

**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-1 β (*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 immuuniteitsgebrek virus (MIV) infeksie en siekteprogressie na die verworwe immuuniteitsgebrek sindroom (VIGS) is aangetoon deur die identifikasie van 'n 32 basis paar (bp) deleisie 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* deleisie 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-1 β (*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

Glu (E)	glutamic acid
Gly (G)	glycine
gp	glycoprotein
GTP	guanosine 5'-triphosphate
HCl	hydrochloric acid
His (H)	histidine
HIV	human immunodeficiency virus
HIV –	HIV seronegative
HIV +	HIV seropositive
HLA	human leucocyte antigen
IL	interleukin
Ile (I)	isoleucine
kb	kilobase
kDa	kilodaltons
Leu (L)	leucine
LTNP	long-term non-progressor
Lys (K)	lysine
MBP	mannose binding protein
MCP	monocyte chemotactic protein
Met (M)	methionine
MIP	macrophage inflammatory protein

M-tropic	macrophage tropic
NaAc	sodium acetate
NaOH	sodium hydroxide
NRAMP1	natural resistance-associated macrophage protein 1
NSI	non-syncytium inducing
ORF	open reading frame
PCR	polymerase chain reaction
Phe (F)	phenylalanine
Pro (P)	proline
RANTES	regulated on activation normal T cell expressed and secreted
RFLP	restriction fragment length polymorphism
SDF1	stromal derived factor-1
Ser (S)	serine
SI	syncytium inducing
SNP	single nucleotide polymorphism
SSCP	single-stranded conformation polymorphism
Th	T helper
Thr (T)	threonine
T lymphocytes	thymus-derived lymphocytes
TM	transmembrane

Trp (W)	tryptophan
T-tropic	T cell line tropic
Tyr (Y)	tyrosine
UF	urea/formamide
3' UTR	3' untranslated region
Valine (V)	valine
VDR	vitamin D receptor
V loop	hypervariable loop

CONTENTS

Summary	iii
Opsomming	v
Acknowledgements	vii
List of abbreviations	ix
<u>Chapter 1: General Introduction</u>	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
<u>Chapter 2: Study aim and overview</u>	23
2.1. Aim of the study	24
2.2. Host susceptibility to HIV-1 infection and/or AIDS	25
2.1.1. CCR5	27
2.1.2. CCR2	34
2.1.3. 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 (<i>AIDS</i> 15:171-177, 2001)	47
--	-----------

Chapter 4: CCR2

Novel single nucleotide polymorphisms identified using a comprehensive CCR2-denaturing gradient gel electrophoresis assay (<i>Submitted</i>)	64
--	-----------

Chapter 5: SDF1

Analysis of the <i>SDF1</i> -3' A single nucleotide polymorphism (SNP) within a South African population.	80
---	-----------

Chapter 6: Discussion **91**

Chapter 7: References **98**

Appendix I: HIV seropositive study cohort **133**

Table 1: Clinical information regarding the HIV seropositive study cohort	134
---	------------

Table 2: SNPs analysed individually in the HIV seropositive study cohort	135
--	------------

Table 3: SNPs analysed collectively in the HIV seropositive study cohort	136
--	------------

Appendix II: Chemical Stock Solutions and Kits **138**

A) Stock solutions used for agarose gel electrophoresis and DGGE	139
--	------------

B) Kits used for DNA extraction and DNA sequencing	140
--	------------

<u>Appendix III: Statistical analysis</u>	141
1) Allelic frequency	142
2) Fisher exact test	143
3) Hardy - Weinberg equilibrium (HWE)	151

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].

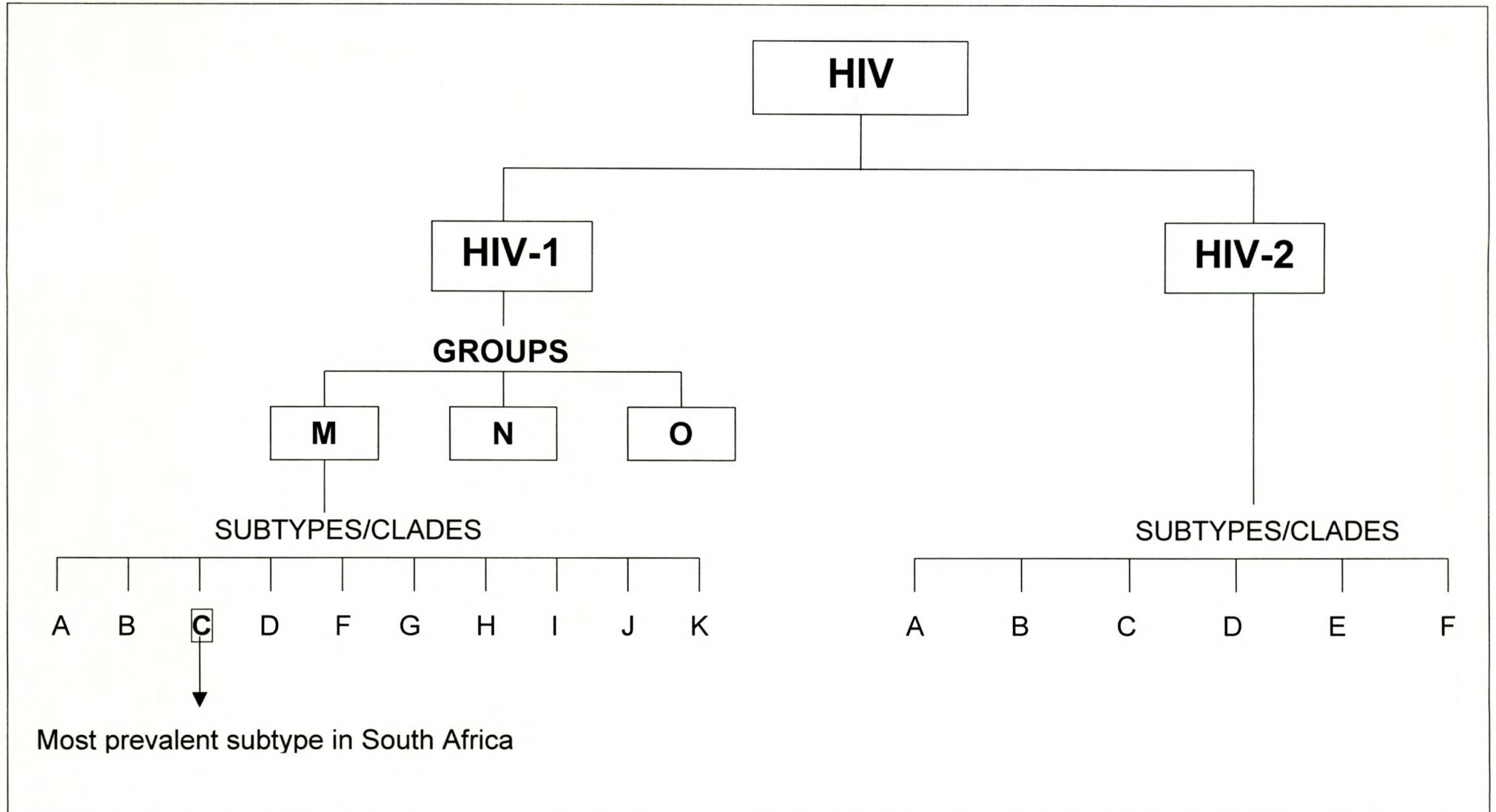


Figure 1. HIV-1 is divided into three groups M, N and O and Group M is further subdivided into 10 subtypes (A – D, F – K). The most prevalent HIV-1 subtype in South Africa is subtype C. HIV-2 is comprised of six diverse subtypes (A – F).

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 Bharmarapavati, 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].

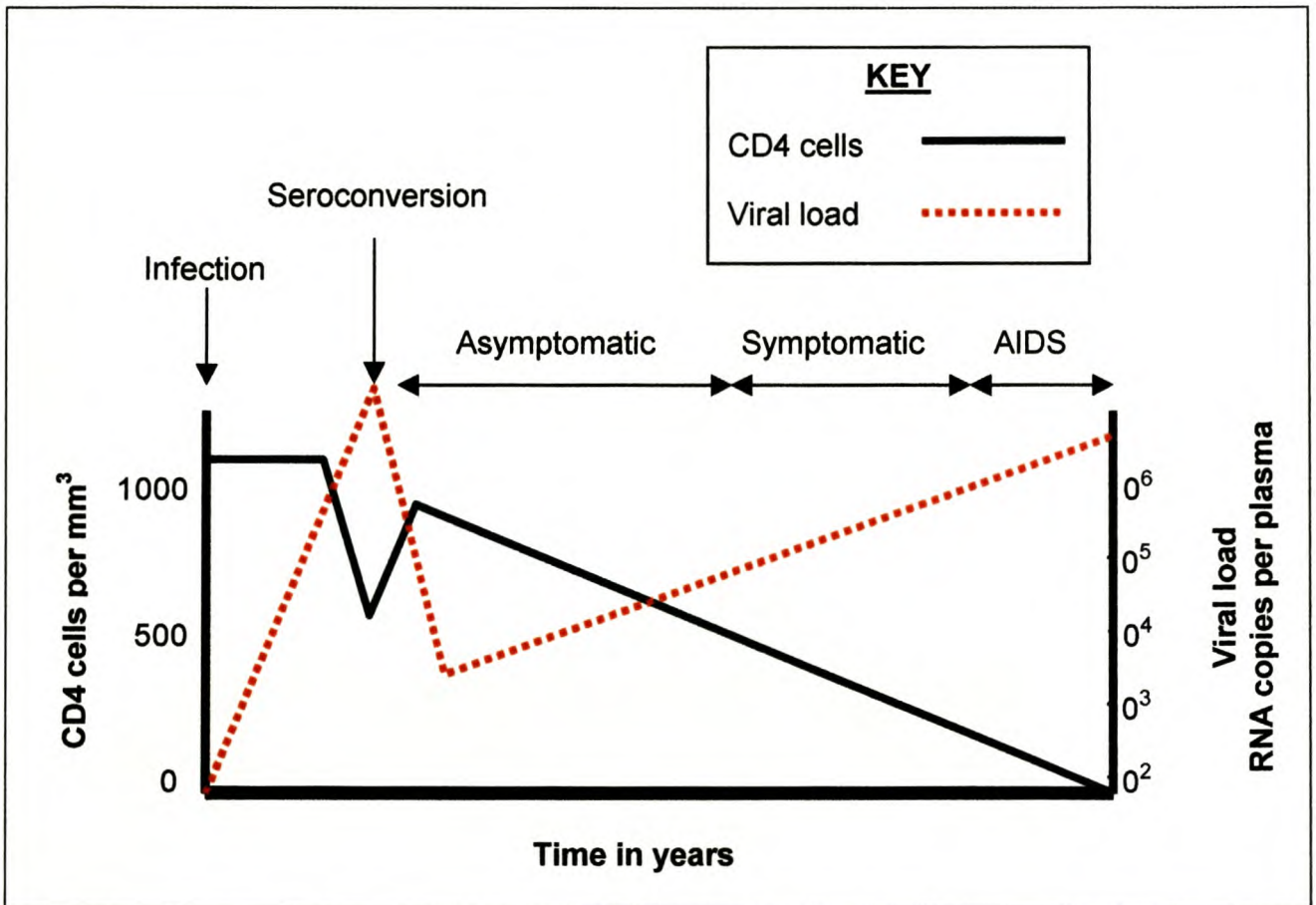


Figure 2. The CD4+ cell count and/or the amount of viral load is an indication of the individual's disease progression to AIDS.

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 Basle, 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 neutrophils, 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) [Broxmeyer *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].

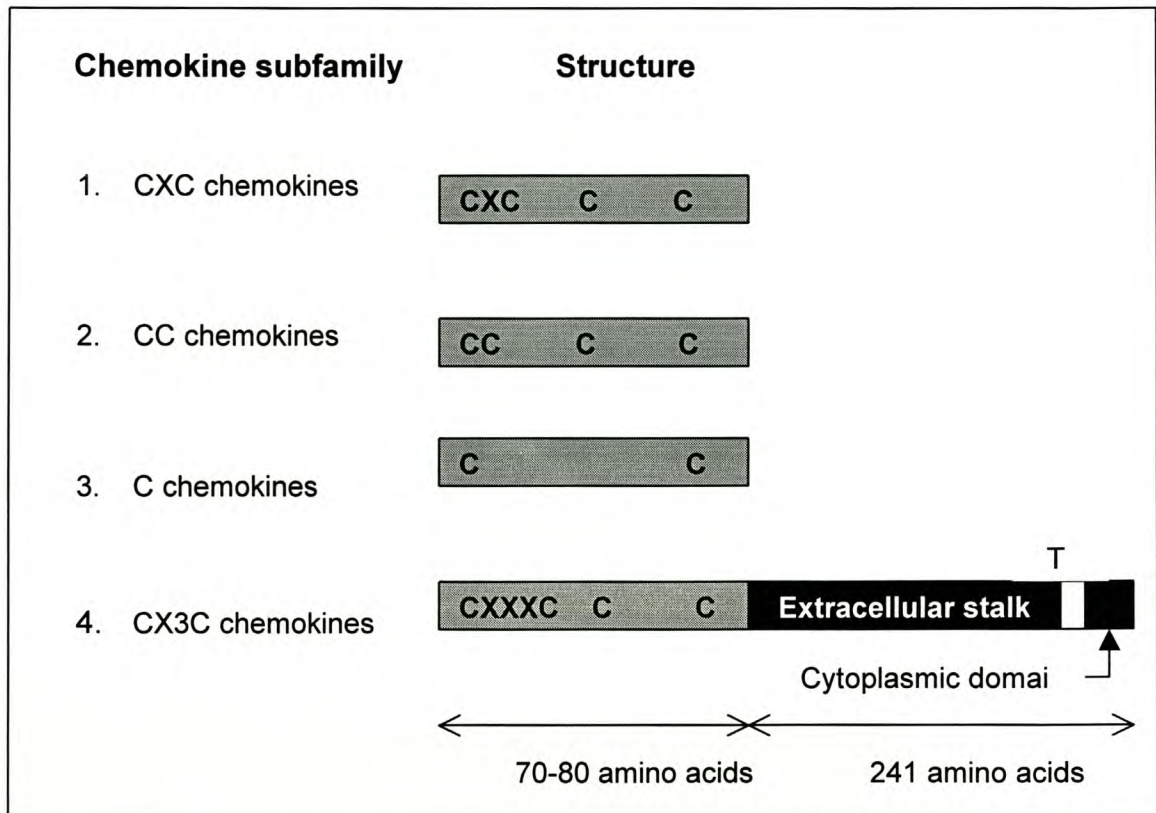


Figure 3. Chemokines are grouped into four subfamilies based on the arrangement of the first two of four highly conserved amino-terminal cysteine residues (Adapted from Ward and Westwick, 1998). Chemokines are not drawn to scale.

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; Baggiolini *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 [Baggiolini *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; Baggiolini *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; Baggiolini *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; Baggiolini *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-1 α (macrophage inflammatory protein alpha) (CCL3) and MIP-1 β (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 [Dalglish *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).

