome 10 [Shirozu et al., 1995] (Figure 4).

\begin{center}
\textbf{\textit{SDF1} (chromosome 10q11)}
\end{center}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{sdf1_chromosome_diagram.png}
\caption{\textit{SDF1} has 2 gene transcripts, \textit{SDF1}α and \textit{SDF1}β. The coding regions for \textit{SDF1}α and \textit{SDF1}β are found within exons 1 to 3 and exons 1 to 4, respectively. Black blocks represent the non-coding regions and the grey blocks represent the coding regions. The lengths of intron 1 and 2 are unknown. Exons and introns are not drawn to scale.}
\end{figure}

A common SNP of the \textit{SDF1}β gene transcript, designated \textit{SOF1}-3′ A, has been identified at position +801 (counting from the ATG start codon) in the 3′ UTR and involves a G to A transition [Winkler et al., 1998]. The \textit{SDF1-3′} A SNP has been observed in different population groups at allelic frequencies ranging from 0.06 to 0.26, occurring more frequently in Caucasians when compared to Africans [Mummidi et al., 1998; Winkler et al., 1998; Williamson et al., 2000]. The presence of the \textit{SOF1}-3′ A SNP in the homozygous state has been associated with delaying the rate of disease progression to AIDS, while the \textit{SDF1-3′} A SNP in the heterozygous state
has no effect on disease progression to AIDS [Hendel et al., 1998; Martin et al., 1998; Winkler et al., 1998]. It was thus hypothesised that the SNP, in a potential regulatory region, upregulates the biosynthesis of SDF1, making the protein more highly available to compete with HIV-1 for binding to the CXCR4 chemokine receptor, and thereby blocking infection of the host cells by the T-tropic/SI viruses [Winkler et al., 1998]. This hypothesis was tested in vitro and the results indicated that the SDF1-3' A SNP does not affect SDF1 RNA synthesis or translation of the SDF1 RNA, which suggests that the 3'UTR region does not regulate SDF1 expression [Arya et al., 1999].

Various studies have however shown that the SDF1-3' A SNP in the homozygous state is not consistently associated with delaying the progression to AIDS, but rather rapid disease progression to death [Mummidi et al., 1998]; prolonged [van Rij et al., 1998] or decreased [Brambilla et al., 2000] survival after AIDS is diagnosed; low CD4 cell counts [Balotta et al., 1999]; and no effect on disease progression [Meyer et al., 1999]. The SDF1-3' A SNP in the heterozygous state has also recently been associated with an increase in HIV-1 vertical transmission from mother to child [John et al., 2000].

Further studies of large and informative study cohorts are thus important to determine specific associations between the SDF1-3' A SNP and susceptibility to HIV pathogenesis, which could ultimately result in a better understanding of the functioning of SDF1. A pilot study was performed to determine the allelic frequencies of the SDF1-3' A SNP within a South African population and is presented in chapter 5.
2.3. Mutation Detection

The identification of mutations that influences host genetic susceptibility of an individual to develop a particular disease has resulted in the need for mutation detection, which today forms the main focus of many research studies. A variety of specific mutation detection methods are available for use, but they differ in a number of ways, most importantly being their reliability and sensitivity to detect genetic variants. The ideal mutation detection method should have 100% sensitivity and specificity with no false positives or negatives, allow for the screening of large DNA sequences, not involve the use of hazardous reagents, not require time-consuming or intensive manual labour and be cost-effective.

DNA sequencing [Sanger et al., 1977; Maxam and Gilbert, 1977] is a common direct screening method that is used for the identification of genetic variants and although it has a mutation detection rate of virtually 100%, it involves intensive labour and great expenses when performing mutation detection on a large scale. There are however several indirect pre-screening mutation detection methods available for the identification of genetic variants. Direct sequencing is only then used to determine the precise location of mutations detected using an indirect pre-screening method.

Future advances in mutation detection include DNA chip technology [Lipshutz et al., 1995; Ginot, 1997], which involves the hybridisation of labelled test single-strand DNA to a microarray of known oligonucleotides on glass or silicon chips. The presence of a mutation is signalled by differing patterns of hybridisation between the wild-type and test DNA. This technique requires great expertise and presently it is very costly due to the use of specific equipment.
2.3.1. Pre-screening methods of mutation detection

Polymerase chain reaction (PCR), which involves the amplification of DNA fragments and the formation of heteroduplexes between the wild-type and mutant DNA strands, usually forms the basis of indirect pre-screening mutation detection methods [Mullis and Faloona, 1987 Saiki et al., 1988]. A few commonly used pre-screening mutation detection methods with their main respective advantages and disadvantages are shown in Table 1.

2.3.1.1. Denaturing gradient gel electrophoresis (DGGE)

Denaturing gradient gel electrophoresis (DGGE) is the pre-screening mutation detection method we used for the identification of previously reported and novel genetic variants in the entire coding regions of the CCR5 and CCR2 (CCR2A and CCR2B) genes and for the partial analysis of the 3' UTR of the SDF1β gene transcript. Our reasons for selecting this particular pre-screening mutation detection method was based on it's main advantages (discussed below) when compared to the advantages of other pre-screening mutation detection methods (Table 1).

DGGE, developed in 1983 by Fischer and Lerman, is a PCR-based method and involves the differential melting behaviour of double-stranded DNA molecules in an increasing concentration gradient of denaturants (urea and formamide, UF) at a fixed and elevated temperature. The melting behaviour is highly sequence-dependant and is thus determined by the composition and order of nucleotide base pairs within a DNA fragment. The principles of DGGE are shown in Figure 5. As the double-stranded DNA fragment passes through the denaturing gel, it will melt and undergo a conformational change, which results in it's electrophoretic mobility being reduced as
Table 1: A few of the indirect pre-screening mutation detection methods that are commonly used for the identification of genetic variants.

<table>
<thead>
<tr>
<th>Pre-screening method</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction fragment length polymorphism (RFLP) analysis:</td>
<td>Simplicity in use and cost-effective. Allows rapid mutation detection. Requires the use of equipment available in most laboratories.</td>
<td>Incomplete digestion of the restriction endonucleases may give false positive or negative results. Can only be used for the detection of specific sequence variants.</td>
<td>Parry et al., 1990; Reviewed in Pourzand and Cerutti, 1993.</td>
</tr>
<tr>
<td>Single-stranded conformation polymorphism (SSCP) analysis:</td>
<td>Versatile and does not require any expertise. Allows rapid mutation detection. Requires the use of equipment available in most laboratories.</td>
<td>Only 80% mutation detection rate for DNA fragments up to 300bp in length. Efficiency of method depends on a number of experimental conditions.</td>
<td>Orita et al., 1989; Reviewed in Cotton et al., 1998.</td>
</tr>
<tr>
<td>Chemical cleavage of mismatch (CCM):</td>
<td>Sensitivity for mutation detection is +/- 99%. Can be used to analyse DNA fragments which are more than 1kb in length (up to 1.5 – 1.8 kb)</td>
<td>Involves the use of hazardous chemicals and the % mutation detection rate can be reduced if low quality chemicals are used.</td>
<td>Cotton et al., 1988; Youil et al., 1995; Reviewed in Ellis et al., 1998.</td>
</tr>
<tr>
<td>Denaturing gradient gel electrophoresis (DGGE) analysis:</td>
<td>100% mutation detection rate for fragments up to 500bp in length. Rapid mutation detection and clear visualisation of mutations. Theory for the entire concept.</td>
<td>Standardising of assays is a time-consuming procedure. Extra costs for the use of with GC-clamped primers. Expertise is required.</td>
<td>Fischer and Lerman, 1983; Reviewed in van Orsouw, 1998; Wu, 1999; Hayes, 1999b;</td>
</tr>
<tr>
<td>Denaturing high-performance liquid chromatography (DHPLC) analysis:</td>
<td>Sensitivity and specificity of mutation detection exceeds 96%. DNA fragments as large as 1.5kb in length can be analysed.</td>
<td>Sensitivity of the method is dependant on temperature. Under completely denaturing conditions, it is not possible to resolve C to G transversions. Homozygous mutations cannot be detected indirectly.</td>
<td>Underhill et al., 1997; Liu et al., 1998; Reviewed in Xiao and Oefner, 2001.</td>
</tr>
</tbody>
</table>
depicted on the schematic representation of a DGGE time-travel gel (a gel that is loaded with the same amplified product at hourly intervals) [Figure 5B]. The introduction of a guanine and cytosine (GC)-rich fragment (GC-clamp) to the 5' end of one of the primers prevents total strand dissociation during fragment amplification [Myers et al., 1985a]. The GC-clamp also allows for the detection of all mutations, including single base changes, which makes DGGE a powerful pre-screening method with virtually 100% mutation detection sensitivity [Sheffield et al., 1989; Abrams et al., 1990]. A heteroduplexing step, which involves denaturation and renaturation of the wild-type and mutant DNA and thus the formation of mismatched heteroduplexes, also further enhances the mutation detection sensitivity of DGGE [Myers et al. 1985b; Figure 5A]. Mismatched heteroduplexes have a lower stability than the homoduplexes and thus always melt earlier within the denaturing gel. This results in heterozygous mutations being visualised as four bands (two heteroduplexes and two homoduplexes), while homozygous mutations are visualised as a single "shifted" DGGE band [Figure 5C]. DGGE has an advantage over other electrophoretic-based pre-screening methods by having a mutation detection rate of 100% for DNA fragments up to 500bp in length. However, it is at a disadvantage to the cleavage-based pre-screening methods, which allow for the detection of mutations in DNA fragments more than 1kb in length.

The design of a successful DGGE-assay is thus dependant on the melting profile of the DNA, the chosen DGGE primers, application of heteroduplexing and the type of gel system used. Using the melt 87 computer program [Lerman and Silverstein, 1987] and considering the conditions for improved DGGE mutation detection previously described by Wu et al., 1998, DGGE primers are designed and a single
melting domain is determined for the DNA fragment. In some cases, the "single" melting domain of the DNA fragment can only be achieved by certain primer modifications, involving the addition of T/A or G/C tails and changing the position and length of the GC-clamp. These adaptations were included in our assay designs. The time-consuming procedure that is used to design the DGGE assay and the extra costs for the GC-clamped primers may be regarded as a disadvantage.

The amount of denaturant (UF) that is required for optimal melting of the DNA fragment can be theoretically calculated using the following formula: \( \% \text{UF} = \frac{[\text{melting temperature} - \text{buffer temperature}] \times 100}{32} \). Additional factors such as the buffer composition and concentration, and the electrophoretic voltage used, are not taken into consideration when applying the formula above, but can also influence the melting behaviour of the DNA fragment. The use of a specific gel system and the choice of a single set of experimental conditions (e.g., gel composition, temperature and voltage), as opposed to SSCP which requires more than one gel condition, could be based on conditions previously described for improving DGGE analysis [Hayes et al., 1999a].

Heteroduplexing is important for the detection of mutations involving C/G to G/C transversions and 1 bp deletions or insertions as homoduplexes may have the same/similar melting behaviour and result in mutations being missed [Myers et al. 1985b]. Using DGGE, it thus possible to identify the different mutations by their specific banding pattern. Verification of commonly occurring mutations can be achieved by the mixing of samples with similar DGGE banding patterns, followed by heteroduplexing and electrophoresis on a denaturing gel. Only samples showing
additional heteroduplex bands are subjected to sequencing and therefore the repeated sequencing of a specific mutation is avoided [Guldberg and Guttler, 1993]. This is an important advantage for SNP analysis in large study cohorts.

Considering the advantages of DGGE, which include high mutation detection sensitivity; a theoretical framework that explains the principles of the method; the easy visualisation of mutations; the possibility to confirm commonly occurring mutations by mixing and heteroduplexing and the application of this method in large-scale analysis, it was chosen to perform comprehensive mutation detection analysis for the respective studies discussed in chapters 3, 4 and 5.
PCR amplification is performed using one GC-clamped primer.

Heteroduplexing step includes: denaturation and reannealing.

Heteroduplexes

Double stranded DNA

Partially melted DNA

Single-stranded DNA held together by GC-clamp

Figure 5. A. PCR amplification of DNA fragments for DGGE using a GC-clamped primer, followed by a heteroduplexing step, which involves denaturation and reannealing to form a wild-type homoduplex, a mutant homoduplex and heteroduplexes. B. The principles of DGGE are depicted as a time-travel DGGE-gel (Myers et al., 1987). C. The detection of mutations by electrophoresis through a increasing denaturing gradient, where the wild-type and mutant DNA have different melting profiles. 1) wild-type. 2) homozygous mutant. 3) heterozygous mutant (Adapted from Hayes. 1999b).
Chapter 3

Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis (DGGE) assay

Desiree C. Petersen¹, Maritha J. Kotze¹, Michele D. Zeier², Ashraf Grimwood³, Deon Pretorius⁴, Eftyhia Vardas⁵, Estrelita Janse van Rensburg⁴ and Vanessa M. Hayes⁴,⁶

Departments of Human Genetics¹, Internal Medicine², Medical Virology⁴, and Urology⁶, University of Stellenbosch, Tygerberg Medical School; Chapel Street Community Health Clinic, Woodstock³; and the AIDS programme, Medical Research Council, Congella, Durban⁵, South Africa.

AIDS (2001) 15:171-177
Objectives: Previous studies have indicated that most CC chemokine receptor 5 (CCR5) gene mutations are either relatively specific to different population groups or rarely observed in Africans. The objective of this study was therefore, 1) to develop a comprehensive mutation detection assay for the entire coding region of the CCR5 gene and 2) to identify novel mutations, which may play a role in genetic susceptibility to human immunodeficiency virus type 1 (HIV-1) infection, within the diverse South African population.

