

https://dspace.org/

Impact of haptoglobin gene variation on HIV resistance and the rate of disease progression in the South African black population

Skhosana, Lindiwe

2006-11-15T09:33:01Z
http://hdl.handle.net/10539/1733

Downloaded from a DSpace repository. https://dspace.org/

IMPACT OF HAPTOGLOBIN GENE VARIATION ON HIV RESISTANCE AND THE RATE OF DISEASE PROGRESSION IN THE SOUTH AFRICAN BLACK POPULATION



Lindiwe Skhosana

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfillment of the requirements for the degree of Master of Science

Johannesburg, March 2005

DECLARATION

I declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

……..

(Signature of candidate)

……day of ……2005.

ABSTRACT

Genetic variation in haptoglobin, a plasma protein, has been reported to be associated with susceptibility to and the rate of HIV/AIDS progression. The purpose of this study was to investigate the influence of haptoglobin polymorphism on HIV/AIDS in black South Africans. Polymorphism in the coding region of the haptoglobin gene was detected by direct DNA and allele-specific amplification. Polymorphism in the coding region of the gene was detected by amplification of DNA and by polyacrylamide gel electrophoresis of plasma protein. A statistically significant association was observed between allele -61C and resistance to HIV infection. The Hp0 phenotype, in which no haptoglobin protein is detected, was associated with HIV status and some promoter genotypes. Since in our study population there were a few samples with usable clinical data , further investigations need to be done to confirm the association of the -61C allele and the Hp0 phenotype with the risk of HIV infection.

DEDICATION

To the Skhosana family

ACKNOWLEDGMENTS

I would like to extend my sincere gratefulness to my supervisor Prof Tracy Mclellan for her patience, support and guidance when everything seemed bleak.

This study would not have been possible without the participants who donated blood samples – thank you all.

I would also thank the staff of Johannesburg General Hospital, especially Dr. Francois Venter, for their assistance, understanding and patience during sample collection.

I am also grateful to Dr Clive Gray and Debbie from the NICD for providing us with the DNA samples.

I would also like to thank the National Research Foundation for financial assistance. Without it studying my MSc would have been a dream beyond my reach.

Above all, I thank God for strength, guidance, good health and sanity He gave me throughout my years of study.

LIST OF FIGURES

Figure 1.1	The mechanism of escape mutation by HIV……	3
Figure 1.2	Phylogenetic tree of HIV/SIV …….....	6
Figure 1.3	HIV/AIDS prevalence is South Africa ……	8
Figure 1.4	HIV/AIDS prevalence in South Africa: Provincial statistics...	9
Figure 1.5	Activation of haptoglobin gene transcription by IL-6 …….	17
Figure 1.6A	Haptoglobin monomers and polymers…….	19
Figure 1.6B	Structural differences between haptoglobin types……..	19
Figure 1.7	Formation of Hp^2 allele ……...	21
Figure 1.8	A schematic representation of the Hp^{del} breakpoints …….	23
Figure 1.9	Schematic representation of NF-κB activation ……...	31
Figure 2.1.	Nucleotide sequence of the promoter region of the haptoglobin gene ……	41
Figure 2.2.	Structure of Hp^1 and Hp^2 alleles ……..	44
Figure 3.1.1	A PCR product of the haptoglobin promoter region ……	53
Figure 3.1.2	Schematic diagram representing locations of the SNPs found by direct sequencing …….	53
Figure 3.2.1.	Allele-specific PCR products of the -55A/G variation …….	56
Figure 3.2.2.	Allele-specific PCR products of the -61A/C variation……...	57
Figure 3.2.3.	Allele-specific PCR products of the -104A/T variation…….	58
Figure 3.2.4.	Allele-specific PCR products of the -295T/C variation…….	59
Figure 3.3.1.1.	Hp1 and Hp2 allele-specific PCR products ……..	62
Figure 3.3.2.1.	Haptoglobin phenotypes on the polyacrylamide gel……..	64

LIST OF TABLES

Table 1.1	Distribution of haptoglobin alleles in black South African populations ……	34
Table 2.1	Primer sequences used in ASA……..	41
Table 3.1.1	Genotype and allele frequencies for the 6 polymorphic sites found by direct sequencing……	54
Table 3.2.1.	Genotype and allele frequencies for the four common polymorphisms in the whole study population……	60
Table 3.3.1.1	Genotype and allele frequencies of the coding region in the whole study population……	63
Table 3.3.2.1	Haptoglobin phenotype distribution in the whole study population……	65
Table 3.3.2.2	Genotype distribution in Hp0 phenotype and the other phenotypes……	66
Table 3.4.1	Hp^1 allele distribution in South African ethnic groups…	67
Table 3.5.1.1.	Genotype and phenotype distribution among HIV positive and general populations……	69
Table 3.5.1.2.1.	Genotype and phenotype frequencies between asymptomatic and symptomatic subjects five years after diagnosis with HIV infection…….	73
Table 3.5.1.2.2	Genotype and phenotype frequencies in rapid progressors and long-term nonprogressors……	74
Table 3.5.1.2.3	Phenotype and genotype distribution in groups with a	

	$CD4^+$ cell count of below and above 200 cells/ml^3 five years after diagnosis……...	75
Table 3.5.1.3	Genotype and phenotype frequencies in groups with and without TB…….	77
Table 3.6.1	Haplotype frequencies in the study population……..	79
Table 3.6.2	Linkage disequilibrium between haptoglobin alleles…….	80
Table 3.6.3.1	Haplotype distribution in HIV positive and the general populations……	81
Table 3.6.3.2.1	Comparisons of haplotype frequencies between rapid progressors and long-term nonprogressors……	82
Table 3.6.3.2.2	Haplotype frequencies and the rate of $CD4^+$ cell count depletion……	82
Table 3.6.3.2.3	Comparison of haplotypes in symptomatic and asymptomatic groups……..	82
Table 3.6.3.3	Haplotype distribution in groups with and without TB…..	83

CONTENTS

DECLARATION …… ii

ABSTRACT …… iii

DEDICATION……... iv

ACKNOWLEGDEMENTS …… v

LIST OF FIGURES …… vi

LIST OF TABLES ……. vii

CHAPTER 1 - INTRODUCTION

1.1 Genetic evolution and infectious diseases …….. 1

1.2 HIV/AIDS ……... 4

1.2.1 HIV types and its origins ……. 4

1.2.2 Epidemiology …….. 7

1.2.3 How does the virus work? …… 10

1.2.4 Factors affecting HIV/AIDS progression ……..... 11

1.3 Genes involved in HIV/AIDS infection and disease progression ……... 13

1.4 Haptoglobin ……. 16

1.4.1 Haptoglobin structure and types…… 18

1.4.2 Variation in the haptoglobin promoter region ……. 25

1.4.3 Functions of the haptoglobin …… 25

1.4.3.1 Haemoglobin binding …….. 26

1.4.3.2 Bacteriostatic effect ……. 26

1.4.3.3 Agglutination effect of the haptoglobin …….. 26

1.4.3.4 Inhibition of prostaglandin synthesis …….. 27

1.4.3.5 Angiogenesis ……... 27

1.5 Involvement of haptoglobin types with HIV/AIDS and other diseases…. 27

1.6 Haptoglobin frequencies in South Africans …….. 32

1.7 Problem identification …….. 35

CHAPTER 2 – MATERIALS AND METHODS

2.1. Subjects……... 36

2.2. DNA isolation …… 37

2.3. Detection of variation in the promoter region of the haptoglobin gene by direct sequencing…….. 37

2.4. Alternative methods for detecting single nucleotide polymorphism….. 39

2.5. Protein polymorphism …… 43

2.5.1. Haptoglobin genotype determination…….. 43

2.5.2. Haptoglobin phenotype determination…… 44

2.6. Data analysis……... 45

2.6.1. Determination of allele and genotype frequencies…….. 45

2.6.2. Hardy-Weinberg equilibrium…….. 46

2.6.3. Linkage equilibrium and disequilibrium……. 47

2.6.4. Haplotype analysis……... 49

2.6.5. Determination of genotype and phenotype/disease status

association …….. 49

2.6.6. Determination of haplotype/disease status association ……... 51

CHAPTER 3 - RESULTS

3.1. Direct sequencing…….. 52

3.2. Allele-specific amplification……. 55

3.2.1. Genotype and allele frequencies…… 59

3.3. Protein polymorphism……... 60

3.3.1 Haptoglobin protein polymorphism genotype determination……. 60

3.3.2 Phenotype determination……. 63

3.4. Distribution of haptoglobin alleles in Black South African ethnic groups…….. 66

3.5. Association of genotypes and phenotypes with HIV/AIDS disease status ……. 67

3.5.1.1. Genotype and phenotype distribution in HIV positive and the general populations……. 68

3.5.1.2. Haptoglobin polymorphism and the rate of the HIV/AIDS progression…… 70

3.5.1.3. Genotype and phenotype frequencies in TB groups……. 76

3.6. Haplotype analysis…… 78

3.6.1 Haplotypes in the study population…….. 78

3.6.2. Pairwise allelic linkage disequilibrium…… 79

3.6.3. Haplotype/Disease status association…….. 80

3.6.3.1. HIV status/Haplotype association…….. 80

3.6.3.2. Disease progression rate/haplotype association……. 81

3.6.3.3. TB status/Haplotype association…… 83

CHAPTER 4 – DISCUSSION AND CONCLUSIONS

Discussion…… 84

Conclusions…… 95

APPENDIX…….. 97

REFERENCES……. 103

CHAPTER 1

INTRODUCTION

1.1 Genetic evolution and infectious diseases

Numerous parasitical, viral and bacterial infectious agents initiate responses of the immune system of the host. Infectious agents are capable of influencing evolution of the host species. The evolutionary changes in the host may cause shifting selective pressures on the infecting agent (McMichael and Klenerman, 2002; Borghans *et al*, 2004). Adaptation system can be demonstrated by human immunodeficiency virus (HIV) and the cells of the human host's immune system. After infection by HIV, the virus replicates at the host makes antibodies against the virus. However, because the virus mutates at a very high rate during reverse transcription, mutants arise the escape the antibodies, and attack the $CD4^+$ cells. In this process, HIV kills $CD4^+$ cells and selects against the host in general; $CD4^+$ cells attack specific strains and select against non-mutated strains and HIV strains that have a low replication rates (Plumelle, 2003).

Host and pathogens exert strong selective pressures on each other. These pressures lead to the two organisms affecting each other's evolution, the process called coevolution. An example of coevolutionary system is a human host gene CCR5 and a poxvirus. The CC-chemokine receptor 5 (CCR5) is one of the coreceptor proteins for HIV. The CCR5-Δ32 deletion confers resistance to HIV infection and the rate of the disease progression (section 1.3) (Dean and O'Brien, 1997; O'Brien and Moore, 2000). The deletion allele was selected in Europe by incessant smallpox epidemics more than 2000 years ago (Hopkins, 2002; Galvani and

Slatkins, 2003). The CCR5Δ32 allele age is estimated between 275 and 1 875 years (Stephen *et al*, 1998). To date, the frequency of this allele is estimated at an average of 10% in European populations (Dean *et al*, 1996). Geographic distribution of smallpox correlates with the frequencies of CCR5Δ32 in Europe. Moreover, both poxvirus and HIV use chemokine receptors to enter cells. Thus, CCR5Δ32 could confer resistance to both poxvirus and HIV (Galvani and Slatkin, 2003).

The ability of the HIV to mutate rapidly and escape immune recognition makes it difficult to control HIV infection. The escape mutations occur at critical positions. One escape mutation is the amino acid substitution within the HLA-restricted cytotoxic T lymphocyte (CTL) epitope that hinders epitope-HLA binding. In that way, the mutants escape recognition by HLA (Fig 1.1) (Moore *et al*, 2002). The polymorphisms in HIV are a result of selective pressure from HLA and are HLA class I allele-specific (Moore *et al*, 2002; McMichael and Klenerman, 2002). Other HLA types select epitopes that for structural and functional reasons are harder to mutate without compromising virus survival (Fig 1.1). HLA B27 and HLA B57 select such epitopes. Individuals with these alleles fight the HIV infection better than individuals carrying other HLA alleles (Moore *et al*, 2002; McMichael and Klenerman, 2002).

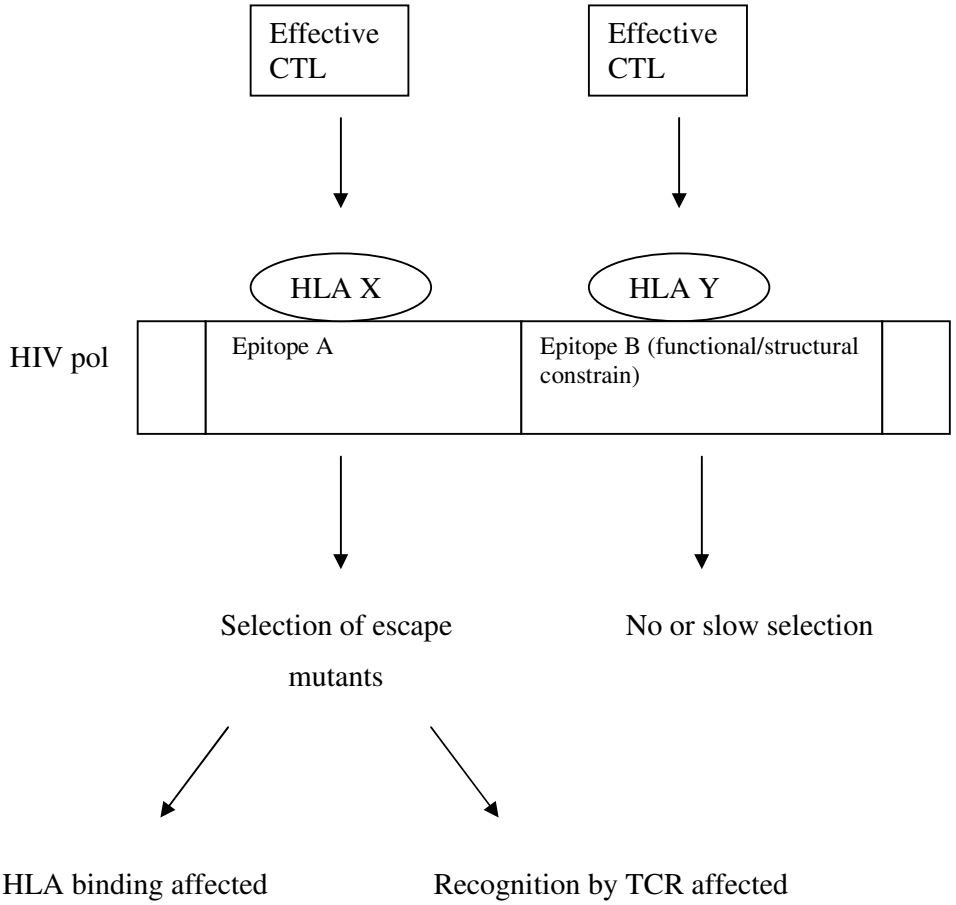


Fig 1.1. HLA molecules on the surface of the CTLs direct the CTL response to specific epitopes within the HIV pol protein. Some of these CTL responses exert selective pressures on the virus which lead to viral escape mutants that are either not recognized by HLA or fail to interact with T cell receptors (left). Other responses are directed against epitope regions that are structurally or functionally constrained which may select slowly or not select at all (right) (McMichael and Klenerman, 2002).

Selective pressures modifying genes to confer a better adaptation to environment may select for alleles that enhance susceptibility (Bottini *et al*, 1999) or resistance to other diseases (Galvani and Slatkin, 2003). Identification of the genetic changes that provide protection against diseases could lead to more effective treatment methods.

1.2 HIV/AIDS

1.2.1 HIV types and its origins

AIDS, or Acquired Immunodeficiency Syndrome, was first recognised in the early 1980's, and has grown to be a devastating global epidemic. AIDS is caused by Human Immunodeficiency Virus or HIV (Barré-Sinoussi *et al*, 1983; Gallo *et al*, 1984; Levy *et al*, 1984). HIV belongs to a family of viruses called lentiviruses, a subfamily of the retroviruses, which have genes composed of ribonucleic acid (RNA) molecules (Gaynor, 1992). Lentiviruses other than HIV have been found in nonhuman primates, chimpanzees. In these primates the virus is called Simian Immunodeficeincy Virus or SIV (Gao *et a*l, 1999).

There are two types of HIV: HIV-1 and HIV-2. HIV-1 is responsible for global AIDS, while HIV-2 is largely confined to West Africa and has tended to spread to other countries with strong links to this region of Africa, such as France (O'Brien and Dean, 1997; Matheron *et al*, 2003).

HIV-1 is a highly variable virus which mutates very readily. Therefore there are many genetically different strains of HIV-1, which are classified according to groups and subtypes. There are three groups: M, N and O (Yang *et al*, 2000).

Within the HIV-1 M-group there are ten HIV-1 subtypes: subtypes A to J. The subtypes are unevenly distributed throughout the world. The most common subtype in South Africa is subtype C (Swanson *et al*, 2003).

It is now generally accepted that HIV is a descendent of SIV (Gao *et al,* 1999; Georges-Courbot *et al*, 1998). The virus crossed over from chimpanzees to humans perhaps as a result of exposure to the blood during killing and eating them for food. This process of species crossing is called zoonosis and is not uncommon a source for novel pathogenesis in humans (Gao *et al,* 1999; Weber and Alcorn, 2000).

HIV-1 is most similar in sequence and genomic organisation to viruses found in chimpanzees (Fig 1.2). However, some subspecies of chimpanzees harbour an SIVcpz strain that is genetically divergent from HIV-1. The chimpanzee subspecies that harbours SIVcpz closely related to HIV-1 is *Pan troglodytes troglodytes.* The natural habitat of *Pan troglodytes troglodytes* coincides with areas of group M, N and O epidemicity. Thus, the genetic relatedness and geographic coincidence suggest that *Pan troglodytes troglodytes* is a reservoir for HIV-1 (Gao *et al*, 1999).

HIV-2 is phylogenetically closely related to SIV harboured by sooty mangabey monkeys, SIVsm (Fig 1.2). The epicenter of the HIV-2 coincides with the natural habitat of wild-living sooty mangabey monkeys. Thus, the primate reservoir of HIV-2 has been clearly identified as the sooty mangabey monkey (Gao *et al*, 1999; Georges-Courbot *et al*, 1998).