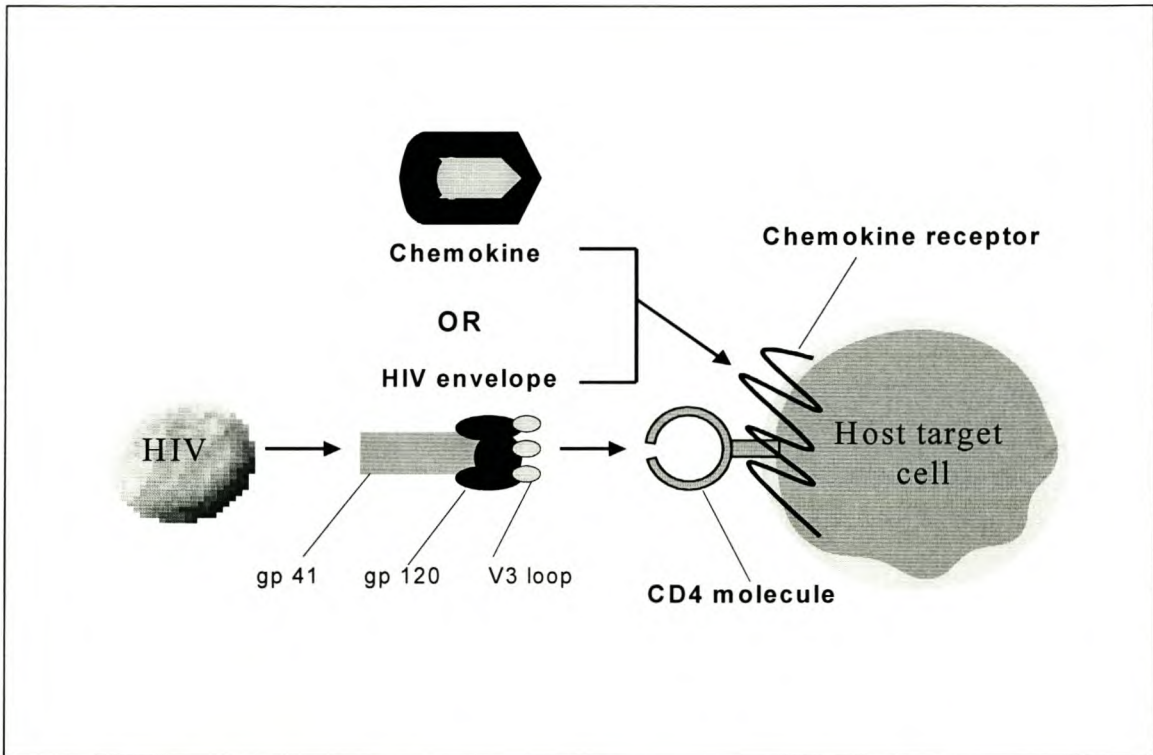


Figure 4. HIV entry is initiated by the interaction of the virion envelope glycoproteins, gp120 and gp41, with two cellular host receptors, of which one is a CD4 molecule and the other, a chemokine receptor, whose natural ligands are specific chemokines.

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.

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

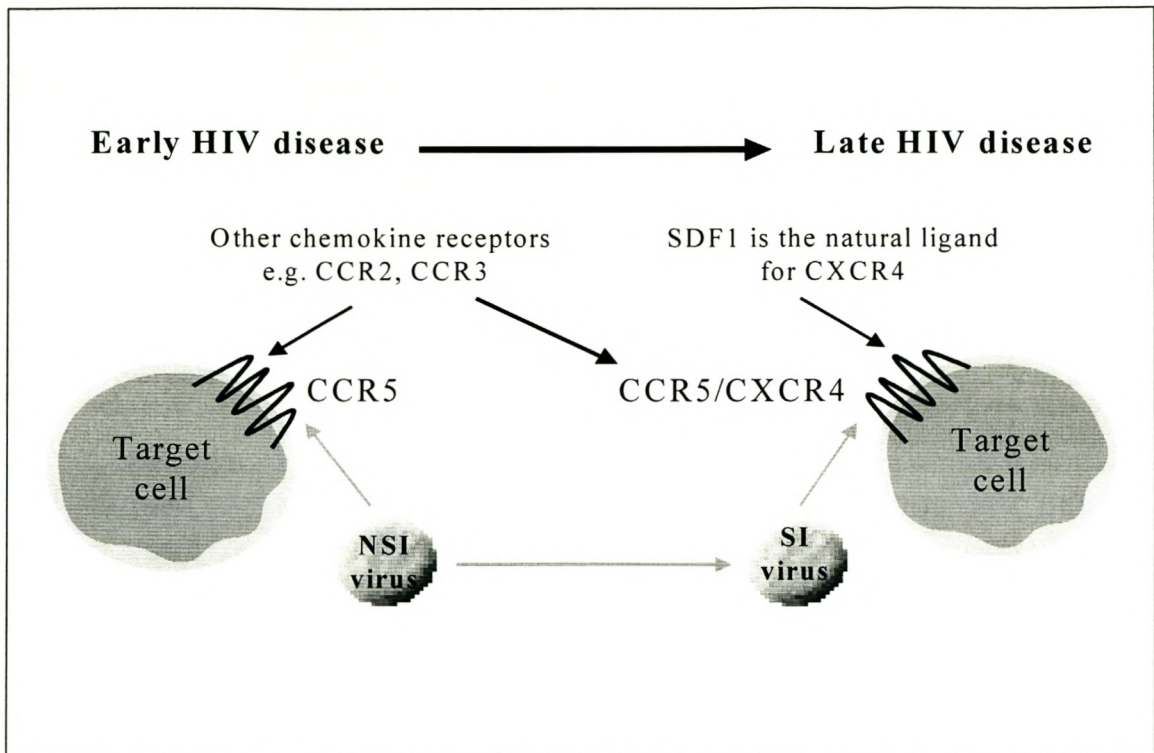


Figure 5. Schematic illustration of the current model for co-receptor usage by different viruses.

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; Langede-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-1 α (CCL3) and MIP-1 β (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 MIP1 β (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 these 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-1 α (CCL3) and MIP-1 β (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 its 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

Previously identified mutations in the *CCR5* gene, include a 32bp deletion (*CCR5* Δ 32), 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 *CCR5* Δ 32 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 *CCR5* Δ 32 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 utilises CXCR4 as a co-receptor [Biti *et al.*, 1997; Balotta *et al.*, 1997 O'Brien *et al.*, 1997b; Theodorou *et al.*, 1997]. Heterozygosity for *CCR5* Δ 32 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].

CCR5 Δ 32 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 *CCR5* Δ 32 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* Δ 32 in the Caucasian population [Stephens *et al.*, 1998]. It has also been hypothesised that the Vikings disseminated *CCR5* Δ 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* Δ 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* Δ 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* Δ 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.

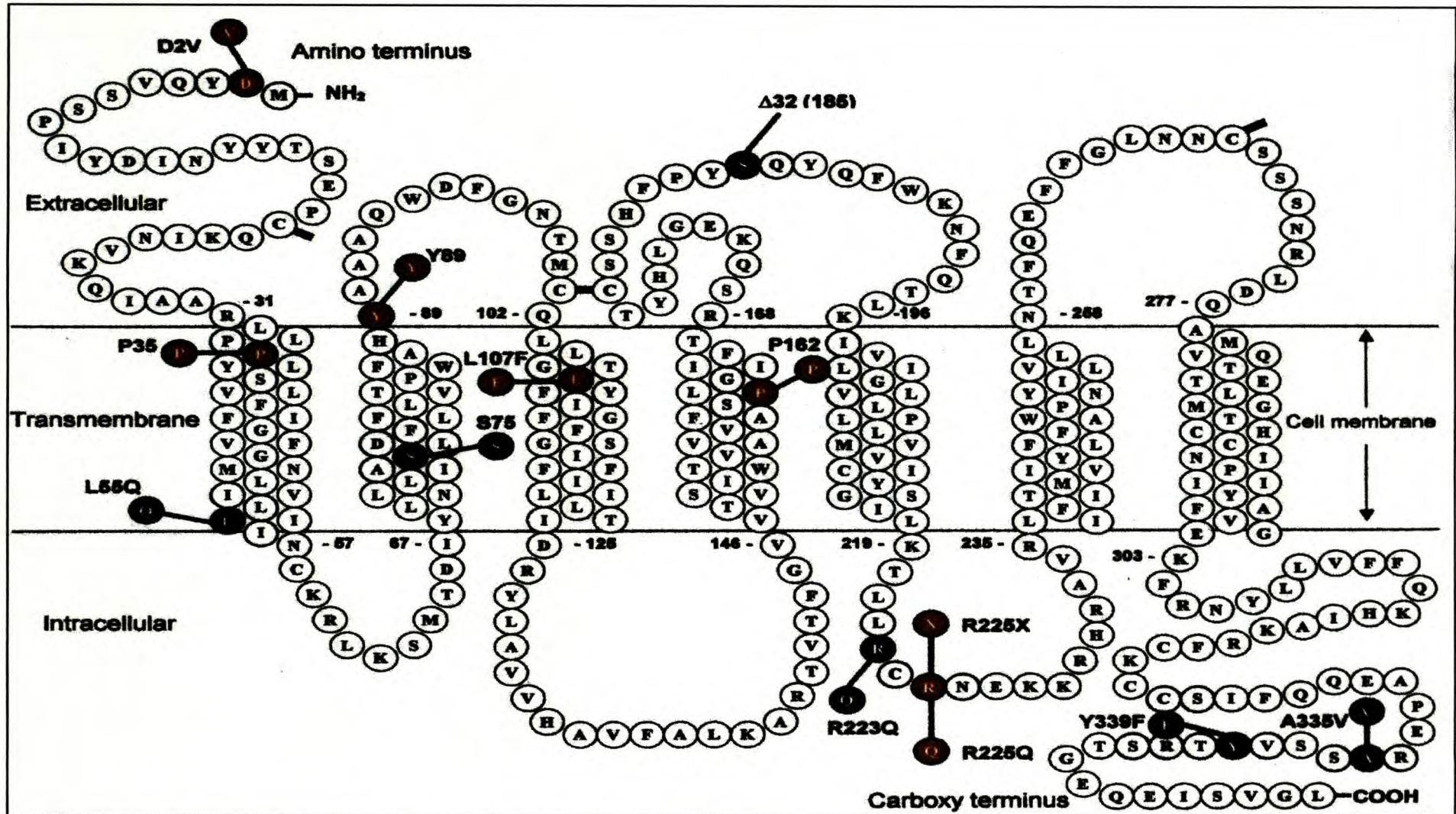


Figure 2. Schematic illustration of the human CCR5 protein. A recent study by our group resulted in the identification of previously reported (indicated by the blue shaded amino acid residues) and novel (indicated by the red shaded amino acid residues) CCR5 mutations within an African-based population 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, Δ K228, 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 [Montecclaro and Charo, 1996; Montecclaro 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 its 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 its 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 *CCR2B* is found within exon 2 [Wong *et al.*, 1997] (see Figure 3). The 5' untranslated region of the *CCR2* gene plays an essential role in transcriptional activation and tissue specific expression of *CCR2* [Yamamoto *et al.*, 1999].

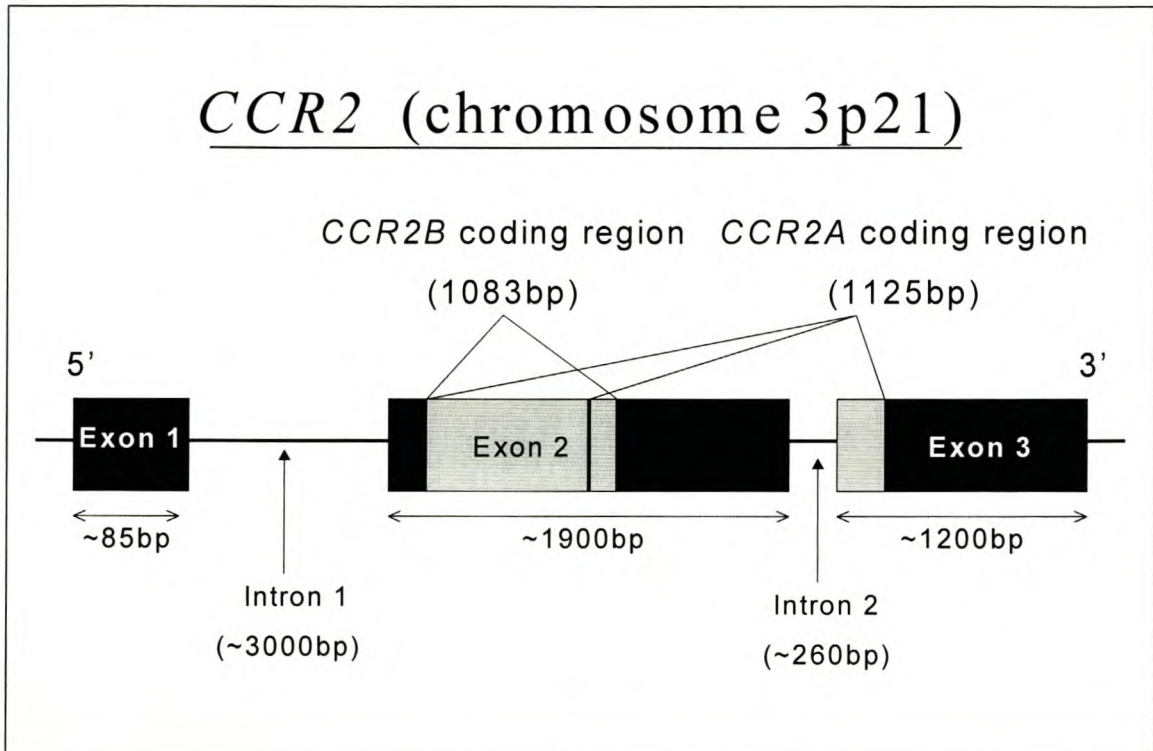


Figure 3. *CCR2* has two gene transcripts. The coding region for *CCR2A* is found in exon 2 and part of exon 3, while exon 2 contains the entire coding region for *CCR2B*. Black blocks represent the non-coding regions and the grey blocks represent the coding regions. Exons and introns are not drawn to scale.

A common genetic variant has been identified in the first transmembrane domain of the *CCR2* gene. It results in a conservative amino acid change from valine to isoleucine at codon 64 (*CCR2V64I*) [Smith *et al.*, 1997a]. Although the mechanism by which the *CCR2V64I* SNP influences HIV pathogenesis still needs to be elucidated [Lee *et al.*, 1998; Mariani *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; Rizzardì *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 its 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 SDF1 α (89 amino acids) and SDF1 β (93 amino acids). The SDF1 β 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 SDF1 α and SDF1 β proteins form an amino acid-cleaved signal peptide and a SDF1 α form processed at the carboxy terminus to generate a 67 residue protein has been purified [Bleul *et al.*, 1996b]. The SDF1 α and SDF1 β 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 *SDF1* gene is an exception and its location is at band q11 of chromosome 10 [Shirozu *et al.*, 1995] (Figure 4).

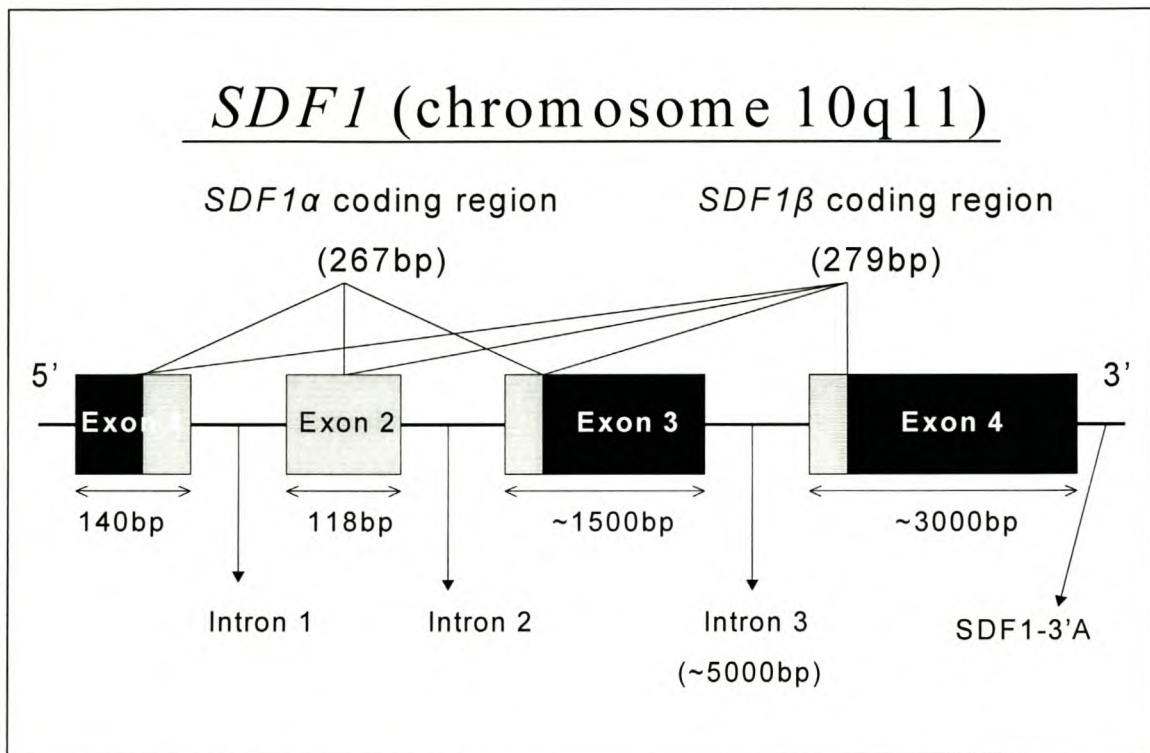


Figure 4. *SDF1* has 2 gene transcripts, *SDF1* α and *SDF1* β . The coding regions for *SDF1* α and *SDF1* β are found within exons 1 to 3 and exons 1 to 4, respectively. Black blocks represent the non-coding regions and the grey blocks represent the coding regions. The lengths of intron 1 and 2 are unknown. Exons and introns are not drawn to scale.