Design: The study cohort consisted of 103 HIV seropositive patients and 146 HIV seronegative controls of predominantly African descent.

Methods: We designed a mutation detection assay for the entire coding region of the CCR5 gene, which included amplification of part of the coding region of the CCR2 gene. The assay, based on denaturing gradient gel electrophoresis (DGGE), allows for the complete analysis of 10 individuals per denaturing gel.

Results: The use of the CCR5-DGGE assay led to the identification of seven novel and six previously reported mutations. All novel mutations, including a common polymorphism at codon 35, occurred exclusively in non-Caucasians, indicating possible African origin.

Conclusion: We present a comprehensive DGGE mutation detection assay for the entire coding region of the CCR5 gene. Application of this assay resulted in the identification of novel CCR5 mutations, which may have a significant effect on the normal functioning of CCR5 and thus contribute to host variability and susceptibility to HIV-1 infection and/or progression to acquired immune deficiency syndrome (AIDS) within this population.

Keywords: CCR5, DGGE, novel mutations, HIV-1 susceptibility, South Africa
Introduction

Various chemokine receptors have been identified as co-receptors necessary for the cellular infection of the human immunodeficiency virus (HIV) [Deng et al., 1996; Dragic et al., 1996; Doranz et al., 1996; Feng et al., 1996]. The CC-chemokine receptor 5 (CCR5), a seven transmembrane G-coupled protein, consists of 352 amino acids and binds the β-chemokines RANTES, MIP-1α and MIP-1β [Samson et al., 1996a]. It is also recognised as the principle co-receptor for the macrophage-tropic (M-tropic) strains or nonsyncytium-inducing (NSI) strains of HIV-1 and facilitates fusion of the viral envelope protein with the CD4+ molecules during the asymptomatic phase of infection [Deng et al., 1996; Dragic et al., 1996; Alkhatib et al., 1996; Choe et al., 1996]. The CCR5 cell surface expression may therefore have a direct influence on the individual's variability and susceptibility to HIV-1 infection.

The CCR5 gene, located at band p21 of chromosome 3 [Liu et al., 1996], consists of 4 exons and 2 introns, of which exon 4 contains the entire open reading frame (ORF) [Mummidi et al., 1997]. The most widely studied CCR5 mutation is a 32 base pair deletion (CCR5Δ32) within the coding region, which results in premature termination of translation and the formation of a truncated protein [Liu et al., 1996; Samson et al., 1996b]. Individuals homozygous for the CCR5Δ32 mutation have been shown to display protection against HIV-1 infection, although this protection is not absolute. Heterozygous carriers display partial protection against HIV-1 infection and evidence exists that a single deletion mutant in HIV seropositive individuals may slow progression to acquired immune deficiency syndrome (AIDS) [Liu et al., 1996; Samson et al., 1996b; Dean et al., 1996; Huang et al., 1996; Zimmerman et al., 1997]. Population surveys have shown that the CCR5Δ32 mutation is largely
confined to Caucasians (allele frequency of 0.092) and is extremely rare in Africans [Samson et al., 1996b; Huang et al., 1996; Martinson et al., 1997]. This suggests the presence of other possibly protective/causative mutations in the African populations and underscores the importance of comprehensive CCR5 mutation analysis in the diverse South African populations. Several other genetic variants in the CCR5 coding region have been described. However, their role in HIV-1 infection or progression to AIDS could not be deduced, due to the low allelic frequencies of these mutations in the population groups studied [Dean et al., 1996; Ansari-Lari et al., 1997; Carrington et al., 1997; Quillent et al., 1998; Carrington et al., 1999b].

In this study, we describe a comprehensive CCR5 mutation detection assay for the entire coding region of the gene, using denaturing gradient gel electrophoresis (DGGE). DGGE, developed by Fischer and Lerman in 1983, is a polymerase chain reaction (PCR)-based method and is believed to be the most powerful of the pre-screening methods of mutation detection currently available. The technique involves the differential melting of double stranded DNA molecules in a gradient with an increasing concentration of urea and formamide (UF). The addition of a guanine and cytosine (GC)-rich fragment (GC-clamp), introduced during fragment amplification, prevents total strand dissociation and allows for the detection of single base mutations, making DGGE virtually 100% sensitive [Sheffield et al., 1989; Abrams et al., 1990].

Using this assay, 103 HIV seropositive patients and 146 healthy controls were screened for mutations in the coding region of the CCR5 gene. Our results obtained in the unique South African population are presented in this study.
Methods

Sample population

Blood samples were drawn from 103 HIV seropositive patients (35 male; 68 female) residing in the Western Cape of South Africa (Tygerberg Hospital and Woodstock Chapel Street Community Health Clinic). Disease progression of the majority of these individuals was unknown. Blood samples from 146 HIV seronegative controls (56 male; 91 female) were obtained from the Western Province Blood Transfusion Service, South Africa. The study cohort consisted of Africans, predominantly Xhosa (70 HIV+ and 64 HIV-), Coloureds (26 HIV+ and 72 HIV-), and to a lesser degree Caucasians (7 HIV+ and 2 HIV-) and Asians (8 HIV-). An additional nine seronegative “high-risk” commercial sex workers of Zulu descent were obtained from KwaZulu-Natal, South Africa. In this study, “African” refers to South Africans of central African descent; “Coloured” refers to individuals of mixed ancestral descent, including San, Khoi, African Negro, Madagascan, Javanese and European origin; and “Caucasian” refers to South Africans of European descent, mainly Dutch, French, German and British origin. Informed consent for the study was obtained from all participants. The Ethics Review Committee of the University of Stellenbosch approved the study protocol.

Primer Design

DGGE PCR primers (Table 1) were designed for the entire coding region, including the donor/acceptor splice site of intron 3/exon 4, of the CCR5 gene, using the melt 87 computer program [Lerman and Silverstein et al., 1987] and conditions described by Wu et al, 1998. The region to be analysed was divided into six overlapping PCR fragments (A to F). In order to prevent complete strand dissociation during
amplification, a GC-rich-fragment (GC-clamp) was added to the 5' end of one of the primers in each primer set. An additional stretch of 10 GC or AT nucleotides were added to either the 5' end of the non-clamped primer (fragment B) or between the GC-clamp and the primer (fragments C and D), respectively, to ensure a single melting domain and thus optimal mutation detection of the fragments (Table 1).

Table 1. CCR5 primer sets and experimental conditions for PCR amplification and DGGE.

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Amplimers, 5'-3'</th>
<th>Size (bp)</th>
<th>Melting</th>
<th>Annealing</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[40GC]TGGAGGGCAACTAAATACAT CGATTTGCTTCACATTGATT</td>
<td>196</td>
<td>67</td>
<td>54</td>
</tr>
<tr>
<td>B</td>
<td>[10GC]ATTATACATCGAGGGGCTGC [40GC]AGCATAATGAGCCGAGAAGG</td>
<td>280</td>
<td>74</td>
<td>60</td>
</tr>
<tr>
<td>C</td>
<td>[40GC][10AT]CTGACCATCTCTGACCTGTGTT GATGATTTCTGAGAGAGAG</td>
<td>332</td>
<td>73</td>
<td>60</td>
</tr>
<tr>
<td>D</td>
<td>[40GC][10AT]ACTTGGGCTGTTGCTGTGTGTT CATTTCGACACCGGAAGCAG</td>
<td>276</td>
<td>72</td>
<td>60</td>
</tr>
<tr>
<td>E</td>
<td>TCAATGCGTCTCTGCTAGTCG [40GC]GGTGTCAGGAGAAGCGAAGGACAA</td>
<td>192</td>
<td>72</td>
<td>58</td>
</tr>
<tr>
<td>F</td>
<td>[40GC]CTCTCTCTGCTGGCTCTGCCTAC GTCACCAGCCCCACTTGAGTC</td>
<td>390</td>
<td>74</td>
<td>60</td>
</tr>
</tbody>
</table>

bp, base pair
GC-clamps used were as follows: [40GC] CGCCCCCGCGGCCCCCGCCGGCCGGGCGGCCCCCCCGCCGCGCCGCGGCGCCGCGCCGCGG
[10GC] CGCCGCGCG
AT-stretch used was as follows: [10AT] TATAATATTA
DNA amplification

Genomic DNA was extracted using conventional methods and amplified using the DGGE primer sets listed in Table 1. Each PCR reaction mixture of 50 μl in total volume contained 100 ng of genomic DNA, 0.1 mM of each deoxyribonucleoside triphosphate (dNTP), 20 pmol of each primer, 2.5 mM of a 10 x Mg⁺ reaction buffer and 0.5 units of DNA Taq polymerase (Boehringer Mannheim). Amplification was performed using the following cycling conditions; an initial denaturation at 96°C for 3 minutes, followed by 32 cycles of denaturation at 96°C for 45 seconds, annealing for 1 minute (annealing temperatures shown in Table 1), and elongation at 72°C for 1 minute. The last cycle was followed by an additional extension step of 72°C for 10 minutes. Amplicons to be subjected to DGGE analysis required an additional heteroduplexing step, which involves denaturation at 96°C for 10 minutes, followed by renaturation for 45 minutes at the annealing temperature of the amplification. The amplified products were checked by electrophoresis of 5 μl (10%) of each sample in a 2% agarose gel.

Denaturing gradient gel electrophoresis (DGGE)

The DGGE conditions were optimised using conditions previously described for broad-range DGGE analysis [Hayes et al., 1999a]. The six amplicons were pooled into three groups (group 1: fragments A, D and C; group 2: fragments E and B and group 3: fragment F) and electrophoresed in a 9% polyacrylamide gel containing a 30% to 70% urea and formamide (UF) denaturing gradient (100% UF = 7mol/L urea per 40% deionised formamide), at 59°C for 110 volts overnight. Gels were stained with ethidium bromide and photographed under an UV transilluminator. The optimally designed CCR5-DGGE assay allows for the complete analysis of 10 patients per denaturing gel (Figure 1).
Figure 1. DGGE banding pattern of 6 pooled amplicons of 10 patients covering the entire coding region of the CCR5 gene. The 30%-70% UF/9% PAA denaturing gel was electrophoresed at 59°C for 110 volts overnight and visualised by staining with ethidium bromide. Lanes 1-10 contains group 1 (fragments A, D and C); lanes 11-20, group 2 (fragments E and B) and lanes 21-30, group 3 (fragment F). The multiple bands depicted for fragment E are explained in further detail in fig. 2a. Individuals in lanes 13 and 19 are heterozygous for the codon 35 polymorphism (fragment B) and the individual in lane 24 is heterozygous for the codon 335 polymorphism (fragment F).

DNA sequencing and mutation confirmation

Amplified products showing aberrant DGGE banding patterns were subjected to automated sequencing using a non-GC-clamped primer and the dye terminator sequencing kit of Applied Biosystems (www.appliedbiosystems.com). Confirmation of commonly occurring polymorphisms were performed by mixing samples showing similar DGGE banding patterns, followed by a heteroduplexing step before electrophoresis on a denaturing gel [Guldberg and Gutler, 1993]. Samples showing additional heteroduplex bands were subjected to sequencing.
Statistical analysis

Allele frequencies were determined by allele counting. Testing for significance of heterogeneity in mutation frequencies among HIV seropositive and HIV seronegative subjects were based on the Fisher’s exact test. The Hardy – Weinberg principle was applied to measure the maintenance of the allelic frequency for the CCR5 mutations in the closely matched HIV seropositive and HIV seronegative African population. The statistical calculations are shown in appendix 3.

Results

Application of the CCR5-DGGE assay presented in this study, resulted in the identification of seven novel point mutations and six previously reported mutations as listed in Table 2, according to the codon in which they occur.

Of the seven novel mutations detected in this study, four may ultimately result in structural changes of the CCR5 protein. The first, a nonsense mutation at codon 225 (CGA-TGA), results directly in the formation of a truncated protein due to the conversion of the amino acid Arginine to a premature stop codon. Both the codon 2 (GAT-GTT) and codon 225 (CGA-CAA) mutations result in a non-conservative amino acid change (replacement of an amino acid by another with different chemical properties), from an Aspartic acid to a Valine, and an Arginine to a Glutamine, respectively. The fourth mutation at codon 107 (CTC-TTC) although resulting in a conservative amino acid change (Leucine to Phenylalanine), involves the inclusion of an aromatic side chain, which may have structural and/or functional implications. All the individuals who presented with the codon 107 (CTC-TTC) mutation, also presented with the codon 225 (CGA-TGA) mutation and no individual was found to
have only one of the two mutations. The remaining three novel mutations were all silent mutations occurring at codons 35, 89 and 162, respectively. A high allelic frequency for the codon 35 (CCG-CCA) polymorphism was detected in both the HIV seropositive and HIV seronegative individuals from African and Coloured descent and occurred in a homozygous state in a single HIV seropositive Coloured female. This novel polymorphism was totally absent in Caucasians. Due to the low numbers of Caucasian individuals in the study cohort, further screening of 28 Caucasians for the codon 35 polymorphism confirmed the absence of this mutation in this population group. All the above mentioned novel mutations were found exclusively in individuals from African or Coloured ethnic background.