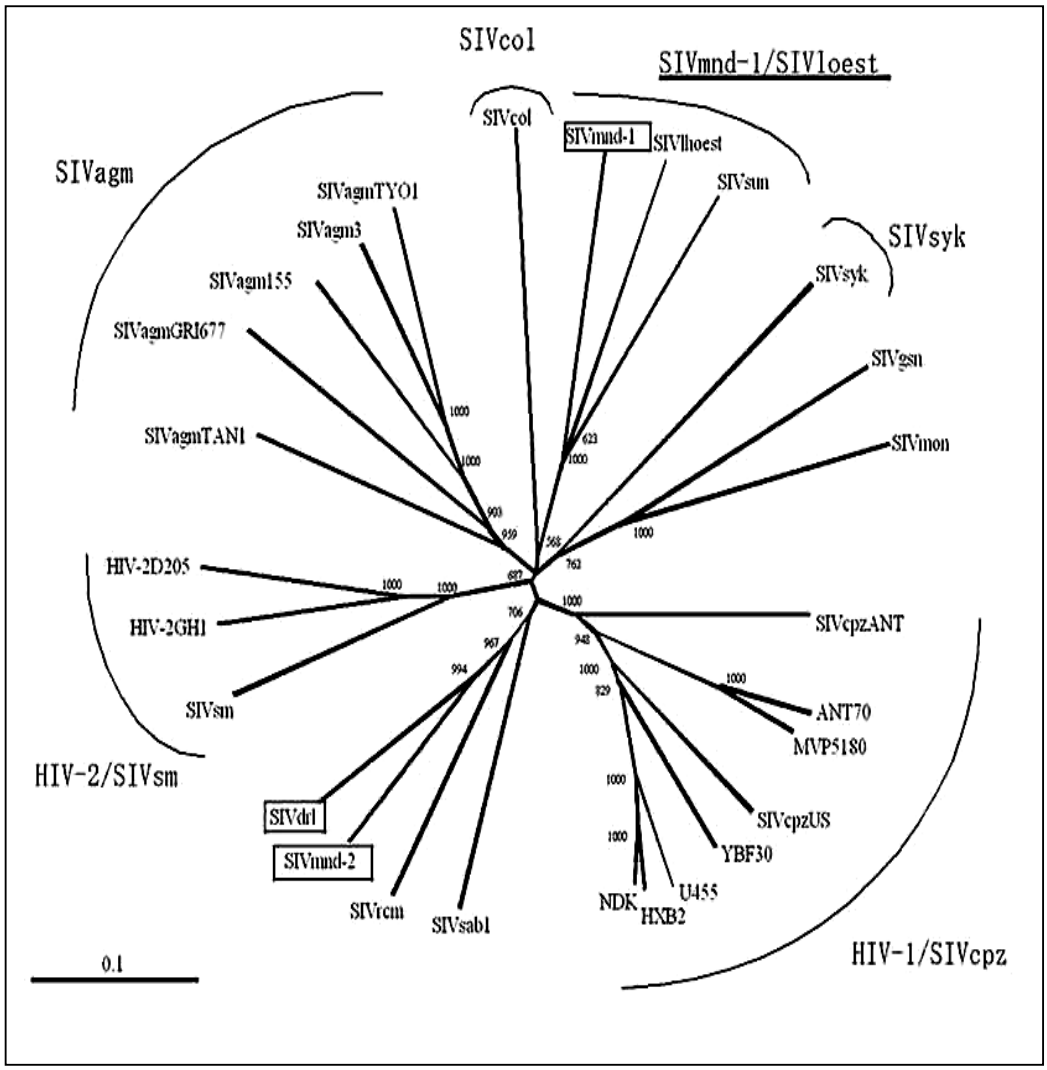


Fig 1.2. A phylogenetic tree of HIV/SIV constructed from RT region (pol). All the HIV-1 strains formed a cluster with SIVcpz. HIV2 clustered with SIVsm. The numbers at nodes indicated bootstrap values in 1000 replication. Six primate lentivirus lineages are indicated by brackets (Takemura and Hayami, 2004).

1.2.2 Epidemiology

AIDS is one of the major causes of death worldwide since the outbreak of the disease. Globally, about 40 million people were living with HIV/AIDS in 2003; 5 million of them were newly infected in 2003 (UNAIDS 2003). In 2003 alone, an estimated 3 million deaths were reported to have been caused by AIDS-related diseases. A total of 27.8 million people died of AIDS-related diseases from the beginning of the epidemic until the end of 2001, of whom 4.3 million are children less than 15 years of age (UNAIDS 2003). Sub-Sahara Africa is the world's most affected region. About 26.6 million people were living with the virus in this region in 2003. HIV prevalence considerably varies across the continent and the islands that fall under it; the range is between 1% (in Mauritius) and about 40% (in Botswana) (UNAIDS 2001, 2002).

In South Africa, the first cases of HIV were reported in 1982. These cases were mainly amongst white gay men (Ras *et al*, 1983). Since then, the number of HIV cases has been increasing. Between 1990 and 2002, the prevalence has increased from 0.8% to 26.5% (Fig 1.3) (Shisana and Simbayi, 2002). In 2001, approximately 5 million people in South Africa were infected by HIV. AIDS claimed the lives of 360 000 people in South Africa in 2001 (Dorrington *et al*, 2001; UNAIDS 2002).

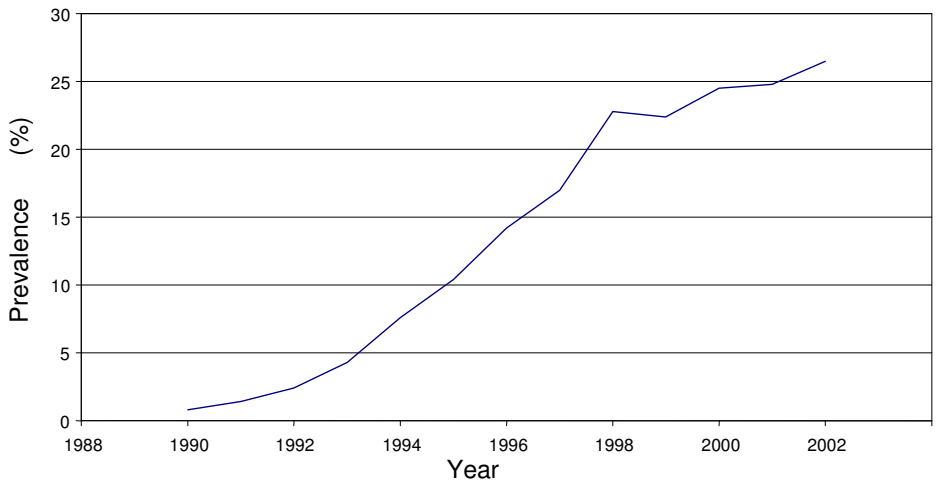


Fig 1.3. HIV prevalence rate in South Africa. The epidemic has increased from 0.8 to 26.5% between 1990 and 2002 (Shisana and Simbayi, 2002; UNAIDS 2002).

Two surveys were conducted across all nine provinces of South Africa in 2002. The first study was based on data taken from the Department of Health. The study contacted women attending antenatal clinics in all the provinces. The figures showed that 26.6% women who attended antenatal clinics nationally in 2002 (Fig 1.4) were infected. Kwa-Zulu Natal and Gauteng were leading with 36.5 and 31.6%, respectively (UNAIDS, 2002).

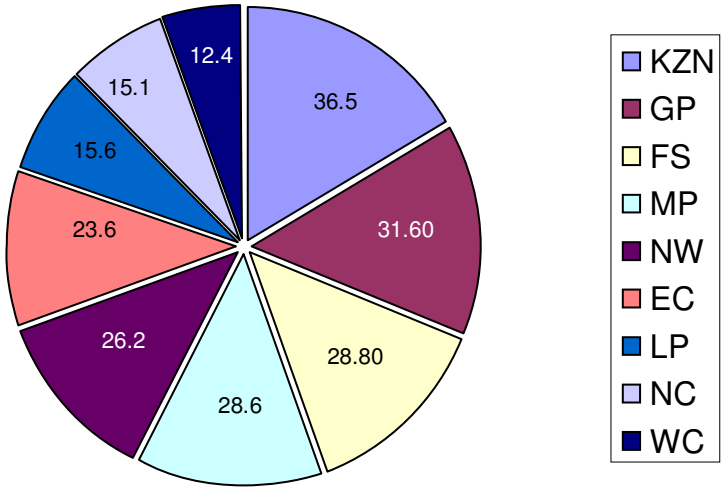


Fig 1.4. HIV/AIDS prevalence in South Africa: Provincial statistics (UNAIDS, 2002).

These figures, show prevalence amongst sexually active women only; the conclusions cannot be applied to non-sexually active women, the elderly, men and children. Sampling in the second study was different from the one above. The Nelson Mandela study looked at proportional cross-section of society in all nine provinces (Shisana and Simbayi, 2002). The results of this study suggested that KwaZulu-Natal did not have the highest HIV prevalence in 2002 as shown by the Health department study, but rather the Free State and Gauteng. The Nelson Mandela study found that 11.4% of overall population in South Africa was infected with HIV in 2002 (Shisana and Simbayi, 2002). Despite the difference in figures between the two studies, one thing clear is that South Africa has high HIV prevalence, and it is still increasing. There is a huge challenge facing the fields of prevention, care and science.

1.2.3 How does the virus work?

The healthy body's immune system fights off infections with a combination of cellular and chemical responses. HIV infects the key components of the immune system, the white blood cells called $CD4^+$ lymphocytes or T-helper cells. This infection progressively depletes the $CD4^+$ cells in the blood; eventually causing AIDS when the immune system becomes compromised and it can no longer fight off diseases. This leaves the host subject to infection from a wide array of infectious agents, many of which do not usually adversely affect healthy people (Dean *et al*, 1996). Often, when someone is said to have died of AIDS, the real cause of death is usually TB, pneumonia, or some other disease which took hold because of the disabled immune system (Nissapatorn *et al*, 2003).

Like all viruses, HIV uses the host's cellular machinery to enter the cell, replicate and produce disease (Saah, *et al*, 1998; Gaynor, 1992; Bednarik and Folks, 1992). HIV infection begins when the viral particle enters the cell. This requires CD4+ receptors and other ligands, which are part of the host (O'Brien and Dean, 1997). After the virus has gained entry into the cell, the viral RNA is converted to DNA by HIV reverse transcriptase, an enzyme coded for by retroviruses. Then the newly made HIV DNA moves to the cell's nucleus, where it is spliced into the host's DNA with the help of HIV integrase. The integrated HIV DNA is now called provirus (Bednarik and Folks, 1992). The integration of retroviral DNA is important for the transcription of new viral copies in the form of messenger RNA (mRNA). The mRNA-making process involves host cell's own enzymes. Cytokine proteins that are involved in the normal regulation of the immune

system, may also regulate transcription. Molecules such as tumor necrosis factor (TNF)-α and interleukin-6 (IL-6), secreted in elevated levels by the cells of HIV-infected individuals may, help to activate HIV proviruses (Bednarik and Folks, 1992; Molina *et al*, 1990). After mRNA has been synthesized in the nucleus, it is transported to the cytoplasm for viral protein synthesis (translation). Again, the host's protein-making machinery plays an essential role in the process (Bednarik and Folks, 1992).

1.2.4 Factors affecting HIV/AIDS progression

HIV is slow to cause illness in comparison to other viral infections like influenza, and many people may be infected for years without knowing it. Time from infection with HIV to AIDS varies from a few months (Isaksson *et al*, 1988) to 20 years (Muzon *et al*, 1995; Matheron *et al*, 2003). This large survival difference is dependent on a number of factors: pathogenecity of the infecting virus (O'Brien and Dean, 1997; Matheron *et al*, 2003), CD4+ lymphocytes in the peripheral blood (Carré *et al* (1998), behaviour (Farzadegan *et al*, 1996), anti-immune response (Ullum *et al*, 1999), viral load (Mellors *et al*, 1995; Mellors *et al*, 1996; Carré *et al*, 1998), and genetic make-up (Haynes *et al*., 1996; Ullum *et al*., 1999). These factors are probably involved both alone and in combinations to affect the rate at which HIV/AIDS progresses.

The progression of HIV to AIDS is characterised by gradual depletion of $CD4^+$ cells. A $CD4^+$ cell count of ≤ 200/μl is predictive of AIDS or the onset of AIDS (Carré *et al*, 1998). Viral load (the number of viral particles in the blood) is highly

predictive of the rate of disease progression: individuals with low viral load progress more slowly than those with higher level of the virus in their blood. High viral load in the plasma ($>10^5$ copies /ml) in the first year after seroconversion is associated with more rapid progression to AIDS (Mellors *et al*, 1995; Mellors *et al*, 1996), but little is known of its predictive value for AIDS in the later stage when the $CD4^+$ threshold is reached (Farzadegan *et al*, 1996). Viral load is inversely proportional to the $CD4^+$ cell count in a given host. This is, in turn, related to the ability of the host to contain the infection (Saah *et al*, 1998). It has also been shown that viral load is predictive of the risk of transmission. Individuals with fewer than 1 500 copies/ml rarely transmit the virus to a partner (Quinn *et al*, 2000).

Another factor that may affect the rate of HIV/AIDS progression is behaviour. Farzaden e*t al* (1996) found that in the population they studied, rapid progressors (those who developed AIDS less than 3 years after infection) had more lifetime sexual partners than the non-rapid progressors. These observations suggested that sexually transmitted infections, or maybe an unknown factor, may affect HIV disease progression.

Host genetic polymorphisms also account for differences in susceptibility to HIV infection and disease progression. These polymorphisms influence various steps of the HIV lifecycle (Williamson *et al*, 2002) and they may either retard (McDermott *et al*, 1998) or accelerate (Martin *et al*, 1998) progression to AIDS.

1.3 Genes involved in HIV/AIDS infection and disease progression

Variation in some human genes is associated with different responses to infections. Genes that are involved in HIV infection work at different aspects of infection. There are those whose products are receptors or co-receptors of HIV when it enters the cell. Others produce proteins that act on the virus after it has gained entry, and those that play a role in the immune system.

Polymorphisms in the following genes CCR5 (Liu *et al*., 1999; McDermott *et al*., 2000; Martin *et al*, 1998), HLA (Kaslow *et al*, 1996; Moore *et al*, 2002; McMichael and Klenerman, 2002), NRAMP1 (Searle and Blackwell, 1998), Toll-like receptor 4 (Arbour *et al*, 2000), mannose-binding protein (Garred *et al*., 1997), Interleukin-4 (Nakayama *et al*, 2002), and haptoglobin (Delanghe *et al*., 1998; Quaye *et al*., 2000; Pulgiese *et al*, 2002) have been related to variation in susceptibility to a number of infections, including HIV. These genes may be acting singly or in combination; their effect may be independent, additive or multiplicative. These may not be the only genes that are related to susceptibility to HIV/AIDS. It is possible that some more genes or genetic variants remain undetected.

As mentioned earlier, CC-chemokine receptor 5 (CCR5) is one of the co-receptor proteins for HIV (O'Brien and Dean, 1997). The most studied variation in this gene is a deletion of 32bp in the coding region (CCR5Δ32). This deletion shifts the open reading frame to create a truncated protein. This shortened protein version fails to reach the cell surface in homozygous individuals, leading to resistance to infection by HIV (O'Brien and Dean, 1997; O'Brien and Moore,

2000). Individuals who are heterozygous for this variant have reduced levels of CCR5 protein on the cell surface. This results in retarded rate of the disease progression (Williamson *et al*, 2000; Dean and O'Brien, 1997; O'Brien and Moore, 2000). This genetic variant is extremely rare in Africans (Williamson *et al*, 2000; Martison *et al*, 1997). There is another less common variant, CCR5m303, which also results in the introduction of premature stop codon (Quillent *et al*, 1998). This variant was not found in the South African population (Williamson *et al*, 2000).

Mutations at the promoter region of CCR5 have been found to either accelerate or retard the rate of the disease progression. CCR5 59029-G/G and CCR5P1/P1 genotypes respectively retard and accelerate progression to AIDS (Martin *et al*, 1998; McDermott *et al*, 1998).

CCR2 is another HIV co-receptor. A common variant is CCR2-641, which substitutes an isoluecine for a valine in the first transmembrane domain of CCR2 was found to delay AIDS by 2 to 4 years (Michael, 1997). This is the only known chemokine co-receptor variant that was significantly higher in Africans compared with Caucasians, and there was no significant association between CCR2-641 variant and HIV status in Africans (Williamson *et al*, 2000). Strong linkage disequilibrium has been demonstrated between CCR5Δ32 and CCR2-64I (Struyf *et al*, 2000) and between CCR2-64I and CCR5 59029A alleles (Hancock, 2002). CCR5P1/P1 and CCR5 59029A/A genotypes which have been shown to independently accelerate HIV/AIDS in Caucasians, were also found to be in complete linkage disequilibrium (An *et al*, 2000).

In an epidemiological study done to address validity of overdominant selection at the HLA class I loci entailed an analysis of HIV positive individuals (Carrington *et al*., 1999), it was found that heterozygosity at HLA class I loci confers relative resistance to AIDS progression because individuals who are homozygous at HLA-A, HLA-B and HLA-C, for example, present a limited repertoire of antigenic epitopes compared to individuals who are heterozygous at these loci. It has also been shown that individuals carrying HLA B27 and HLA B57 alleles have better prognosis after HIV infection than individuals carrying other HLA alleles (Moore *et al*, 2002; McMichael and Klenerman, 2002).

Mannose-binding protein (mannose-binding lectin) is a member of the collectin family of proteins and acts in the first line of defence against various bacterial, viral and parasitic infections, before the establishment of the adaptive immune protection by B and T cells (Turner, 1996). Polymorphisms in the promoter region of the mannose-binding protein gene (Madsen *et al*., 1995) and the first exon of this gene (Lipscombe *et al*., 1992) have been shown to affect serum concentration of mannose-binding protein. Low serum levels of this protein are associated with opsonozation defects and impaired phagocytosis (Super *et al*., 1989). A study done by Ezekowits *et al* (1989) demonstrated that mannose-binding protein was able to inhibit HIV.

This present study was focusing on the genetic polymorphisms in the haptoglobin gene which has been shown to be associated with the rate of the disease progression.