A common SNP of the *SDF1* β gene transcript, designated *SDF1*-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 *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 *SDF1*-3' A SNP in the homozygous state has been associated with delaying the rate of disease progression to AIDS, while the *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 *SDF1 β* 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.

Pre-screening method	Advantages	Disadvantages	References
<u>Restriction fragment length polymorphism (RFLP) analysis:</u> Involves the use of restriction endonucleases with recognition sites in DNA sequences for the detection of mutations.	Simplicity in use and cost-effective. Allows rapid mutation detection. Requires the use of equipment available in most laboratories.	Incomplete digestion of the restriction endonucleases may give false positive or negative results. Can only be used for the detection of specific sequence variants.	Parry <i>et al.</i> , 1990; Reviewed in Pourzand and Cerutti, 1993.
<u>Single-stranded conformation polymorphism (SSCP) analysis:</u> Is based on the fact that a mutation results in a conformational change of the single-stranded DNA, which is then observed as an electrophoretic mobility shift.	Versatile and does not require any expertise. Allows rapid mutation detection. Requires the use of equipment available in most laboratories.	Only 80% mutation detection rate for DNA fragments up to 300bp in length. Efficiency of method depends on a number of experimental conditions.	Orita <i>et al.</i> , 1989; Reviewed in Cotton <i>et al.</i> , 1998.
<u>Chemical cleavage of mismatch (CCM):</u> Involves the use of chemical agents for cleavage and thus detection of single base mismatches in the target DNA, which is hybridised to a single-stranded radiolabelled DNA probe. An adaptation of this method is enzyme mismatch cleavage (ECM), which utilises T4 endonuclease VII for the detection of mutations.	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)	Involves the use of hazardous chemicals and the % mutation detection rate can be reduced if low quality chemicals are used.	Cotton <i>et al.</i> , 1988; Youil <i>et al.</i> , 1995; Reviewed in Ellis <i>et al.</i> , 1998.
<u>Denaturing gradient gel electrophoresis (DGGE) analysis:</u> Is based on the differential melting behaviour of the wild-type and normal DNA fragments within an urea/formamide-denaturing gradient gel. Less used adaptations of this method include, constant denaturing gel electrophoresis (CDGE), temperature gradient gel electrophoresis (TGGE) and two-dimensional (2-D) DNA electrophoresis.	100% mutation detection rate for fragments up to 500bp in length. Rapid mutation detection and clear visualisation of mutations. Theory for the entire concept.	Standardising of assays is a time-consuming procedure. Extra costs for the use of with GC-clamped primers. Expertise is required.	Fischer and Lerman, 1983; Reviewed in van Orsouw, 1998; Wu, 1999; Hayes, 1999b;
<u>Denaturing high-performance liquid chromatography (DHPLC) analysis:</u> Involves detection of mutations by chromatography and thus the observation of differential retention signals for the wild-type and mutant DNA.	Sensitivity and specificity of mutation detection exceeds 96%. DNA fragments as large as 1.5kb in length can be analysed.	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.	Underhill <i>et al.</i> , 1997; Liu <i>et al.</i> , 1998; Reviewed in Xiao and Oefner, 2001.

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: $\% \text{ UF} = [(\text{melting temperature} - \text{buffer temperature}) \times 100 / 32]$. 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 (egs. 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 is 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.

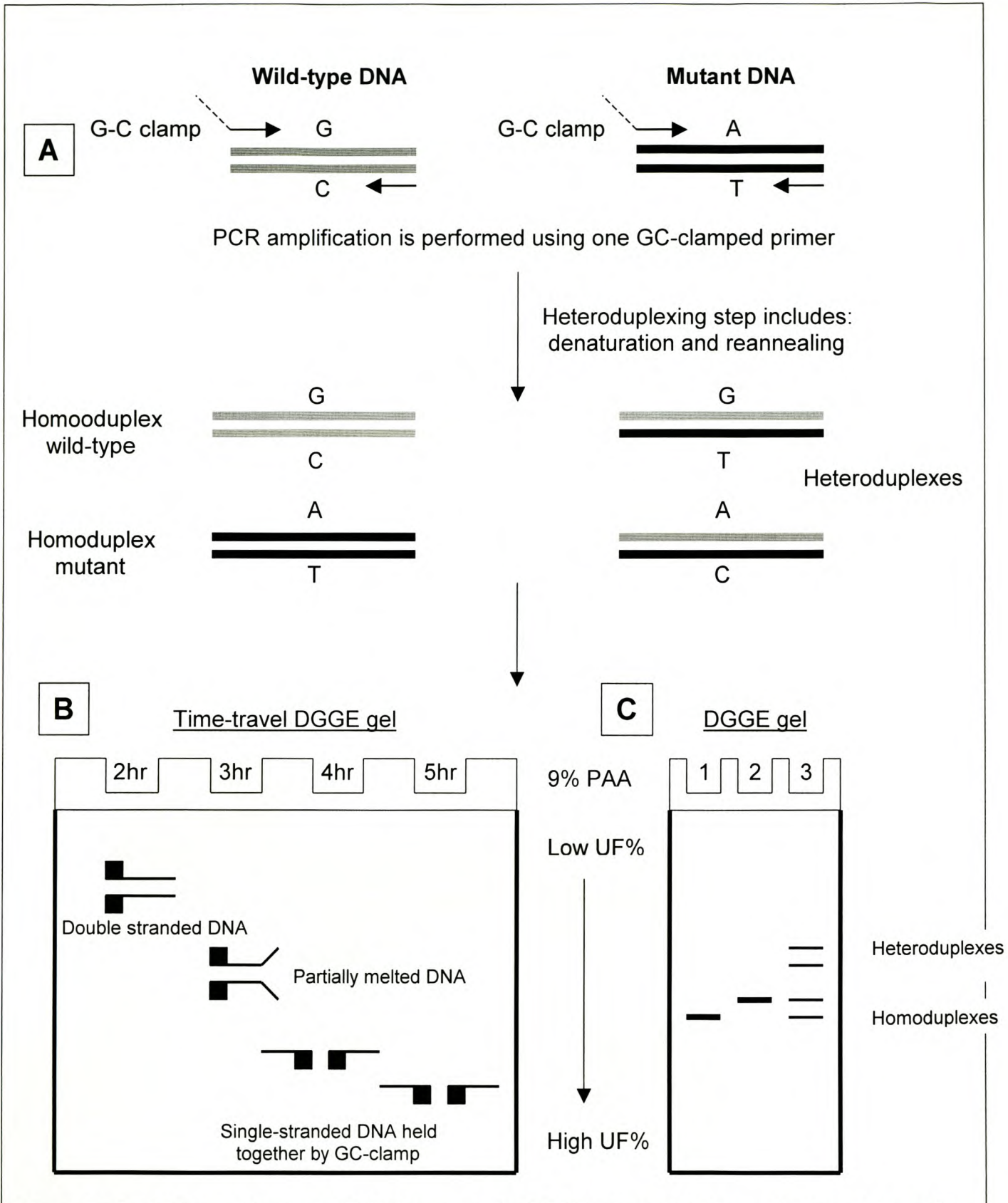


Figure 5. **A.** PCR amplification of DNA fragments for DGGE using a GC-clamped primer, followed by a heteroduplexing step, which involves denaturation and reannealing to form a wild-type homoduplex, a mutant homoduplex and heteroduplexes. **B.** The principles of DGGE are depicted as a time-travel DGGE-gel (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 Haves, 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. Hayes^{4,6}**

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-1 α and MIP-1 β [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 (*CCR5* Δ 32) 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 *CCR5* Δ 32 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 *CCR5* Δ 32 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.

Fragment	Amplimers, 5'-3'	Size (bp)	Temperature (°C)	
			Melting	Annealing
A	[40GC]TGGAGGGCAACTAAATACAT CGATTTGCTTCACATTGATT	196	67	54
B	[10GC]ATTATACATCGGAGCCCTGC [40GC]AGCATAGTGAGCCCAGAAGG	280	74	60
C	[40GC][10AT]CTGGCCATCTCTGACCTGTT GATGATTCTGGGAGAGACG	332	73	60
D	[40GC][10AT]ACTTGGGTGGTGGCTGTGTT CATTTGACACCGAAGCAGA	276	72	60
E	TCATGGTCATCTGCTAGTCG [40GC]GGTGTTCCAGGAGAAGGACAA	192	72	58
F	[40GC]TTCTCTTCTGGGCTCCCTAC GTCACCAGCCCCTTGAGTC	390	74	60

bp, base pair

GC-clamps used were as follows: [40GC] CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCGCCCCG
[10GC] CGCCGCCGCG

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 μ l 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 μ l (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).

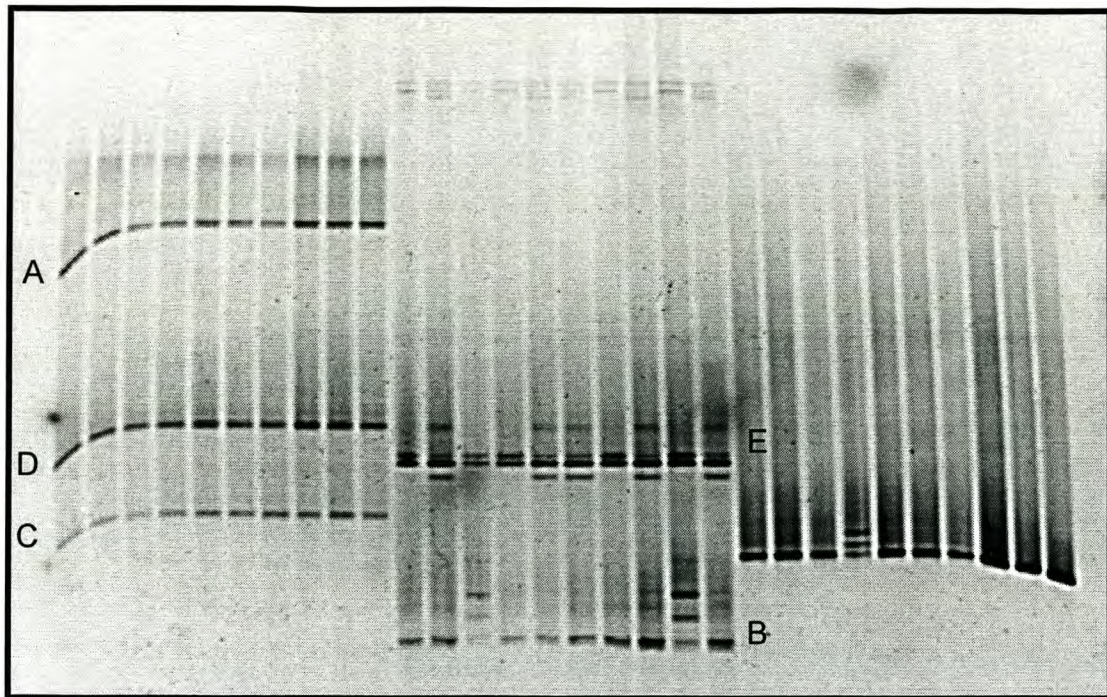


Figure 1. DGGE banding pattern of 6 pooled amplicons of 10 patients covering the entire coding region of the *CCR5* gene. The 30%-70% UF/9% PAA denaturing gel was electrophoresed at 59°C for 110 volts overnight and visualised by staining with ethidium bromide. Lanes 1-10 contains group 1 (fragments A, 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 *CCR5*Δ32 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.

Mutation	Base change	Fragment	Allele frequency					
			Africans		Coloureds		Caucasians	
			HIV + (n = 140)	HIV - (n = 128)	HIV + (n = 52)	HIV - (n = 144)	HIV + (n = 14)	HIV - (n = 4)
D2V*	GAT-GTT	A	1(.007)	0	0	1(.007)	0	0
P35*	CCG-CCA	B	6(.043)	9(.070)	9(.173)	10(.069)	0	0
L55Q	CTG-CAG	B	0	0	0	0	1(.071)	0
S75	TCT-TCC	B	1(.007)	0	0	0	0	0
Y89*	TAT-TAC	C	0	0	0	1(.007)	0	0
L107F*#	CTC-TTC	C	1(.007)	2(.016)	1(.019)	0	0	0
P162*	CCA-CCG	D	1(.007)	0	0	0	0	0
Δ32(185)	-	D	0	0	1(.019)	5(.035)	1(.071)	0
R223Q	CGG-CAG	E	0	0	0	1(.007)	0	0
R225X*#	CGA-TGA	E	1(.007)	2(.016)	1(.019)	0	0	0
R225Q*	CGA-CAA	E	0	0	0	1(.007)	0	0
A335V	GCA-CTA	F	4(.029)	2(.016)	1(.019)	2(.014)	0	0
Y339F	TAC-TTC	F	0	1(.008)	0	0	0	0

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 *CCR5*Δ32 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.