The six previously reported mutations observed in this study, include the most commonly studied CCR5Δ32 mutation at codon 185. This deletion mutation was observed heterozygously in one HIV seropositive Coloured and one Caucasian and in five HIV seronegative Coloureds, while it was absent in the Africans studied. Three non-conservative mutations at codons 55 (Leucine to Glutamine), 223 (Arginine to Glutamine) and 339 (Tyrosine to Phenylalanine), all previously reported by Ansari-Lari et al, 1997, were observed in one HIV seropositive Caucasian, one seronegative Coloured and one seronegative African, respectively. The silent mutation at codon 75, previously reported by Carrington et al, 1997, was found to be present in one HIV seropositive African. The codon 335 polymorphism, involving an amino acid change from Alanine to Valine, has also been previously reported by Ansari-Lari et al, 1997 and in our study we detected this polymorphism in four HIV seropositive and four seronegative individuals of African and Coloured descent.
Table 2. CCR5 mutations detected in 103 HIV seropositive patients and 146 HIV seronegative controls, ordered according to the codon in which they occurred.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Base change</th>
<th>Fragment</th>
<th>Africans</th>
<th>Coloureds</th>
<th>Caucasians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIV + (n = 140)</td>
<td>HIV - (n = 128)</td>
<td>HIV + (n = 52)</td>
</tr>
<tr>
<td>D2V*</td>
<td>GAT-GTT</td>
<td>A</td>
<td>1 (.007)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P35*</td>
<td>CCG-CCA</td>
<td>B</td>
<td>6 (.043)</td>
<td>9 (.070)</td>
<td>9 (.173)</td>
</tr>
<tr>
<td>L55Q</td>
<td>CTG-CAG</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S75</td>
<td>TCT-TCC</td>
<td>B</td>
<td>1 (.007)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Y89*</td>
<td>TAT-TAC</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L107F*#</td>
<td>CTC-TTC</td>
<td>C</td>
<td>1 (.007)</td>
<td>2 (.016)</td>
<td>1 (.019)</td>
</tr>
<tr>
<td>P162*</td>
<td>CCA-CCG</td>
<td>D</td>
<td>1 (.007)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>△32(185)</td>
<td>-</td>
<td>D</td>
<td>0</td>
<td>0</td>
<td>1 (.019)</td>
</tr>
<tr>
<td>R223Q</td>
<td>CGG-CAG</td>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R225X*#</td>
<td>CGA-TGA</td>
<td>E</td>
<td>1 (.007)</td>
<td>2 (.016)</td>
<td>1 (.019)</td>
</tr>
<tr>
<td>R225Q*</td>
<td>CGA-CAA</td>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A335V</td>
<td>GCA-CTA</td>
<td>F</td>
<td>4 (.029)</td>
<td>2 (.016)</td>
<td>1 (.019)</td>
</tr>
<tr>
<td>Y339F</td>
<td>TAC-TTC</td>
<td>F</td>
<td>0</td>
<td>1 (.008)</td>
<td>0</td>
</tr>
</tbody>
</table>

n, number of alleles; HIV+, seropositive; HIV-, seronegative

* Novel mutation identified in this study

# Mutations occurring together in patients

No mutations detected in 8 Asian HIV seronegative controls
In fragment E, as depicted in Figure 2a, all HIV seropositive and seronegative individuals presented with additional DGGE bands, lower (L) and/or upper (U), in combination with the normal (N) CCR5 band. Heteroduplex bands at a low percentage of urea and formamide (UF) were also noted. Excision of these aberrant bands from the gel, followed by direct sequencing, revealed 11 nucleotide variations occurring in the lower band and an additional twelfth variation was included in the upper band. Blasting the mutant sequence, using the Genbank database (www.ncbi.nlm.nih.gov), revealed that this sequence forms part of the chemokine receptor 2 (CCR2) gene, including codons 217 to 267 (Figure 2b). The polymorphism at codon 260 (AAC to AAT) of the CCR2 gene showed an allelic frequency of 0.62 for the T nucleotide and 0.38 for the C nucleotide within our population group. All mutations occurring within fragment E were confirmed as being CCR5 mutations by recognition of the DGGE band intensities and confirmation by excision of additional heteroduplex bands from the gel followed by direct sequencing.

**Discussion**

In this study, we describe a comprehensive and efficient mutation detection assay for the entire coding region of the CCR5 gene. This assay, based on DGGE with its virtual 100% sensitivity, allows for the complete analysis of 10 patients per denaturing gel. We applied this assay to screen for possible novel CCR5 sequence variants in a predominantly African and Coloured HIV seropositive and HIV seronegative cohort from South Africa. Most studies to date have restricted their analysis to the CCR5Δ32 mutation, which although fairly common in Caucasians, is extremely rare in the African populations. Comprehensive analysis of the CCR5 gene is, therefore, of vital importance in the diverse South African population.
Figure 2. Fragment E of the CCR5 gene. **A**) DGGE banding pattern of the CCR5 fragment E, in combination with codons 217 to 267 of the CCR2 gene. All samples (lanes 1-3) are homozygous normal (N band) for the CCR5 gene. Samples were either homozygous or heterozygous for the CCR2 codon 260 polymorphism. Lane 1, homozygous 260-T (U band); lane 2, homozygous 260-C (L band); and lane 3, heterozygous 260-CIT (U and L homoduplex and heteroduplex bands). Additional CCR5/CCR2 heteroduplex bands melt at a low percentage of denaturant in the DGGE gel, due to the high number of nucleotide mismatches. **B**) CCR5 and CCR2 gene sequences amplified using DGGE fragment E primer set (arrows). The codon 260 (AAC/AAT) polymorphism of CCR2 is in indicated in bold.
Using the described assay, seven novel CCR5 mutations were identified in the African and Coloured populations. No novel mutations were identified in the Caucasian or Asian populations, although numbers were small. Novel mutations at codons 107 and 225 (CGA-TGA), which occur simultaneously, and at codons 2 and 225 (CGA-CAA), may affect the functioning of CCR5 and thus provide possible protection against HIV infection and/or progression to AIDS. One cannot, however, exclude the possibility that the three novel 'silent' mutations (codons 35, 89 and 162) detected in this study, affect disease progression by altering regulatory elements that affect RNA splicing [D'Souza et al., 1999, Lorson et al., 1999]. The novel codon 35 polymorphism (CCG to CCA), occurring at an allelic frequency of 0.06 and 0.1 in the African and Coloured populations, respectively, and its absence in Caucasians, indicates this polymorphism has a definite African origin. The closely matched HIV seropositive and HIV seronegative African population are in Hardy – Weinberg equilibrium for the codon 35 polymorphism. No significant association was observed for both the African ($P = 0.4273$) and Coloured ($P = 0.0516$) HIV seropositives compared to the HIV seronegatives. Sample numbers are however small and the significance of the codon 35 polymorphism thus warrants further investigation. The Coloured female homozygous for the polymorphism showed normal disease progression (progression to AIDS within 8 to 10 years after HIV infection). Due to the lack of clinical information regarding disease progression of majority of the HIV seropositive patients, the potential consequences of the different novel mutations could not be evaluated and no significant associations could be made. It is, therefore, necessary to obtain updated reports on the disease progression of all the HIV seropositive patients.
The CCR5Δ32 mutation, generally restricted to Caucasians, was found to be absent in the 134 Africans studied, while it occurred at an allelic frequency of 0.03 in the Coloured population. The presence of this deletion mutation in the Coloured population may be a reflection of the possible admixture with people of Caucasian descent [Martinson et al., 1997]. The presence of the CCR5Δ32 mutation in Coloureds also provides evidence that no genetic incompatibility between ethnic groups exists for this mutation [Kantor and Gershoni, 1999]. The two HIV seropositive individuals, heterozygous for the CCR5Δ32 mutation, include an asymptomatic (seven years after infection) Coloured male and an asymptomatic Caucasian male with long-term non-progression (15 years since date of infection).

The polymorphism at codon 335 was only observed in Africans (allelic frequency of 0.02) and Coloureds (allelic frequency of 0.02). This supports previous studies, which suggest that the polymorphism has an African origin with an allelic frequency of approximately 0.03 and is rarely observed in Caucasian populations [Ansari-Lari et al., 1997; Carrington et al., 1997; Carrington et al., 1999b]. The mutations at codons 55, 75, 223 and 339 were found at low allelic frequencies.

Within the nine "high-risk" seronegative commercial sex workers of Zulu descent, no possibly protective mutations were found within the coding region of the CCR5 gene. As this assay does not include the promoter region, the remaining 5' and 3' end untranslated regions and the intronic sequences, we cannot exclude the possibility that some significant mutations occurring in these regions may have been missed. Although this study cohort is small, our findings suggest that other factors (including the possibility of alternative gene involvement) may provide protection against HIV-infection within this population group.
Due to the high degree of sequence homology between the CCR5 and CCR2 genes, part of CCR2 was simultaneously amplified using the CCR5 fragment E primer set. Therefore, this assay also allows for comprehensive analysis of codons 217 to 267 of the CCR2 gene. In our study, the T-allele of codon 260 (AAC to AAT) was found to occur more frequently (0.62) than the commonly reported C-allele (0.38). No statistically significant differences in allelic frequencies for this polymorphism were observed and no additional CCR2 sequence variants were detected.

The relatively high frequency of novel mutations observed in the African and Coloured populations demonstrates the effectiveness of the CCR5-DGGE assay and importance of comprehensive CCR5 gene analysis in populations where the CCR5Δ32 mutation is rare. The recently admixed Coloured population of South Africa may, therefore, represent a valuable candidate population for the identification of genes/mutations underlying susceptibility to HIV/AIDS within the African context. Future analysis on the effect of the novel mutations on the functioning of CCR5, will result in a better understanding of this chemokine receptor and may contribute to the development of HIV therapeutic and preventative measures that focus on the interaction of HIV with the host proteins.

Acknowledgements
Thanks to Dr John Gasson for co-ordination of blood specimens from the Western Province Blood Transfusion Service; Tammy Cashmore, Roshan Hendricks, Tracey-Lee Smith and Louise Wilsdorf for technical assistance; Chris Miller for statistical analysis; and to all the study participants.
Sponsorship: This study was supported by the South African AIDS Vaccine Initiative (SAAVI), the Poliomyelitis Research Foundation, the Medical Research Council (MRC) and Syfrets (GT Searle).
Chapter 4

Novel single nucleotide polymorphisms identified using a comprehensive CCR2-denaturing gradient gel electrophoresis assay

Desiree C. Petersen\textsuperscript{1}, Annette Laten\textsuperscript{1}, Michele D. Zeier\textsuperscript{2}, Ashraf Grimwood\textsuperscript{3}, J.H. (Bill) Louw\textsuperscript{4}, Estrelita Janse van Rensburg\textsuperscript{1} and Vanessa M. Hayes\textsuperscript{5}

Departments of Medical Virology\textsuperscript{1}, Internal Medicine\textsuperscript{2}, Genetics\textsuperscript{4} and Urology\textsuperscript{5}, University of Stellenbosch and Chapel Street Community Health Clinic, Woodstock\textsuperscript{3}, South Africa.

\textit{(Submitted)}
**Background:** A single nucleotide polymorphism (SNP) at codon 64 in the CC chemokine receptor 2 gene (CCR2V64I), has been associated with a dominant effect of delaying disease progression from human immunodeficiency virus-1 (HIV-1) infection to acquired immunodeficiency syndrome (AIDS).

**Objectives:** To design a comprehensive mutation detection assay for the entire coding region of the CCR2A and CCR2B gene transcripts and to identify novel mutations and SNPs, within our predominantly African-based population, which could influence an individual's susceptibility to HIV-1 infection and/or progression to AIDS.

**Design:** The study cohort consisted of 102 HIV seropositive patients and 144 HIV seronegative controls from the diverse South African population.

**Methods:** A mutation detection assay, based on denaturing gradient gel electrophoresis (DGGE), was designed for the entire coding region of both the CCR2A and CCR2B transcripts, including all relevant splice site junctions. The assay allows for the complete analysis of 5 individuals per denaturing gel.

**Results:** Application of the CCR2-DGGE assay resulted in the detection of two previously reported CCR2 polymorphisms, namely CCR2V64I and CCR2N260, and 11 novel mutations, including seven SNPs occurring at high allelic frequencies within specific population groups of South Africa.

**Conclusion:** The large number of novel mutations / SNPs identified, using the CCR2-DGGE assay, indicates the importance for comprehensive analysis of all candidate genes in host susceptibility to HIV-1 infection, specifically in the under-studied African-based populations.

**Key words:** CCR2, novel SNPs, novel mutations, DGGE, HIV-1 susceptibility, South Africa
INTRODUCTION

The discovery of specific chemokine receptors acting together with CD4 molecules as cellular co-receptors for human immunodeficiency virus (HIV) entry [Deng et al., 1996; Dragic et al., 1996, Doranz et al., 1996; Feng et al., 1996], has led to major interest in host factors involved in HIV-1 infection. Furthermore, the study of host genetic factors has been advanced by the identification of mutations and single nucleotide polymorphisms (SNPs) in genes encoding chemokines and chemokine receptors, which seem to be associated with susceptibility to HIV-1 infection and / or progression to acquired immune deficiency syndrome (AIDS) [reviewed in O'Brien and Moore, 2000; Hogan and Hammer, 2001].