1.4 Haptoglobin

Haptoglobin is a type II acute phase plasma α_2-sialoglycoprotein (Langlios and Delanghe, 1996; Aucan *et al*, 2002). Haptoglobin is synthesized by hepatocytes (Smithies and Walker, 1995; Bowman, 1993) and there is evidence which suggests that haptoglobin originates from the organs of reproductive system including human uterus (O'Bryan *et al*, 1997; Olson *et al*, 1997; Sharpe-Timms *et al*, 2002). Transcription of the haptoglobin gene is induced by interleukin-6 (IL-6) in human hepatoma cells. This involves several nucleoprotein-DNA complexes associated with specific regions in the promoter region: -157, -111 and -61. These protein complexes are 1, 2, 3, 4, 5, 6 and V (Oliviero and Cortese, 1989). In the absence of IL-6, proteins responsible for the formation of protein complexes 1, 2, 3, 5 and/or 6 are bound around and including nucleotide positions -157 and -61, flanking the complex V at position -111 (Fig 1.5a). This conformation causes poor transcription of the haptoglobin gene. In the presence of IL-6, the IL-6-dependent proteins responsible for the formation of complex 4 substitutes the other proteins on site -157 and -61 (Fig 1.5b). This conformation results in the activation of transcription (Oliviero and Cortese, 1989; Teye *et al*, 2003).

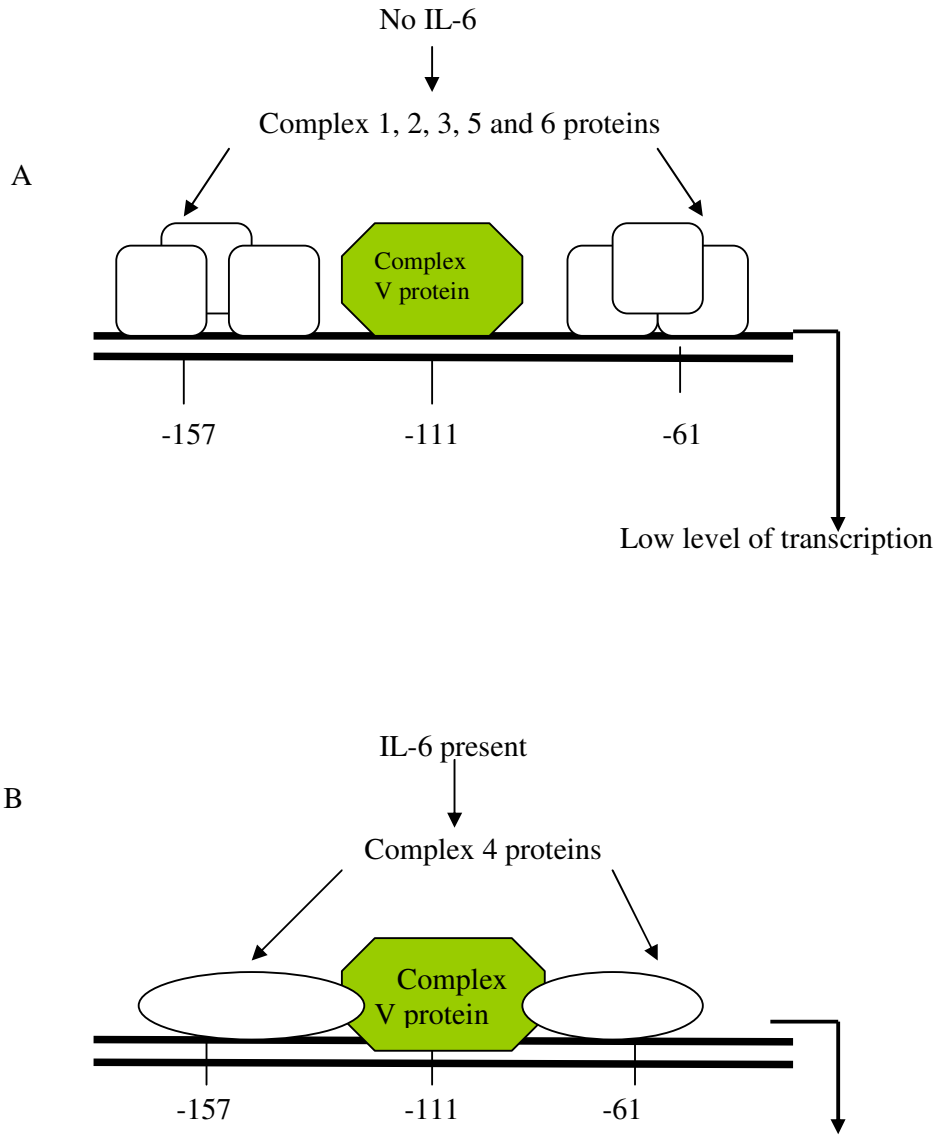


Fig 1.5. (A) In the absence of IL-6, haptoglobin gene is transcribed at a low level. (B) The presence of IL-6 triggers the substitution of the protein complexes at positions -157 and -61 by complex 4. This results in active transcription (Oliviero and Cortese, 1989).

1.4.1 Haptoglobin structure and types

The haptoglobin protein has a tetrameric structure consisting of 2 α and 2 β chains encoded by a single gene on chromosome 16q22.3. These chains are generated by posttranslational cleavage from a single polypeptide (Yang *et al*, 1983; Ranyuei *et al*, 1983). The β chains are identical in all individuals, while the α chains are polymorphic (Smithies *et al*, 1962; Langlois and Delanghe, 1996) and found only in humans (Black and Dixon, 1968; Teye *et al*, 2004; Langlois and Delanghe, 1996). Although haptoglobin is found in serum of all mammals, this polymorphism is only found in humans (Bowman, 1993). The protein polymorphism is due to two codominant alleles Hp^1 and Hp^2 which result to three common genotypes Hp1-1, Hp2-1 and Hp2-2. These genotypes give rise to structurally and functionally distinct phenotypes: Hp1-1, Hp2-1 and Hp2-2. Hp1-1 has α1 chains, Hp2-2 has α2 chains whereas Hp2-1 contains both chains (Fig 1.6A and 1.6B) (Koda *et al*, 2000; Schultze, 1996, Teye *et al*, 2002; Melamed-Frank, 2001). Hp^1 is further divided into Hp^{1S} and Hp^{1F}. These alleles encode polypeptides of equal length, Hp1S and Hp1F respectively (Giblett, 1969; Maeda and Smithies, 1985; Asakawa *et al*, 1999). The two polypeptides differ in the number of charged amino acids that make them migrate slowly (Hp1S) or fast (Hp1F) during electrophoresis with an acidic buffer. Lysine in Hp1F is substituted by glutamic acid in Hp1S at position 54 (Giblett, 1969; Asakawa *et al*, 1999).

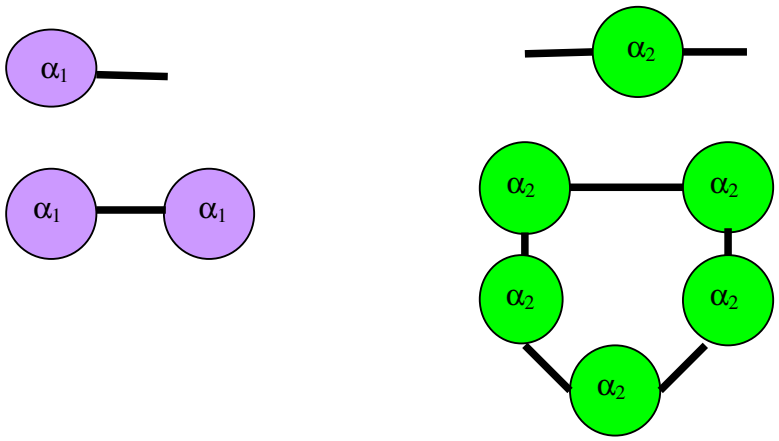


Fig 1.6A: Schematic representations of haptoglobin monomers and polymers. Hp1 monomer is a monovalent, meaning it can form a bond with one other monomer to form a dimer. Hp2 monomer is bivalent and can associate with 2 different monomers, which could result in circular structure in homozygous Hp2 individuals.

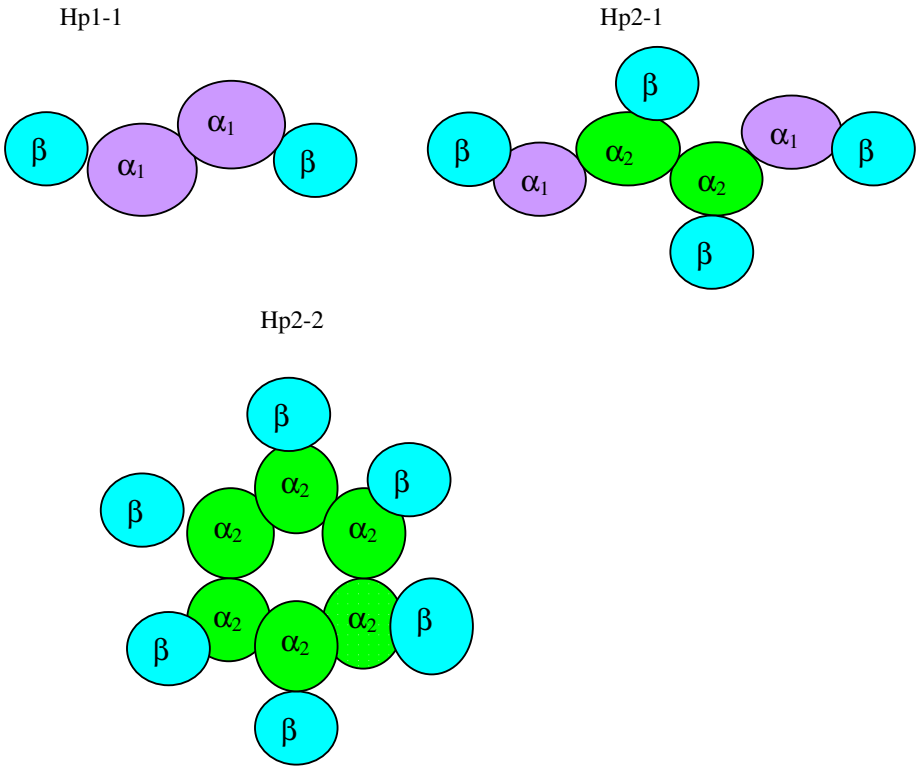


Fig 1.6B. Structural differences between haptoglobin types. The haptoglobin chains are held together by disulfide bonds (Melamed-Frank, 2001).

The α2 chain contains two free cysteine residues compared to one in the α1 chain, leading to polymerization in Hp2-1 and Hp2-2, while Hp1-1 is a small monomer (Fig 1.6A) (Wuyts *et al*, 2002).

The main difference between Hp^1 and Hp^2 alleles is the duplication of a 1 700 bp segment in Hp^2 but not in Hp^1 (Maeda and Smithies, 1986; Maeda *et al*, 1984). Each of the copies includes two of the exons encoding the α-chain of the haptoglobin (Koch *et al*, 2002). This duplication was formed by non-homologous cross-over between the Hp^{1S} and Hp^{1F} alleles (Fig 1.7) in a heterozygote during meiosis (Wuyts *et al*, 2002; Asakawa *et al*, 1999; Maeda *et al*, 1984; Maeda and Smithies, 1986). The crossing over occurred between the fourth intron of Hp^{1F} allele and the second intron of the Hp^{1S} allele, between regions 480 of Hp^{1S} and region 1271 of Hp^{1F} (Asakawa *et al*, 1999).

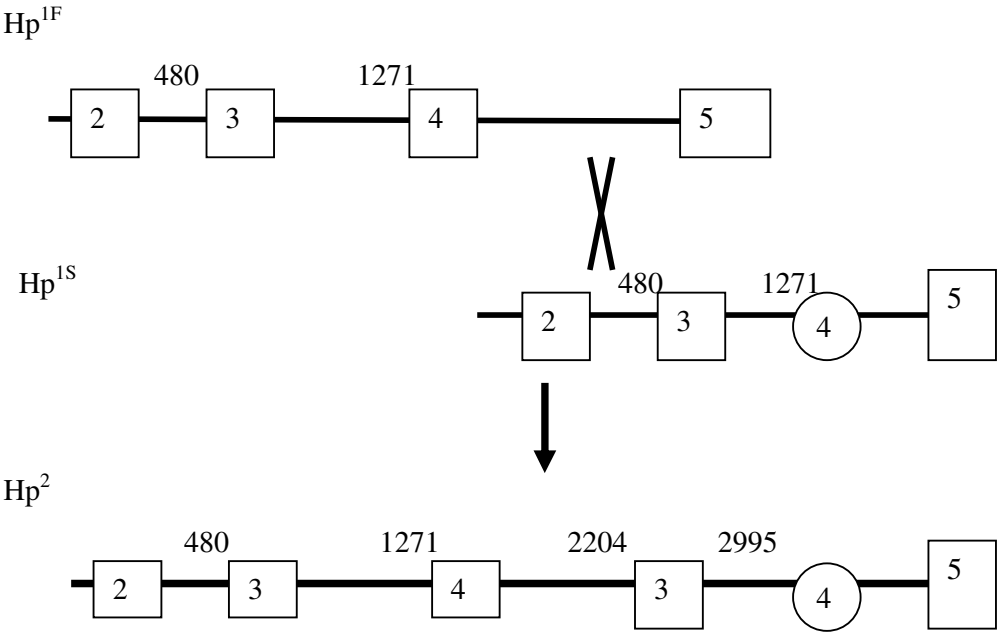


Fig 1.7: A diagrammatic representation of formation of Hp^2 allele by nonhomologous cross-over between Hp^{1F} and Hp^{1S}. The numbered blocks and circle represent exons. The numbers show position in the gene (Asakawa *et al*, 1999).

The haptoglobin phenotypes are distinguished by their band pattern on starch (Smithies, 1955) or polyacrylamide (Linke, 1984) gel electrophoresis. The homozygote Hp1-1 shows a single fast-migrating band of 86 kDa. The homozygote Hp2-2 has a series of slower-migrating bands, whose size is between 170 to 900 kDa. The heterozygote Hp2-1 displays another series of slow-moving bands and weak Hp1-1 band, the sizes of the bands range between 86 and 300 kDa (Wuyts *et al*, 2002; Michel *et al.*, 1996; Smithies, 1955). The slow-moving bands

are found only in humans. In other animals including higher primates, haptoglobin shows only a single band corresponding to the human Hp1-1 (Langlois and Delanghe, 1996).

Another haptoglobin variant, Hp2-1 modified or Hp2-1 mod, is common is black Americans, but it is also found in other races (Maeda *et al*, 1991). The variant contains either Hp1F or Hp1S chain and an α chain that is not distinguishable from Hp α2 chain, except by its fainter staining (Giblett, 1969). Hp2-1 mod is formed when the amount of Hp2 polypeptide synthesized in the Hp^2/Hp^1 heterozygote is less than that of Hp1 polypeptide. This unequal expression of the polypeptides is the result of polymorphism in the promoter region of haptoglobin gene. The polymorphism is a single base pair substitution (A/C) at position -61 in one of the interluekin-6 responsive elements of the haptoglobin gene (Maeda, 1991). The polymorphism has also been reported in Africans (Teye *et al*, 2003).

The fourth phenotype, Hp0, represents hypohaptoglobinemia or anhaptoglobinemia (Giblett, 1969). The Hp0 and Hp1 phenotypes in subSaharan African countries have been reported to be linked to a selection pressure by malaria parasite (Trape and Fribourg-Blanc, 1988). In Hp0 phenotype, the expression of haptoglobin is either absent (anhaptoglobinemia) or very low (less than 15 to 20 mg/100ml) to be detected by gel electrophoresis (hypohaptoglobinemia) (Giblett, 1969; Delanghe *et al*, 1998). Although Hp0 could be caused by pathological states such as liver dysfunction and haemolytic disorders, there is evidence that this phenotype has a genetic origin (Koda *et al*,

1998; Teye *et al*. 2003, Teye *et al*, 2004). In Asian populations Hp0 occurs by an allelic deletion. The deletion which is 28 kb (Fig 1.8) extends from the promoter region of the haptoglobin gene to the exon 5 of the Hpr. The corresponding allele is Hp^{del} (Koda *et al*, 1998).

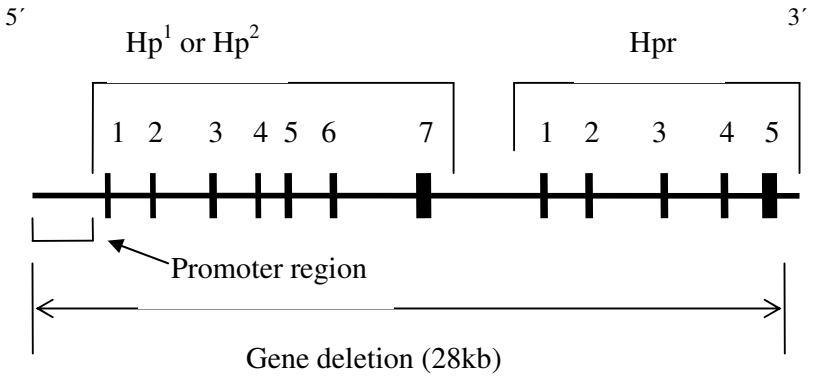


Fig 1.8: A schematic representation of the Hp^{del} breakpoints. The deletion extends from the promoter region to exon 5 of Hpr gene. The blocks represent exons and their numbers (Koda *et al*, 1998).

The homozygous gene deletion (Hp^{del}/Hp^{del}) is associated with haptoglobin deficiency or anhaptoglobinemia, whereas the heterozygous genotype, Hp^{2}/Hp^{del} - but not Hp^{1}/Hp^{del}, is associated with low haptoglobin concentrations or hypohaptoglobinemia (Koda *et al*, 1998).

Hp0 occurs at frequencies of 10 to 40% in SubSahara Africa (Constants *et al*, 1981). Unlike in Asian populations, the cause of Hp0 is not Hp^{del}, since this allele is not found in Africans (Koda *et al*, 2000; Teye *et al*, 2003). In Ghanaians the Hp^2 allele was found to be associated with anhaptoglobinemia. The absence of haptoglobin expression in Hp^2 allele was reported to be caused by a T to C point mutation in the exon 7 at position 6802, leading to the alteration of the codon 247 from ATT to ACT. The result of this mutation was an I247T substitution of a nonpolar amino acid isoleucine to a polar amino acid threonine (Teye *et al*, 2004). None of the anhaptoglobinemic individuals had an Hp^1/Hp^1 genotype. Hp^1/Hp^1 was present in the hypohaptoglobinemic individuals, predominantly as Hp^{1S} (Teye *et al*, 2003). Teye *et al* (2003) also reported that promoter sequences are associated with Hp0: -61C and -101G single nucleotide polymorphisms were associated with anhaptoglobinemia and hypohaptoglobinemia, respectively. The promoter sequence -101G was found in Ghanaians (Teye *et al*, 2003), but not in Americans (Maeda, 1991).