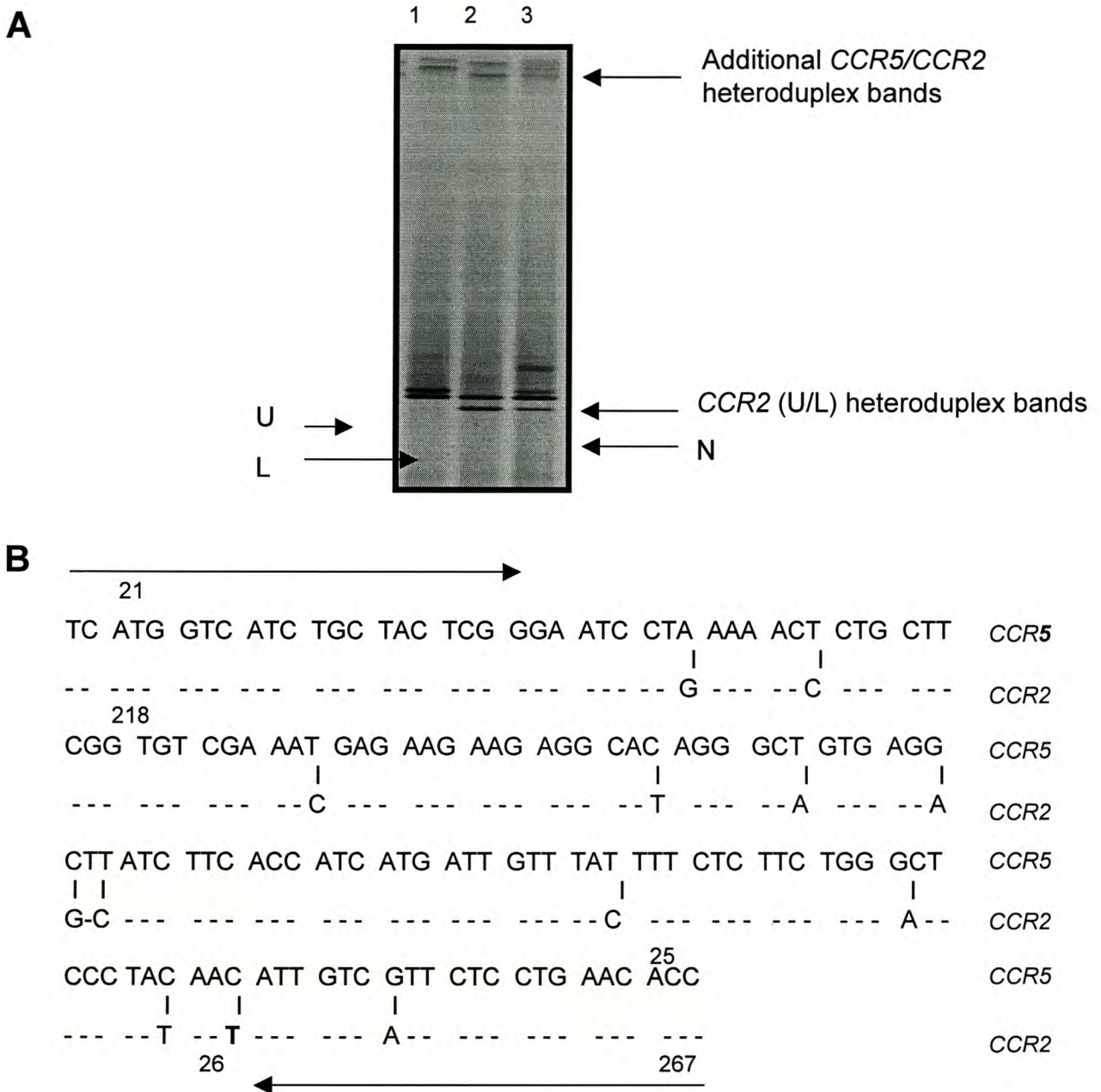


Figure 2. Fragment E of the *CCR5* gene. **A)** DGGE banding pattern of the *CCR5* fragment E, in combination with codons 217 to 267 of the *CCR2* gene. All samples (lanes 1-3) are homozygous normal (N band) for the *CCR5* gene. Samples were either homozygous or heterozygous for the *CCR2* codon 260 polymorphism. Lane 1, homozygous 260-T (U band); lane 2, homozygous 260-C (L band); and lane 3, heterozygous 260-C/T (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 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 $CCR5\Delta32$ 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 $CCR5\Delta32$ 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 $CCR5\Delta32$ 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 *CCR5* Δ 32 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. Petersen¹, Annette Laten¹, Michele D. Zeier²,
Ashraf Grimwood³, J.H. (Bill) Louw⁴, Estrelita Janse van
Rensburg¹ and Vanessa M. Hayes⁵**

Departments of Medical Virology¹, Internal Medicine², Genetics⁴ and Urology⁵,
University of Stellenbosch and Chapel Street Community Health Clinic, Woodstock³,
South Africa.

(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 understudied 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 its 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

(www.ncbi.nlm.nih.gov). 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.

Fragment	Amplimers, 5'-3'	Size (bp)	Temperature (°C)	
			Melting	Annealing
A	[40GC]TGCTTATGTGGTGCCAGACT TGA ACACCAGCGAGTAGAGC	335	72	58
B	[6GC]TGATTA TGATTACGGTGCTCC [40GC]CGATTGTCAGGAGGATGATG	384	72	58
C	[40GC]GCTGTATCACATCGGTTATT [8GC]GCCACAGACATAAACAGAAT	268	73	55
D	[40GC][10AT]TGGCTGTGTTTGCTTCTGT CGAGTAGCAGATGACCATGA	220	70	55
E	[5GC]CCACACA ATA ATGAGGAACA [40GC]TGGTGCTTTCACAGTTACTC	284	73	55
F	ACCTTCCAGGAATTCTTCG [40GC]ACAATCAAAGTCTCCTCGT	346	74	55
G	TGTCTGGATCTGAGCTGGTT [40GC]TCCAAAGCAGAGATCTGTCAT	333	73	58

bp, base pair

GC-clamps used were as follows: [40GC], CGCCCGCCGCGCCCGCGCCCGGCCCGCCGCCCGCCCGCCCG

[8GC], CGCCGCCG

[6GC], CGCCGC

[5GC], CGCCG

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 50µl, 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 Mg²⁺ 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 96⁰C for 3 minutes, followed by 32 cycles of denaturation at 96⁰C for 45 seconds, annealing for 1 minute (annealing temperatures shown in Table 1), and elongation at 72⁰C for 1 minute 20 seconds. The last cycle was followed by an additional extension step of 72⁰C for 7 minutes. For optimal DGGE analysis, amplification was followed by a heteroduplexing step, which involves denaturation at 96⁰C 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 59.5⁰C 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 [Guldborg 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.

Mutation	Base change	DGGE Fragment	Allele frequency						
			Africans		Coloureds		Caucasians		Asians
			HIV + (n = 138)	HIV - (n = 124)	HIV + (n = 52)	HIV - (n = 144)	HIV + (n = 14)	HIV - (n = 4)	HIV - (n = 16)
Int1 -57a/g ^{*,a,b}	A - G	A	14 (0.101)	13 (0.105)	3 (0.058)	10 (0.069)	0	0	0
Int1 -43g/a ^{*,a,b}	G - A	A	3 (0.022)	5 (0.040)	3 (0.058)	3 (0.021)	0	0	0
V52 ^{*,a,b}	GTG - GTT	B	0	0	1 (0.019)	1 (0.007)	0	0	0
V63 ^{*,a,b}	GTC - GTT	B	1 (0.007)	1 (0.008)	5 (0.096)	5 (0.035)	2 (0.143)	0	1 (0.063)
V64I ^{a,b}	GTC - ATC	B	18 (0.130)	21 (0.169)	1 (0.019)	25 (0.174)	0	1 (0.250)	2 (0.125)
S223 ^{*,a,b}	TCG - TCA	E	0	0	0	1 (0.007)	0	0	0
R233Q ^{*,a,b}	CGA - CAA	E	0	0	0	2 (0.007)	0	0	0
N260 ^{a,b}	AAC - AAT	E	75 (0.543)	79 (0.637)	30 (0.577)	97 (0.674)	10 (0.714)	2 (0.500)	11 (0.688)
L283 ^{*,a,b}	CTG - CTT	F	6 (0.043)	3 (0.024)	1 (0.019)	4 (0.028)	0	0	0
T287M ^{*,a,b}	ACG - ATG	F	0	4 (0.032)	1 (0.019)	0	0	0	0
P339 ^{*,a}	CCA - CCG	G	21 (0.152)	18 (0.145)	2 (0.038)	5 (0.035)	0	0	0
T348 ^{*,b}	ACG - ACA	F	0	0	1 (0.019)	6 (0.042)	1 (0.071)	0	0
G355E ^{*,a}	GGA - GAA	G	2 (0.014)	1 (0.008)	0	0	0	0	0

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*

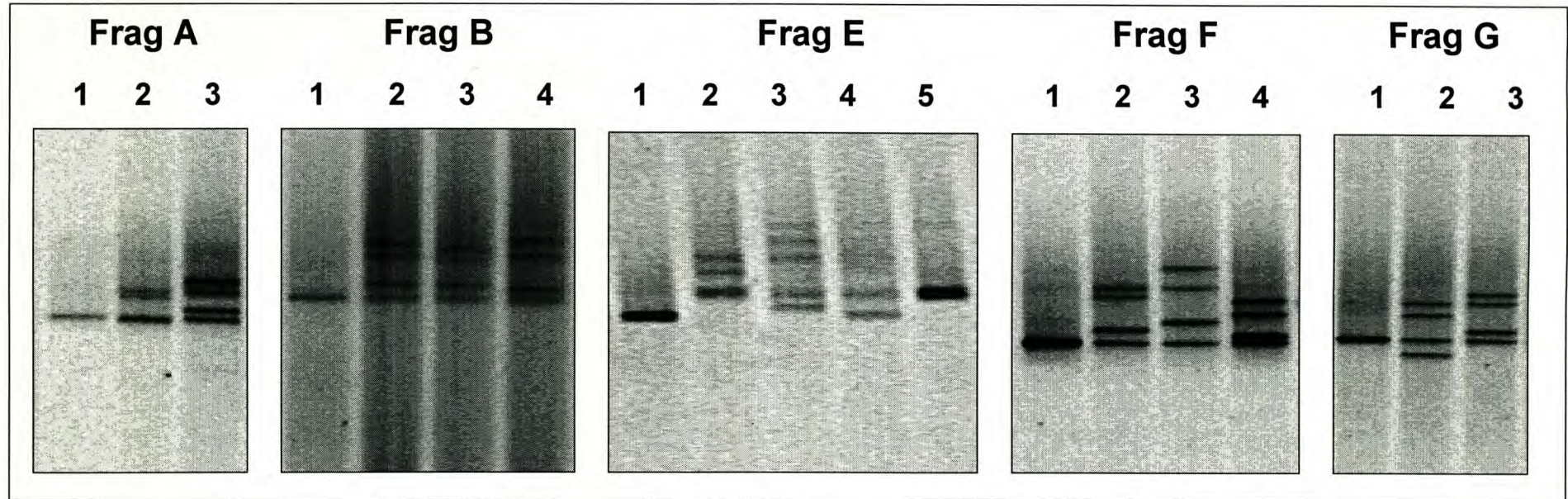


Figure 1. Aberrant DGGE banding patterns for all thirteen *CCR2* gene mutations identified and listed in Table 2, according to the fragment (Frag) in which they were found. No mutations were identified in fragments C and D. Lane 1 of all the fragments represents the DGGE banding pattern of a normal control. 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 *CCR2*V64I SNP; lane 3, heterozygous for the *CCR2*V63 SNP; and lane 4, heterozygous for the *CCR2*V52 mutation; **Frag E:** lane 2, heterozygous for the *CCR2*R233Q mutation; lane 3, heterozygous for the *CCR2*S223 mutation; lane 4, heterozygous for the *CCR2*N260 SNP; and lane 5, homozygous for the *CCR2*N260 SNP; **Frag F:** lane 2, heterozygous for the *CCR2*L283 SNP; lane 3, heterozygous for the *CCR2*T287M SNP; and lane 4, heterozygous for the *CCR2*T348 SNP; **Frag G:** lane 2, heterozygous for the *CCR2*P339 SNP; and lane 3, heterozygous for the *CCR2*G355E 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 its 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 *CCR2*V64I, 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. Petersen¹, Estrelita Janse van Rensburg¹
and Vanessa M. Hayes²**

Departments of Medical Virology¹ and Urology², 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 *SDF1 β* 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 *SDF1 β -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 SDF1 α (89 amino acids) and SDF1 β (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 SDF1 α and SDF1 β proteins form an amino acid-cleaved signal peptide [Bleul *et al.*, 1996b]. Three and four exons contain the coding regions for *SDF1 α* and *SDF1 β* , respectively. The *SDF1 β* 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 *SDF1 β* 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 *SDF1 β -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 β* 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 *SDF1 β* 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***: GTGAAGGCTTCTCTCTGTGG and ***SDF1-3'AR***: [40GC]GTGGACACA CATGATGATGG. 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 μ l 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 96⁰C for 3 minutes, followed by 32 cycles of denaturation at 96⁰C for 45 seconds, annealing at 56⁰C for 1 minute and elongation at 72⁰C for 1 minute 20 seconds. Following the last cycle was an additional extension step of 72⁰C for 7 minutes. For heteroduplex formation, the PCR products were subjected to denaturation at 96⁰C for 10 minutes, followed by renaturation for 45 minutes at 56⁰C. Electrophoresis was used to check the amplified products, where 5 μ l (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 [Guldborg 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 *SDF1* β -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 *SDF1* β mutations detected in 104 HIV seropositive patients and 192 HIV seronegative controls, ordered according to the nucleotide position at which they occurred.

Mutation	Allele frequency							
	Africans		Coloureds		Caucasians		Asians	Unknown
	HIV + (n = 140)	HIV - (n = 126)	HIV + (n = 52)	HIV - (n = 146)	HIV + (n = 14)	HIV - (n = 96)	HIV - (n = 16)	HIV + (n = 2)
nt +801 G-A	7 (0.050)	1 (0.008)	4 (0.076)	27 (0.185)	3 (0.214)	27 (0.281)	2 (0.125)	0
nt +875 A-T *	0	0	0	0	0	0	0	1 (0.500)
Δ 8 (nt +916) *	0	0	0	0	0	1 (0.010)	0	0

* Novel mutation identified in this study

n, number of alleles; nt, nucleotide position; Δ , deletion

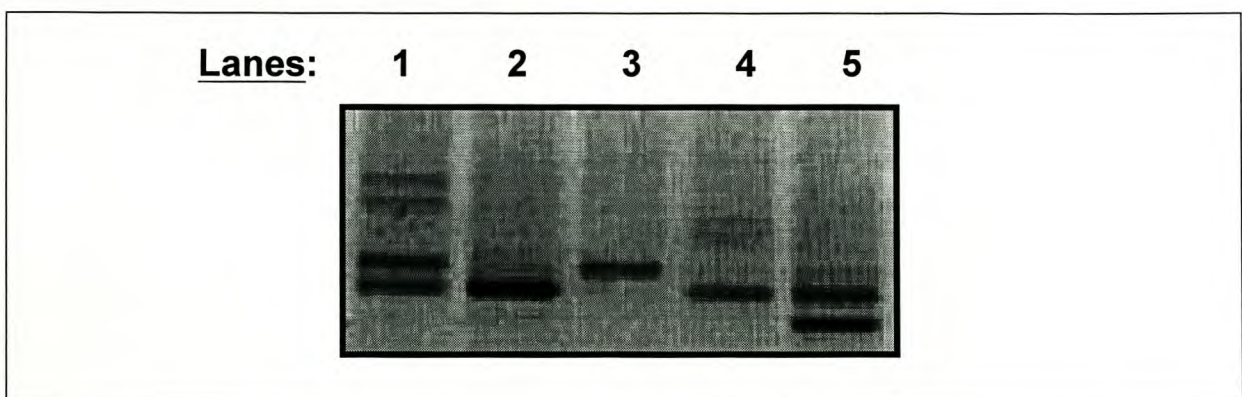


Figure 1. DGGE banding pattern for the *SDF1* β gene mutations detected in this study. Lane 2, normal control; Lane 1 and 3, Individuals heterozygous and homozygous for the *SDF1* β -3' UTR G-A (nt +801) SNP, respectively; Lane 4, Individual heterozygous for the *SDF1* β -3' UTR A - T (nt +875) mutation and Lane 5, Individual heterozygous for the *SDF1* β -3' UTR 8bp deletion (nt +916).

The segment of the 3' UTR of *SDF1 β* 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 *SDF1 β -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 *SDF1 α* and *SDF1 β* 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).