CC-chemokine receptor 2 (CCR2) has two isoforms, namely CCR2A and CCR2B [Charo et al., 1994], which are both functional seven transmembrane G-coupled proteins. CCR2A is found almost exclusively in the cytoplasm, while the predominant isoform, CCR2B, is expressed at both the cell surface and in the cytoplasm [Wong et al., 1997]. The CCR2 protein binds β-chemokines monocyte chemoattractant protein 1 to 5 (MCP-1 to 5) [reviewed in Kalinkovich et al., 1999] and also acts as an additional co-receptor during cellular infection of a few HIV-1 virus strains [Doranz et al., 1996]. Located at band p21 of chromosome 3 [Daugherty and Springer, 1997], the CCR2 gene comprises 3 exons and 2 introns, spanning approximately 7kb. The coding region for CCR2A is found in exon 2 and part of exon 3, while exon 2 contains the entire coding region of CCR2B. The two isoforms thus differ only in their carboxyl tails, as a result of alternative splicing of a single gene [Wong et al., 1997].
A common SNP has previously been reported in the first transmembrane region of CCR2 and involves a G-A transition at codon 64, resulting in a conservative amino acid change from valine to isoleucine [Smith et al., 1997a]. Studies have shown that the CCR2V64I SNP in both the heterozygous and homozygous state does not offer any resistance to HIV-1 infection, but is associated with delaying progression to AIDS by two to four years [Smith et al., 1997a; Smith et al., 1997b; Michael et al., 1997; Kostrikis et al., 1998]. Population surveys indicate that the CCR2V64I SNP occurs at an allelic frequency of 0.10 to 0.25 within specific ethnic groups [Smith et al., 1997a; Michael et al., 1997] and it’s effect seems to be more apparent in Africans than in Caucasians [Mummidi et al., 1998].

Given the fact that CCR2 is only used by a minority of HIV-1 strains for entry into the host cells [Doranz et al., 1996], the underlying mechanism by which the CCR2V64I SNP influences disease progression still needs to be elucidated. Results from various studies indicate that the CCR2V64I SNP does not affect both CCR2 and CCR5 expression or co-receptor activity [Lee et al., 1998; Mariani et al., 1999]. It is therefore possible that the CCR2V64I SNP either exerts a subtle effect on CCR2 function that cannot be detected, or it is linked to an unidentified mutation in a gene that is known or not yet known to be associated with HIV pathogenesis. It has been shown that the CCR2V64I SNP is linked to a specific CCR5 promoter variant (59653-T) [Kostrikis et al., 1998; Martinson et al., 2000] and this finding requires further investigation.

A second CCR2 SNP occurring at codon 260 (AAC-AAT) and resulting in a silent mutation (CCR2N260) has been reported in the Genbank database.
No association studies with HIV-1 infection or disease progression have been performed. In a previous study, our group observed the T allele to occur more frequently than the commonly reported C allele within the South African population [Petersen et al., 2001]. We cannot exclude the possibility that susceptibility to HIV-1 infection and/or disease progression to AIDS could be the result of a combination of common, but weaker genetic events that collectively determine the "risk profile" of an individual. Comprehensive mutation analysis of the CCR2 gene is thus important for identifying novel mutations that may have a possible protective or causative effect, or novel SNPs for inclusion in association studies of large study cohorts.

This study involved the design of a comprehensive mutation detection assay for the entire coding region of both CCR2 gene transcripts (CCR2A and CCR2B), based on denaturing gradient gel electrophoresis (DGGE). DGGE is believed to be the most powerful of the polymerase chain reaction (PCR), gel-based mutation detection assays currently available. The use of the CCR2-assay led to the identification of novel mutations and SNPs in 102 HIV seropositive patients and 144 HIV seronegative controls from a diverse South African population.

**METHODS**

**Blood samples**

The study cohort consisted of 102 HIV seropositive patients (34 male; 68 female) of who most are presently residents in the Western Cape of South Africa (Tygerberg Hospital and Woodstock Chapel Street Community Health Clinic). The disease status was unknown for the majority of these patients. Also forming part of the study...
cohort, were 144 HIV seronegative healthy controls (56 male; 88 female) who are all blood donors for the Western Province Blood Transfusion Service, South Africa. The individuals participating in this study included Africans, predominantly Xhosa (69 HIV+ and 62 HIV-), Coloureds (26 HIV+ and 72 HIV-), and to a lesser degree Caucasians (seven HIV+ and two HIV-) and Asians (eight HIV-). The various population groups have been previously defined in Petersen et al., 2001. Informed consent for the study was obtained from all participants. The Ethics Review Committee of the University of Stellenbosch approved the study protocol.

Primer Design

Using the melt 87 computer program [Lerman and Silverstein, 1987] and conditions for selecting appropriate PCR fragments [Wu et al., 1998], DGGE PCR primers (Table 1) were designed for the entire coding region, including the intron/exon boundaries of both transcripts of the CCR2 gene. The coding region of CCR2B (codons 1 to 361) and most of the coding region of CCR2A (codons 1 to 313), contained in exon 2, was divided into six (A - F) overlapping amplicons. An additional amplicon “G” was designed to include the remaining coding region (exon 3) of CCR2A (codons 314 to 375). The addition of a GC-rich-fragment to the 5' end of one of the primers in each primer set prevents complete strand dissociation during amplification. Additional stretches of GC or AT nucleotides were added to either the 5' end of the non-clamped primer (fragments B, C and E) or between the GC-clamp and the primer (fragment D), to ensure a single melting domain for optimal detection of all mutations (Table 1).
Table 1. CCR2 primer sets and experimental conditions for PCR amplification and DGGE.

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Amplimers, 5'-3'</th>
<th>Size (bp)</th>
<th>Melting</th>
<th>Annealing</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[40GC]TGCTTTATGTGGTGCCAGACT TGA ACACCAGCGAGTAGAGC</td>
<td>335</td>
<td>72</td>
<td>58</td>
</tr>
<tr>
<td>B</td>
<td>[6GC]TGATTA TGATTACGGGTGCTCC [40GC]CGATTTGTCAGGAGGATGAT</td>
<td>384</td>
<td>72</td>
<td>58</td>
</tr>
<tr>
<td>C</td>
<td>[40GC]GCTGTATCACAATCGGTTATT [8GC]GCCACAGACATAAAACAGAAT</td>
<td>268</td>
<td>73</td>
<td>55</td>
</tr>
<tr>
<td>D</td>
<td>[40GC][10AT]TGGCTGTGTTTGCTTCTGT CGAGTAGCAGATGACCATA</td>
<td>220</td>
<td>70</td>
<td>55</td>
</tr>
<tr>
<td>E</td>
<td>[5GC]CCACACA ATA ATGAGGAACA [40GC]TGTTGCTTTTCACAGTTACTC</td>
<td>284</td>
<td>73</td>
<td>55</td>
</tr>
<tr>
<td>F</td>
<td>ACCTTCAGGAATTCTTTCG [40GC]ACAATCAAACGTCTCCTCGT</td>
<td>346</td>
<td>74</td>
<td>55</td>
</tr>
<tr>
<td>G</td>
<td>TGTCTGGATCTGAGCTGTTT</td>
<td>[40GC]TCCAAGCAGAGATCTGTCAT</td>
<td>333</td>
<td>73</td>
</tr>
</tbody>
</table>

bp, base pair
GC-clamps used were as follows: [40GC], CGCCCGCCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGC

AT-stretch used was as follows: [10AT], TATAATATTA

DNA amplification

Genomic DNA was isolated from peripheral blood leukocytes using conventional methods and amplified using DGGE primer sets, specific for each amplicon (A – G) (Table 1). With a total volume of 50 μl, each PCR reaction mixture contained 100 ng of genomic DNA, 0.1 mM of each deoxyribonucleoside triphosphate (dNTP), 20 pmol
of each primer (except for fragment B which required only 10 pmol of each primer), 2.5 mM of a 10 x Mg$^{2+}$ reaction buffer and 0.5 units of DNA Taq polymerase (Boehringer Mannheim, Mannheim, Germany). Amplification was performed using a Perkin Elmer 9600 thermocycler (PE Applied Biosystems) and the PCR cycling conditions were as follows; an initial denaturation at 96°C for 3 minutes, followed by 32 cycles of denaturation at 96°C for 45 seconds, annealing for 1 minute (annealing temperatures shown in Table 1), and elongation at 72°C for 1 minute 20 seconds. The last cycle was followed by an additional extension step of 72°C for 7 minutes. For optimal DGGE analysis, amplification was followed by a heteroduplexing step, which involves denaturation at 96°C for 10 minutes, followed by renaturation for 45 minutes at the annealing temperature of the amplification. The amplified products were checked using electrophoresis, where 5μl (10%) of each sample was resolved on 2% agarose gel.

Denaturing gradient gel electrophoresis (DGGE)

Optimised DGGE conditions were achieved by considering the conditions previously described for the improvements of broad-range DGGE analysis [Hayes et al., 1999a]. DGGE was performed using the Ingeny phorU-2 system (www.ingeny.com). The seven amplicons were electrophoresed in six lanes (fragments C and D were pooled) of a 9% polyacrylamide gel containing a 30% to 70% urea and formamide (UF) denaturing gradient (100% UF = 7mol/L urea per 40% deionised formamide), at 59.5°C for 110 volts overnight. The gels were stained with ethidium bromide and photographed under an UV transilluminator. Thus the CCR2-DGGE allows for the complete analysis of 5 individuals per denaturing gel.
DNA sequencing and mutation confirmation

Automated sequencing of amplified samples showing aberrant DGGE banding patterns was performed using a non-GC-clamped primer and the dye terminator sequencing kit of Applied Biosystems (www.appliedbiosystems.com). The commonly occurring SNPs were verified by mixing samples showing similar DGGE banding patterns, followed by a heteroduplexing step before electrophoresis on a denaturing gel [Guldberg and Gutler, 1993]. Samples showing additional heteroduplex bands were subjected to sequencing for the exact determination of the sequence variants.

Statistical analysis

Manual allele counting was used for calculating allele frequencies. Tests for significance of heterogeneity in the frequencies among HIV seropositive patients and seronegative controls for both the mutations and SNPs were performed by means of Fisher's exact test for 2X2 contingency tables. The Hardy – Weinberg principle was applied to measure the maintenance of the allelic frequency for the CCR2 SNPs in the closely matched HIV seropositive and HIV seronegative African population. The statistical calculations are shown in appendix 3.

RESULTS

The CCR2 primer sets and experimental conditions for PCR amplification and DGGE analysis are shown in Table 1. Using our CCR2-DGGE assay, we identified 11 novel mutations and two previously reported mutations as shown in Figure 1 and listed in Table 2, according to the intron / codon in which they occur.
Table 2. CCR2 mutations detected in 102 HIV seropositive patients and 144 HIV seronegative controls, ordered according to the intron/codon in which they occurred.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Base change</th>
<th>DGGE Fragment</th>
<th>Africa</th>
<th>Coloureds</th>
<th>Caucasians</th>
<th>Asians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HIV+ (n = 138)</td>
<td>HIV- (n = 124)</td>
<td>HIV+ (n = 52)</td>
<td>HIV- (n = 144)</td>
</tr>
<tr>
<td>Int1 -57a/g</td>
<td>A - G</td>
<td>A</td>
<td>14 (0.101)</td>
<td>13 (0.105)</td>
<td>3 (0.058)</td>
<td>10 (0.069)</td>
</tr>
<tr>
<td>Int1 -43g/a</td>
<td>G - A</td>
<td>A</td>
<td>3 (0.022)</td>
<td>5 (0.040)</td>
<td>3 (0.058)</td>
<td>3 (0.021)</td>
</tr>
<tr>
<td>V52 *a,b</td>
<td>GTG - GTT</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>1 (0.019)</td>
<td>1 (0.007)</td>
</tr>
<tr>
<td>V63 *a,b</td>
<td>GTC - GTT</td>
<td>B</td>
<td>1 (0.007)</td>
<td>1 (0.008)</td>
<td>5 (0.096)</td>
<td>5 (0.035)</td>
</tr>
<tr>
<td>V64I *a,b</td>
<td>GTC - ATC</td>
<td>B</td>
<td>18 (0.130)</td>
<td>21 (0.169)</td>
<td>1 (0.019)</td>
<td>25 (0.174)</td>
</tr>
<tr>
<td>S223 *a,b</td>
<td>TCG - TCA</td>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (0.007)</td>
</tr>
<tr>
<td>R233Q *a,b</td>
<td>CGA - CAA</td>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2 (0.007)</td>
</tr>
<tr>
<td>N260 *a,b</td>
<td>AAC - AAT</td>
<td>E</td>
<td>75 (0.543)</td>
<td>79 (0.637)</td>
<td>30 (0.577)</td>
<td>97 (0.674)</td>
</tr>
<tr>
<td>L283 *a,b</td>
<td>CTG - CTT</td>
<td>F</td>
<td>6 (0.043)</td>
<td>3 (0.024)</td>
<td>1 (0.019)</td>
<td>4 (0.028)</td>
</tr>
<tr>
<td>T287M *a,b</td>
<td>ACG - ATG</td>
<td>F</td>
<td>0</td>
<td>4 (0.032)</td>
<td>1 (0.019)</td>
<td>0</td>
</tr>
<tr>
<td>P339 *a</td>
<td>CCA - CCG</td>
<td>G</td>
<td>21 (0.152)</td>
<td>18 (0.145)</td>
<td>2 (0.038)</td>
<td>5 (0.035)</td>
</tr>
<tr>
<td>T348 *a</td>
<td>ACG - ACA</td>
<td>F</td>
<td>0</td>
<td>0</td>
<td>1 (0.019)</td>
<td>6 (0.042)</td>
</tr>
<tr>
<td>G355E *a</td>
<td>GGA - GAA</td>
<td>G</td>
<td>2 (0.014)</td>
<td>1 (0.008)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

n, number of alleles; HIV+, seropositive; HIV-, seronegative; Int, intron

* Novel mutation identified in this study

a Mutation identified in CCR2A

b Mutation identified in CCR2B
Figure 1. Aberrant DGGE banding patterns for all thirteen CCR2 gene mutations identified and listed in Table 2, according to the fragment (Frag) in which they were found. No mutations were identified in fragments C and D. Lane 1 of all the fragments represents the DGGE banding pattern of a normal control. Mutants are depicted as follows; Frag A: lanes 2, heterozygous for the CCR2 Int1 -57a/g SNP; and lane 3, heterozygous for CCR2 Int1 -43g/a SNP; Frag B: lane 2, heterozygous for the CCR2V64I SNP; lane 3, heterozygous for the CCR2V63 SNP; and lane 4, heterozygous for the CCR2V52 mutation; Frag E: lane 2, heterozygous for the CCR2R233Q mutation; lane 3, heterozygous for the CCR2S223 mutation; lane 4, heterozygous for the CCR2N260 SNP; and lane 5, homozygous for the CCR2N260 SNP; Frag F: lane 2, heterozygous for the CCR2L283 SNP; lane 3, heterozygous for the CCR2T287M SNP; and lane 4, heterozygous for the CCR2T348 SNP; Frag G: lane 2, heterozygous for the CCR2P339 SNP; and lane 3, heterozygous for the CCR2G355E mutation.
Seven of the 11 novel mutations occur at an allelic frequency of greater than or equal to 0.01 and thus appear to be SNPs within our unique South African population. Of these seven SNPs, two are found within intron 1 and involve an a to g change at position -57 base pairs (bp) and a g to a change at position -43bp downstream from the acceptor splice site, respectively. Four of the SNPs are silent mutations and are observed at codons 63, 283, 339 (CCR2A) and 348 (CCR2B), while another SNP at codon 287 (ACG - ATG) involves a non-conservative amino acid change (replacement of an amino acid by another with different chemical properties) from Threonine to Methionine, which could result in changing the functional structure of the CCR2 protein. Except for the SNP at codon 63, which was observed in all the different population groups and the SNP at codon 348 (CCR2B), which was present in the Coloured and Caucasian population groups, the rest of the SNPs were found exclusively in Africans and / or Coloureds.