At the 3′-end of the haptoglobin gene cluster lies a haptoglobin-related gene, Hpr (Fig 1.8). It is a result of Hp1 gene duplication on chromosome 16, and is found only in apes and humans (Giblett, 1969; Maeda, 1984). The coding sequence of Hpr gene appears to be normal and to have no frameshift or splicing mutations (Maeda, 1985). Its predicted amino acid sequence differs by 8% from that of the Hp1F. The differences appear to be located on the surface of the protein molecule, and the regions and specific residues considered to be important for haemoglobin

binding are identical in Hp and Hpr proteins. Hpr gene in humans is 2.2kb downstream of Hp gene, and it is not expressed in humans (Bensi *et al*, 1985).

1.4.2 Variation in the haptoglobin promoter region

Some variation in the noncoding region of the haptoglobin gene has been found in African-Americans (Maeda, 1991) and Ghanaians (Teye *et al*, 2003). As genetic variation is higher in Africans than in other populations, Teye *et al* (2003) found three polymorphisms in the promoter region (-242C/T, -191T/G and -101C/G) in addition to those previously reported (-55A/G, -61A/C and -104T/A) (Maeda, 1991). Some of these polymorphisms are linked with certain phenotypes. The -61A/C base substitution is associated with Hp2-1 mod phenotype (Maeda, 1991). As mentioned earlier that position -61 is one of the IL-6 responsive elements, the A/C base substitution explains the unequal synthesis of Hp1 and Hp2 polypeptides in Hp2-1 mod protein. None of the promoter sequences were linked to Hp0 in African-Americans (Maeda, 1991). By contrast, the -61C and -101G promoter sequences showed strong association with ahaptoglobinemia and hypohaptoglobinemia, respectively (Teye *et al*, 2003).

1.4.3 Functions of the haptoglobin

Haptoglobin has several functional properties of biological and pathological importance. Haptoglobin types differ in their functional properties (Langlios and Delanghe, 1996).

1.4.3.1 Haemoglobin binding

The best known function of haptoglobin is haemoglobin binding. After erythrocyte destruction, free haemoglobin is not filtered through the glomeruli because of the bound haptoglobin. This reduces the risk of renal damage and the loss of haemoglobin and iron. The Hb-Hp complex is transported to the liver, where it is broken down in the parenchymal cells by lysosomes. The binding property of haptoglobin also prevents accumulation of free radicals generated by iron Fe^{2+} in the presence of H_20_2 (van Vlierberge *et al*, 2004). The binding of haemoglobin to haptoglobin is dependent of the serum haptoglobin concentration and haptoglobin type. Hp2-2 phenotype clears haemoglobin circulating in the plasma with less efficiency than Hp1-1 and Hp2-1 (Langlois and Delanghe, 1996).

1.4.3.2 Bacteriostatic effect

Due to free haemoglobin capture by haptoglobin, heme iron is unavailable for bacterial growth. This indicates the role of haptoglobin as part of non-specific defence against bacterial invasion (Langlois and Delanghe, 1996). Haptoglobin also plays a role in repairing tissue injuries. Exposure of the lung to chemicals, organic and inorganic dust makes it vulnerable. Haptoglobin synthesized in the lung provides a source of anti-oxidant in the mucous blanket that protects the lung (Dobryszycka, 1997)

1.4.3.3 Agglutination effect of the haptoglobin

Hp2-2 and Hp2-1 proteins have an ability to agglutinate antigen T4 carried by *S. pyogenes* group A. Hp2-2 serum has higher agglutination titer than Hp2-1.

However, haptoglobin is not a true antibody, and it does not activate complement. The agglutination effect of Hp2-1 and Hp2-2 is probably mediated via binding with lectin-like structure. In contrast, Hp1-1 has no agglutination effect (Langlois and Delanghe, 1996).

1.4.3.4 Inhibition of prostaglandin synthesis

Haptoglobin functions as prostaglandin synthetase inhibitor. The inhibitory effect of haptoglobin on prostaglandins synthesis is important in anti-inflammatory action. The inhibitory effect of Hp2-2 and Hp2-1 on prostaglandin synthesis is less strong than that of Hp1-1 phenotype (Dobryszycka, 1997; Langlois and Delanghe, 1996).

1.4.3.5 Angiogenesis

Haptoglobin has been reported to stimulate angiogenesis, the processes leading to the generation of new blood vessels through sprouting from already existing blood vessels. The angiogenic effect of Hp2-2 is more pronounced than that of Hp2-1 and Hp1-1 types (Langois and Delanghe, 1996).

1.5 Involvement of haptoglobin types with HIV/AIDS and other diseases

Haptoglobin types differ in their biological activities, and they may influence the course of a disease (Langlois and Delanghe, 1996). Haptoglobin polymorphism has been reported to be associated with infection by TB (Fedoseeva *et al*, 1993) and outcomes after treatment initiation (Kasvove *et al*, 2000). Hp2-2 phenotype

was found to be associated with recurrent pulmonary TB in comparison with Hp1-1 and Hp2-1 (Eisaev, 1995).

Fedoseeva *et al* (1993) reported that Hp2-2 individuals with TB carrying an HLA-DR2 exhibited large cavities due to tissue destruction, advanced dissemination and fast disease progression. In contrast, a study done in Zimbabwe showed no association between haptoglobin phenotypes and susceptibility to clinical pulmonary TB, however, haptoglobin genotypes had an effect on the outcome of TB after initiation of treatment. A high number of Hp2-2 carriers died compared to the other haptoglobin phenotypes (Kasvove *et al*, 2000).

A high incidence of Hp0 was observed among malaria patients in Indian (Joshi *et al*, 1987) and Republic of the Congo (Trape and Fribourg-Blanc, 1988) populations, suggesting that anhaptoglobinemia is associated with malaria. Unlike in an Indian population, in a Sudanese population Hp1-1 phenotype was found to be associated with susceptibility to malaria and development of severe complications. The other phenotypes were found to confer resistance (Elagib *et al*, 1998). The two above studies show that haptoglobin polymorphism was in some way associated with susceptibility to malaria. By contrast, no significant association between haptoglobin polymorphism and susceptibility to malaria was found in Gambia (Aucan *et al*, 2002).

Haptoglobin variation has been shown to be associated with diabetes, atherosclerosis and cardiovascular diseases (van Vlierberghe *et al*, 2004). Coronary artery lesions and target organ damage in hypertension are commoner

among Hp2-2 carriers. For hypertension treatment, Hp2-2 individuals need a more complex combination of antihypertension drugs than Hp1-1 individuals (van Vlierberghe *et al*, 2004). Diabetic patients who are homozygous for Hp^1 allele are protected against the development of vascular complications. It has been proposed that the specific interaction between diabetes, cardiovascular disease and haptoglobin phenotypes is the result of the impaired clearing capacity of glycosylated haemoglobin-haptoglobin complexes from the subendothelial space. A delay in the clearing of these complexes result in oxidation of low-density lipoproteins to atherogenic oxidized low-density lipoproteins (Asleh *et al*, 2003)

Haptoglobin polymorphism has been reported to be associated with prognosis in HIV infection (Delanghe *et al*, 1996; Quaye *et al*, 2000). This effect of haptoglobin polymorphism on HIV progression to AIDS and death is related to iron stores in the body. As mentioned earlier, haptoglobin functions to capture and clear free haemoglobin from the plasma. Hp2-2 binds haemoglobin with less affinity, resulting in the retention of iron in Hp2-2 individuals (Delanghe *et al*, 1996). All organisms, large and small, require iron to a certain concentration (Weinberg, 1978). In humans excess iron has harmful consequences. During iron overload, excess iron is stored in parenchymal cells and macrophages (Bothwell *et al*, 1960). This causes direct cytotoxicity, enhancement of infection and increased oxidative stress (Weinberg, 1990). Excess iron also accelerates oxidative catabolism of ascorbic acid (vitamin C), leading to deficiency of this vitamin (Kasvove *et al*, 2002). Vitamin C provides antioxidant protection because is a free radical scavenger. In the presence of excess iron, vitamin C has a pro-oxidant

activity. Iron gets trapped within ferritin as Fe^{3+} and enters the pores of the ferritin where it is converted to Fe^{2+}; vitamin C becomes oxidized in the process. Fe^{2+} then leaks out of the ferritin protein and generate free radicals (Delanghe *et al*, 2002). In HIV infection, if free haemoglobin is not removed from the plasma, free radicals generated activate HIV replication through the activation of the nuclear transcription factor NF-κB, an element in the modulatory region of HIV LTR and thus results in increased HIV-RNA in the body (Boelaert *et al*, 1996). The mechanism by which free radicals activate NF-κB (Fig 1.9), is through either degradation or modification of IκB (a cytoplasmic protein that inhibits NF-κB), that results in its dissociation from the p50-p65 complex of proteins that bind and activate NF- κB. After the dissociation of the inhibitor from the p50-p65 complex, the complex is translocated to the nucleus and binds to the NF-κB. This activates the NF-κB factor and subsequently HIV gene expression (Gaynor, 1992).

Harakeh *et al* (1990) found that vitamin C suppresses HIV reverse transcriptase activity and viral replication in HIV-infected cells. Therefore, the combination of oxidative stress and vitamin C deficiency may contribute to a high viral load and poor prognosis in Hp2-2 individuals. Lower vitamin C concentrations have been also observed in healthy individuals with Hp2-2 phenotype compared with the other haptoglobin phenotypes (Delanghe *et al,* 1998). This shows that Hp2-2 carriers are less protected against haemoglobin/iron-driven oxidation.

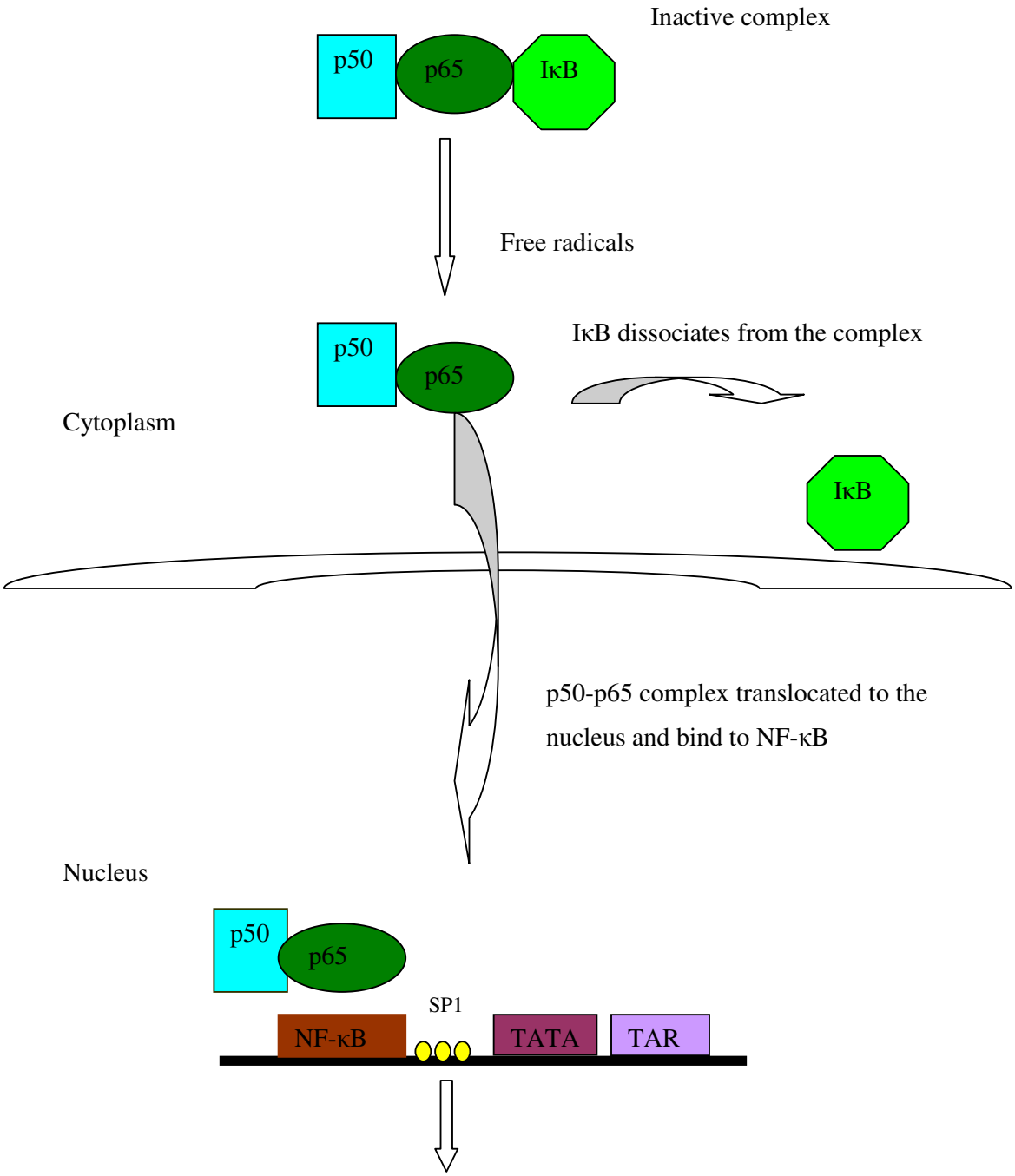


Fig 1.9: Schematic representation of NF-κB activation. Dissociation of IκB from the p50-p65 complex results in NF-κB activation and subsequent activation of HIV gene replication (Gaynor, 1992).

1.6 Haptoglobin frequencies in South Africans

The South African black population has nine different subpopulations which are distributed across all the nine provinces. These groups are Zulu, Swazi, Xhosa, Ndebele (collectively known as Nguni), Tswana, Southern Sotho, Nothern Sotho (known as Sotho-speakers), Venda and Tsonga. Although certain groups dominate in certain provinces, Johannesburg (in Gauteng province) is a home for all nine subpopulations. The South African black ethnic groups originate from Bantu expansion and share a common ancestor (Lane *et al*, 2001); however, there is little genetic differentiation among them. The analysis of autosomal DNA and Y chromosome polymorphisms revealed that subpopulations that fall under Nguni are genetically closely related. The Sotho-speakers formed another cluster, and Venda showed close similarity with Tsonga (Lane *et al*, 2001).

The gene frequencies of haptoglobin show marked geographic differences, with lowest Hp^1 allele frequency in Southeast Asia and the greatest frequency in Africa and South America (Schultze and Heremans, 1966). However, these Hp^1 gene frequencies are significantly different from higher frequency found among West African blacks, and there is evidence suggesting that as one passes south from the Congo through Zambia to South Africa, a decrease in the frequency of this gene might be demonstrable (Jenkins and Steinberg, 1966). A very large number of papers have been written about the haptoglobin allele frequencies in various populations throughout the world. Jenkins and Steinberg (1966) and Barnicot *et al* (1959), found that Hp^1 gene frequency is 0.47 in Cape Coloureds (Barnicot *et al*., 1959); 0.29 and 0.31 in South African Bushmen according to Barnicot *et al* (1959)

and Jenkins and Steinberg (1966), respectively. In a recent study by Koda *et al*., (2000), no Hp^{del} was observed in a group of Xhosa and European-Africans based in Cape Town. Haptoglobin allele frequencies in South African black populations are shown in table 1.1.

Table 1.1. Distribution of haptoglobin alleles in black South African populations.

Population	Hp^1 frequency	Hp^2 frequency	Reference
Zulu	0.505-0.575	0.425-0.495	Jenkins, (1972); McDermid and Vos, (1971a); Hitzeroth and Hummel, (1978); Nurse *et al*, (1974)
Swazi	0.475-0.545	0.525-0.455	Jenkins (1972); Hitzeroth and Hummel (1978)
Ndebele	0.440-0.480	0.520-0.560	Jenkins (1972); Hitzeroth and Hummel (1978)
Xhosa	0.491-0.535	0.509-0.465	Jenkins (1972); Weissman *et al*, (1982); Giblett *et al* (1966)
Venda Urban	0.555-0.586	0.415-0.445	Jenkins (1972); Hitzeroth and Hummel (1978); Jenkins (1972); Nurse *et al* (1985)
Venda Rural	0.541-0.553	0.456-0.447	Jenkins (1972) and Nurse *et al* (1985)
N. Sotho Urban	0.585-0.541	0.415-0.459	Jenkins, (1972) and Nurse *et al*, (1974).
N. Sotho Rural	0.493-0.500	0.507 and 0.500	Jenkins (1972) and Hitzeroth and Hummel (1978)
Tswana	0.493	0.507	Hitzeroth and Hummel (1978)
S. Sotho	0.511-0.545	0.455-0.489	Giblett *et al* (1966); Moullec *et al* (1966)
Tsonga	0.546-0.555	0.445-0.454	Hitzeroth and Hummel (1978); Jenkins (1972); Matznetter and Spielmann (1969)

Although the statistics given above shows that Hp^1 is low in Swazi and Ndebele groups compared with the other South African populations, the difference in the frequencies among South African blacks is small, 0.1 and less. Although their

languages differ, they originate from the Bantu expansion, which explains the genetic similarity. The difference in frequencies between black South Africans and Bushmen is huge. Bushmen show genetic similarity with Khoi. This similarity could be due to that they share a common ancestor or due to genetic flow (Nurse *et al*, 1985).

1.7 Problem identification

Genetic diversity is higher within African populations, than in any other human populations in the world. The diversity is observed even between closely related or located groups (Jorde *et al*, 1997, Tishkoff, 2002). It has been mentioned earlier that variation in the coding region of the haptoglobin gene was found to be associated with the rate of HIV/AIDS progression in Europeans. Similar studies have not been done in black South African population. South Africa has the highest number of people living with HIV/AIDS. In 2001 alone, 360 000 people died of this disease. In this study we looked at the impact of variation in the coding region of the haptoglobin gene on the rate of the disease progression in the black South Africans population. Because of high genetic diversity in Africans, we further detected variation in the upstream noncoding (promoter) region of this gene and determined whether the previously reported and additional polymorphisms found in black South Africans in this region have any impact on the rate of HIV/AIDS progression.