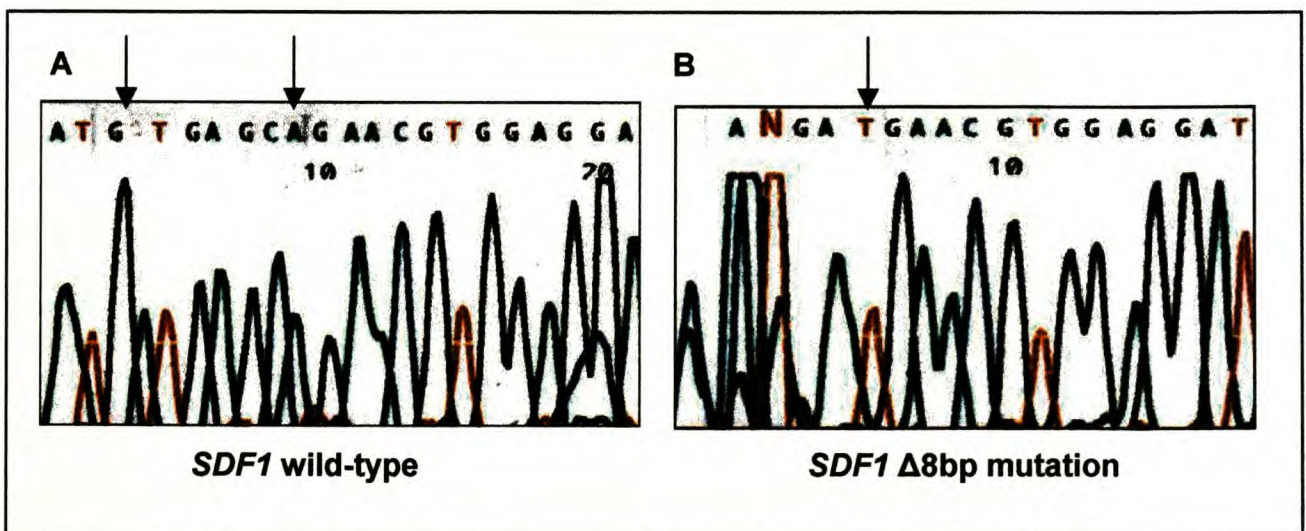


Figure 2. A) Sequence for the *SDF-1* wildtype. The black arrows indicate the 8bp that form the deletion mutation, starting at nucleotide position +961. B) Sequence for the *SDF1* Δ8bp mutation in the homozygous state. The black arrow indicates the start of the deletion mutation.

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 coordination 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 *SDF1 β* 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 *CCR5 Δ 32* 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

- Abrams ES, Murdaugh SE, Lerman LS. **Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC-clamp.** *Genomics* 1990, **7**: 463-475.
- Alkhatib G, Combadiere C, Broder CC *et al.* **CC CKR5: a RANTES, MIP-1 α , MIP-1 β receptor as a fusin cofactor for the macrophage-tropic HIV-1.** *Science* 1996, **272**:1955-1958.
- Alkhatib G, Ahuja SS, Light D, Mummidi S, Berger EA and Ahuja SK. **CC chemokine receptor 5-mediated signalling and HIV-1 co-receptor activity share common structural determinants. Critical residues in the third extracellular loop support HIV-1 fusion.** *J Biol Chem* 1997, **271**:19771-19776.
- Amara A, Gall SL, Schwartz O, *et al.* **HIV coreceptor downregulation as antiviral principle: SDF-1 alpha-dependant internalisation of the chemokine receptor CXCR4 contributes to inhibition of HIV replication.** *J Exp Med* 1997, **186**:139-146.
- Ansari-Lari MA, Liu X, Metzker ML, Rut AR, Gibbs RA. **The extent of genetic variation in the CCR5 gene.** *Nat Genet* 1997, **16**:221-222.
- Anzala AO, Ball TB, Rostron T, O'Brien SJ, Plummer FA, Rowland-Jones SL. **CCR2-64I allele and genotype association with delayed AIDS progression on African women.** *Lancet* 1998, **351**:1632-1633.
- Arnaout RA, Lloyd AL, O'Brien TR, Goedert JJ, Leonard JM, Nowak MA. **A simple relationship between viral load and survival time in HIV-1 infection.** *Proc Natl Acad Sci USA* 1999, **96**:11549-11553.

- Arya SK, Ginsberg CC, Davis-Warren A, D' Costa J. **In vitro phenotype of SDF1 gene mutant that delays the onset of human immunodeficiency virus disease in vivo.** *J Hum Virol* 1999, **2**:133-138.
- Atchison RE, RE, Gosling J, Monteclaro FS *et al.* **Multiple extracellular elements of CCR5 and HIV-1 entry: dissociation from response to chemokines.** *Science* 1996, **274**:1924-1926.
- Baggiolini M, Dewald B, Moser B. **Interleukin-8 and related chemotactic cytokines – CXC and CC chemokines.** *Adv Immunol* 1994, **55**:97-179.
- Baggiolini M, Dewald B, Moser B. **Human chemokines: An Update.** *Annu Rev Immunol* 1997, **15**:675-705.
- Baggiolini M. **Reflections on chemokines.** *Immunol Rev* 2000, **177**:5-7.
- Balotta C, Bagnarelli P, Violin M, *et al.* **Homozygous Δ 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1 infected patient.** *AIDS* 1997, **11**:F67-F71.
- Balotta C, Bagnarelli P, Corvasce S *et al.* **Identification of two subsets of long-term nonprogressors with divergent viral activity by stromal-derived factor 1 chemokine gene polymorphism analysis.** *J Infect Dis* 1999, **180**:285-289.
- Barcellini W, Rizzardi GP, Velati C *et al.* **In vitro production of type 1 and type 2 cytokines by peripheral blood mononuclear cells from high-risk HIV-negative intravenous drug users.** *AIDS* 1995, **9**:691-694.
- Barre-Sinoussi F, Chermann JC, Rey F, *et al.* **Isolation of the T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).** *Science* 1983, **220**:868-871.
- Bazan JF, Bacon KB, Hardiman G, *et al.* **A new class of membrane-bound chemokine with a CX₃C motif.** *Nature* 1997, **385**:640-644.

- Beretta A, Furci L, Burastero S, *et al.* **HIV-1-specific immunity in persistently seronegative individuals at high risk for HIV infection.** *Immunol Lett* 1996, **51**:39-43.
- Berger EA, Doms RW, Fenyo EM, *et al.* **A new classification for HIV-1 [Letter].** *Nature* 1998, **391**:240.
- Berger EA, Murphy PM, Farber JM. **Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism and disease.** *Annu Rev immunol* 1999, **17**:657-700.
- Bernard NF, Yannakis CM, Lee JS, Tsoukas CM. **Human immunodeficiency virus (HIV)-specific cytotoxic T lymphocyte activity in HIV-exposed seronegative persons.** *J Infect Dis*, 1999, **179**:538-547.
- 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.
- Biti R, French R, Young J, Bennets B, Stewart G, Liang T. **HIV-1 infection in an individual homozygous for the CCR5 deletion allele.** *Nat Med* 1997, **3**:252-253.
- Blanpain C, Doranz BJ, Vakili J, *et al.* **Multiple charged and aromatic residues in CCR5 amino-terminal domain are involved in high affinity binding of both chemokines and HIV-1 env protein.** *J Biol Chem* 1999a, **274**:34719-34727.
- Blanpain C, Lee B, Vakili J, *et al.* **Extracellular cysteines of CCR5 are required for chemokine binding, but dispensable for HIV-1 co-receptor activity.** *J Biol Chem* 1999b, **274**:18902-18908.
- Blanpain C, Lee B, Tackoen M, *et al.* **Multiple non-functional alleles of CCR5 are frequent in various human populations.** *Blood* 2000, **96**:1638-1645.

- Bleul CC, Farzan M, Choe H, *et al.* **The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusion and blocks HIV-1 entry.** *Nature* 1996a, **382**:829-833.
- Bleul CC, Fuhlbrigge RC, Casasnovas JM, Aiuti A, Springer TA. **A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1)** *J Exp Med* 1996b, **184**:799-802.
- Brambilla A, Villa C, Rizzardi G *et al.* **Shorter survival of SDF1-3' A/3 'A homozygotes linked to CD4+ T cell decrease in advanced human immunodeficiency virus type 1 infection.** *J Infect Dis* 2000, **182**: 311-315.
- Bredell H, Williamson C, Sonneberg P, Martin DJ, Morris L. **Genetic characterisation of HIV type 1 from migrant workers in three South African gold mines.** *AIDS Res Hum Retroviruses* 1998, **14**:1015-1018.
- Broxmeyer HE, Sherry B, Lu L *et al.* **Myelopoietic enhancing effects of murine macrophage inflammatory proteins 1 and 2 on colony formation in vitro by murine and human bone marrow granulocyte/macrophage progenitor cells.** *J Exp Med* 1989, **170**:1583-1594.
- Cao Y, Qin L, Zhang L, Safrit J, Ho DD. **Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection.** *N Eng J Med* 1995, **332**:201-208.
- Carrington M, Kissner T, Gerrard B, Ivanov S, O'Brien SJ, Dean M. **Novel alleles of the chemokine-receptor gene CCR5.** *Am J Hum Genet* 1997, **61**:1261-1267.
- Carrington M, Nelson GW, Martin MP, *et al.* **HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage.** *Science* 1999a, **283**:1748-1752.

- Carrington M, Dean M, Martin MP, O'Brien SJ. **Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences.** *Hum Molec Genet* 1999b, **8**:1939-1945.
- Centers for Disease Control and Prevention (www.cdc.org).
- Charo IF, Myers SJ, Herman A, Franci C, Connolly AJ, Coughlin SR. **Molecular cloning and functional expression of two monocyte chemoattractant protein 1 receptors reveals alternative splicing of the carboxyl-terminal tails.** *Proc Natl Acad Sci USA* 1994, **91**:2752-2756.
- Chesebro B, Nishio J, Perryman S *et al.* **Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, macrophages.** *J Virol* 1991, **65**:5782-5789.
- Choe H, Farzan M, Sun Y, *et al.* **The β -chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates.** *Cell* 1996, **85**:1135-1148.
- Clark VJ, Metheny N, Dean M, Peterson RJ. **Statistical estimation and pedigree analysis of CCR2-CCR5 haplotypes.** *Hum Genet* 2001, **108**:484-493.
- Clark-Lewis I, Kim KS, Rajarathnam K *et al.* **Structure-activity relationships of chemokines.** *J Leukoc Biol* 1995, **57**:703-711.
- Clavel F, Mansinho K, Chamaret S, *et al.* **Human immunodeficiency virus type 2 infection associated with AIDS in West Africa.** *N Engl J Med* 1987, **316**:1180-1185.
- Clerici M, Giorgi JV, Chou CC, *et al.* **Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative homosexual men with recent sexual exposure to HIV-1.** *J Infect Dis* 1992, **165**:1012-1019.

- Clerici M, Levin JM, Kessler HA, *et al.* **HIV-specific T-helper activity in seronegative health care workers exposed to contaminated blood.** *JAMA* 1994, **271**:42-46.
- Cocchi F, De Vico A, Garzino-Demo A, Arya S, Gallo R, Lusso P. **Identification of RANTES, MIP-1 α , and MIP- β as the major HIV suppressive factors produced by CD8+ T cells.** *Science* 1995, **270**:1811-1815.
- Cohen J. **South Africa's New Enemy.** *Science* 2000a, **288**:2168-2170.
- Cohen J. **Is AIDS in Africa a Distinct Disease?** *Science* 2000b, **288**:2153-2155.
- 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.
- Connor RI, Sheridan KE, Lai C, Zhang L, Ho DD. **Characterization of the functional properties of env genes from long-term survivors of human immunodeficiency virus type 1 infection.** *J Virol* 1996, **70**:5306-5311.
- Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. **Change in coreceptor use correlates with disease progression in HIV-1 infected individuals.** *J Exp Med* 1997, **185**:621-628.
- 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.

- Cotton RGH, Edkins E, Forrest S. **Mutation detection: a practical approach.** Oxford University Press, Oxford 1998.
- Craps L, Basle MD. **Immunology a primer.** Sandoz Pharma Ltd, Basle 1993.
- Crump MP, Gong J-H, Loetscher P *et al.* **Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1.** *EMBO J* 1997, **16**:6996-7007.
- Daar ES. **Virology and immunology of acute HIV type 1 infection.** *AIDS Res Hum Retroviruses* 1998, **14**:S229-S234.
- Dalgeish AG, Beverley PCL, Clapham PR, Crawford DH, Greaves MF, Weiss RA. **The CD4 (T4) antigen is an essential of the receptor of the AIDS retrovirus.** *Nature* 1984, **312**:763-767.
- Daugherty BL, Siciliano SJ, De Martino JA, Malkowitz L, Sirofina A, Springer MS. **Cloning, expression, and characterization of the human eosinophil eotaxin receptor.** *J Exp Med* 1996, **183**:2349-2354.
- Daugherty BL and Springer MS. **The β -chemokine receptor genes CCR1 (CMKBR1), CCR2 (CMKBR2), and CCR3 (CMBRK3) cluster within 285kb on human chromosome 3p21.** *Genomics* 1997, **41**: 294-297.
- Dean M, Carrington M, Winkler C, *et al.* **Genetic restriction of HIV-1 infection by a deletion allele of the *CCR5* structural gene.** *Science* 1996, **273**:1856-1862.
- Deacon NJ, Tsykin A, Solomon A, *et al.* **Genomic structure of an attenuated quasispecies of HIV-1 from a blood transfusion donor and recipients.** *Science* 1995, **270**:988-991.

- Delanghe JR, Langlois MR, Boelaert JR, *et al.* **Haptoglobin polymorphism, iron metabolism and mortality in HIV-infection.** *AIDS* 1998, **12**:1027-1032.
- De Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J. **Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: Analysis of single amino acid substitution.** *J Virol* 1992, **66**:6777-6680.
- De Maria A, Cirillo C, Moretta L. **Occurrence of human immunodeficiency virus type 1 (HIV-1)-specific cytolytic T cell activity in apparently uninfected children born to HIV-1 infected mothers.** *J Infect Dis* 1994, **170**:1296-1299.
- Deng H, Liu R, Ellmeier W, *et al.* **Identification of a major co-receptor for primary isolates of HIV-1.** *Nature* 1996, **381**:661-666.
- Department of Health/HIV/AIDS Directorate. **Release of National HIV Prevalence Survey at Public Sector Antenatal Clinics.** South Africa, 2001.
- Doranz BJ, Rucker J, Yi Y, *et al.* **A dual-tropic primary HIV-1 isolate that uses fusin and the β -chemokine receptors CKR-5, CKR-3, and CKR-2 as fusin cofactors.** *Cell* 1996, **85**:1149-1158.
- Doranz BJ, Lu ZH, Rucker J, *et al.* **Two distinct CCR5 domains can mediate coreceptor usage by human immunodeficiency virus type 1.** *J Virol* 1997, **71**:6305-6314.
- Dragic T, Litwin V, Allaway GP, *et al.* **HIV-1 entry into CD4⁺ cells is mediated by the chemokine receptor CC-CKR-5.** *Nature* 1996, **381**:667-673.
- Dragic T, Trkola A, Lin SW, *et al.* **Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry.** *J Virol* 1998, **72**: 279-285.

- D'Souza I, Poorkaj P, Hong M, *et al.* **Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements.** *Proc Natl Acad Sci USA* 1999, **90**:5598-5603.
- Earl PL, Moss B, Doms RW. **Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type1 envelope protein.** *J Virol* 1991, **65**:2047-2055.
- Ellis TP, Humphrey KE, Smith MJ, Cotton RG. **Chemical cleavage of mismatch: a new look at an established method.** *Hum Mutat* 1998, **11**:345-353.
- Engelbrecht S, Laten JD, Smith T-L, van Rensburg EJ. **Identification of env subtypes in fourteen HIV type 1 isolates from South Africa.** *AIDS Res Hum Retroviruses* 1995, **11**:1296-1271.
- Engelbrecht S, Koulinska I, Smith T-L, Bareto J, van Rensburg EJ. **Variation in the HIV type V3 region env sequences from Mozambique.** *AIDS Res Hum Retroviruses* 1998, **14**:803-805.
- Engelbrecht S, Smith TL, Kasper P, *et al.* **HIV type 1 V3 domain serotyping and genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape provinces of South Africa.** *AIDS Res Hum Retroviruses* 1999, **15**:325-328.
- Esparza J, Bhamarapavati N. **Accelerating the development and future availability of HIV-1 vaccines: Why, when, where, and how?** *Lancet* 2000, **355**:2061-2066.
- 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**

disease progression in a Danish cohort of HIV-infected individuals. Copenhagen AIDS cohort. *J Acquir Immun Defic Syndr Hum Retrovirol* 1998, 18:110-116.