The remaining four novel mutations include two silent mutations at codons 52 and 223, which were found in two Coloureds (one HIV+ and one HIV-) and one HIV seronegative Coloured, respectively, and two non-conservative mutations at codon 233 (CGA - CAA) and codon 355 of CCR2A (GGA – GAA). These non-conservative amino acid changes, could ultimately affect the structure of CCR2. The mutation at codon 233, which was observed homozygously in one HIV seronegative Coloured, involves an amino acid change from Arginine to Glutamine, while the codon 355 (CCR2A) mutation, found in three Africans (two HIV+ and one HIV-), results in an amino acid change from Glycine to Glutamic Acid.
The commonly occurring SNP at codon 64 (GTC - ATC), results in a non-conservative amino acid change from Valine to Isoleucine [Smith et al., 1997a]. The CCR2V64I SNP was observed homozygously in four HIV seronegative controls (one African, two Coloureds and one Asian) and heterozygously in both the HIV seropositive patients and HIV seronegative controls of all the different South African population groups. For the previously reported silent SNP at codon 260 (AAC – AAT), it was found that the T allele occurs at a higher allelic frequency than the C allele in the entire South African study cohort.

DISCUSSION

This study involved the design and use of a comprehensive mutation detection assay for the entire coding region of both gene transcripts of CCR2 (CCR2A and CCR2B), for the identification of novel and previously reported mutations and SNPs in both HIV seropositive and HIV seronegative individuals in a predominantly African-based population from South Africa. The assay, based on DGGE, allows for the complete analysis of 5 patients per denaturing gel. Previous studies have been restricted to the analysis of the CCR2V64I SNP in different population groups and thus comprehensive analysis of CCR2 is important and ideal in our diverse South African study cohort.

Novel mutations and SNPs were identified in all the different population groups represented in this study. The novel mutations at codons 233 and 355 (CCR2A), and the novel SNP at codon 287, all result in non-conservative amino acid changes, which may change the structure of the CCR2 protein and thus affect it’s functioning. The novel silent mutations (codons 52, 223) and SNPs (63, 283, 339 (CCR2A) and
348 (CCR2B)), as well as, the two novel intronic SNPs (Int1 -57a/g and Int1 -43g/a) could all possibly influence gene expression and/or RNA splicing by altering regulatory elements [D'Souza et al., 1999, Lorson et al., 1999]. The significance of all these mutations on the functioning of CCR2 requires further investigation.

The SNP at codon 63 was detected in all the population groups, while the SNP at codon 348 (CCR2B) appears to have a Caucasian origin as it was only observed in Coloureds and Caucasians. Three of the novel mutations (codons 52, 223 and 233) were found exclusively in the Coloured population. Due to the relatively recent admixture within the people of Coloured descent, it is not certain whether these rare mutations have an African or Caucasian-based origin. Although the number of Caucasians included in this study are small, the novel mutation at codon 355 (CCR2A) and the novel SNPs at Int1 -57a/g, Int1 -43g/a and codon 287, seem to be African-related. Due to insufficient information regarding the clinical status of the HIV seropositive patients, no significant associations with the novel mutations or SNPs and HIV-1 susceptibility and/or rates of disease progression to AIDS could be made.

The commonly reported CCR2V64I SNP, which has an allelic frequency ranging from 0.10 to 0.25 within specific populations [Smith et al., 1997a; Michael et al., 1997], was observed in all the South African population groups with allelic frequencies of 0.149 in the Africans; 0.133 in the Coloureds; and although numbers are small, 0.056 in the Caucasians and 0.125 in the Asians. It is known that the presence of the CCR2V64I SNP heterozygously and homozygously is associated with delaying disease progression to AIDS, provided that the patient's date of HIV-1 infection is more or less known [Smith et al., 1997a; Smith et al., 1997b; Michael et al., 1997;
Kostrikis et al., 1998. Most of the HIV seropositive patients forming part of our study cohort do not have known dates of HIV-1 infection and thus no associations could be made. However, a significant decrease in the frequency of the mutant allele was noted in the HIV seropositives, when compared to the HIV seronegative control group for the Coloured population ($P = 0.0034$).

The observation that the T allele of the $CCR2N260$ SNP occurs at higher frequencies within all the population groups in this study, confirms our previous data [Petersen et al., 2001]. This is contrary to recent data, which suggests that the C allele occurs more frequently in Caucasians, Africans and Hispanics residing in America [Clark et al., 2001].

Although expression of $CCR2A$ is mostly restricted to the cytoplasm [Wong et al., 1997], we included the $CCR2A$ transcript in our analysis. The reason for this was the possibility of identifying novel African-related SNPs within $CCR2A$, which although does not singly lead to susceptibility, may in combination with other weaker genetic events lead collectively to determining disease status. The African-associated SNP identified at codon 339 of the $CCR2A$ transcript is currently under investigation by our group in a large cohort of patients of known disease progression to determine the significance of this SNP in combination with other identified African-associated SNPs. The closely matched HIV seropositive and HIV seronegative African population are in Hardy – Weinberg equilibrium for all the African-associated $CCR2$ SNPs.
The large number of novel mutations, especially novel SNPs, identified in our South African study cohort, with its diverse population groups, is an indication of the efficiency of the described CCR2-DGGE assay and also emphasises the importance of screening for mutations, other than the CCR2V64I, in different population groups. Further studies are required to determine the underlying mechanisms of the CCR2 novel mutations and SNPs identified, so that its possible effects on influencing host susceptibility to HIV-1 infection and/or developing AIDS can be understood more clearly.

ACKNOWLEDGEMENTS

Thanks to Dr John Gasson of the Western Province Blood Transfusion Service, for co-ordination of the HIV seronegative control blood samples and to all the study participants.

Sponsorship: This study was supported by the South African AIDS Vaccine Initiative (SAAVI), the Poliomyelitis Research Foundation, the Medical Research Council (MRC), the Harry Crossley Foundation, Syfrets (GT Searle) and Unistel Medical Laboratories.
Chapter 5

Analysis of the SDF1-3'A single nucleotide polymorphism (SNP) within a South African population

Desiree C. Petersen\textsuperscript{1}, Estrelita Janse van Rensburg\textsuperscript{1}

and Vanessa M. Hayes\textsuperscript{2}

Departments of Medical Virology\textsuperscript{1} and Urology\textsuperscript{2}, University of Stellenbosch,
Tygerberg Medical School, South Africa.
Background: The CXC chemokine, stromal derived factor 1 (SDF1), is the natural ligand for the human immunodeficiency virus-1 (HIV-1) co-receptor, CXC chemokine receptor 4 (CXCR4). A single nucleotide polymorphism (SNP), SDF1-3' A, has been previously identified in the 3' untranslated region (3' UTR) of the SDF1β transcript and is associated with influencing disease progression when found homozygously.

Objectives: To determine the allelic frequencies of the SDF1-3' A SNP within the different South African population groups.

Design: The South African study cohort consisted of 104 HIV seropositive patients and 196 HIV seronegative healthy controls.

Methods: A mutation detection assay, based on denaturing gradient gel electrophoresis (DGGE), was designed for part of the SDF1β-3' UTR.

Results: We identified the commonly reported SDF1-3' A SNP as well as 2 novel mutations. The SDF1-3' A SNP was observed in both the HIV seropositive patients and HIV seronegative controls of all the South African population groups, but with a higher allelic frequency in the Caucasians and Coloureds.

Conclusion. Two of the HIV-seropositive patients of whom the disease progression is unknown were homozygous for the SDF1-3' A SNP, while 10 patients with widely variant disease progression were heterozygous for this polymorphism. No association with HIV-1 susceptibility and/or disease progression to AIDS could thus be made. However, the high allelic frequency of the SDF1-3' A SNP in both the HIV seropositive patients and HIV seronegative controls warrants further investigation of larger and informative study cohorts to analyse the effect of the SDF1-3' A SNP on the surface expression and functioning of the SDF1 protein.

Keywords: SDF1, SNP, novel mutations, HIV-1 susceptibility, South Africa
INTRODUCTION

The observation of specific chemokines acting as inhibitors of human immunodeficiency virus-1 (HIV-1) infection and also possibly influencing viral replication [Cocchi et al., 1995], was further underscored by the discovery that these chemokines are the natural ligands for chemokine receptors, which serve as necessary co-factors for HIV-1 entry. It is thus possible that chemokines suppress HIV-1 infection, by either direct competition with the virus for binding to the chemokine receptors, or down-regulation of the chemokine receptors [Bleul et al., 1996a; Oberlin et al., 1996; Samson et al., 1996; Combadiere et al., 1996; Raport et al., 1996].

The CXC - chemokine, stromal cell-derived factor 1 (SDF1 or CXCL12), is the natural ligand for the HIV-1 entry co-factor CXC chemokine receptor 4 (CXCR4) and inhibits infection by T cell line tropic (T-tropic) or SI syncytium-inducing (SI) virus strains [Bleul et al., 1996a; Oberlin et al., 1996] by interfering with the use of CXCR4 by HIV-1 [Amara et al., 1997; Signoret et al., 1997]. It has been found that the SDF1 gene, located at band q11 of chromosome 10, encodes for two isoforms, namely SDF1α (89 amino acids) and SDF1β (93 amino acids), which is the result of alternative splicing of a single gene [Tashiro et al., 1993; Nagasawa et al., 1994; Shirozu et al., 1995]. The first 21 amino acid residues of the SDF1α and SDF1β proteins form an amino acid-cleaved signal peptide [Bleul et al., 1996b]. Three and four exons contain the coding regions for SDF1α and SDF1β, respectively. The SDF1β gene transcript thus has an extra exon, which encodes for four additional amino acids [Shirozu et al., 1995].
A study by Winkler et al. in 1998, led to the identification of a commonly occurring single nucleotide polymorphism (SNP) in the 3’ untranslated region (3’ UTR) of the SDF1β gene transcript, which involves a G to A transition at nucleotide position +801 (counting from the first nucleotide of the ATG start codon). Allelic frequencies for the SNP range from 0.06 to 0.26 within various population groups, occurring more commonly in Caucasian populations [Mummidi et al., 1998; Winkler et al., 1998; Williamson et al., 2000]. The SNP has been shown to be associated with delaying the onset of acquired immune deficiency syndrome (AIDS) when found in the homozygous state [Hendel et al., 1998; Martin et al., 1998; Winkler et al., 1998] and the simplest hypothesis for this recessive protective effect was that the SNP results in the up-regulation of SDF1 biosynthesis, making it more available to compete with HIV-1 for binding to CXCR4 and thus blocking T-tropic or SI variants from emerging [Winkler et al., 1998]. Ayra et al., 1999 has tested this hypothesis and the results indicate that the SNP does not affect the regulation of SDF1 expression.

Other studies however suggest that the SDF1β-3’ UTR-801G-A (abbreviated SDF1-3’ A) SNP, when found homozygously in HIV-infected individuals, is not associated with a protective effect, but rather associated with accelerated progression to death [Mummidi et al., 1998]; prolonged [van Rij et al., 1998] or decreased [Brambilla et al., 2000] survival after AIDS is diagnosed; low CD4 cell counts [Balotta et al., 1999]; and no effect on disease progression [Meyer et al., 1999]. The study cohorts used in the studies mentioned above consisted of predominantly Caucasians. A recent study has also shown an association between the SDF1-3’ A SNP in the heterozygous state and increased vertical HIV-1 transmission from mother to child in an African study cohort [John et al., 2000]. All the findings mentioned above emphasises the
need for further studies, so that a clear correlation can be made between the $SDF1$-3' A SNP and HIV-1 pathogenesis.

This study included the design and application of a comprehensive mutation detection assay for part of the 3' UTR of the $SDF1\beta$ gene transcript, using denaturing gradient gel electrophoresis (DGGE) and allowed for the screening of the $SDF1$-3' A SNP in a South African study cohort, consisting of 103 HIV seropositive patients and 194 HIV seronegative controls. Our results are presented in this study.