CHAPTER 2

MATERIALS AND METHODS

2.1 Subjects

The study population comprises of 163 black South Africans with proven HIV infection and 52 samples from a general population taken regardless of HIV status. One hundred HIV positive and two HIV negative blood samples along with clinical data were collected from willing participants from Johannesburg General Hospital. Clinical data collected was self reported and included an estimate of number of years after HIV infection, most recent $CD4^+$ cell count, whether the participant had ever had TB or not or any other HIV-related diseases. About 80% of our participants were woman who had tested positive when they were pregnant. They were asked their HIV status on their previous pregnancies or HIV tests to get a close estimate of the number of years after infection. Sixty three DNA samples from HIV positive people with viral load and $CD4^+$ count data were provided by Dr Clive Gray from NICD. Forty blood samples were collected regardless of HIV status from willing black students and staff members of the University of the Witwatersrand. Ten DNA samples with unknown HIV status were provided by Prof Himla Soodyall from the NHLS. Written informed consent to participate in this study was obtained from all participants. The study protocol was approved by the University of the Witwatersrand Committee for Research on Human Subjects, protocol M040221 (Appendix). Blood was aseptically collected into ethylenediaminetetraacetic acid (EDTA) tubes.

2.2 DNA isolation

The blood tubes were centrifuged at 80 rpm for 10 minutes to separate plasma, buffy coats (leukocytes) and whole blood. Plasma was stored at -70^{o}C, and later used for determination of protein polymorphism. DNA was extracted from the leukocytes of the blood samples using QIAmp® Blood DNA kit according to manufacturer's instructions (Qiagen). DNA extraction was carried out in QIAamp spin columns which have DNA-adsorbing silica-gel membrane. RNase was used to eradicate any traces of RNA or virus that might be in the DNA. The DNA adsorbed onto the membrane was eluted in Tris-EDTA (TE) buffer consisting of 10 mM Tris and 1.0 mM EDTA and stored at -20^{o}C. After purification, the concentration and size of the DNA was checked by electrophoresis on 0.8% agarose gel in TBE buffer at 70V for 45 minutes. The TBE buffer contained 0.9 M Tris, 0.89 M boric acid and 25 mM EDTA. This DNA was used for characterization of variation in the haptoglobin promoter and coding regions.

2.3 Detection of variation in the promoter region of haptoglobin gene by direct sequencing

Variation in the promoter region of the haptoglobin gene was detected in 52 samples collected from 42 HIV positive and 10 individuals with unknown HIV status. The detection was done by PCR amplification and direct sequencing of a 645 bp fragment of the promoter region. This sample size was large enough to detect genetic variation that occurs frequently in the population.

Polymerase chain reaction was carried out in a 50µl reaction volume as described by Maeda (1991), except that amplification programmes were slightly modified.

The reaction mixture contained 1.25 units of Taq polymerase, 200μM of dNTPs, 1.5mM $MgCl_2$, 0.2μM of each of the primers and 200ng of template DNA. The primers used have the following sequences:

5'-ACTATAAAACCATGAGAACCAC-3' (forward primer) and

5'-CCTCATCTTGGTTGGTCTTGC-3' (reverse primer). The primers were synthesized by Inqaba Biotec, and were dissolved in Tris-EDTA consisting of 10 mM Tris and 1.0 mM EDTA.

The amplification consisted of predenaturation at 94^{o}C for 5 minutes, followed by 35 cycles of denaturation at 94^{o}C for 30 seconds, annealing at 55^{o}C for 45 seconds and extension at 72^{o}C for 60 seconds, followed by final extension at 72^{o}C for 5 minutes. The presence of the desired product was checked on 1% agarose gel in TBE buffer at 70V for 45 minutes. The 645 bp PCR product was purified and sequenced in both directions by Inqaba Biotec. Sequencing primers used have the following sequences:

5'-CATGAGAACCACTGCCATTG-3' (forward) and

5'-CTTGCCTCTGGAAGAGCAG-3' (reverse). Sequences were aligned with the haptoglobin reference sequence from GeneBank using computer software Sequencher™ 4.0.

2.4 Alternative methods for detecting single nucleotide polymorphism

Allele-specific amplification (ASA) was used for the detection of single nucleotide polymorphism (SNP) found by direct sequencing, which is applicable to a large number of samples and less expensive than direct sequencing.

The method uses one set of primers for each allele of the single nucleotide polymorphism (Okayama *et al*, 1989). The primers are designed such that the 3'-end of one primer in a primer set matches allele A, and another primer in the other set with the 3'-end matching B allele (Fig 2.1). Homozygous samples for A (or B) will yield a PCR product only with A-specific (or B-specific) primer. Heterozygous samples (AB) will yield a product with both primers. PCR was ran at stringent annealing temperatures for specific binding (Okayama *et al*, 1989). The ASA was optimized by using samples with known nucleotide sequences. A control was included during optimization in order to conclude with certainty that the PCR amplification was allele-specific under those amplification conditions. For the -295T allele, optimization did not include a control as only TT and TC genotypes were found by direct sequencing. In that case, the amplification was done at the highest possible annealing temperature.

Based on the polymorphisms obtained by direct sequencing, allele-specific primers for the four most common polymorphisms were designed (Fig 2.1). PCR was carried out in a 20 µl reaction volume containing 1.25 units of Taq polymerase, 200 µM of dNTPs, 1.5 mM MgCl, 0.2 µM of each of the primers and 1 µg of template DNA. The thermocycling procedure consisted of predenaturation at 94°C for 5 minutes, 35 cycles of denaturation at 94°C for 1 min, annealing for 45 seconds at different temperatures due different GC contents (Table 2.1) and extension at 72°C for 1 minute, followed by a final extension at 72°C for 7 min. The presence of the PCR products was verified on 1 to 1.7% agarose gels, depending on the size of the expected fragment. Electrophoresis was carried out in TBE buffer for 45 minutes at a constant voltage of 75V. The gels were then

visualised under the UV light using the image analyser.

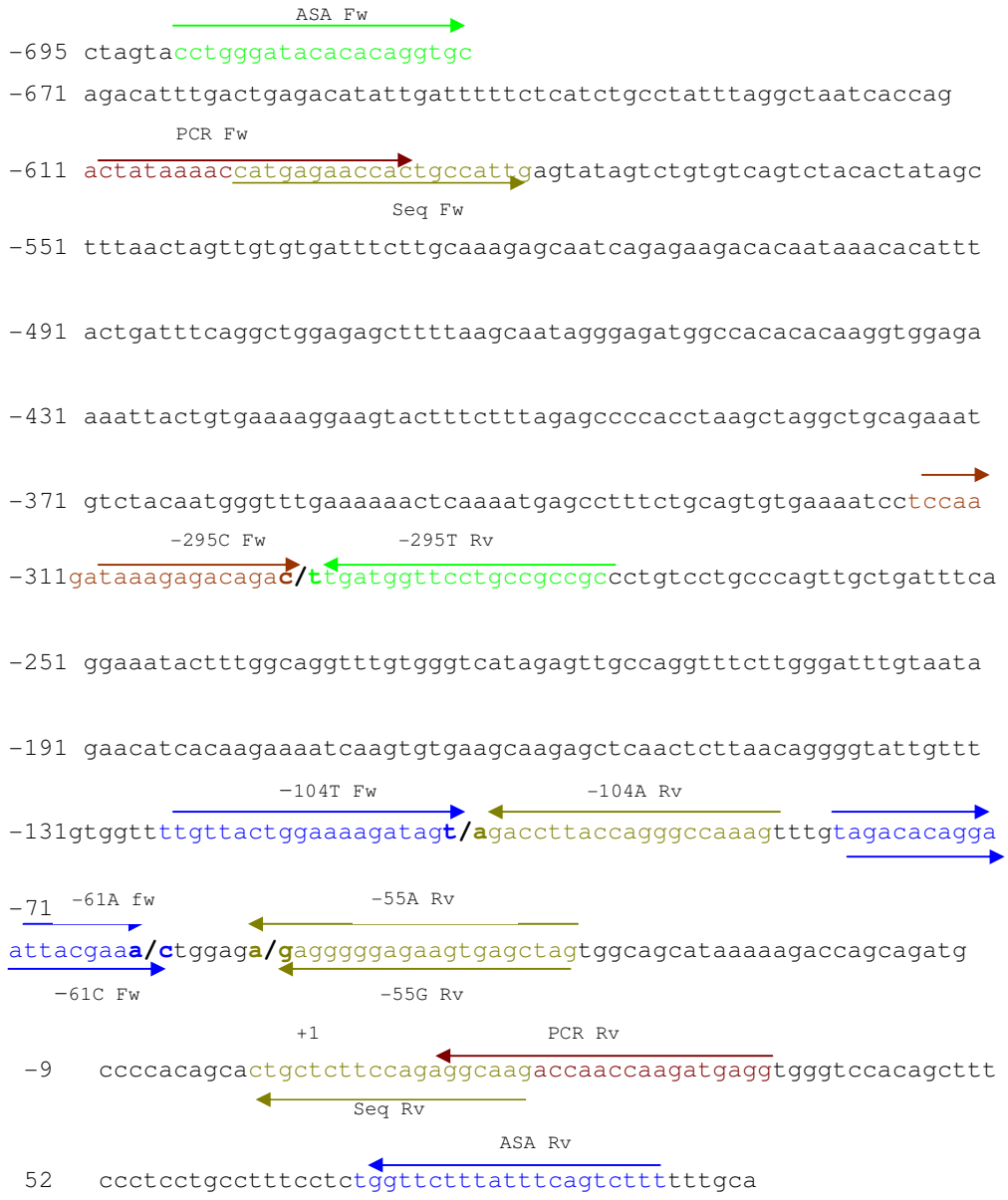


Fig 2.1. Nucleotide sequence of the promoter region of the haptoglobin gene. The arrows show the binding sites of the primers and the directions in which they extend. The nucleotide bases in bold are the single nucleotide polymorphisms.

Table 2.1. Primer set sequences used in ASA, annealing temperatures for the primers and sizes of the products. The nucleotide bases in bold at the 3´-end are allele-specific.

Primer name	**Primer sequence**	**SNP**	**Annealing temp**	**Product size**
-295T Rv	5'GCGGCGCAGGAACCATCA**A**-3'	-295T/C	68°C	416 bp
ASA Fw	5'-CCTGGGATACACACAGGTGC -3'			
-295C Fw,	5'-TCCAAGATAAAGAGACAGA**C**-3'	-295T/C	56°C	350 bp
PCR Rv	5'-CTTGCCTCTGGAAGAGCAG-3'			
-104A Rv	5'-CTTTGGCCCTGGTAAGGTC**T** -3'	-104A/T	60 °C	517 bp
Seq Fw	5'- CATGAGAACCACTGCCATTG -3'			
-104T Fw	5'-TTGTTACTGGAAAAGATAG**T**-3'	-104A/T	55°C	212 bp
ASA Rv	5'- AAAGACTGAAATAAAGAACCA 3'			
-61A Fw	5'-TAGACACAGGAATTACGAA**A**-3'	-61A/C	54.4°C	168 bp
ASA Rv	5'- AAAGACTGAAATAAAGAACCA - 3'			
-61C Fw,	5'-TAGACACAGGAATTACGAA**C** -3'	-61A/C	60°C	168 bp
ASA Rv	5'- AAAGACTGAAATAAAGAACCA -3'			
-55A Rv	5'-CTAGCTCACTTCTCCCCCT**T**-3'	-55A/G	60 °C	547 bp
Seq Fw	5'- CATGAGAACCACTGCCATTG -3'			
-55G Rv	5'-CTAGCTCACTTCTCCCCCT**C**-3'	-55A/G	61 °C	547 bp
Seq Fw	5'- CATGAGAACCACTGCCATTG -3'			

2.5. Protein polymorphism

2.5.1 Haptoglobin genotype determination

Exploiting the known size difference between Hp^1 and Hp^2 alleles (Fig 2.2), specific alleles were amplified by PCR with one pair of primers (Koch *et al*, 2002). The primers used were 5'-GAGGGGAGCTTGCCTTTCCATTG-3' (primer A) and 5'-GAGATTTTTGAGCCCTGGCTGGT-3' (primer B). The primers bind to sites that are not specific to the alleles, upstream and downstream of the allele-specific regions (Fig 2.2).

Depending on the genotype presented by the template, an Hp^1/Hp^1 product of 1757 bp, a 3481 bp product of Hp^2/Hp^2, or both bands in Hp^2/Hp^1 were generated by PCR using primer set AB. In most cases, the Hp^2-specific 3481 bp product was very faint or invisible in the presence of the 1757 bp band. In such cases another set of primers, CD, was used to amplify a 349 bp Hp^2 allele-specific sequence to verify the absence of the 3481 bp product (Fig 2.2). The primers have the following sequences: 5'-CCTGCCTCGTATTAACTGCACCAT-3' (primer C) and 5'-CCGAGTGCTCCACATAGCCATGT-3' (primer D). The primers were synthesized by Inqaba Biotec and were diluted in TE. PCR was carried out in a 20 µl reaction volume containing 1.25 units of Taq polymerase, 200 µM of dNTPs, 1.5 mM MgCl, 0.2 µM of each of the primers and 1 µg of template DNA. After initial denaturation at 95°C for 2 min, the thermocycling procedure consisted of denaturation at 95°C for 1 minute, annealing 69°C for 2 min and extension at 69°C for 2 min, repeated for 35 cycles, and followed by a final extension at 72°C for 5 min (Koch *et al*, 2002). The PCR products were separated on 1% agarose gel containing 0.5 µg/ml ethidium bromide in TBE buffer at 70V for 1 hour.

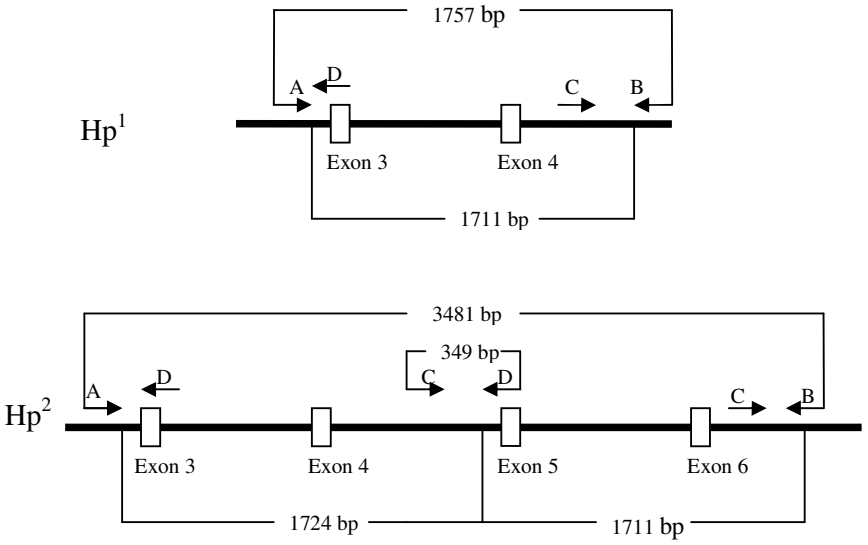


Fig 2.2. Structure of haptoglobin alleles Hp^1 and Hp^2. The arrows show the direction in which the primers extend. The letters A, B, C and D are the names of the primers; they show the sites where they bind (Koch *et al*, 2002)

2.5.2 Phenotype determination

The haptoglobin genotypes were confirmed by phenotype determination. The plasma samples were first treated with nevirapine (0.02mg/L) to deactivate any HIV that might be in the samples. Haptoglobin phenotypes were determined from 12 µl of haemoglobin-enriched plasma by polyacrylamide gel electrophoresis (PAGE) followed by peroxidase staining (Linke, 1984; Koch *et al*, 2002). The haemoglobin stock solution (10%) was prepared from an HIV negative sample. After the removal of plasma, the blood sample was spun at 2500 rpm for 10 minutes. The sedimented blood cells were washed five times, each time in 10 ml of phosphate buffer saline (PBS) containing 0.1 M sodium phosphate and 0.075 M

NaCl, pH 7.2. One millilitre of packed cells was lysed for 30 minutes with 9 ml of distilled water. The solution was centrifuged at 10 000 g for 1 hour, and the supernatant (haemoglobin solution) was kept in aliquots at -70^{o}C. Ten microliters of plasma was mixed with 2 µl of 10% haemoglobin solution, and the samples were incubated at room temperature for 5 minutes to permit haptoglobin-haemoglobin (Hp-Hb) complexes to form. An equal volume of sample buffer containing 0.125 mol/L Tris (pH 6.8), 200 g/L glycerol and 0.01g/L bromophenol blue was added to each sample before loading and electrophoresis. The Hp-Hb complexes were resolved by electrophoresis on continuous 4.7% polyacrylamide gel in Tris-Borate buffer containing 0.05 M Tris and 0.023 M boric acid, pH 8.7, using Tris-Borate buffer. Electrophoresis was performed at a constant voltage of 250 for 2.5 hours.

After electrophoresis, the Hp-Hb complexes were visualised by soaking the gel in the fresh peroxidase staining solution containing 0.14 M phosphate –citrate buffer, 0.3% guaiacol and 0.03% hydrogen peroxide for about 15 minutes.

2.6. Data analysis

2.6.1. Determination of allele and genotype frequencies

Genotype and allele frequencies were determined by gene counting. This was done by counting the total number of each genotype and allele in the population, and determined the proportion of each in the population (Hartl and Clark, 1989).

If *x, y* and *z* respectively mean the number of individuals carrying AA, AB and BB genotypes in a population, and the total sample size is *n*, then:

$$\text{Frequency of AA} = \frac{x}{n}$$

$$\text{Frequency of BB} = \frac{y}{n}$$

$$\text{Frequency of AB} = \frac{z}{n}$$

The A and B alleles' numbers in the population were deduced from the number of individuals carrying genotypes AA, AB and BB, and the frequencies were calculated:

Number of A alleles = $2x + y$

Number of B allele = $2z + y$

The total number of alleles is $2n$ because each individual carries two alleles, then:

$$\text{A freq} = \frac{2x + y}{2n} = p$$

$$\text{B freq} = \frac{2z + y}{2n} = q$$

2.6.2. Hardy-Weinberg equilibrium

To determine whether the population was in Hardy-Weinberg equilibrium, expected and observed genotypes numbers were compared. The expected genotypes numbers were calculated from allele frequencies using the Hardy-Weinberg equation: $p^2 + 2pq + q^2 = 1$

If the population is in Hardy-Weinberg equilibrium, AA, AB and BB frequencies would be p^2, 2pq and q^2 respectively (Crow, 1986; Hartl and Clark, 1989).

The observed and the expected numbers were compared using the chi-squared (χ^2) test:

$$\chi^2 = \Sigma \frac{(o-e)^2}{e}$$

where o = observed number

e = expected number.

The χ^2 test determines whether observed numbers are significantly different from the expected numbers (Crow, 1986). A P value of below 0.01 was considered to indicate a lack of significant deviation from the predicted Hardy-Weinberg distribution.