- Farzan M, Choe H, Vaca L, *et al.* **A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5. *J Virol* 1998, 72:1160-1164.**
- Feng Y, Broder CC, Kennedy PE, Berger E. **HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996, 272:827-877.**
- Frade JMR, Llorente M, Mellado M, *et al.* **The amino-terminal domain of the CCR2 chemokine receptor acts as a coreceptor for HIV-1 infection. *J Clin Invest* 1997, 100:497-502**
- Fischer SG, Lerman LS. **DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc Natl Acad Sci USA* 1983, 80: 1579-1583.**
- Forster R, Emrich T, Kremmer E, Lipp M. **Expression of the G-protein-coupled receptor BLR1 defines mature, recirculating B cells and a subset of T-helper memory cells. *Blood* 1994, 84:830-840.**
- Fouchier RA, Groenink M, Koostra NA, Tersmette M, Huisman HG, Miedema F, Schuitemaker H. **Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* 1992, 66:3183-3187.**

- Fowke KR, Nagelkerke NJ, Kimani J, *et al.* **Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya.** *Lancet* 1996, **348**:1347-1351.
- Gallo RC, Sarin PS, Gelmann EP, *et al.* **Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).** *Science* 1983; **220**:865-867.
- Garred P, Madsen HO, Balslev U, *et al.* **Susceptibility to HIV infection and progression to AIDS in relation to variant alleles of mannose-binding lectin.** *Lancet* 1997, **349**:236-240.
- Gelmann EP, Popovic M, Blayney D *et al.* **Proviral DNA of a retrovirus, human T-cell leukemia virus, in two patients with AIDS.** *Science* 1983, **220**:862-865.
- Genbank database (www.ncbi.nlm.nih.gov).
- Ginot F. **Oligonucleotide micro-arrays for identification of unknown mutations: how far from reality?** *Hum Mutat* 1997, **10**:1-10.
- Goh WC, Markee J, Akridge RE, *et al.* **Protection against human immunodeficiency virus type 1 infection in persons with repeated exposure: evidence for T cell immunity in the absence of inherited CCR5 coreceptor defects.** *J Infect Dis* 1999, **179**:548-557.
- Grant AD, Djomand G, De Cock KM. **Natural history and spectrum of disease in adults with HIV/AIDS in Africa.** *AIDS* 1997, **11**:S43-S54.
- Gray IC, Campbell DA, Spurr NK. **Single nucleotide polymorphisms as tools in human genetics.** *Hum Molec Genet* 2000, **9**:2403-2408.
- Gulberg P, Guttler F. **A simple method for identification of point mutations using denaturing gradient gel electrophoresis.** *Nucleic Acids Res* 1993, **21**:2261-2262.

- Han KI, Green SR, Tangirala RK, Tanaka S, Quehenberger O. **Role of the first extracellular loop in the functional activation of CCR2.** *J Biol Chem* 1999, **274**:32055-32062.
- 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.
- Haynes BF, Pantaleo G, Fauci AS. **Toward an understanding of the correlates of protective immunity to HIV infection.** *Science* 1996, **271**:324-328.
- Hendel H, Henon N, Lebuanec H, *et al.* **Distinctive effects of CCR5, CCR2, and SDF1 genetic polymorphisms in AIDS progression.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1998, **19**:381-386.
- Hendel H, Caillat-Zucman S, Lebuanec H *et al.* **New class I and II HLA alleles strongly associated with opposite patterns of disease progression to AIDS.** *J Immunol* 1999, **162**:6942-6946.
- Hill AVS. **The immunogenetics of human infectious diseases.** *Annu Rev Immunol* 1998, **16**:593-617.
- Horuk R. **Molecular properties of the chemokine receptor family.** *Trends Pharmacol Sci* 1994, **15**:159-165.
- Hogan CM and Hammer SM. **Host determinants in HIV infection and disease, Part 2: Genetic factors and implications for antiretroviral therapeutics.** *Ann Intern Med* 2001, **134**: 978-996.

- Howard OMZ, Shirakawa A-K, Turpin JA, *et al.* **Naturally occurring CCR5 extracellular and transmembrane domain variants affect HIV-1 co-receptor and ligand binding function.** *J Biol Chem* 1999, **274**:16228-16234.
- Huang Y, Paxton WA, Wolinsky SM, *et al.* **The role of a mutant CCR5 allele in HIV-1 transmission and disease progression.** *Nat Med* 1996, **2**:1240-1243.
- Huang Y, Zhang L, Ho DD. **Characterization of gag and pol sequences from long-term survivors of human immunodeficiency virus type 1 infection.** *Virology* 1998, **240**:36-49.
- Huttley GA, Smith MW, Carrington M and O'Brien SJ. **A scan for linkage disequilibrium across the human genome.** *Genetics* 1999, **152**:1711-1722.
- Hwang SS, Boyle TJ, Lyerly, HK, Cullen BR. **Identification of the envelope V3 loop as the primary determinant of cell tropism.** *Science* 1991, **253**:71-74.
- Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. **Identification of envelope V3 loop as a major determinant of CD4 neutralization sensitivity of HIV-1.** *Science* 1992, **257**:535-537.
- Ioannidis JP, O'Brien TR, Rosenberg PS, Contopoulos-Ioannidis DG, Goedert JJ. **Genetic effects on HIV disease progression [letter].** *Nat Med* 1998, **4**:536.
- Iversen AK, Shpaer EG, Rodrgio AG. **Persistence of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual.** *J Virol* 1995, **69**:5743-5753.
- John GC, Rousseau C, Dong T *et al.* **Maternal SDF1 3' A polymorphism is associated with increased perinatal human immunodeficiency virus type 1 transmission.** *J Virol* 2000, **74**:5736-5739.
- Kalinkovich A, Weisman Z, Bentwich Z. **Chemokine and chemokine receptors: role in HIV infection.** *Immunol Lett* 1999, **68**:281-287.

- Kanki PJ, De Cock KM. Epidemiology and natural history of HIV-2. *AIDS* 1994, **8**:S85-S93.
- Kantor R, Gershoni JM. **Distribution of the CCR5 gene 32-base pair deletion in Israeli ethnic groups.** *J Acquir Immune Defic Syndr* 1999, **20**:81-84.
- Kelner GS, Kennedy J, Bacon KB *et al.* **Lymphotactin: A cytokine that represents a new class of chemokine.** *Science* 1994, **266**:1395-1399.
- Kirchhoff F, Greenough TC, Brettler DB, Sullivan JL, Desrosiers RC. **Absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection.** *N Engl J Med* 1995, **332**:228-232.
- Klatzman D, Champagne E, Chamaret S, *et al.* **T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV.** *Nature* 1984, **312**:767-768.
- Kostrikis LG, Huang Y, Moore JP, *et al.* **A chemokine receptor CCR2 allele delays HIV-1 disease progression and is associated with a CCR5 promoter mutation.** *Nat Med* 1998, **4**: 350-355.
- Kuhmann SE, Platt EJ, Kozak SL, Kabat D. **Polymorphisms in the CCR5 genes of African monkeys and mice implicate specific amino acids in infections by simian and human immunodeficiency viruses.** *J Virol* 1997, **71**:8642-8656.
- Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and neutralizing human antibody.** *Nature* 1998, **393**:648-659.
- Lander ES and Schork NJ. **Genetic dissection of complex traits.** *Science* 1994, **265**:2037-2047.

- Langlade-Demoyen P, Ngo-Giang-Huong N, Ferchal F, Oksenhendler E. **Human immunodeficiency virus (HIV) nef-specific cytotoxic T lymphocytes in noninfected heterosexual contact of HIV-infected patients.** *J Clin Invest* 1994, **93**:1293-1297.
- Lapham CK, Ouyang J, Chandrasekhar B, Nguyen NY, Dimitrov DS, Golding H. **Evidence for cell-surface association between fusin and the CD4-gp12 complex in human cell lines.** *Nature* 1996, **274**:602-605.
- Lasky LA, Nakamura G, Smith DH, *et al.* **Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor.** *Cell* 1987, **50**:975-985.
- Learmont JC, Geczy AF, Mills J, *et al.* **Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1. A report from the Sydney Blood Bank Cohort.** *N Engl J Med* 1999, **340**:1715-1722.
- Lee B, Doranz BJ, Rana S, *et al.* **Influence of the CCR2-V64I polymorphism on the human immunodeficiency virus type 1 coreceptor activity and on the chemokine receptor function of CCR2b, CCR3, CCR5 and CXCR4.** *J Virol* 1998, **72**: 7450-7458.
- Lee B, Montaner LJ. **Chemokine immunobiology in HIV-1 pathogenesis.** *J Leukoc Biol* 1999, **65**:552-565.
- Lerman LS, Silverstein K. **Computational simulation of the DNA melting and its application to denaturing gel electrophoresis.** *Meth Enzymol* 1987, **155**:482-501.
- Liao F, Rabin RL, Yanneili JT, Leonidas GK, Vanguri P, Farber JM. **Human mig chemokine: biochemical and functional characterization.** *J Exp Med* 1995, **182**:1301-1314.

- Libert F, Cochaux P, Beckman G, *et al.* **The Delta ccr5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe.** *Hum Mol Genet* 1998, **7**:399-406.
- Lifson AR, Buchbinder SP, Sheppard HW, *et al.* **Long-term human immunodeficiency virus infection in asymptomatic homosexual and bisexual men with normal CD4+ lymphocyte counts: immunologic and virologic characteristics.** *J Infect Dis* 1991, **163**:959-965.
- Lindley IJD, Westwick J, Kunkel SL. **Nomenclature announcement – the chemokines.** *Immunol Today* 1993, **14**:24
- Lipschutz RJ, Morris D, Chee M *et al.* **Using oligonucleotide probe arrays to access genetic diversity.** *Biotechniques* 1995, **19**:442-447.
- Liu R, Paxton WA, Choe S *et al.* **Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply exposed individuals to HIV-1 infection.** *Cell* 1996, **86**:367-377.
- Liu SL, Schacker T, Musey L, *et al.* **Divergent patterns of progression to AIDS after infection from the same source: Human immunodeficiency virus type 1 evolution and antiviral responses.** *J Virol* 1997, **71**:4284-4295.
- Liu W, Smith DI, Rechtzigel KJ, Thibodeau SN, James CD. **Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations.** *Nucleic Acids Res* 1998, **26**:1396-1400.
- Loetscher P, Moser B, Baggiolini M. **Chemokines and their receptors in lymphocyte traffic and HIV infection.** *Adv Immunol* 2000, **74**:127-180.
- 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).
- Lucotte G, Mercier G. **Distribution of the CCR5 gene 32bp deletion in Europe.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1998, **19**:174-177.
- Lucotte G. **Distribution of the CCR5 gene 32-basepair deletion in West Europe. A hypothesis about the possible dispersion of the mutation by the Vikings in historical times.** *Hum Immunol* 2001, **62**:933-936.
- Luo J, Luo Z, Zhou N, Hall JW, Huang Z. **Attachment of C-terminus of SDF-1 enhances the biological activity of its N-terminal peptide.** *Biochem Biophys Res Commun* 1999, **264**:42-47.
- Maddon PJ, Dalglish AG, McDougal JS, Clapham PR, Weiss RA, Axel R. **The T4 encodes the AIDS virus receptor and is expressed in the immune system and the brain.** *Cell* 1986, **47**:333-348.
- Mader SS. **Biology.** Fifth edition. Times Mirror Higher Education Group Inc, Dubuque 1996.
- Magierowska M, Theodorou I, Debre P, *et al.* **Combined genotypes of CCR5, CCR2, SDF1, and HLA genes can predict the long-term-nonprogressor status in human immunodeficiency virus-1-infected individuals.** *Blood* 1999, **93**:936-941.
- Mariani R, Wong S, Mulder LCF, *et al.* **CCR2-64I polymorphism is not associated with altered CCR5 expression of coreceptor function.** *J Virol* 1999, **73**: 2450-2459.
- Marlink RG, Ricard B, M' Boup S *et al.* **Clinical, hematologic and immunologic cross-sectional evaluation of individuals exposed to human immunodeficiency virus type 2 (HIV-2).** *AIDS Res Hum Retroviruses* 1988, **4**: 137-148.

- Marlink R, Kanki P, Thior I, *et al.* **Reduced rate of disease development after HIV-2 infection as compared to HIV-1.** *Science* 1994, **265**:1587-1590.
- Markowitz DM. **Infection with the human immunodeficiency virus type 2.** *Ann Intern Med* 1993, **118**:211-218.
- Marquet S, Sanchez FO, Arias M, *et al.* **Variants of the human NRAMP1 gene and altered human immunodeficiency virus infection susceptibility.** *J Infect Dis* 1999, **180**:1521-1525.
- Martin MP, Dean M, Smith MW, *et al.* **Genetic acceleration of AIDS progression by a promoter variant of CCR5.** *Science* 1998, **282**:1907-1911.
- Martinson JJ, Chapman NH, Rees DC, Liu YT, Clegg JB. **Global distribution of the CCR5 gene 32-basepair deletion.** *Nat Genet* 1997, **16**:100-103.
- Martinson JJ, Hong L, Karanickolas R, Moore JP, Kostrikis LG. **Global distribution of the CCR2-64I/CCR5-59653T HIV-1 disease-protective haplotype.** *AIDS* 2000, **14**:483-489.
- Maxam AM, Gilbert W. **A new method for sequencing DNA.** *Proc Natl Acad Sci* 1977, **74**:560-564.
- Mazzoli S, Trabattoni D, Lo Caputo S, *et al.* **HIV-specific mucosal and cellular immunity in HIV-seronegative partners of HIV-seropositive individuals.** *Nat Med* 1997, **3**:1250-1257.
- Merigan TC, Hirsch RL, Fischer AC, Meyerson LA, Goldstein G, Winters MA. **The prognostic significance of serum viral load, codon 215 reverse transcriptase mutation and CD4+ T cells on progression of HIV disease in double-blind study of thymopentin.** *AIDS* 1996, **10**:159-165.

- Meyer L, Magierowska M, Hubert JB, *et al.* **CC-chemokine receptor variants, SDF-1 polymorphism, and disease progression in 720 HIV-infected patients. SEROCO cohort [letter].** *AIDS* 1999, **13**:624-626.
- McDermott DH, Zimmerman PA, Guignard F, Kleeberger CA, Leitman SF, Murphy PM. **CCR5 promoter polymorphism and HIV-1 disease progression.** *Lancet* 1998, **352**:866-870.
- Michael NL, Chang G, D'arcy LA, *et al.* **Defective accessory genes in a human immunodeficiency type 1-infected long-term survivor lacking recoverable virus.** *J Virol* 1995, **69**: 4228-4236.
- Michael NL, Louie LG, Rohrbaugh AL, *et al.* **The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and disease progression.** *Nat Med* 1997, **3**:1160-1162.
- Miller MD, Krangel MS. **Biology and biochemistry of the chemokines: a family of chemotactic and inflammatory cytokines.** *CRC Crit Rev Immunol* 1992, **12**:17-46.
- Monteclaro FS, Charo IF. **The amino-terminal extracellular domain of the MCP-1 receptor, but not the RANTES/MIP-1 α receptor, confers chemokine selectivity.** *J Biol Chem* 1996, **271**:19084-19092.
- 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.
- Moodley D, Smith T-L, van Rensburg EJ, Moodley J and Engelbrecht S. **HIV type 1 V3 region subtyping in Kwazulu Natal, a high-seroprevalence South African region.** *AIDS Res Hum Retroviruses* 1998, **14**:1015-1018.