**METHODS**

**Study cohort**

104 HIV seropositive individuals (35 male; 69 female) formed part of the study cohort and they are all patients of either Tygerberg Hospital or Woodstock Chapel Street Community Health Clinic, both in the Western Cape of South Africa. The disease progression for most of these patients remains unknown. The study cohort also consisted of 192 HIV seronegative healthy individuals (78 male; 114 female) and these control samples were obtained from the Western Province Blood Transfusion Service, South Africa. The population groups represented in this study include Africans, predominantly Xhosa (70 HIV+ and 63 HIV-), Coloureds (26 HIV+ and 73 HIV-), Caucasians (seven HIV+ and 48 HIV-) and Asians (eight HIV-). There was also one HIV seropositive female of whom the population group is unknown. The different population groups have been previously defined in Petersen *et al.*, 2001. Informed consent was obtained from all the study participants. The Ethics Review Committee of the University of Stellenbosch approved the study protocol.
Primer Design

A DGGE PCR primer set was designed for part of the 3’ UTR of the SDF1β gene transcript, using the melt 87 computer program [Lerman and Silverstein, 1987] and considering conditions previously described for the improvements of selecting appropriate DGGE PCR primers [Wu et al., 1998]. Primers used were as follows; SDF1-3’AF: GTGAAGGCTTCTCTGTGG and SDF1-3’AR: [40GC]GTGGACACA CATGATGATG. To prevent complete strand dissociation during amplification and ensure a single melting domain for optimal detection of all mutations, a GC-rich-fragment was added to the 5’ end of the reverse primer (40GC-clamp as previously published in Petersen et al., 2001).

DNA amplification

Using conventional methods, genomic DNA was extracted and amplified using a single DGGE primer set. Each PCR reaction mixture had a final volume of 50 µl and contained 100 ng of genomic DNA, 0.1 mM of each deoxyribonucleoside triphosphate (dNTP), 10 pmol of each primer, 2.5 mM of a 10 x Mg2+ reaction buffer and 0.5 units of DNA Taq polymerase (Boehringer Mannheim). Amplification was performed using a Perkin Elmer 9600 thermocycler (PE Applied Biosystems) and the following PCR cycling conditions; an initial denaturation at 96°C for 3 minutes, followed by 32 cycles of denaturation at 96°C for 45 seconds, annealing at 56°C for 1 minute and elongation at 72°C for 1 minute 20 seconds. Following the last cycle was an additional extension step of 72°C for 7 minutes. For heteroduplex formation, the PCR products were subjected to denaturation at 96°C for 10 minutes, followed by renaturation for 45 minutes at 56°C. Electrophoresis was used to check the amplified products, where 5µl (10%) of each sample was resolved on 2% agarose gel.
Denaturing gradient gel electrophoresis (DGGE)

Optimal DGGE analysis was achieved using previously described conditions for broad-range mutation detection by DGGE [Hayes et al., 1999a]. Amplified PCR products were electrophoresed in a 9% polyacrylamide gel with a denaturing gradient of 45% to 85% urea and formamide (UF) (100% UF = 7mol/L urea per 40% deionised formamide), at 60°C for 110 volts overnight, using the Ingeny phorU-2 system (www.ingeny.com). The gels were stained with ethidium bromide and photographed under an UV transilluminator.

DNA sequencing and mutation verification

Automated DNA sequencing was performed for amplified samples showing aberrant DGGE banding patterns, using a non-GC-clamped primer and the dye terminator sequencing kit of Applied Biosystems (www.appliedbiosystems.com). To verify the commonly occurring SDF1-3’ A SNP, samples with similar DGGE banding patterns were mixed and subjected to heteroduplex formation before electrophoresis on a denaturing gel [Guldberg and Guttler, 1993]. Samples showing additional heteroduplex bands were subjected to sequencing.

Statistical analysis

The allelic frequencies of the SDF1-3’ A SNP were determined by manual allele counting. Testing for significance of heterogeneity in the frequencies among HIV seropositive patients and HIV seronegative controls for the SDF1-3’ A SNP was based on Fisher’s exact test for 2x2 contingency tables. The Hardy – Weinberg principle was applied to measure the maintenance of the allelic frequency for the SDF1-3’ A SNP in the closely matched HIV seropositive and HIV seronegative African population. The statistical calculations are shown in appendix 3.
RESULTS AND DISCUSSION

The previously reported SDF1-3' A SNP and two novel mutations were detected using the SDF1β-3' UTR-DGGE assay and are shown in Table 1 and in Figure 1, according to the nucleotide position at which they occur.

Table 1. Allelic frequencies of SDF1β mutations detected in 104 HIV seropositive patients and 192 HIV seronegative controls, ordered according to the nucleotide position at which they occurred.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Africans</td>
</tr>
<tr>
<td></td>
<td>HIV+</td>
</tr>
<tr>
<td>nt +801 G-A</td>
<td>7 (0.050)</td>
</tr>
<tr>
<td>nt +875 A-T *</td>
<td>0</td>
</tr>
<tr>
<td>Δ8 (nt +916) *</td>
<td>0</td>
</tr>
</tbody>
</table>

* Novel mutation identified in this study
n, number of alleles; nt, nucleotide position; Δ, deletion

Figure 1. DGGE banding pattern for the SDF1β gene mutations detected in this study. Lane 2, normal control; Lane 1 and 3, Individuals heterozygous and homozygous for the SDF1β-3' UTR G-A (nt +801) SNP, respectively; Lane 4, Individual heterozygous for the SDF1β-3' UTR A - T (nt +875) mutation and Lane 5, Individual heterozygous for the SDF1β-3' UTR 8bp deletion (nt +916).
The segment of the 3’ UTR of SDF1β screened in this study is highly conserved in sequence and the extent of conservation suggests that it may serve as a target for cis-acting factors, which could influence transcription, biosynthesis, transport, stability and splicing [Winkler et al., 1998]. Thus the previously described SDF1-3’ A SNP as well as the 2 novel mutations, in a potential regulatory region, could have an effect on the expression or functioning of the SDF1 protein. The assay used in this study only includes part of the SDF1β-3’ UTR, so therefore we cannot exclude the possibility that novel significant mutations could occur within the remaining 5’ and 3’ UTRs, the coding region and the intronic sequences of both SDF1α and SDF1β transcripts.

The first novel mutation identified in this study involves an A to T change at nucleotide position +875 (counting from the first nucleotide of the ATG start codon) and was found in a HIV seropositive female of whom the population group is unknown. The second novel mutation, detected in a HIV seronegative Caucasian female, involved an eight base pair (8bp) deletion starting at nucleotide position +916 (Figure 2).

![Fig 2](image_url) **Figure 2.** A) Sequence for the SDF-1 wildtype. The black arrows indicate the 8bp that form the deletion mutation, starting at nucleotide position +961. B) Sequence for the SDF1 Δ8bp mutation in the homozygous state. The black arrow indicates the start of the deletion mutation.
The SDF1-3’ A SNP was observed in both the HIV seropositive patients and HIV seronegative controls from all the different population groups. Two of the HIV seropositive patients, an African female and a Caucasian male, were homozygous for the SNP, but the disease status is unknown for both these individuals. The SDF1-3’ A SNP was found heterozygously in 10 HIV seropositive patients of which two, three, one and four had slow, normal, fast and unknown disease progression, respectively. Nine HIV seronegative controls were homozygous for the SDF1-3’ A SNP, while the SDF1-3’ A SNP was found heterozygously in 39 HIV seronegative controls.

The SDF1-3’ A SNP was found to occur at a much higher allelic frequency in the Caucasians (0.273), Coloureds (0.157) and Asians (0.125) compared to the Africans (0.030). These findings are similar to the allelic frequencies previously reported [Winkler et al., 1998; Mummidi et al., 1998; Williamson et al., 2000]. A higher allelic frequency for the SDF1-3’ A SNP was however observed in the African HIV seropositive patients (0.050) when compared to the closely matched African HIV seronegative controls (0.008) and the significance of this finding will have to be investigated further. Both the HIV seropositive and HIV seronegative African population are in Hardy – Weinberg equilibrium for the SDF1-3’ A SNP. The high allelic frequency of the SDF1-3’ A SNP in the Coloured population group may be the result of admixture with persons of Caucasian descent [Martinson et al., 1997].

The two HIV seropositive individuals who have the SDF1-3’ A SNP in the homozygous state both lack clinical information, while the 10 HIV seropositive individuals who are heterozygous for the SDF1-3’ A SNP display widely variant clinical outcomes (slow, normal and fast disease progression). Thus no significant
associations with disease progression could be made. However, the identification of the SDF1-3' A SNP at significant allelic frequencies in both the HIV seropositive patients and HIV seronegative controls of different population groups, raises the importance for determining the underlying mechanism of the SDF1-3' A SNP and how it influences susceptibility to HIV-1 infection and/or disease progression to AIDS. Due to the inconsistent results observed between various studies, it is necessary that larger cohorts with detailed clinical information are screened to not only determine the effect of the SDF1-3' A SNP, but to more fully understand the functioning of SDF1, which could contribute to HIV therapy and vaccine development.

ACKNOWLEDGEMENTS

Thanks to Dr Michele D. Zeier, Dr Ashraf Grimwood and Dr Eftyhia Vardas for the provision of the HIV seropositive blood samples; Dr John Gasson for the co-ordination of the HIV-seronegative blood samples; Annette Laten for performing DNA sequencing; and to all the study participants.

Sponsorship: This study was supported by the South African AIDS Vaccine Initiative (SAAVI), the Poliomyelitis Research Foundation, the Medical Research Council (MRC), the Harry Crossley Foundation, Syfrets (GT Searle) and Unistel Medical Laboratories.
Chapter 6

Discussion
DISCUSSION

Host susceptibility to HIV-1 pathogenesis varies widely amongst individuals [Liu et al., 1997] and is determined by certain parameters, which include viral characteristics and host immunological and genetic factors. The existence of individuals remaining uninfected, despite repeated exposure to HIV-1 and the observation of HIV-1 infected long-term non-progressors (see chapter 2.2.) suggested that genetic factors play an important role in determining an individual's susceptibility to HIV-1.

Linkage analysis and association studies are two experimental methods used to identify and investigate genetic variants in human disease. Linkage analysis involves the study of family pedigrees to compare inheritance patterns of genetic variants, which could serve as markers for identifying the location of a specific disease gene. It is often quite difficult to collect large families to effectively apply linkage analysis. The application of linkage analysis to complex traits (not determined by a single gene locus) could also be problematic as a pedigree model (diagram showing ancestral relationships and transmission of genetic traits in a family) that explains the pattern of inheritance will be hard to find. Association studies involve the use of a case-control or cohort strategy of unrelated individuals to indicate correlations of genetic variants with a disease phenotype. It is however fundamental that the study cohort consists of matched and well-defined ethnic groups to observe significant associations that are not bias due to continuous population admixture [Lander and Schork, 1994; Taylor et al., 2001].

Several association studies based on the comparison of unrelated HIV-1 infected and uninfected individuals from specific population groups have resulted in the
identification of genetic variants, which are associated with influencing host susceptibility to HIV-1 infection and/or disease progression to AIDS (see chapter 2.2.). To date, the analysis of these variants in genes such as CCR5, CCR2 and SDF1, have been largely restricted to Caucasian populations. The South African population consists of a number of diverse well-defined African ethnic groups (described in chapter 3), which is ideal for identifying previously reported and novel CCR5, CCR2 and SDF1 mutations and determining whether certain mutations are relatively specific and rarely or commonly observed within different population groups.

Comprehensive mutation detection assays, based on DGGE (see chapter 2.3), were thus designed for the entire coding region of the CCR5 gene (see chapter 3), the entire coding region for both transcripts of the CCR2 gene (CCR2A and CCR2B) (see chapter 4), and part of the 3’ UTR region of the SDF1β gene transcript (see chapter 5). The CCR5, CCR2 and SDF1 DGGE assays, allowed for the complete analysis of 10, 5 and 30 individuals per denaturing gel, respectively. DGGE, which is believed to be the most powerful of the gel-based pre-screening mutation detection methods currently available, is highly sensitive and allowed for the detection of both previously reported and novel genetic variants of the CCR5 (see chapter 3), CCR2 (see chapter 4) and SDF1 (see chapter 5) genes within a diverse South African population.

Although an association between a single gene and host susceptibility to HIV-1 infection and disease progression to AIDS has been shown with the CCR5Δ32 mutation, which is largely restricted to Caucasian populations [Liu et al., 1996;
Samson et al., 1996b], it is likely that host susceptibility to HIV-1 infection and/or disease progression to AIDS in other population groups may not be the result of a single genetic variant, but due to a large number of polymorphic sites (SNPs) that collectively determine an individual's "risk profile". The identification of these SNPs (single base mutations occurring at an allelic frequency of 0.01 or greater), which are relatively stable due to a low rate of recurrent mutation, are thus important as they may contribute to elucidating the role of host genetic factors in HIV-1 pathogenesis within various population groups [Lander and Schork, 1994; Gray et al., 2000; Sachidanandam et al., 2001].