2.6.3. Linkage equilibrium and disequilibrium

Linkage disequilibrium measures association of specific alleles at different loci on the same chromosome (Nei, 1987). Pairwise linkage disequilibrium was analysed using a computer program called Linkage Disequilibrium Analyzer (LDA) version 1.0 (Keyue *et al*, 2003).

If there are two alleles on the same chromosome:

Allele 1 ——— A_1 ——— B_1 ———

Allele 2 ——— A_2 ——— B_2 ———

Let the allele frequencies of A_1, A_2, B_1 and B_2 be y_1, y_2, y_3 and y_4, respectively. The possible genotypes are A_1B_1, A_1B_2, A_2B_1 and A_2B_2. If x_1, x_2, x_3 and x_4 are the frequencies of these genotypes, respectively, then:

The frequency of gametes that carry $A_1 = x_1 + x_2$

The frequency of gametes that carry $B_1 = x_1 + x_3$

The frequency of gametes that carry $A_2 = x_3 + x_4$

The frequency of gametes that carry $B_2 = x_2 + x_4$

Linkage disequilibrium was calculated using the formula that compares the gametic frequency with the product of allele frequencies (Nei, 1987; Hartl and Clark, 1989):

$$D = x_1 - y_1y_3$$

Where $x_1 = A_1B_1$ frequency, $y_1 = A_1$ frequency and $y_3 = B_1$ frequency.

If $D > 0$, allele A_1 is associated with B_1 and A_2 with B_2

If $D = 0$, different alleles are randomly associated at different loci, and the alleles are in linkage equilibrium

If $D < 0$, A_1 is associated with B_2 and A_2 is associated with B_1 (Nei, 1987; Lewontin, 1988; Hartl and Clark, 1989).

The degree of the non-random association of the alleles was measured using the formula:

$$D' = \frac{D}{D_{max}}$$

where D = linkage disequilibrium, D_{max} is the largest value D can take with given marginals (Nei, 1987; Lewontin, 1988; Hartl and Clark, 1989).

The largest value D can be is either y_1y_4 or y_2y_3, while the most negative value can be either y_1y_3 or y_2y_4, thus:

$$D_{max} = \min (y_1y_3, y_2y_4) \text{ when } D<0$$

and

$$D_{max} = \min (y_1y_4, y_2y_3) \text{ when } D>0$$

The value of D´ ranges between -1 and 1 (Lewontin, 1988).

2.6.4. Haplotype analysis

A haplotype is a set of SNP alleles along a region of a chromosome. If a chromosome has two SNP alleles, A/G and C/T, the four possible haplotypes for these alleles are AC, AT, GC and GT (Tregouet *et al*, 2004). Haplotypes and haplotype frequencies were determined using Arlequin version 2.000 software which implements expectation maximization (EM) algorithm (Schneider *et al*, 2000).

2.6.5 Determination of genotype and phenotype /disease status association

To examine whether variation in the coding and noncoding regions of the haptoglobin gene are associated with susceptibility to infection by HIV and TB, and the rate of HIV/AIDS disease progression, the study population was divided into groups and subgroups.

For susceptibility to HIV and TB, the study population was divided into:

(i) HIV positive and general population (with unknown HIV status)

(ii) With TB and without TB.

To examine whether the polymorphisms have an impact on the rate of HIV/AIDS disease progression, the HIV positive group was divided into the following groups:

(i) A group that was diagnosed with HIV 5 years or less prior to the collection of the samples and were symptomatic and the group that had been HIV positive for more than 5 years and were asymptomatic

(ii) A group that was diagnosed with HIV 5 years or less prior to the collection of the samples and had a $CD4^+$ cell count of $\leq$200 cells/ml^3 and the group that had been HIV positive for more than 5 years and had a $CD4^+$ cell count of greater than 200 cells/ml^3.

(iii) Rapid progressors (those who were diagnosed with HIV 5 years or less prior the collection of the samples and were symptomatic) and the long-term nonprogressors (those who had been living with the virus for 8 to 14 years).

Frequencies of genotypes and phenotypes were determined amongst the subgroups. Because of the small values in the subgroups, Fisher's exact test was used to determine whether the numbers between the subgroups differed significantly. The Fisher's exact test procedure calculates an exact probability value for the relationship between two variables (Sokal and Rohlf, 1981; Rosner B, 1990). Statistical Analysis System (SAS) software was used to perform the Fisher's exact test.

2.7.6 Haplotype/Disease status association

We examined whether haplotypes of the haptoglobin gene have an impact on susceptibility to HIV and TB infection. The study population was categorized as in section 2.7.5. Haplotype frequencies between the subgroups were determined using Arlequin software, and compared between the subgroups. Statistical Analysis System (SAS) software was used to perform the Fisher's exact test.

CHAPTER 3

RESULTS

3.1. Direct sequencing

Sequencing of both strands of the 645 bp fragment of the promoter region (Fig 3.1) showed polymorphism at six sites (Fig 3.2), which resulted in 34 different sequences in 52 samples from 42 HIV positive subjects and 10 subjects with unknown HIV status. In addition to the four previously known base substitutions, *i.e*., -55A/G, -61A/C, -104T/A (Maeda, 1991; Teye *et al*, 2003) and -242T/C (Teye *et al*, 2003), two which were not reported previously were found at positions -103 and -295. The polymorphisms were G/T and T/C, respectively. None of the individuals was homozygous for C at position -295; only -295TT and TC were found. Homozygous genotypes TT for polymorphisms -103G/T and -242C/T were not found. At position -61, only one individual was C homozygous. Genotype and allele frequencies for all the polymorphic sites were determined (Table 3.1.1). As in Ghanaians, -242T/C occurred at a very low frequency (0.02) in the study population. The polymorphism at site -103 was rare as well, with frequency of 0.02. The C alleles at sites -61 and -295, respectively, were found at low frequencies in the population: 0.15 and 0.11 respectively.

The χ^2 test for goodness of fit showed that the population did not deviate significantly from the expectations of the Hardy-Weinberg equilibrium at any of the six sites (Table 3.1.1).

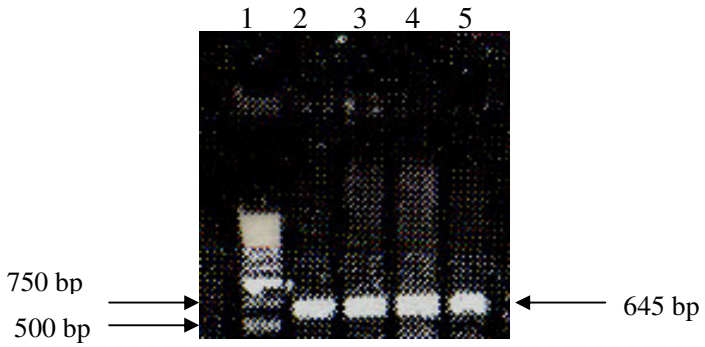


Fig 3.1.1. An agarose gel of the 645 bp promoter region. In lane 1 is a molecular weigh marker VI.

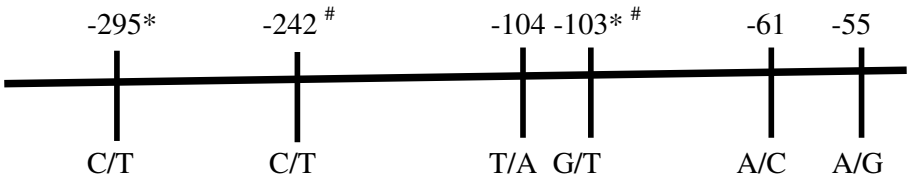


Fig 3.1.2. Schematic diagram showing locations of the six SNPs found by direct sequencing of the 645 bp fragment of the upstream noncoding region of the haptoglobin gene. The new ones are indicated by *, the ones that occur at low frequencies in the population are indicated by #.

Table 3.1.1. Genotype and allele frequencies for the 6 polymorphic sites found by direct sequencing and the χ^2 test for the goodness of fit between the genotypes for each site. There was no significant deviation from the expectations of Hardy-Weinberg equilibrium (P<0.01).

Site	Genotype	n	Genotype frequency	Allele	n	Allele frequency	χ^2_{1df}
-55	AA	18	0.35	A	57	0.55	
	AG	21	0.40				
	GG	13	0.25	G	47	0.45	
Total		52			104		1.80
-61	AA	37	0.71	A	88	0.85	
	AC	14	0.27				
	CC	1	0.02	C	16	0.15	
Total		52			104		0.09
-103	GG	51	0.98	G	103	0.99	
	TG	1	0.02				
	TT	0	0.00	T	1	0.01	
Total		52			104		0.00
-104	AA	6	0.11	A	35	0.34	
	AT	23	0.44				
	TT	23	0.44	T	69	0.66	
Total		52			104		0.005
-242	CC	51	0.98	C	103	0.99	
	TC	1	0.02				
	TT	0	0.00	T	1	0.01	
Total		52			104		0.00
-295	TT	41	0.79	T	93	0.89	
	TC	11	0.21				
	CC	0	0.00	C	11	0.11	
Total		52			104		0.73

3.2. Allele-specific amplification

Allele-specific amplification was applied to a large population to detect the four most common SNPs found by direct sequencing. The presence or absence of the PCR product is dependent on the allele or genotype presented by the DNA template. Samples homozygous for a certain allele yielded a product in the presence of the primer specific for that allele, and not with the other allele-specific primer. Heterozygous samples gave products with both sets of primers. The sizes of PCR products for different alleles are shown in figures 3.2.1, 3.2.2, 3.2.3 and 3.2.4.

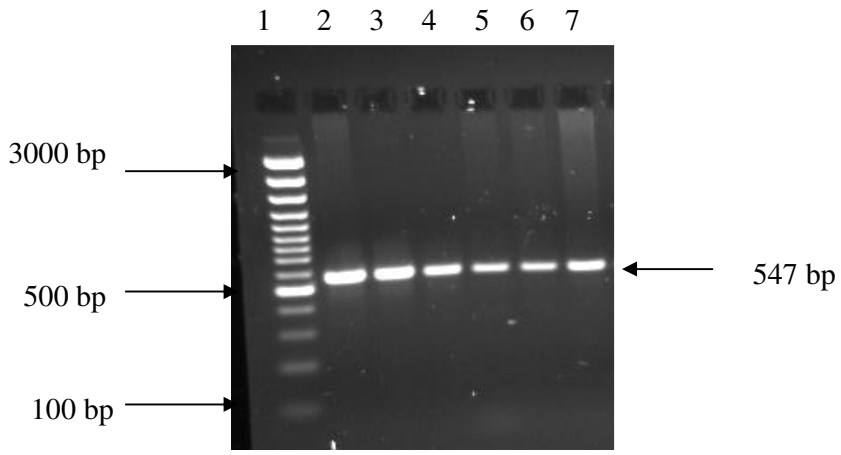


Fig 3.2.1. An agarose gel showing allele-specific PCR products for the -55A/G polymorphism. Lanes 2-4 show -55A allele-specific product, lanes 4-7 show -55G allele-specific product. The PCR products are of the same size (547 bp) for both alleles because the allele-specific primers differ by one nucleotide base at the 3'-end, they extend in the same direction and they were used with the same allele-nonspecific forward primer. The reactions for each allele were done separately (Fig 2.1). The samples were run against 100bp molecular weight marker (lane 1). Samples which are genotype -55AA produced a -55A allele PCR products only, -55GG samples gave -55G allele PCR products only, and heterozygous samples yielded a product with both allele-specific reactions.

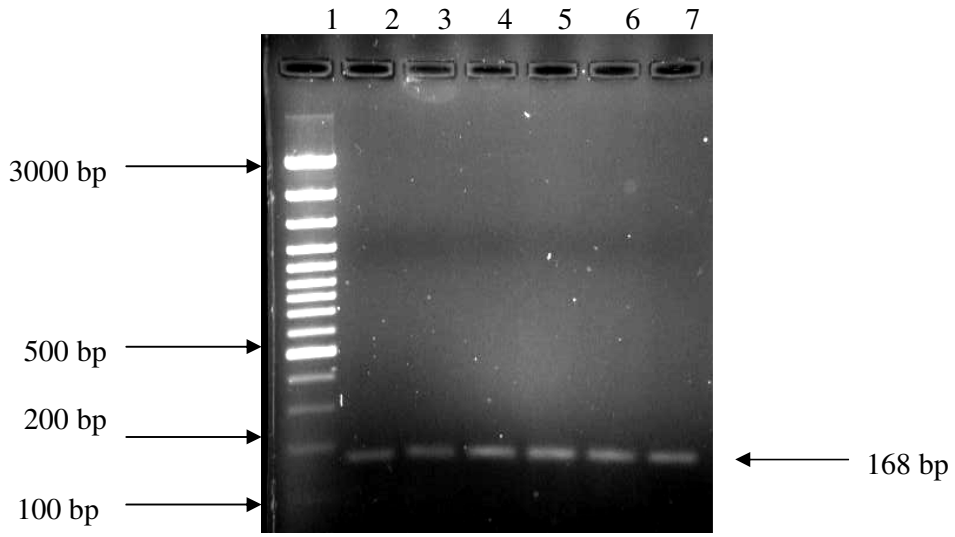


Fig 3.2.2. A 1.7% agarose gel showing PCR products of allele-specific amplification for -61A (lanes 2-4) and -61C (lanes 5-7) alleles run against a 100 bp molecular weight marker (lane 1). The size of the product was 168 bp for both alleles because the allele-specific primers differ by one nucleotide base at the 3'-end, they extend in the same direction and they were used with the same allele-nonspecific forward primer. The reactions for each allele were done separately (Fig 2.1). Samples which were homozygous for A yielded a -61A allele product only, homozygous C yielded a -61C allele product only. Heterozygous samples showed PCR products in both reactions.

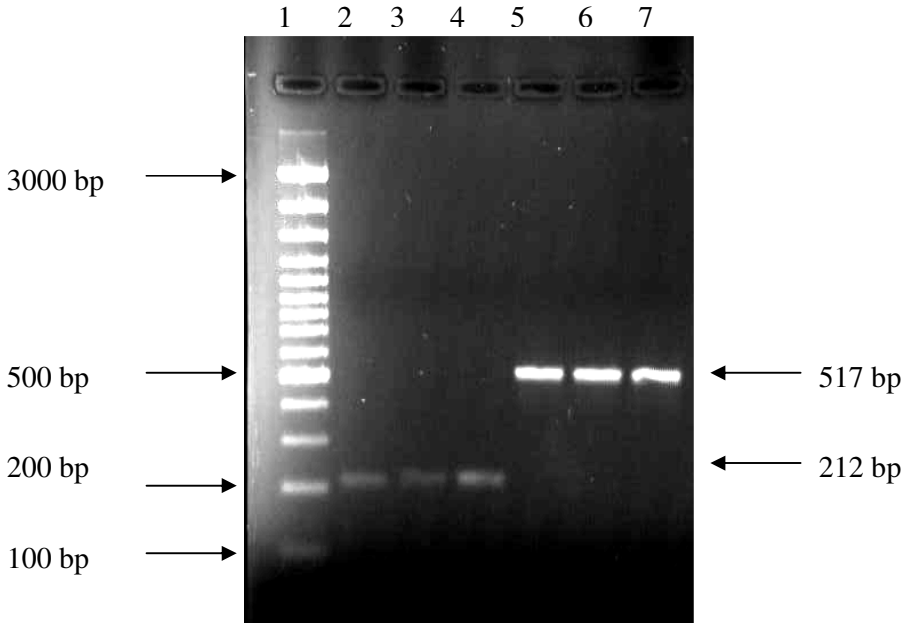


Fig 3.2.3. An agarose gel showing allele-specific PCR products for the A/T polymorphism at site -104. Allele T was characterized by a 212 bp PCR product (lanes 2-4). Allele T was characterized by a 517 bp PCR product (lanes 5-7). The samples were run against the 100 bp ladder (lane 1). Samples which were homozygous for genotypes AA and TT yielded 212- and 517-bp bands, respectively. The AT genotype was characterized by the presence of both the 212- and 517-bp products.

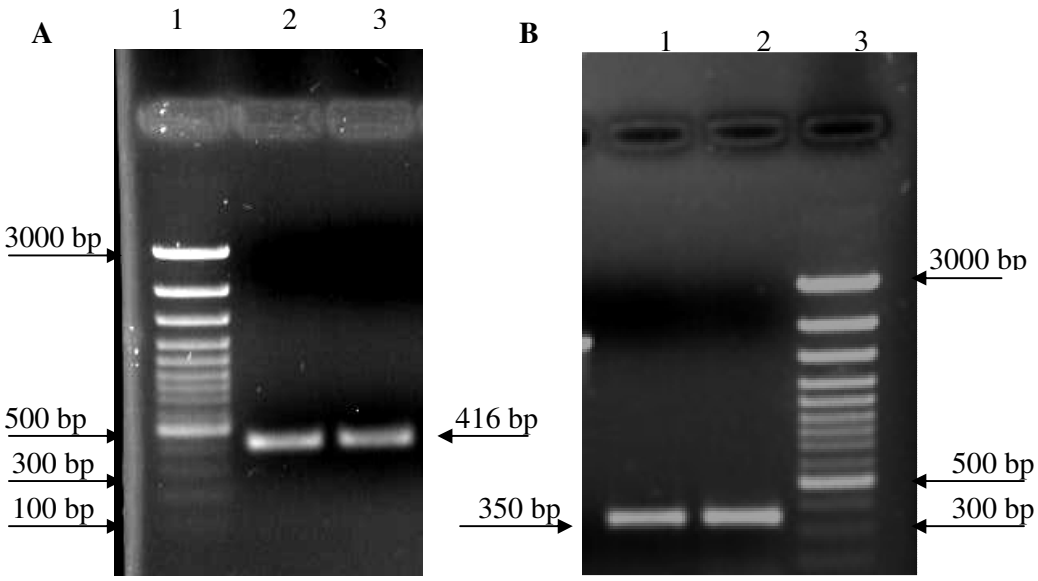


Fig 3.2.4. An agarose gel showing allele-specific PCR products for -295T (A) and -295C (B) alleles. The TT genotype gave a 416 bp PCR product (A, lanes 2 and 3), the CC genotype gave a 350 bp product (B, lanes 1 and 2), and the heterozygote TC was distinguished by the presence of both 416 and 350 bp PCR products.

3.2.1. Genotype and allele frequencies

Genotype and allele frequencies in the whole study population were determined, and the χ^2 test demonstrated that the population did not deviate significantly from the expectations of the Hardy-Weinberg equilibrium at any of the four sites at $P<0.01$ (Table 3.2.1).