- Moore J, Jameson B, Weiss R, Sattentau Q. **The HIV-cell fusion reaction.** *In Viral Fusion Mechanisms* (J. Bentz, ed.), Boca Raton, FL: CRC Press 1993, 233-289.
- *Morbidity and Mortality Weekly Report* 1981a, **30:250-252. Pneumocystis pneumonia – Los Angeles.**
- *Morbidity and Mortality Weekly Report* 1981b, **30:305-308. Kaposi's sarcoma and pneumocystis pneumonia among homosexual men – New York City and California.**
- Mullis KB, Faloona FA. **Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.** *Methods Enzymol* 1987, **155:335-350.**
- Mummidi S, Ahuja SS, McDaniel BL, Ahuja SK. **The human CC chemokine receptor 5 (CCR5) gene.** *J Biol Chem* 1997, **272:30662-30671.**
- Mummidi S, Ahuja SS, Gonzalez SA, *et al.* **Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression.** *Nat Med* 1998, **4:786-793.**
- Munoz A, Kirby AJ, He YD, *et al.* **Long-term survivors with HIV-1 infection: incubation period and longitudinal patterns of CD4+ lymphocytes.** *J Acquir Immune Defic Syndr Hum Retrovirol* 1995, **8:496-505.**
- Murphy PM. **The molecular biology of leucocyte chemoattractant receptors.** *Annu Rev Immunol* 1994, **12:593-633.**
- Myers RM, Fischer SG, Lerman LS, Maniatis T. **Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis.** *Nucleic Acids Res* 1985a, **13:3131-3145.**

- Myers RM, Lumelsky N, Lerman LS, Maniatis T. **Detection of single base substitutions in total genomic DNA.** *Nature* 1985b, **313**:495-497.
- Nagasawa T, Kikutani H, Kishimoto T. **Molecular cloning and structure of a pre-B-cell growth-stimulating factor.** *Proc Natl Acad Sci USA* 1994, **91**:2305-2309.
- Nagasawa T, Hirota S, Tachibana K, *et al.* **Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF1.** *Nature* 1996, **382**:635-638.
- Nakayama EE, Hoshino Y, Xin X, *et al.* **Polymorphism in the interleukin-4 promoter affects acquisition of human immunodeficiency virus type 1 syncytium-inducing phenotype.** *J Virol* 2000, **74**:5452-5459.
- Oberlin E, Amara A, Bachelier F, *et al.* **The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1.** *Nature* 1996, **382**:833-835.
- O'Brien TR, Dean M. **In search of AIDS-resistance genes.** *Sci Am* 1997a, **277**:44-51.
- O'Brien TR, Winkler C, Dean M, *et al.* **HIV-infection in a man homozygous for CCR5-Δ32.** *Lancet* 1997b; **349**:1219.
- O'Brien SJ and Moore JP. **The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS.** *Immunol Rev* 2000, **177**: 99-111.
- Oppenheim JJ, Zachariae COC, Mukaida N, Matsushima, K. **Properties of the novel proinflammatory supergene "intercrine" cytokine family.** *Annu Rev Immunol* 1991, **9**:617-648.

- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. **Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms.** *Proc Natl Acad Sci USA* 1989, **86**:2766-2770.
- Pan Y, Lloyd C, Zhou H, *et al.* **Neurotactin, a membrane-anchored chemokine upregulated in brain inflammation.** *Nature* 1997, **387**:611-617.
- Pantaleo G, Menzo S, Vaccarezza M *et al.* **Studies in subjects with long-term nonprogressive human immunodeficiency virus infection.** *N Engl J Med* 1995, **332**:209-216.
- Parry J, Shamsheer M, Skibinski D. **Restriction site mutation analysis, a proposed methodology for the detection and study of DNA base changes following mutagen exposure.** *Mutagenesis* 1990, **5**:209-212.
- Paxton WA, Martin SR, Tse D, *et al.* **Relative resistance to HIV-1 infection of CD4 lymphocytes from persons who remain uninfected despite multiple high-risk sexual exposure.** *Nat Med* 1996, **2**:412-417.
- Paxton WA, Liu R, Kang S, *et al.* **Reduced HIV-1 infectibility of CD4+ lymphocytes from exposed-uninfected individuals: association with low expression of CCR5 and high production of β -chemokines.** *Virology* 1998, **244**:66-73.
- Peakman M, Vergani D. **Basic and clinical immunology.** Pearson Professional Limited, London 1997.
- Petersen DC, Kotze MJ, Zeier MD, *et al.* **Novel mutations identified using a comprehensive CCR5-denaturing gradient gel electrophoresis assay.** *AIDS* 2001, **15**:171-177.

- Peterson AC, Di Rienzo A, Lehesjoki A-E, de la Chapelle A Slatkin M, Freimer NB. **The distribution of linkage disequilibrium over anonymous genome regions.** *Hum Molec Genet* 1995, **4**:887-894.
- Phair J, Jacobson L, Detels R *et al.* **Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: the Multicenter AIDS Cohort Study.** *J Acquir Immune Defic Syndr* 1992, **5**:490-496.
- Phair JP. **Determinants of the natural history of human immunodeficiency virus type 1 infection.** *J Infect Dis* 1999, **179**:S384-S386.
- Pinto LA, Sullivan J, Berzofsky JA, *et al.* **ENV-specific cytotoxic T lymphocyte responses to HIV seronegative health care workers occupationally exposed to HIV-contaminated body fluids.** *J Clin Invest* 1995, **96**:867-876.
- Picard L, Simmons G, Power CA, Meyer A, Weiss RA, Clapham PR. **Multiple extracellular domains of CCR-5 contribute to human immunodeficiency type 1 entry and fusion.** *J Virol* 1997, **71**:5003-5011.
- Playfair JHL, Chain BM. **Immunology at a glance.** Seventh Edition. Blackwell Science Ltd, Great Britain 2001.
- Plummer FA, Ball TB, Kimani J, Fowke KR. **Resistance to HIV-1 infection among highly exposed sex workers in Nairobi: what mediates infection and why does it develop?** *Immunol Lett* 1999, **66**:27-34.
- Ponath PD, Qin S, Post TW, *et al.* **Molecular cloning and characterization of a human eotaxin receptor expressed selectively on eosinophils.** *J Exp Med* 1996, **183**:2437-2448.
- Pourzand C, Cerutti P. **Genotypic analysis by RFLP/PCR.** *Mutat Res* 1993, **288**:113-121.

- Quillent C, Oberlin E, Braun J, *et al.* **HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of the CCR5 gene.** *Lancet* 1998, **351**:14-18.
- Rabut GE, Konner JA, Kajumo F, Moore JP, Dragic T. **Alanine substitutions of polar and non-polar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage and dualtropic isolates of human immunodeficiency virus type 1.** *J Virol* 1998, **72**:3464-3468.
- Raport C, Gosling J, Schweickart V, Gray P, Charo I. **Molecular cloning and functional characterization of a novel human CC chemokine receptor (CCR5) for RANTES, MIP-1beta and, and MIP-1 alpha.** *J Biol Chem* 1996, **271**:17161-17166.
- Rhodes DI, Ashton L, Solomon A, *et al.* **Characterization of three *nef*-defective human immunodeficiency type 1 strains associated with long-term nonprogression.** *J Virol* 2000, **74**:10581-10588.
- Rinaldo C, Huang XL, Fan ZF *et al.* **High levels of anti-human immunodeficiency virus type 1 (HIV-1) memory cytotoxic T-lymphocyte activity and low viral load are associated with a lack of disease in HIV-infected long-term nonprogressors.** *J Virol* 1995, **69**:5838-5842.
- Rizzardì GP, Morawetz RA, Vicenzi E *et al.* **CCR2 polymorphism and HIV disease. Swiss HIV cohort.** *Nat Med* 1998, **4**:252-253.
- Rizzuto CD, Wyatt R, Hernandez-Ramos N, *et al.* **A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding.** *Science* 1998, **280**:1949-1953.
- 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.
- Rowland-Jones SL, Nixon DF, Aldhous MC, *et al.* **HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant.** *Lancet* 1993, **341**:860-861.
- Rowland-Jones S, Sutton S, Ariyoshi K *et al.* **HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women.** *Nat Med* 1995, **1**:59-64.
- Rucker J, Samson M, Doranz BJ *et al.* **Regions in beta-chemokine receptors CCR5 and CCR2b that determine HIV-1 cofactor specificity.** *Cell* 1996, **87**:437-446.
- Rucker J, Edinger AL, Sharron M, *et al.* **Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses.** *J Virol* 1997, **71**:8999-9007.
- Sachidanandam R, Weissman D, Schmidt SC, *et al.* **A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.** *Nature* 2001, **409**:928-933.
- Saiki RK, Gelfand DH, Stoffel S, *et al.* **Primer-directed enzymatic amplification of DNA with thermostable DNA polymerase.** *Science* 1988, **239**:487-491.
- Salkowitz JR, Purvis SF, Meyerson H, *et al.* **Characterization of high-risk HIV-1 seronegative hemophiliacs.** *Clin Immunol* 2001, **98**:200-211.

- Salvi R, Garbuglia AR, Di Caro A, Pulciani S, Montella F, Benedetto A. **Grossly defective *nef* gene sequences in a human immunodeficiency virus type 1-seropositive long-term nonprogressor.** *J Virol* 1998, **72**:3646-3657.
- Samson M, Labbe O, Mollereau C, Vassart G, Parmentier M. **Molecular cloning and functional expression of a new human CC-chemokine receptor gene.** *Biochemistry* 1996a, **35**:3362-3367.
- Samson M, Libert F, Doranz BJ, *et al.* **Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene.** *Nature* 1996b, **382**:722-725.
- Samson M, La Rosa G, Libert F, *et al.* **The second extracellular loop of CCR5 is the major determinant of ligand specificity.** *J Biol Chem* 1997, **272**:24934-24941.
- Sanger F, Nicklen S, Coulson AR. **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci USA* 1977, **74**:5463-5467.
- Sattentau OJ, Zolla-Pazner S, Poignard P. **Epitope exposure on functional, oligomeric HIV-1 gp41 molecules.** *Virology* 1995, **206**:713-717.
- 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, Kostra 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 monocyctotropic 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.
- Sheppard HW, Lang W, Ascher MS, Vittinghoff E, Winkelstein W. **The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels.** *AIDS* 1993, **7**:1159-1166.
- Shioda T, Levy JA, Cheng-Mayer C. **Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1.** *Proc Natl Acad Sci USA* 1992, **89**:9434-9438.
- Shirozu M, Nakano T, Inazawa J, *et al.* **Structure and chromosomal localization of the human stromal cell-derived factor (SDF1) gene.** *Genomics* 1995, **28**:495-500.
- Siciliano S, Kuhmann SE, Weng Y, *et al.* **A critical site in the core of the CCR5 chemokine receptor required for binding and infectivity of human immunodeficiency virus type 1.** *J Biol Chem* 1999, **274**:1905-1913.
- Signoret N, Oldridge J, Pelchen-Matthews A, *et al.* **Phorbol esters and SDF-1 induce rapid endocytosis and down modulation of the chemokine receptor CXCR4.** *J Cell Biol* 1997, **139**:651-664.
- 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.

- Smith MW, Dean M, Carrington M, *et al.* **Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression.** *Science* 1997a, **277**: 959-965.
- Smith MW, Carrington M, Winkler C, *et al.* **CCR2 chemokine receptor and AIDS progression.** *Nat Med* 1997b, **3**:1052-1053.
- Sozzani S, Allavena PD, Amico G, *et al.* **Differential regulation of chemokine receptors during dendritic cell maturation: a model for their trafficking properties.** *J Immunol* 1998, **161**:1083-86.
- Speck RF, Wehrly K, Platt EJ *et al.* **Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop.** *J Virol* 1997, **71**:7136-7139.
- Stanford SA, Skurnick J, Louria D *et al.* **Lack of infection in HIV-exposed individuals is associated with a strong CD8(+) cell noncytotoxic anti-HIV response.** *Proc Natl Acad Sci USA* 1999, **96**:1030-1035.
- 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.
- Stephens JC, Reich DE, Goldstein DB *et al.* **Dating the origin of the CCR5-Delta 32 AIDS-resistance allele by the coalescence of haplotypes.** *Am J Hum Genet* 1998, **62**:1507-1515.
- Strader CD, Fong TM, Tota MR, Underwood D, Dixon RA. **Structure and function of G protein-coupled receptors.** *Annu Rev Biochem* 1994, **63**:101-132.

- Sullivan AD, Wigginton J, Kirschner D. **The coreceptor mutation CCR5 Δ 32 influences the dynamics of HIV epidemics and is selected for by HIV.** *Proc Natl Acad Sci USA* 2001, **98**:10214-10219.
- Tashiro K, Tada H, Heilker R, Shirozu M, Nakano T and Honjo T. **Signal sequence trap: A cloning strategy for secreted protein and type I membrane proteins.** *Science* 1993, **261**:600-603.
- Taub DD, Lloyd AR, Conlon K, *et al.* **Recombinant human interferon-inducible protein 10 is a chemoattractant for human monocytes and T lymphocytes and promotes T cell adhesion to endothelial cells.** *J Exp Med* 1993, **177**:1809-1814.
- Taub DD, Sayers TJ, Carter CRD, Ortaldo JR. **α and β chemokines induce NK cell migration and enhance NK-mediated cytotoxicity.** *J Immunol* 1995, **155**:3877-3888.
- Taylor JG, Choi E-H, Foster CB, Chanock SJ. **Using genetic variation to study human disease.** *Trends Molec Med* 2001, **7**:507-512.
- Tersmette M, deGoede REY, Ai BJM *et al.* **Differential syncytium-inducing capacity of human immunodeficiency virus isolates: Frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency virus syndrome (AIDS) and AIDS-related complex.** *J Virol* 1988, **62**:2026-2032.
- Tersmette M, Gruters R, Wolf F *et al.* **Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency virus syndrome: Studies on sequential isolates.** *J Virol* 1989, **63**:2118-2125.

- Theodorou I, Meyer L, Magierowska M, Katlama C, Rouzioux C. **HIV-1 infection in an individual homozygous for CCR5-Δ32.** *Lancet* 1997, **349**:1219-1220.
- Tiffany HL, Lautens LL, Gao JL, *et al.* **Identification of CCR8: a human monocyte and thymus receptor for the CC chemokine I-309.** *J Exp Med* 1997, **186**:165-170.
- Treurnicht FK, Smith T-L, Engelbrecht S, *et al.* **Genotypic and Phenotypic analysis of the env gene from South African HIV-1 subtype B and C isolates.** *J Medical Virology* (in press).
- Trkola A, Dragic T, Arthos J *et al.* **CD4-dependant antibody-sensitive interactions between HIV-1 and its co-receptor CCR5.** *Nature* 1996, **384**:184-187.
- Ullum H, Cozzi Lepri A, Victor J, *et al.* **Production of beta-chemokines in human immunodeficiency (HIV) infection: evidence that high levels of macrophage inflammatory protein-1beta are associated with a decreased risk of HIV disease progression.** *J Infect Dis* 1998, **177**:331-336.
- UNAIDS (www.unaids.org). **Report on the global HIV/AIDS epidemic, December 2000.**
- Underhill PA, Jin L, Lin AA, *et al.* **Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography.** *Genome Research* 1997, **7**:996-1005.
- van Harmelen JH, Wood R, Lambrick M, Rybicki EP, Williamson A-L, Williamson C. **An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa.** *AIDS* 1997, **11**:81-87.