Association studies, with a large sample size, where cases of disease are defined and compared to matched controls from the same population, are the most ideal for detecting the possible effects of genetic variants. Thus an important factor to consider when performing association studies is population admixture (Coloured population), particularly when there are great differences in both the ethnic admixture of cases of disease and controls and allelic frequencies based upon ethnicity. Determining the role of host genetic factors in susceptibility to HIV-1 infection and/or disease progression to AIDS thus involves the investigation of whether a particular allele occurs at a significantly higher frequency among HIV seropositive individuals with known disease progression when compared to preferably "high risk" uninfected population-matched individuals. Taking into consideration the difficulty of selecting a population-matched control group, an "internal control", consisting of HIV-1 infected individuals and their families, should also be considered for determining specific allelic distributions and possible associations within a population [Lander and Schork, 1994; Taylor et al., 2001].
The South African sample cohort, analysed for the studies presented in chapters 3, 4 and 5, consisted of approximately 100 HIV seropositive patients (see appendix I), 144 to 196 seronegative healthy controls and nine seronegative “high-risk” commercial sex workers, all of predominantly Coloured (mixed ancestry) and African (Xhosa) descent. The majority of HIV seropositive individuals of the sample cohort had incomplete clinical information as only 43 HIV-infected individuals had informative clinical data regarding sexuality, CD4 count, disease progression and clinical staging (see appendix I – Table 1). The clinicians involved in the study, considered CD4 counts and clinical staging for the classification of these 43 HIV seropositive patients into specific groups based on their rate of disease progression to AIDS [World Health Organisation (www.who.int)] (see chapter 1.1.2). The HIV seronegative controls of the sample cohort are all blood donors and it is not known whether any of these individuals have been previously exposed to HIV-1. Due to confidentiality, information regarding the health status and age of the HIV seronegative controls is also unknown.

It is evident that the study cohort forms the all-important basis for association studies and that the sample group representing the diverse South African population groups in the studies presented in chapters 3, 4 and 5 is not ideal. These studies should therefore be considered as pilot investigations for the identification of mutations and SNPs that should be further analysed in larger African ethnic groups, consisting of HIV-seropositive patients with known disease status and preferably “high-risk” HIV-seronegative controls.
Future prospects

A spectrum of SNPs have been identified in the pilot studies presented chapters 3, 4 and 5 and they provide a suitable basis for further investigation of larger and informative African-based study cohorts, to develop a possible model of polygenic contribution (small additive effects of multiple mutations and/or SNPs) to HIV/AIDS susceptibility. Due to the lack of clinical information regarding the majority of our HIV seropositive individuals, no significant disease associations could be made with any of the SNPs identified in chapters 3, 4 and 5 when analysed both individually (see appendix I – Table 2) and collectively (see appendix I – Table 3). Future intentions are to screen at least 300 HIV seropositive patients (including slow, normal and fast progressors) from diverse African ethnic groups of South Africa and to compare the allelic distribution of specific SNPs in a number of genes that are associated with HIV-1 pathogenesis to a population matched HIV-seronegative control group. The selected SNPs include: CCR5P35, CCR5Y335 (see chapter 3), CCR2 Int1-57a/g, CCR2 Int1-43g/a, CCR2V63, CCR2V64I, CCR2L283, CCR2AP339, CCR2BT348 (see chapter 4) and SDF1-3'A (see chapter 5).

It is possible that a few of the mutations and/or SNPs identified in the studies presented in this dissertation (see chapters 3, 4 and 5) may be in linkage disequilibrium with each other or with genetic variants of other known and/or unknown candidate genes involved in influencing host susceptibility to HIV-1 infection and disease progression to AIDS. Linkage disequilibrium is the tendency of specific combinations of alleles at linked loci to occur together on the same chromosome more frequently than would be expected by chance and is the result of selection for specific alleles, mutation, random genetic drift (chance variation of allelic frequencies
from one generation to the next) and population admixture. The significance of linkage disequilibrium is shown using chromosomal haplotype analysis, which identifies combinations of alleles as genetic markers associated with susceptibility to infection or disease. SNPs with no functional consequences, such as CCR5P35, CCR5Y339 (see chapter 3), CCR2V63, CCR2V64I, CCR2L283, CCR2AP339, CCR2BT348 (see chapter 4), can thus also act as genetic markers for identifying possible candidate genes, which are directly involved in determining susceptibility for infection and disease [Peterson et al., 1995; Huttley et al., 1999; Clark et al., 2001; Sachidanandam et al., 2001]. Further studies involving the identification of informative haplotypes for susceptibility to HIV-1 infection and disease progression to AIDS within different population groups is thus essential.

Besides association studies, research efforts to investigate and determine the underlying mechanisms of identified genetic variants, which may influence host susceptibility to HIV-1 infection and/or disease progression to AIDS, are also required [Howard et al., 1999; Blanpain et al., 2000; Lee et al., 1998; Mariani et al., 1999]. Our group has recently started with studies determining the functional consequences of a few naturally occurring CCR5 mutations. These mutations include CCR5D2V, CCR5L107F, CCR5R225X and CCR5R225Q (see chapter 3). The possibility that specific mutations/SNPs within genes encoding for chemokines and chemokine receptors may play a regulatory role in HIV-1 pathogenesis will provide a clear understanding of the host response to HIV-1 and raises the prospect of therapeutic intervention by targeting and disrupting the complex interaction between HIV-1 and the host target cells during viral entry.
Chapter 7

References
REFERENCES


• Bieniasz PD, Fridell RA, Aramori L, Ferguson SS, Caron M, Cullen BR. HIV-1-induced cell fusion is mediated by multiple regions within both the viral envelope and the CCR-5 co-receptor. *EMBO Journal* 1997, 16:2599-2609.


• Centers for Disease Control and Prevention (www.cdc.org).


• Cohen J. *South Africa’s New Enemy*. *Science* 2000a, **288**:2168-2170.


• Combadiere C, Ahuja S, Tiffany H, Murphy P. Cloning and functional expression of CC CKR5, a human monocyte CC chemokine receptor selective for MIP-1 (alpha), MIP-1 (beta), and RANTES. *J Leukoc Biol* 1996, **60**:147-152.

• Connor RI, Mohri H, Cao Y, Ho DD. Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J Virol* 1993, **67**:1772-1777.


• Cotton RGH, Rodrigues NR, Campbell RD. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide
and it's application to the study of mutations. Proc Natl Acad Sci USA 1988, 85:4397-4401.


• Eugen-Olsen J, Iversen AK, Benfield TL, Koppelhus U, Garred P. Chemokine receptor CCR2b 64I polymorphism and it's relation to CD4 cell counts and


• Hayes VM, Wu Y, Osinga J *et al.* Improvements in gel composition and electrophoretic conditions for broad-range mutation analysis by denaturing gradient gel electrophoresis. *Nucleic Acids Res* 1999a, **27**:e29.

• Hayes VM. *Mutation screening of the TP53 and KRAS genes in tumors of the colon, lung, brain and cervix.* University of Stellenbosch, PhD Thesis 1999b.


• Lander ES and Schork NJ. Genetic dissection of complex traits. Science 1994, 265:2037-2047.


• Lindley IJD, Westwick J, Kunkel SL. Nomenclature announcement – the chemokines. *Immunol Today* 1993, 14:24


• Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci USA* 1999, 90:6307-6311.
• Los Alamos HIV database (hiv-web.lanl.gov).


• Monteclaro FS, Charo IF. The amino terminal domain of CCR2 is both necessary and sufficient for high affinity binding of monocyte chemoattractant protein 1. J Biol Chem 1997, 272: 23186-23190.


• Phair JP. Determinants of the natural history of human immunodeficiency virus type 1 infection. *J Infect Dis* 1999, 179:S384-S386.


• Rollins BJ. Chemokines. *Blood* 1997, **90**:909-928.
• Roos MTL, Lange JMA, Goede REYd et al. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. *J Infect Dis* 1992, **165**:427-432.

• Ross TM, Bieniasz PD, Cullen BR. Multiple residues contribute to the inability of murine CCR-5 to function as a coreceptor for macrophage-tropic human immunodeficiency virus type 1 isolates. *J Virol* 1998, **72**:1918-1924.


• Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977, **74**:5463-5467.


• Schall TJ. Biology of the RANTES/SIS cytokine family. *Cytokine* 1991, **3**:165-183.

• Schall TJ, Bacon KB. Chemokines, leukocyte trafficking, and inflammation. *Curr Opin Immunol* 1994, **6**:865-873.

• Schuitemaker H, Koot M, Koogra NA *et al.* Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: Progression of disease is associated with shift from monocytotropic to T-cell tropic populations. *J Virol* 1992, **66**:1354-1360.
• Sheffield VC, Cox DR, Lerman LS, Myers RM. Attachment of a 40 base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single base changes. *Proc Natl Acad Sci USA* 1989, 86:232-236.


• Simmons G, Wilkinson D, Reeves JD et al. Primary, syncytium-inducing human immunodeficiency virus type 1 isolates are dual-tropic and most can use either Lestr or CCR5 as coreceptors for virus entry. *J Virol* 1996, 70:8355-8360.


• Starcich BR, Hahn BH, Shaw GM et al. Identification and characterization of conserved and variable regions of the envelope gene HTLV-III/LAV, the retrovirus of AIDS. Cell 1986, 45:637-648.


• Sullivan AD, Wigginton J, Kirschner D. The coreceptor mutation CCR5Δ32 influences the dynamics of HIV epidemics and is selected for by HIV. *Proc Natl Acad Sci USA* 2001, *98*:10214-10219.


• Verrier F, Borman AM, Brand D, Girard M. Role of the HIV type 1 glycoprotein 120 V3 loop in determining coreceptor usage. *AIDS Res Hum Retroviruses* 1999, **15**: 731-743.

• Vicari AP, Figueroa DJ, Hedrick JA et al. TECK: a novel CC chemokine specifically expressed by thymic dendritic cells and potentially involved in T cell development. *Immunity* 1997, **7**: 291-301.


• World Health Organisation (www.who.int).
• Wiley RL, Bonifacino JS, Potts BJ, Martin MA, Klausner RD. **Biosynthesis, cleavage and degradation of the human immunodeficiency virus type 1 envelope protein gp160.** *Proc Natl Acad Sci USA* 1988, 85:9580-9584.


• Wu Y. **Design and application of DGGE-based mutation detection systems for genes involved in HNPCC.** *University of Groningen, Phd Thesis* 1999.


Appendix I

HIV seropositive cohort
Table 1. The HIV seropositive cohort for the respective studies presented in this dissertation consisted of 103 individuals from three different South African population groups. Limited clinical information regarding the HIV seropositive individuals was provided by clinicians and is summarised below.

<table>
<thead>
<tr>
<th>Population groups</th>
<th>African (n = 70)</th>
<th>Coloured (n = 26)</th>
<th>Caucasian (n = 7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Female</td>
<td>49</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td><strong>Sexuality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>45</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>Homosexual</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Unknown</td>
<td>24</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td><strong>CD4 count</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 200 (cells/mm$^3$)</td>
<td>15</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>200 to 500</td>
<td>23</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>&gt; 500</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>26</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Disease progression</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slow</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Normal</td>
<td>10</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Fast</td>
<td>13</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>44</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td><strong>Clinical stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>10</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Stage II</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Stage III</td>
<td>11</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Stage IV</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Unknown</td>
<td>41</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

$n =$ Number of HIV seropositive individuals representing each of the population groups.
Table 2. The HIV seropositive cohort for the respective studies presented in this dissertation consisted of 43 individuals with known disease progression. SNPs identified in these 43 HIV seropositive individuals of different population groups are shown below for determining possible disease associations.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Africans (n = 26)</th>
<th>Coloureds (n = 14)</th>
<th>Caucasians (n = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slow</td>
<td>Normal</td>
<td>Fast</td>
</tr>
<tr>
<td>CCR5P35</td>
<td>1 - Hetero</td>
<td>1 - Hetero</td>
<td>1 - Hetero</td>
</tr>
<tr>
<td>CCR5Y335</td>
<td>-</td>
<td>-</td>
<td>1 - Hetero</td>
</tr>
<tr>
<td>CCR2 Int1-57a/g</td>
<td>-</td>
<td>1 - Hetero</td>
<td>1 - Hetero</td>
</tr>
<tr>
<td>CCR2 Int1-43g/a</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CCR2V63</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CCR2V64I</td>
<td>1 - Hetero</td>
<td>3 - Hetero</td>
<td>4 - Hetero</td>
</tr>
<tr>
<td>CCR2L283</td>
<td>-</td>
<td>1 - Hetero</td>
<td>1 - Hetero</td>
</tr>
<tr>
<td>CCR2AP339</td>
<td>1 - Homo</td>
<td>4 - Hetero</td>
<td>1 - Homo</td>
</tr>
<tr>
<td>-CCR2BT348</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SDF1-3'A</td>
<td>-</td>
<td>-</td>
<td>3 - Hetero</td>
</tr>
</tbody>
</table>

n, number of individuals with known disease progression; Int, intron; Homo, Homozygous for SNP; Hetero, Heterozygous for SNP.
Table 3. 43 HIV seropositive individuals with known disease progression formed part of the HIV seropositive cohort for the respective studies presented in this dissertation. The analysis of SNPs occurring collectively within the 43 HIV seropositive individuals of different population groups are shown below for determining possible disease associations.