Table 3.2.1. Genotype and allele frequencies for the four common polymorphisms in the whole study population and the χ^2 test.

Site	Genotype	n	Frequency	Allele	n	Allele frequency	χ^2_{1df}
-55	AA	66	0.30	A	243	0.56	
	AG	111	0.52				
	GG	38	0.18	G	187	0.43	
Total		215			430		0.51
-61	AA	178	0.83	A	392	0.91	
	AC	36	0.17				
	CC	1	0.005	C	38	0.09	
Total		215			430		0.53
-104	AA	42	0.19	A	183	0.43	
	AT	99	0.46				
	TT	74	0.34	T	247	0.57	
Total		215			430		0.70
-295	TT	134	0.62	T	341	0.79	
	TC	73	0.34				
	CC	8	0.04	C	89	0.21	
Total		215			430		0.17

3.3. Protein polymorphism

Polymorphism in the protein coding region of the haptoglobin gene was determined by amplifying allele-specific fragments according to Koch *et al* (2002). The genotypes were then confirmed by phenotype determination on polyacrylamide gel.

3.3.1 Haptoglobin protein polymorphism genotype determination

In PCR, primer set AB gave 1757- and 3481-bp products for Hp1-1 and Hp2-2 genotypes, respectively. Heterozygote genotype Hp2-1 was characterized by the presence of both 1757 and 3481 bp bands (Fig 3.3.1.1 A). In some cases, it was

not possible to determine with certainty the presence of the 3481-bp Hp2-specific band in Hp2-1 samples. In such cases, primer set CD was used to confirm the presence or absence of Hp2 band in the samples that produced a 1757-bp band with primer set AB. Hp2-1 genotype gave a 349-bp Hp2-specific band in the presence of primer set CD (Fig 3.3.1.1 B). Genotype and allele frequencies were calculated. The Hp2-1 genotype occurred at a high frequency, 0.567. The homozygous genotypes occurred at fairly similar frequencies: 0.22 and 0.21 for Hp1-1 and Hp2-2, respectively. The Hp^1 and Hp^2 allele frequencies were fairly similar in the population, 0.502 and 0.498, respectively (Table 3.3.1.1). The χ^2 test illustrates that the population does not deviate significantly from the expectations of Hardy-Weinberg equilibrium ($P<0.01$).

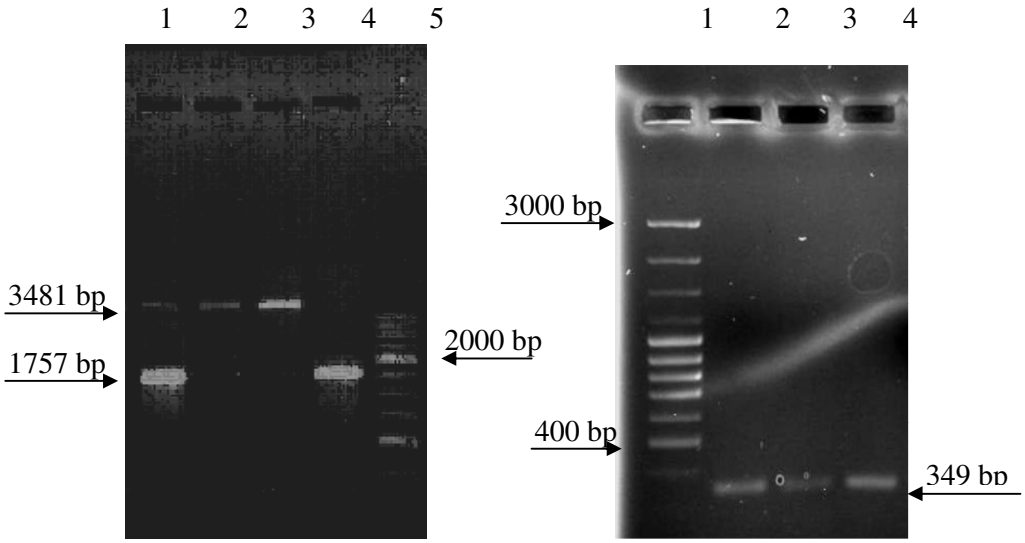


Fig 3.3.1.1. Hp^1 and Hp^2 allele-specific PCR products. (A) Depending on the genotype presented by the template, an Hp2-2 product of 3481bp (lanes 2 and 3), a 1757bp band of Hp1-1 (lane 4) or both bands in Hp2-1 were amplified (lane 1) using primer set AB. Lane 5 is a molecular weight marker. (B) A 349bp product of primer set CD yielded by Hp2-1 samples in the confirmation of the absence or presence of the Hp2-allele-specific fragment. In lane 1 is a 100bp molecular weight marker ladder.

Table 3.3.1.1. Genotype and allele frequencies of the coding region in the whole study population.

Genotype	n	Genotype frequency	Allele	n	Allele frequency	χ^2_{1df}
Hp1-1	47	0.22	Hp1	216	0.50	
Hp2-1	122	0.57				
Hp2-2	46	0.21	Hp2	214	0.50	
Total	215			430		3.65

3.3.2 Phenotype determination

Polyacrylamide gel electrophoresis of plasma gave a band pattern that differentiated between Hp1-1, Hp2-1 and Hp2-2 protein phenotypes (Fig 3.3.2.1). Samples which are phenotypically Hp0 were identified by showing no protein bands on the polyacrylamide gel. The Hp1-1 phenotype was characterized by a single fast-migrating band, Hp2-2 by a series of slow-migrating bands, Hp2-1 by the presence of both slow- and fast-migrating bands (Fig 3.3.2.1).

The proportion of each phenotype in the whole study population was determined. Hp0 phenotype occurred at the frequency of 0.303 (Table 3.3.2.1). This frequency is in accordance with findings of Constants *et al* (1981), who found that in Sub-Sahara Africa the frequency of the Hp0 phenotype is between 10 to 40%.

The Hp0 phenotype was not exclusively found in one genotype, but distributed among all three genotypes, Hp1-1, Hp2-1 and Hp2-2. The Hp1-1 and Hp2-1 genotypes were found in equal frequencies in Hp0 individuals (0.42). This difference in frequencies was statistically significant (P=0.0001) (Table 3.3.2.2).

Among the polymorphisms of the promoter region, the Hp0 phenotype was most common in individuals who carried a -55GG and -104AA genotypes. The

frequency difference was statistically significant for both polymorphic sites ($P<0.05$). Previous studies have reported -61C allele to be a cause of Hp0 phenotype in Africans (Teye *et al*, 2003). In this study, there was only one individual who was homozygous for this allele. The individual was Hp0.

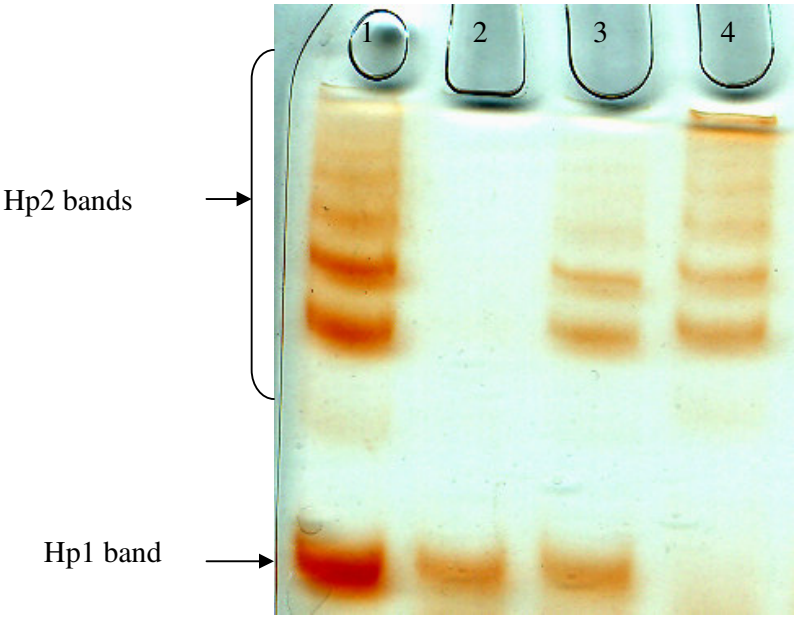


Fig 3.3.2.1. Haptoglobin phenotypes on the polyacrylamide gel. The band pattern distinguished between the phenotypes. Hp1-1 showed a single fast-moving band (lane 2). Hp2-2 was characterized by a series of slow-moving bands (lane 4). Hp2-1 phenotype showed both Hp1 and Hp2 bands (lanes 1 and 3).

Table 3.3.2.1. Haptoglobin phenotype distribution in the whole study population.

Phenotype	**n**	**Phenotype frequency**
Hp0-0	43	0.30
Hp1-1	12	0.08
Hp2-1	64	0.45
Hp2-2	23	0.16
Total	142	

Table 3.3.2.2. Genotype distribution in Hp0 phenotype and the other phenotypes

		Hp0 (n=42)		**With Hp protein (n=100)**		**P**
Promoter region	Genotype	n	Freq	n	Freq	
-55	AA	6	0.14	32	0.34	0.002
	AG	18	0.43	54	0.54	
	GG	18	0.43	14	0.14	
-61	AA	35	0.83	86	0.86	0.426
	AC	6	0.14	14	0.14	
	CC	1	0.02	0	0	
-104	AA	16	0.38	12	0.12	0.0005
	AT	20	0.48	52	0.52	
	TT	6	0.14	36	0.36	
-295	TT	25	0.59	57	0.57	0.853
	TC	15	0.36	38	0.38	
	CC	1	0.02	5	0.05	
Coding region	Genotype					
	Hp1-1	18	0.43	11	0.11	0.0001
	Hp2-1	18	0.43	68	0.68	
	Hp2-2	6	0.14	21	0.21	

P = Fisher's probability

3.4. Distribution of haptoglobin alleles in Black South African ethnic groups

In our study population of 136 Black South Africans with information on family origin, some groups were represented by sample sizes too small for frequency distribution analysis. Venda and Ndebele groups were each represented by one individual, Swazi group by 5 individuals (Table 3.4.1). Twenty five percent of the individuals in the study population had parents or grandparents from different ethnic groups, or from outside South Africa.

The frequency of Hp^1 allele in Zulu and Sotho groups was lower than the previously reported frequencies, while in Tswana and Tsonga groups it was higher (Table 3.4.1) (Nurse *et al*, 1974). The allele distribution in Xhosa and Pedi groups was in accordance with values published (Jenkins, 1972; Nurse *et al*, 1974). The frequencies in Zulu, Xhosa and Tswana groups were similar. The allele's distribution in Pedi and Tsonga groups showed very similar frequencies. The Southern Sotho group showed a very low Hp^1 frequency, lower than the reported frequency (Nurse *et al*, 1974). These distributions are comparable to the ones previously published (Nurse *et al*, 1974), where Zulu, Swazi, Xhosa, Tswana and Sotho groups showed very similar frequencies.

Table 3.4.1. Hp^1 allele distribution in South African ethnic groups.

Ethnic group	n	Hp^1 frequency
Zulu	60	0.47
Xhosa	16	0.50
Pedi	11	0.59
Tswana	19	0.53
Sotho	14	0.43
Tsonga	9	0.60

3.5. Association of genotypes and phenotypes with HIV/AIDS disease status

Even though the size of the HIV positive population was 163, there were only a few subjects with usable clinical data for studying association between the rate of the disease progression and the haptoglobin gene polymorphism. Thirty percent of the samples collected from Johannesburg General Hospital were collected from individuals who were diagnosed with HIV infection between the years 2000 and 2003, and they were still asymptomatic at the time of sample collection. Sixty three samples from the NICD had no dates of diagnosis. However, these samples were useful in studying susceptibility to infection by HIV.

Genotype and phenotype frequencies were determined and compared between the HIV positive population and the general population, between a TB+ and TB- groups, and between the groups defined by fast and slow rate of the disease progression.

3.5.1.1. Genotype and phenotype distribution in HIV positive and the general populations

Neither the HIV positive population nor the general population deviated significantly from Hardy-Weinberg equilibrium at all the four polymorphic sites of the promoter region at significance level $P<0.05$ (Table 3.5.1.1). There was significant difference between the HIV positive and the general population at only one site. A significant difference ($P=0.04$) was observed at site -61, with a higher frequency of AC genotype in the general population. No significant association was observed between any of the other polymorphisms and the risk of HIV infection.

An excess of the heterozygous genotype Hp2-1 was observed in the HIV positive population. The frequencies of the homozygote genotypes were similar in the HIV positive population. The χ^2 test showed no significant deviation from the expectations of Hardy-Weinberg in both the HIV positive and the general populations for haptoglobin coding region genotypes ($P<0.01$). There was no association between the haptoglobin genotypes or alleles and susceptibility to HIV infection.

Hp0 phenotype occurred at a frequency of 0.43 in the general population, and 0.25 in the HIV positive population. These frequencies are in accordance with the findings of Quaye *et al* (2000), where the Hp0 phenotype was higher in HIV negative individuals than in the HIV positive individuals. Despite the difference in Hp0 phenotype frequency distribution observed in HIV positive and the general populations, the Fisher's exact test showed no significant difference between the distribution of the four phenotypes in these two populations (Table 3.5.1.1).

Table 3.5.1.1. Genotype and phenotype distribution among HIV positive and general populations.

		HIV positive			General population			P
Promoter region	Genotype	n	Freq	χ^2_{1df}	n	Freq	χ^2_{1df}	
-55	AA	53	0.32	0.03	13	0.25	1.246	0.551
	AG	81	0.50		30	0.58		
	GG	29	0.18		9	0.17		
Total		163			52			
-61	AA	140	0.86	0.01	38	0.73	1.36	0.045
	AC	22	0.13		14	0.27		
	CC	1	0.01		0	0.00		
Total		163			52			
-104	AA	36	0.22	0.90	6	0.12	0.18	0.212
	AT	74	0.45		25	0.48		
	TT	53	0.33		21	0.40		
Total		163			52			
-295	TT	104	0.64	0.23	30	0.58	3.7	0.118
	TC	51	0.31		22	0.42		
	CC	8	0.05		0	0.00		
Total		163			52			
Coding region	Genotype							
	Hp1-1	35	0.21	4.81	12	0.23	0.0	0.435
	Hp2-1	96	0.59		26	0.50		
	Hp2-2	32	0.20		14	0.27		
Total		163			52			
Protein	Phenotype							
	Hp0-0	25	0.25		18	0.43		0.169
	Hp1-1	9	0.09		3	0.07		
	Hp2-1	50	0.50		14	0.33		
	Hp2-2	16	0.16		7	0.17		
Total		100		-	42		-	

P = Fisher's probability

3.5.1.2. Haptoglobin polymorphism and the rate of the HIV/AIDS progression

To determine whether the haptoglobin polymorphism has an impact on the rate of HIV/AIDS progression, phenotype and genotype frequencies were compared between the groups below, and the Fisher's exact test was used to determine whether the frequencies differ significantly among the groups:

(i) A group that was symptomatic 5 years or less prior the sample collection and the group that had been HIV positive for more than 5 years and were asymptomatic (Table 3.5.1.2.1)

(ii) Rapid progressors and long-term nonprogressors (Table 3.5.1.2.2).

(iii) A group that was diagnosed with HIV 5 years or less prior the collection of the samples and had a $CD4^+$ cell count of $\leq$200 cells/ml^3 and the group that had been HIV positive for more than 5 years and had a $CD4^+$ cell count of greater than 200 cells/ml^3 (Table 3.5.1.2.3).

The χ^2 test showed none of the groups in all the categories deviated significantly from the expectations of the Hardy-Weinberg equilibrium at all the polymorphic sites of the promoter region at $P<0.01$.

A high frequency of the genotype Hp2-1 was observed in both the symptomatic and asymptomatic groups. The symptomatic group did not deviate significantly from the expectations of Hardy-Weinberg equilibrium at $P<0.01$. The asymptomatic group showed a deviation from the Hardy-Weinberg equilibrium at $P<0.01$, and no deviation at $P<0.001$ (Table 3.5.1.2.1). Both groups in the other

two categories did not show a significant deviation from the expectations of Hardy-Weinberg equilibrium for the coding region polymorphism ($P<0.01$).

The genotype and the phenotype frequencies did not differ significantly from each other in the symptomatic/asymptomatic groups, and the rapid progressors and the long-term nonprogressors. There was no significant difference between the genotype and phenotype distribution (Table 3.5.1.2.1; Table 3.5.1.2.2).

In the $CD4^+$cell depletion rate category, a statistically insignificant but noteworthy difference in frequency distribution between the groups was observed at site -104 of the promoter region, the coding region and the phenotypes (Table 3.5.1.2.3).

At site -104, there was a high frequency of a heterozygous genotype AT in the group with a $CD4^+$ cell count of below 200 (0.72) compared with 0.31 in the group with a count of more than 200 cells/ml^3. The frequency of homozygote AA at the same site was considerably higher in the group with a $CD4^+$ cell count of above 200 (0.38) than the group with a count of below 200 cells/ml^3 (0.11).

The frequency of Hp2-2 genotype was higher in the group with a count of below 200 cells/ml^3 (0.39) than Hp1-1 genotype of the same group (0.11), and higher than the frequency of Hp2-2 genotype in the group with a count of above 200 cells/ml^3 (0.15).

The phenotype frequencies of Hp0 and Hp2-2 differed between the two groups. A high frequency of Hp0 phenotype was observed in the group with a $CD4^+$ count of above 200 cells/ml^3. This phenotype was found at a low frequency in the other group. Hp2-2 phenotype was higher in the group with a $CD4^+$ count of below 200 cells/ml^3 and lower in the other group. The frequencies for Hp0-0 were 0.17 and

0.38 for the group with below and above 200 cells/ml^3, respectively. For Hp2-2 the frequencies were 0.39 and 0.15 for the group with a count of below and above 200, respectively.

The high frequency of Hp2-2 in the group with the faster rate of $CD4^+$ cell depletion is in agreement with previous studies where it was reported that Hp2-2 is associated with a faster rate of $CD4^+$ cell depletion (Quaye *et al*, 2000), and the rate of HIV/AIDS progression (Delanghe *et al*, 1996; Quaye *et al*, 2000). Although these frequency differences were statistically insignificant, further investigations need to be done on a larger population size with a more reliable clinical data.

Table 3.5.1.2.1. Genotype and phenotype frequencies between asymptomatic and symptomatic subjects five years after diagnosis with HIV infection.