- van Harmelen JH, van der Ryst E, Loubser AS *et al.* **A predominantly HIV type 1 C-restricted epidemic in South African Urban populations.** *AIDS Res Hum Retroviruses* 1999, **15**: 395-398
- van Orsouw NJ. **Design and applications of mutation detection in large genes by two-dimensional gene scanning.** *University of Groningen, Phd Thesis* 1998.
- van Rij RP, Broersen S, Goudsmit J, Coutinho RA, Schuitemaker H. **The role of a stromal cell-derived factor-1 chemokine gene variant in the clinical course of HIV-1 infection.** *AIDS* 1998, **12**:F85-F90.
- 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.
- Vlahov D, Graham N, Hoover D, *et al.* **Prognostic indicators for AIDS and infectious diseases death in HIV-infected injection drug-users: plasma viral load and CD4+ cell count.** *JAMA* 1998, **279**:35-40.
- Wang Z, Lee B, Murray JL, *et al.* **CCR5 HIV-1 coreceptor activity.** *J Biol Chem* 1999, **274**:28413-28419.
- Ward SG, Westwick J. **Chemokines: understanding their role in T-lymphocyte biology.** *Biochem J* 1998, **333**:457-470.
- 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.
- Williamson C, Engelbrecht S, Lambrick M, *et al.* **HIV-1 subtypes in different risk groups in South Africa [letter].** *Lancet* 1995, **346**:782.
- Williamson C, Loubser SA, Brice B *et al.* **Allelic frequencies of host genetic variants influencing susceptibility to HIV-1 infection and disease in South African populations.** *AIDS* 2000, **14**: 449-451.
- Winkler C, Modi W, Smith MW, *et al.* **Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant.** *Science* 1998, **279**:389-393.
- Witkin SS, Linhares IM, Gerber S, Caetano ME, Segurado AC. **Interleukin-1 receptor antagonist gene polymorphism and circulating levels of human immunodeficiency virus type 1 RNA in Brazilian women.** *J Virol* 2001, **75**:6242-6244.
- Wong L-M, Myers SJ, Tsou C-L, Gosling J, Arai H, Charo IF. **Organization and differential expression of the human monocyte chemoattractant protein 1 receptor gene.** *J Biol Chem* 1997, **272**: 1038-1045.
- Wu L, Gerard NP, Wyatt R, *et al.* **CD4-induced interaction of primary HIV-1 gp120 hglcoproteins with the chemokine receptor CCR-5.** *Nature* 1996, **384**:179-183.
- Wu Y, Hayes VM, Osinga J, *et al.* **Improvement of fragment and primer selection for mutation detection by denaturing gradient gel electrophoresis.** *Nucleic Acids Res* 1998, **26**:5432-5440.
- Wu Y. **Design and application of DGGE-based mutation detection systems for genes involved in HNPCC.** *University of Groningen, Phd Thesis* 1999.

- Xiao W, Oefner PJ. **Denaturing high-performance liquid chromatography: A review.** *Hum Mutat* 2001, **17**:439-474.
- Yamamoto K, Takeshima H, Hamada K, *et al.* **Cloning and functional characterization of the 5'-flanking region of the human monocyte chemoattractant protein-1 receptor (CCR2) gene.** *J Biol Chem* 1999, **274**:4646-4654.
- Yoshie O, Imai T, Nomiyama H. **Novel-lymphocyte specific CC chemokines and their receptors.** *J Leukoc Biol* 1997, **62**:634-644.
- Yoshie O, Imai T, Nomiyama H. **Chemokines in Immunity.** *Adv Immunol* 2001, **78**:57-110.
- Youil R, Kemper B, Cotton RGH. **Screening for mutations by enzyme cleavage of mismatch using T4 endonuclease VII.** *Proc Natl Acad Sci USA* 1995, **92**:87-91.
- Zagury D, Lachgar A, Chams V, *et al.* **C-C chemokines, pivotal in protection against HIV type 1 infection.** *Proc Natl Acad Sci USA* 1998, **95**:3857-3861.
- Zhou N, Luo Z, Hall JW, Luo J, Han X, Huang Z. **Molecular modelling and site-directed mutagenesis of CCR5 reveal residues critical for chemokine binding and signal transduction.** *Eur J Immunol* 2000, **30**:164-173.
- Zhu T, Mo H, Wang N, *et al.* **Genotypic and phenotypic characterization of HIV-1 in patients with primary infection.** *Science* 1993, **261**:1179-1181.
- Zimmerman PA, Buckler-White A, Alkhatib G, *et al.* **Inherited resistance to HIV-1 conferred by an inactivating mutation in the CC chemokine receptor 5: studies in populations with contrasting clinical phenotypes, defined racial background and quantified risk.** *Molec Med* 1997, **3**:23-26.

- Zlotnik A, Morales J, Hedrick JA. **Recent advances in chemokines and chemokine receptors.** *Crit Rev immunol* 1999, **19**:1-47.
- Zlotnik A, Yoshie O. **Chemokines: A new classification system and their role in immunity.** *Immunity* 2000, **12**:121-127.

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.

	Population groups		
	African (n = 70)	Coloured (n = 26)	Caucasian (n = 7)
Gender	Male - 21 Female - 49	Male - 8 Female - 18	Male - 6 Female - 1
Sexuality	Heterosexual - 45 Homosexual - 1 Unknown - 24	Heterosexual - 21 Homosexual - 1 Unknown - 4	Heterosexual - 3 Homosexual - 3 Unknown - 1
CD4 count (cells/mm³)	< 200 - 15 200 to 500 - 23 > 500 - 6 Unknown - 26	< 200 - 12 200 to 500 - 6 > 500 - 8 Unknown - 2	< 200 - 2 200 to 500 - 4 > 500 - 0 Unknown - 1
Disease progression	Slow - 3 Normal - 10 Fast - 13 Unknown - 44	Slow - 5 Normal - 5 Fast - 4 Unknown - 12	Slow - 1 Normal - 0 Fast - 2 Unknown - 4
Clinical stage	Stage I - 10 Stage II - 2 Stage III - 11 Stage IV - 6 Unknown - 41	Stage I - 5 Stage II - 3 Stage III - 12 Stage IV - 4 Unknown - 2	Stage I - 0 Stage II - 3 Stage III - 2 Stage IV - 0 Unknown - 2

n = Number of HIV seropositive individuals representing each of the population groups.

Clinical staging - World Health Organisation (WHO) (www.who.int).

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.

Mutation	Population groups								
	Africans (n = 26)			Coloureds (n = 14)			Caucasians (n = 3)		
	Disease progression								
	Slow	Normal	Fast	Slow	Normal	Fast	Slow	Normal	Fast
<i>CCR5P35</i>	1 - Hetero	1 - Hetero	1 - Hetero	2 - Hetero	2 - Hetero	1 - Hetero	-	-	-
<i>CCR5Y335</i>	-	-	1 - Hetero	1 - Hetero	-	-	-	-	-
<i>CCR2 Int1-57a/g</i>	-	1 - Hetero	1 - Hetero	-	2 - Hetero	-	-	-	-
<i>CCR2 Int1-43g/a</i>	-	-	-	-	-	1 - Homo	-	-	-
<i>CCR2V63</i>	-	-	-	-	-	1 - Hetero	-	-	-
<i>CCR2V64I</i>	1 - Hetero	3 - Hetero	4 - Hetero	-	-	-	-	-	-
<i>CCR2L283</i>	-	1 - Hetero	1 - Hetero	-	-	-	-	-	-
<i>CCR2AP339</i>	1 - Homo	4 - Hetero	1 - Homo 3 - Hetero	-	1 - Hetero	1 - Hetero	-	-	-
<i>-CCR2BT348</i>	-	-	-	1 - Hetero	-	-	1 - Hetero	-	-
<i>SDF1-3'A</i>	-	-	3 - Hetero	2 - Hetero	1 - Hetero	-	-	-	-

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.

HIV patient	Population group	Disease progression	CCR5 P35	CCR5 Y335	CCR2 Int -57a/g	CCR2 Int -43g/a	CCR2 V63	CCR2 V64I	CCR2 L283	CCR2A P339	CCR2B T348	SDF1 3'A
1	African	Slow	+	-	-	-	-	-	-	-	-	-
2	Coloured	Slow	-	-	-	-	-	-	-	-	-	+
3	Coloured	Slow	+	-	-	-	-	-	-	-	-	-
4	Coloured	Slow	+	-	-	-	-	-	-	-	+	+
5	Coloured	Slow	-	-	-	-	-	-	-	-	-	-
6	African	Slow	-	-	-	-	-	-	-	++	-	-
7	Caucasian	Slow	-	-	-	-	-	-	-	-	+	-
8	African	Slow	-	-	-	-	-	+	-	-	-	-
9	Coloured	Slow	-	+	-	-	-	-	-	-	-	-
10	African	Normal	-	-	-	-	-	-	-	+	-	-
11	Coloured	Normal	-	-	-	-	-	-	-	+	-	-
12	African	Normal	-	-	-	-	-	+	-	-	-	-
13	Coloured	Normal	+	-	-	-	-	-	-	-	-	-
14	African	Normal	-	-	-	-	-	-	-	-	-	-
15	Coloured	Normal	-	-	+	-	-	-	-	-	-	+
16	Coloured	Normal	-	-	-	-	-	-	-	-	-	-
17	Coloured	Normal	+	-	+	-	-	-	-	-	-	-
18	African	Normal	-	-	-	-	-	-	-	-	-	-
19	African	Normal	-	-	+	-	-	-	-	+	-	-
20	African	Normal	-	-	-	-	-	+	-	-	-	-
21	African	Normal	-	-	-	-	-	+	-	-	-	-

HIV patient	Population group	Disease progression	CCR5 P35	CCR5 Y335	CCR2 Int -57a/g	CCR2 Int -43g/a	CCR2 V63	CCR2 V64I	CCR2 L283	CCR2A P339	CCR2B T348	SDF1 3'A
22	African	Normal	-	-	-	-	-	-	+	+	-	-
23	African	Normal	+	-	-	-	-	-	-	-	-	-
24	African	Normal	-	-	-	-	-	-	-	+	-	-
25	African	Fast	-	-	-	-	-	-	-	+	-	-
26	Caucasian	Fast	-	-	-	-	-	-	-	-	-	-
27	Caucasian	Fast	-	-	-	-	-	-	-	-	-	-
28	Coloured	Fast	+	-	-	-	-	-	-	-	-	-
29	African	Fast	-	-	-	-	-	-	-	+	-	-
30	African	Fast	-	-	-	-	-	-	-	+	-	-
31	Coloured	Fast	-	-	-	-	-	-	-	+	-	-
32	African	Fast	-	-	-	-	-	-	-	++	-	-
33	African	Fast	-	-	-	-	-	+	-	-	-	-
34	Coloured	Fast	-	-	-	-	+	-	-	-	-	-
35	Coloured	Fast	-	-	-	++	-	-	-	-	-	-
36	African	Fast	-	-	-	-	-	-	-	-	-	+
37	African	Fast	-	-	-	-	-	+	-	-	-	-
38	African	Fast	-	+	-	-	-	+	-	-	-	-
39	African	Fast	+	-	-	-	-	+	-	-	-	-
40	African	Fast	-	-	-	-	-	-	+	-	-	+
41	African	Fast	-	-	-	-	-	-	-	-	-	-
42	African	Fast	-	-	-	-	-	-	-	-	-	+
43	African	Fast	-	-	+	-	-	-	-	-	-	-

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.

	Allele genotype			Total
	MM	MN	NN	
No of individuals	64	5	1	70
No of N alleles	128	5	0	133
No of M alleles	0	5	2	7
Total no of alleles	128	10	2	140

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:

a	b	a + b
c	d	c + d
a + c	b + d	n

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)

	African population		<i>Total</i>
	HIV +	HIV -	
CCR5 wild-type	134	119	253
CCR5P35	6	9	15
<i>Total</i>	140	128	268

n = no. of alleles

 $P = 0.4273$ (not significant)

	Coloured population		<i>Total</i>
	HIV +	HIV -	
CCR5 wild-type	43	134	177
CCR5P35	9	10	19
<i>Total</i>	52	144	196

n = no. of alleles

 $P = 0.0516$ (not significant)

Example 2: CCR5A335V SNP (see chapter 3)

	African population		<i>Total</i>
	HIV +	HIV -	
CCR5 wild-type	136	126	262
CCR5A335V	4	2	6
<i>Total</i>	140	128	268

n = no. of alleles

 $P = 0.6859$ (not significant)

	Coloured population		<i>Total</i>
	HIV +	HIV -	
CCR5 wild-type	51	142	193
CCR5A335V	1	2	3
<i>Total</i>	52	144	196

n. = no. of alleles

 $P = 1.0000$ (not significant)

Example 3: CCR2V64I SNP (see chapter 4)

	African population		Total
	HIV +	HIV -	
CCR2 wild-type	120	103	223
CCR2V64I	18	21	39
<i>Total</i>	138	124	262

n = no. of alleles

$P = 0.3906$ (not significant)

	Coloured population		Total
	HIV +	HIV -	
CCR2 wild-type	51	119	170
CCR2V64I	1	25	26
<i>Total</i>	52	144	196

n = no. of alleles

$P = 0.0034$ (very significant)

	Caucasian population		Total
	HIV +	HIV -	
CCR2 wild-type	14	3	17
CCR2V64I	0	1	1
<i>Total</i>	14	4	18

n = no. of alleles

$P = 0.2222$ (not significant.)

Sample numbers for Caucasian HIV+ patients and HIV - controls are however small.

Example 4: CCR2N260 SNP (see chapter 4)

	African population		<i>Total</i>
	HIV +	HIV -	
CCR2 wild-type	63	45	108
CCR2N260	75	79	154
<i>Total</i>	138	124	262

n = no. of alleles

$P = 0.1331$ (not significant)

	Coloured population		<i>Total</i>
	HIV +	HIV -	
CCR2 wild-type	22	47	69
CCR2N260	30	97	127
<i>Total</i>	52	144	196

n = no. of alleles

$P = 0.2374$ (not significant)

	Caucasian population		<i>Total</i>
	HIV +	HIV -	
CCR2 wild-type	4	2	6
CCR2N260	10	2	12
<i>Total</i>	14	4	18

n = no. of alleles

$P = 0.5686$ (not significant)

Sample numbers for Caucasian HIV+ patients and HIV - controls are however small.

Example 5: CCR2P339 SNP (see chapter 4)

	African population		<i>Total</i>
	HIV +	HIV -	
CCR2 wild-type	117	106	223
CCR2P339	21	18	39
<i>Total</i>	138	124	262

n = no. of alleles

$P = 1.0000$ (not significant)

	Coloured population		<i>Total</i>
	HIV +	HIV -	
CCR2 wild-type	50	139	189
CCR2P339	2	5	7
<i>Total</i>	52	144	196

n = no. of alleles

$P = 1.0000$ (not significant)

Example 6: *SDF1*-3'A SNP (see chapter 5)

	African population		<i>Total</i>
	HIV +	HIV -	
<i>SDF1</i> wild-type	133	125	258
<i>SDF1</i> -3'A	7	1	8
<i>Total</i>	140	126	266

n = no. of alleles

$P = 0.0692$ (not significant)

	Coloured population		<i>Total</i>
	HIV +	HIV -	
<i>SDF1</i> wild-type	48	119	167
<i>SDF1</i> -3'A	4	27	31
<i>Total</i>	52	146	198

n = no. of alleles

$P = 0.0768$ (not significant)

	Caucasian population		<i>Total</i>
	HIV +	HIV -	
<i>SDF1</i> wild-type	11	69	80
<i>SDF1</i> -3'A	3	27	30
<i>Total</i>	14	96	110

n = no. of alleles

$P = 0.7544$ (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).

<u>Observed (n = 70)</u>	<u>Expected (n = 70)</u>
N/N = 64	N/N = 64.11($p^2 \times 70$)
N/M = 6	N/M = 5.76 ($2pq \times 70$)
M/M = 0	MM = 0.13 ($q^2 \times 70$)

n = number of individuals

N = wildtype allele; M = mutant allele

Incidence of wild-type allele = $2(NN) + NM / \text{Total no. of alleles}$

$$= [2(64) + 6]/140$$

$$p = 0.957$$

Incidence of mutant allele = $2(MM) + NM / \text{Total no. of alleles}$

$$= [2(0) + 6]/140$$

$$q = 0.043$$

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

Further reading

- Miller SL. **Introductory statistics for dentistry and medicine.** Reston Publishing Company Inc, Virginia 1981.
- Barnard C, Gilbert F, McGregor P. **Asking questions in Biology.** Longman Group Limited, Harlow 1993.
- Mould RF. **Introductory medical statistics.** Third edition. IOP Publishing Ltd, London 1998.
- Rees DG. **Essential statistics.** Third edition. Chapman and Hall, London 1998.
- Mueller RF, Young ID. **Emery's elements of Medical Genetics.** Tenth Edition. Churchill Livingstone, London 1998.