<table>
<thead>
<tr>
<th>HIV patient</th>
<th>Population group</th>
<th>Disease progression</th>
<th>CCR5 P35</th>
<th>CCR5 Y335</th>
<th>CCR2 Int -57a/g</th>
<th>CCR2 Int -43g/a</th>
<th>CCR2 V63</th>
<th>CCR2 V64I</th>
<th>CCR2 L283</th>
<th>CCR2 P339</th>
<th>CCR2A</th>
<th>CCR2B T348</th>
<th>SDF1 3'A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>African</td>
<td>Slow</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Coloured</td>
<td>Slow</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Coloured</td>
<td>Slow</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>Coloured</td>
<td>Slow</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>Coloured</td>
<td>Slow</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>African</td>
<td>Slow</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>Caucasian</td>
<td>Slow</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>African</td>
<td>Slow</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Coloured</td>
<td>Slow</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>Coloured</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Coloured</td>
<td>Normal</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Coloured</td>
<td>Normal</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>16</td>
<td>Coloured</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>Coloured</td>
<td>Normal</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>HIV patient</td>
<td>Population group</td>
<td>Disease progression</td>
<td>CCR5 P35</td>
<td>CCR5 Y335</td>
<td>CCR2 Int -57a/g</td>
<td>CCR2 Int -43g/a</td>
<td>CCR2 V63</td>
<td>CCR2 V64I</td>
<td>CCR2 L283</td>
<td>CCR2 P339</td>
<td>CCR2A</td>
<td>CCR2B</td>
<td>T348</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>22</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>African</td>
<td>Normal</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>African</td>
<td>Normal</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>Caucasian</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>Caucasian</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>Coloured</td>
<td>Fast</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>29</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>Coloured</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>33</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>34</td>
<td>Coloured</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>35</td>
<td>Coloured</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>36</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>37</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>38</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>African</td>
<td>Fast</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>40</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>41</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>42</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>43</td>
<td>African</td>
<td>Fast</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Int, intron; ++, homozygous for SNP; +, heterozygous for SNP; -, no SNP.
Appendix II

Chemical stock solutions and Kits
A) Stock solutions used for agarose gel electrophoresis and DGGE

• **Agarose electrophoresis**

1) **5X TBE - buffer (5 Litres)**
   
   272.5 grams Tris  
   139 grams Boric Acid  
   14.5 grams EDTA  
   
   Fill up to 5 litres with distilled water.

2) **1X gel electrophoresis loading buffer (25 millilitres)**
   
   22.7 millilitres Formamide  
   0.168 grams EDTA  
   0.125 grams Xylene cyanol  
   0.125 grams Bromophenol blue  
   
   Fill up to 25 millilitres with distilled water.

• **DGGE**

1) **20X TAE - buffer (10 litres)**
   
   969 grams Tris - HCl  
   544 grams NaAc  
   74.5 grams EDTA  
   
   Adjust pH to 8.0 with NaOH (300-350ml) and fill up to 10 litres with distilled water.

2) **9% polyacrylamide (9% PAA) (1 litre)**
   
   25 millilitres 20 X TAE  
   225 millilitres 40% polyacrylamide (acryl: bisacryl = 37.5: 1)  
   
   Fill up to 1 litre with distilled water.
3) 90% UF / 9% polyacrylamide (1 litre)

- 25 millilitres 20 X TAE
- 225 millilitres 40% polyacrylamide (acryl: bisacryl = 37.5: 1)
- 378 grams Urea
- 360 millilitres Formamide (deionised)

Fill up to 1 litre with distilled water.

4) 10X gel electrophoresis loading buffer (500 millilitres)

- 1.25 grams Bromophenol blue
- 125 grams Ficoll 400
- 10 millilitres 0.5M EDTA

Fill up to 500 millilitres with distilled water

B) Kits used for DNA extraction and DNA sequencing

- **DNA extraction**
  
  1) QiAamp DNA mini kit (www.qiagen.com)

- **DNA sequencing**
  
  1) QiAquick PCR purification kit (www.qiagen.com)
  
  2) DyeEx spin column kit (www.qiagen.com)
Appendix III

Statistical analysis
**Statistical analysis**

Statistics is the science of collecting, analysing, presenting and interpreting mathematical data. As geneticists it is important to consider the mathematical aspects of the ways in which genes are inherited and to determine how 'normal' and 'mutant' genes are distributed in various population groups. Statistical analysis was performed for the respective studies presented in this dissertation and involved the calculation of allelic frequencies and the use of the Fischer exact tests for 2x2 contingency tables to test for a possible significance of heterogeneity in frequencies among HIV seropositive and HIV seronegative individuals for both mutations and SNPs.

1) **Allelic frequency**

Individuals can carry only two different alleles of a given gene. Therefore, a group of individuals in various populations can carry a large number of different alleles, giving rise to a reservoir of genetic diversity. The allelic frequency is thus the measurement of the proportion of individuals in a population group carrying a particular allele. The formula for calculating allelic frequencies is as follows:

\[
\text{Allelic frequency} = \frac{\text{The total no of a specific allele in a population}}{\text{The total no of alleles in a population}}
\]

The allelic frequencies have been calculated for all mutations and SNPs identified in both the HIV seropositive and HIV seronegative sample groups for the respective studies presented in chapters 3, 4 and 5.
Example 1: SDF1-3’A SNP (see chapter 5)

The HIV seropositive African population group consists of 70 individuals, where 64 individuals are homozygous for the wild-type allele (NN), 5 individuals are heterozygous for the wild-type and mutant alleles (MN) and 1 individual is homozygous for the mutant allele.

<table>
<thead>
<tr>
<th>Allele genotype</th>
<th>MM</th>
<th>MN</th>
<th>NN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of individuals</td>
<td>64</td>
<td>5</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>No of N alleles</td>
<td>128</td>
<td>5</td>
<td>0</td>
<td>133</td>
</tr>
<tr>
<td>No of M alleles</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Total no of alleles</td>
<td>128</td>
<td>10</td>
<td>2</td>
<td>140</td>
</tr>
</tbody>
</table>

Allelic frequency of the N allele in the population = 133/140 = 0.95
Allelic frequency of the M allele in the population = 7/140 = 0.05

2) Fisher exact test

The Fisher exact test is a commonly used test to determine statistical significance. The application of the Fisher exact test for a 2x2 contingency table is a test of association between mutually exclusive categories of one variable (given in the rows of table) and mutually exclusive categories of another variable (given in the columns of table). This specific statistical method thus tests the null hypothesis (no association exists between the two variable categories) against a two-sided (both variable extremes are specified) alternative hypothesis (the two variables are associated) by calculating the exact probabilities of all 2x2 contingency table of observed frequencies where marginal totals are kept the same. A derived probability
level for the test is then calculated by summation. It is always necessary to consider a threshold of significance, which is the risk taken in rejecting the null hypothesis. The usual value chosen for a significance level is 0.05. An explanation of the test is as follows:

\[
\begin{array}{ccc|cc}
& & & a +b & \\
& a & b & & \\
& c & d & c +d & \\
& a +c & b +d & n & \\
\end{array}
\]

Consider a 2x2 contingency table of observed frequencies with marginal (row and column) totals, and \( n \) is the sum of the observed frequencies, \( a + b + c + d \). The probability is calculated using the formula:

\[
p = \frac{(a + b)! \ (c + d)! \ (a + c)! \ (b + d)!}{N! \ a! \ b! \ c! \ d!}
\]

The formula is used repeatedly for all 2x2 contingency tables with the same marginal totals. All the probabilities thus obtained which are less than or equal to the initial probability, including the initial probability itself, are summed to obtain the total derived probability. The null hypothesis is rejected if this total derived probability is less than the significance level (usually 0.05) that has been chosen.

Using the GraphPad InStat computer program, the Fisher exact test for 2x2 contingency tables was applied to test the significance of \( CCR5, CCR2 \) and \( SDF1 \) mutations and SNPs within the African, Coloured and Caucasian population groups. The degree of significance is defined as: not significant, when the \( P \) value is greater than 0.05; significant, when the \( P \) value is 0.01 – 0.05; very significant, when the \( P \) value is 0.001 – 0.01; extremely significant, when the \( P \) value is less than 0.001.
Example 1: CCR5P35 SNP (see chapter 3)

<table>
<thead>
<tr>
<th></th>
<th>African population</th>
<th></th>
<th>Coloured population</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HIV +</td>
<td>HIV -</td>
<td>HIV +</td>
<td>HIV -</td>
</tr>
<tr>
<td>CCR5 wild-type</td>
<td>134</td>
<td>119</td>
<td>43</td>
<td>134</td>
</tr>
<tr>
<td>CCR5P35</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>140</td>
<td>128</td>
<td>52</td>
<td>144</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.4273 \text{ (not significant)} \]

\[ P = 0.0516 \text{ (not significant)} \]
Example 2: CCR5A335V SNP (see chapter 3)

### African population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5 wild-type</td>
<td>136</td>
<td>126</td>
<td>262</td>
</tr>
<tr>
<td>CCR5A335V</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>140</td>
<td>128</td>
<td>268</td>
</tr>
</tbody>
</table>

\( n = \) no. of alleles

\( P = 0.6859 \) (not significant)

### Coloured population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5 wild-type</td>
<td>51</td>
<td>142</td>
<td>193</td>
</tr>
<tr>
<td>CCR5A335V</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>144</td>
<td>196</td>
</tr>
</tbody>
</table>

\( n = \) no. of alleles

\( P = 1.0000 \) (not significant)
Example 3: CCR2V64I SNP (see chapter 4)

<table>
<thead>
<tr>
<th>African population</th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2 wild-type</td>
<td>120</td>
<td>103</td>
<td>223</td>
</tr>
<tr>
<td>CCR2V64I</td>
<td>18</td>
<td>21</td>
<td>39</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138</td>
<td>124</td>
<td>262</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.3906 \text{ (not significant)} \]

<table>
<thead>
<tr>
<th>Coloured population</th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2 wild-type</td>
<td>51</td>
<td>119</td>
<td>170</td>
</tr>
<tr>
<td>CCR2V64I</td>
<td>1</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>144</td>
<td>196</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.0034 \text{ (very significant)} \]

<table>
<thead>
<tr>
<th>Caucasian population</th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2 wild-type</td>
<td>14</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>CCR2V64I</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14</td>
<td>4</td>
<td>18</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.2222 \text{ (not significant.)} \]

Sample numbers for Caucasian HIV+ patients and HIV - controls are however small.
Example 4: CCR2N260 SNP (see chapter 4)

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>African population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCR2 wild-type</td>
<td>63</td>
<td>45</td>
<td>108</td>
</tr>
<tr>
<td>CCR2N260</td>
<td>75</td>
<td>79</td>
<td>154</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138</td>
<td>124</td>
<td>262</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.1331 \text{ (not significant)} \]

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coloured population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCR2 wild-type</td>
<td>22</td>
<td>47</td>
<td>69</td>
</tr>
<tr>
<td>CCR2N260</td>
<td>30</td>
<td>97</td>
<td>127</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>144</td>
<td>196</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.2374 \text{ (not significant)} \]

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Caucasian population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCR2 wild-type</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CCR2N260</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14</td>
<td>4</td>
<td>18</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.5686 \text{ (not significant)} \]

Sample numbers for Caucasian HIV+ patients and HIV - controls are however small.
**Example 5: CCR2P339 SNP (see chapter 4)**

### African population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2 wild-type</td>
<td>117</td>
<td>106</td>
<td>223</td>
</tr>
<tr>
<td>CCR2P339</td>
<td>21</td>
<td>18</td>
<td>39</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>138</td>
<td>124</td>
<td>262</td>
</tr>
</tbody>
</table>

\( n = \) no. of alleles

\[ P = 1.0000 \text{ (not significant)} \]

### Coloured population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR2 wild-type</td>
<td>50</td>
<td>139</td>
<td>189</td>
</tr>
<tr>
<td>CCR2P339</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>144</td>
<td>196</td>
</tr>
</tbody>
</table>

\( n = \) no. of alleles

\[ P = 1.0000 \text{ (not significant)} \]
Example 6: SDF1-3'A SNP (see chapter 5)

### African population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDF1 wild-type</td>
<td>133</td>
<td>125</td>
<td>258</td>
</tr>
<tr>
<td>SDF1-3'A</td>
<td>7</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>140</td>
<td>126</td>
<td>266</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.0692 \text{ (not significant)} \]

### Coloured population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDF1 wild-type</td>
<td>48</td>
<td>119</td>
<td>167</td>
</tr>
<tr>
<td>SDF1-3'A</td>
<td>4</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>146</td>
<td>198</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.0768 \text{ (not significant)} \]

### Caucasian population

<table>
<thead>
<tr>
<th></th>
<th>HIV +</th>
<th>HIV -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDF1 wild-type</td>
<td>11</td>
<td>69</td>
<td>80</td>
</tr>
<tr>
<td>SDF1-3'A</td>
<td>3</td>
<td>27</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14</td>
<td>96</td>
<td>110</td>
</tr>
</tbody>
</table>

n = no. of alleles

\[ P = 0.7544 \text{ (not significant)} \]
3) **Hardy - Weinberg equilibrium (HWE)**

The Hardy - Weinberg principle is applied to measure the maintenance of allele frequencies in a large population with random mating (not considering partner's genotype) and the absence of selection for a specific genotype, mutation, migration (diffusion of alleles across racial and geographical boundaries) and random genetic drift (the increased transmission of a specific allele to offspring by chance). Determining whether a specific population is in Hardy – Weinberg equilibrium (HWE) is thus important for association studies conducted in large study cohorts.

**Example 1: Investigating whether the 70 African HIV seronegative controls are in HWE for the CCR5P35 SNP (see chapter 3).**

<table>
<thead>
<tr>
<th>Observed (n = 70)</th>
<th>Expected (n = 70)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/N = 64</td>
<td>N/N = 64.11 (p^2 x 70)</td>
</tr>
<tr>
<td>N/M = 6</td>
<td>N/M = 5.76 (2pq x 70)</td>
</tr>
<tr>
<td>M/M = 0</td>
<td>MM = 0.13 (q^2 x 70)</td>
</tr>
</tbody>
</table>

n = number of individuals  
N = wildtype allele; M = mutant allele

Incidence of wild-type allele = \( \frac{2(NN) + NM}{\text{Total no. of alleles}} \)  
\[ = \frac{2(64) + 6}{140} \]  
\[ p = 0.957 \]

Incidence of mutant allele = \( \frac{2(MM) + NM}{\text{Total no. of alleles}} \)  
\[ = \frac{2(0) + 6}{140} \]  
\[ q = 0.043 \]

The observed and the expected values correspond closely and therefore the population is in equilibrium.
Further reading