		≤5yrs since diagnosis, symptomatic			**>5yrs since diagnosis, asymptomatic**			**P**
Promoter region	Genotype	n	Freq	χ^2_{1df}	n	Freq	χ^2_{1df}	
-55	AA	5	0.26	0.05	6	0.22	0.04	1.000
	AG	9	0.47		14	0.52		
	GG	5	0.26		7	0.26		
Total		19			27			
-61	AA	17	0.89	0.10	23	0.85	3.73	1.000
	AC	2	0.10		3	0.11		
	CC	0	0.00		1	0.04		
Total		19			27			
-104	AA	4	0.21	0.05	7	0.26	0.03	1.000
	AT	10	0.53		13	0.48		
	TT	5	0.26		7	0.26		
Total		19			27			
-295	TT	10	0.53	0.20	16	0.59	0.13	0.712
	TC	7	0.37		10	0.37		
	CC	2	0.10		1	0.04		
Total		19			27			
Coding region	Genotype							
	Hp1-1	4	0.21	2.79	2	0.07	8.67	0.502
	Hp2-1	13	0.68		21	0.78		
	Hp2-2	2	0.10		4	0.15		
Total		19			27			
Protein	Phenotype							
	Hp0-0	5	0.26		9	0.33		0.575
	Hp1-1	3	0.16		1	0.04		
	Hp2-1	9	0.47		14	0.52		
	Hp2-2	2	0.10		3	0.11		
Total		19		-	27		-	

P = Fisher's probability

Table 3.5.1.2.2. Genotype and phenotype frequencies in rapid progressors and long-term nonprogressors

		Rapid progressors			Long-term nonprogressors			P
Promoter region	Genotype	n	Freq	χ^2 1df	n	Freq	χ^2 1df	
-55	AA	5	0.26	0.05	4	0.18	0.01	0.845
	AG	9	0.47		11	0.50		
	GG	5	0.26		7	0.32		
Total		19			22			
-61	AA	17	0.89	0.10	20	0.91	0.00	1.000
	AC	2	0.10		2	0.09		
	CC	0	0.00		0	0.00		
Total		19			22			
-104	AA	4	0.21	0.05	5	0.23	0.00	1.000
	AT	10	0.53		11	0.50		
	TT	5	0.26		6	0.27		
Total		19			22			
-295	TT	10	0.53	0.20	13	0.59	0.00	0.893
	TC	7	0.37		8	0.36		
	CC	2	0.10		1	0.04		
Total		19			22			
Coding region								
	Hp1-1	4	0.21	2.79	2	0.09	6.57	0.599
	Hp2-1	13	0.68		17	0.77		
	Hp2-2	2	0.10		3	0.14		
Total		19			22			
Protein	Phenotype							
	Hp0-0	5	0.26		9	0.35		0.645
	Hp1-1	3	0.16		1	0.07		
	Hp2-1	9	0.47		10	0.45		
	Hp2-2	2	0.10		2	0.13		
Total		19		-	22		-	

P = Fisher's probability

Table 3.5.1.2.3. Phenotype and genotype distribution in groups with a $CD4^+$ cell count of below and above 200 cells/ml^3 five years after diagnosis

		$CD4^+ \leq 200$, ≤ 5 yrs since diagnosis			**$CD4^+ > 200$, > 5yrs since diagnosis**			**P**
Promoter region	Genotype	n	Freq	χ^2 1df	n	Freq	χ^2 1df	
-55	AA	6	0.33	0.48	3	0.23	0.05	0.499
	AG	10	0.56		6	0.46		
	GG	2	0.11		4	0.31		
Total		18			13			
-61	AA	16	0.89	0.10	12	0.92	0.00	1.000
	AC	2	0.11		1	0.08		
	CC	0	0.00		0	0.00		
Total		18			13			
-104	AA	2	0.11	3.58	5	0.38	1.93	0.669
	AT	13	0.72		4	0.31		
	TT	3	0.17		4	0.31		
Total		18			13			
-295	TT	11	0.61	0.02	7	0.54	1.19	0.832
	TC	6	0.33		6	0.46		
	CC	1	0.06		0	0.00		
Total		18			13			
Coding region	Genotype							
	Hp1-1	2	0.11	0.09	1	0.08	3.82	0.365
	Hp2-1	9	0.50		10	0.77		
	Hp2-2	7	0.39		2	0.15		
Total		18			13			
Protein	Phenotype							
	Hp0-0	3	0.17	-	5	0.38	-	0.238
	Hp1-1	2	0.11		0	0.00		
	Hp2-1	6	0.33		6	0.46		
	Hp2-2	7	0.39		2	0.15		
Total		18			13			

P = Fisher's probability

3.5.1.3. Genotype and phenotype frequencies in TB groups

Both the TB+ and TB- groups did not show a significant deviation from the Hardy-Weinberg at any of the four polymorphic sites of the promoter region (Table 3.5.1.3).

Excess of the Hp2-1 genotype was observed in both groups, and thus, deviation from the expectations of Hardy-Weinberg equilibrium at the statistical level $P<0.001$. The frequencies of the genotypes were comparable between the groups.

Genotypes for all the polymorphic sites were comparable in the TB+ and TB- groups, and there was no significant difference between the two groups. These findings do not agree with the previous report on association of Hp2-2 with susceptibility to infection by TB (Fedoseeva *et al*, 1993), but are consistent with another study (Kasvosve *et al*, 1998) where the haptoglobin polymorphism was not associated with susceptibility to TB.

Table 3.5.1.3. Genotype and phenotype frequencies in groups with and without TB.

		With TB			Without TB			P
Promoter region	Genotype	n	Freq	χ^2_{1df}	n	Freq	χ^2_{1df}	
-55	AA	4	0.27		23	0.29		0.873
	AG	7	0.46		39	0.49		
	GG	4	0.27		17	0.21		
Total		15		0.07	79		0.00	
-61	AA	14	0.93		67	0.85		0.733
	AC	1	0.07		11	0.14		
	CC	0	0.00		1	0.01		
Total		15		0.00	79		0.59	
-104	AA	3	0.20		16	0.20		0.929
	AT	7	0.47		42	0.53		
	TT	5	0.33		21	0.27		
Total		15		0.00	79		0.37	
-295	TT	10	0.67		47	0.59		0.901
	TC	4	0.26		27	0.34		
	CC	1	0.07		5	0.06		
Total		15		0.42	79		0.18	
Coding region	Genotype							
	Hp1-1	3	0.20		10	0.13		0.502
	Hp2-1	11	0.73		55	0.70		
	Hp2-2	1	0.07		14	0.17		
Total		15		4.41	79		11.83	
Protein	Phenotype							
	Hp0-0	5	0.33		20	0.25		0.815
	Hp1-1	1	0.07		8	0.10		
	Hp2-1	8	0.53		38	0.48		
	Hp2-2	1	0.07		13	0.16		
Total		15		-	79		-	

P = Fisher's probability

3.6. Haplotype analysis

3.6.1 Haplotypes in the study population

Haplotypes for all five polymorphic sites were analyzed using Arlequin software. The frequencies of the haplotypes for all five polymorphic sites were determined in the whole study population.

A total of 32 haplotypes was found in 215 samples. The most common haplotype in the 20 most common ones, Hap14, occurred at the frequency of 0.215 in the study population. Hap10, Hap 19 and Hap 20 were the least common haplotypes in the population, with frequencies that was between 0.002 and 0.003 (Table 3.6.1).

Table 3.6.1. Haplotype frequencies in the study population

Haplotype name	Frequency	Haplotype
Hap 1	0.013	AAAC1
Hap 2	0.035	AAAT1
Hap 3	0.042	AAAT2
Hap 4	0.007	AATC1
Hap 5	0.063	AATC2
Hap 6	0.183	AATT1
Hap 7	0.139	AATT2
Hap 8	0.011	ACAT2
Hap 9	0.021	ACTC2
Hap 10	0.003	ACTT1
Hap 11	0.048	ACTT2
Hap 12	0.047	GAAC1
Hap 13	0.045	GAAC2
Hap 14	0.215	GAAT1
Hap 15	0.018	GAAT2
Hap 16	0.023	GATC1
Hap 17	0.028	GATT1
Hap 18	0.054	GATT2
Hap 19	0.002	GCTT1
Hap 20	0.003	GCTT2

3.6.2. Pairwise allelic linkage disequilibrium

Pairwise allelic linkage analysis of the five polymorphic sites showed that all the alleles were nonrandomly associated (linkage disequilibrium). The specific associations were deduced from the value of D, whether it was positive or negative (Table 3.6.2.). The degree of nonrandom association was very strong between the haplotypes of -55/-61 and -61/coding region, 0.88 and 0.84, respectively. The haplotypes of -55/-104, -61/-104 and -61/-295 sites also showed strong linkage disequilibrium. The degree of nonrandom association was weak between the other haplotypes, with |D'| values of 0.46 and less (Kidd *et al,* 1998).

Table 3.6.2. Linkage disequilibrium between haptoglobin alleles.

Sites	D	Specific associations	|D'|	Std deviation
-55/-61	-0.42	-55G with -61A, -55A with -61C	-0.88	0.03
-55/-104	0.56	-55A with -104T, -55G with -104A	0.57	0.02
-55/-295	0.22	-55G with -295C, -55A with -295T	0.31	0.02
-61/-104	-0.32	-61A with -104A, -61C with -104T	-0.69	0.02
-61/-295	-0.16	-61A with -295C, -61C with -295T	-0.73	0.03
-104/-295	0.16	-104A with-295C, -104T with -295T	0.22	0.02
-55/Coding region	-0.41	-55A with Hp^2, -55G with Hp^1	-0.46	0.01
-61/ Coding region	0.48	-61A with Hp^1, -61C with Hp^2	0.84	0.03
-104/Coding region	-0.38	-104T with Hp^2, -104A with Hp^1	-0.44	0.01
-295/Coding region	-0.19	-295T with Hp^2, -295C with Hp^1	-0.31	0.01

3.6.3. Haplotype/Disease status association

Given our small sample size, the haplotype/disease status association determination was done on the three promoter region polymorphic sites that are found within the IL-6 response segment, sites -55, -61 and -104. The sites are close together and showed strong linkage disequilibrium (Table 3.6.2). The Fisher's exact test was used to test for statistical significance.

3.6.3.1. HIV status/Haplotype association

The most common haplotype in both the HIV positive and the general populations was Hap2 (AAT). This haplotype occurred at a frequency of 0.41 and 0.37 in HIV

positive population and the general population, respectively (Table 3.6.3.1). The least common haplotype in both groups was Hap1 with frequencies of 0.1 and 0.05 in HIV positive group and the general population, respectively. Overall, the haplotype frequencies were comparable between the groups, and the there was no significant difference in the distribution of frequencies between the two groups.

Table 3.6.3.1. Haplotype distribution in HIV positive and the general populations

Haplotype name	Haplotype	HIV positive (n=163)	General population (n=52)
Hap1	AAA	0.10	0.05
Hap2	AAT	0.41	0.37
Hap3	ACA	0.01	0.01
Hap4	ACT	0.06	0.11
Hap5	GAA	0.33	0.29
Hap6	GAT	0.09	0.15
Hap7	GCT	0.00	0.01

3.6.3.2. Disease progression rate/haplotype association

Haplotype frequencies were compared between the groups that define the rate of HIV/AIDS disease progression (Table 3.6.3.2.1, Table 3.6.3.2.2 and Table 3.6.3.2.3). The most common haplotype in all three rate-defining categories in both the compared groups were AAT and GAA. The frequency distribution among the study groups and there was no significant association between the haplotypes and the rate of the disease progression.

Table 3.6.3.2.1. Comparisons of haplotype frequencies between rapid progressors and long-term nonprogressors

Haplotype name	Haplotype	Rapid progressors (n=19)	Long-term nonprogressors (n=22)
Hap1	AAA	0.09	0.02
Hap2	AAT	0.38	0.36
Hap3	ACT	0.03	0.05
Hap4	GAA	0.39	0.45
Hap5	GAT	0.09	0.12
Hap6	GCT	0.02	0.00

Table 3.6.3.2.2. Haplotype frequencies and the rate of $CD4^+$ cell count depletion.

Haplotype name	Haplotype	$CD4^+ \leq$ 200, ≤5 yrs since diagnosis (n=18)	$CD4^+ >200$, > 5 yrs since diagnosis (n=13)
Hap1	AAA	0.09	0.04
Hap2	AAT	0.46	0.38
Hap3	ACA	0.02	0.0
Hap4	ACT	0.04	0.04
Hap5	GAA	0.36	0.50
Hap6	GAT	0.03	0.04

Table 3.6.3.2.3. Comparison of haplotypes in symptomatic and asymptomatic groups

Haplotype name	Haplotype	Symptomatic, ≤5 yrs since diagnosis (n=19)	Asymptomatic, >5 yrs since diagnosis (n=27)
Hap1	AAA	0.09	0.06
Hap2	AAT	0.38	0.33
Hap3	ACT	0.03	0.09
Hap4	GAA	0.39	0.44
Hap5	GAT	0.10	0.08
Hap6	GCT	0.02	0.0

3.6.3.3. TB status/Haplotype association

In six haplotypes found in TB+ and TB- study groups, the most common ones were Hap2 and Hap4 (Table 3.6.3.3). The overall distribution of the haplotypes between the study groups was similar, no statistical significance was found.

Table 3.6.3.3. Haplotype distribution in TB+ and TB- groups.

Haplotype name	**Haplotype**	**With TB (n=15)**	**Without TB (n=79)**
Hap1	AAA	0.07	0.07
Hap2	AAT	0.43	0.39
Hap3	ACT	0.00	0.08
Hap4	GAA	0.36	0.40
Hap5	GAT	0.11	0.06
Hap6	GCT	0.03	0.00

CHAPTER 4

DISCUSSION

PCR amplification and direct sequencing of the 645 bp promoter fragment revealed six polymorphisms in 52 samples. This sample size was large enough to detect common polymorphisms that occur at a frequency of more than 2% in the population. The study population did not show any significant deviation from the expectations of Hardy-Weinberg equilibrium at any of the six polymorphic sites in the 52 sequenced samples. The indirect methods for detection of the four most common SNPs were successful. There was no significant deviation from the expectations of Hardy-Weinberg equilibrium at any of the four polymorphic sites. This showed that all the alleles of the polymorphic sites under investigation were accounted for.

Sequencing of a 647 bp promoter region fragment in Ghanaians showed six polymorphisms as well. When put together, there are eight polymorphisms in a promoter fragment of about 650 bp in Africans, compared with three in African-Americans. This high genetic diversity in these two African populations emphasizes that non-African populations have a subset of genetic variation found in Africans (Jorde *et al*, 1997; Tishkoff, 2002).

The Hp^1 and Hp^2 alleles were found in equal frequencies in the study population. In Africans, Hp^1 allele frequency ranges between 0.40 and 0.87 (Giblett, 1969). In South African populations, the frequency is between 0.44 and 0.58 (Nurse *et al*, 1985). The frequency calculated for our study population was within this range.

An excess of the heterozygous genotype, Hp2-1 was observed at a frequency of 0.57, nonetheless, the population did not deviate significantly from the expectations of the Hardy-Weinberg equilibrium.

Using polyacrylamide gel electrophoresis, we were able to characterize the four major protein phenotypes in the population: Hp0, Hp1-1, Hp2-1 and Hp2-2. The Hp0 phenotype was characterized by the lack of protein bands on the gel. It was not investigated whether the Hp0 phenotype was because the protein was expressed at very low concentrations to be detected by gel electrophoresis (hypohaptoglobinemia), or due to a complete lack of protein expression (anhaptoglobinemia). And so, we did not discriminate between the two conditions and use Hp0 to describe both conditions.

The genetic origin of the Hp0 phenotype has been reported to be deletion in Asians (Koda *et al*, 1998). But in Africans, where the frequency of the Hp0 phenotype is high, the deletion was not found (Koda *et al*, 1998; Teye *et al*, 2003). The genetic basis for this phenotype in Ghanaians has been reported to be a single nucleotide base substitution in one of the functional sites of the promoter region, A to C at -61 (Teye *et al*, 2003). Another haptoglobin mutation within the β-chain of Hp^2 allele in Ghanaians was reported to be associated with Hp0 phenotype (Teye *et al*, 2004). In this study, there was only one individual homozygous for C at this site, and was found to be Hp0. Only 17% (7 out of 42) of Hp0 individuals had a C at site -61 in this study. Overall, there was no significant association

between the A to C substitution at -61 with phenotype Hp0. Rather, the Hp0 phenotype was associated with -55G and -104A alleles.

A high frequency of Hp0 individuals who were homozygous for G at -55 polymorphic site was observed, even though the frequency of this genotype was low in the whole study population. The frequency of the G allele was 0.64, and the frequency of the homozygous genotype -55GG was 0.43. In 61% (11 out of 18) individuals who had Hp0, the -55GG genotype occurred with the -104AA genotype, and only 1 in 18 occurred with -104TT genotype. The -55 polymorphic site lies six base pairs downstream of the -61 site, which is one of the responsive sites to IL-6 (Oliviero and Cortese, 1989). The -55 polymorphic site is within the 216 bp segment (between positions -186 and +30) of the haptoglobin promoter that responds to induction by IL-6 (Oliviero and Cortese, 1989). It has been reported (Maeda, 1991) that a G at position -55 can negatively affect haptoglobin protein expression. The combination of the promoter polymorphisms -55G and -104A has been observed in Hp0 in black Americans (Maeda, 1991), although no conclusion on association between the promoter polymorphism and Hp0 phenotype was drawn in that study because of small sample size. The significant association between -55GG and -104AA genotypes suggests that there is a relationship between these genotypes and the Hp0 phenotype: carriers of one or both these genotypes are more likely to be Hp0.

The Hp0 phenotype was also found to be more common in individuals who were genotypically Hp1-1 than in Hp2-2 individuals, and most common in Hp2-1 individuals. The high frequency of the heterozygous genotype was not unexpected, since in the whole study population there was an excess of this

genotype. When comparing the distribution of the homozygous genotypes and the alleles in those with the Hp0 phenotype, Hp1-1 genotype and Hp^1 allele frequencies were high, even though they were comparable in the whole study population. The frequency of Hp^1 allele was 0.64, and Hp1-1 was 0.36. The distribution of haptoglobin genotypes in Hp0 individuals was significantly different. Teye *et al* (2003) found an association of Hp^2 allele with anhaptoglobinemia, and Hp^1 allele with hypohaptoglobinemia. In this study, an association between the Hp^1 allele and the Hp0 phenotype was observed. Other studies have found that Hp0 phenotype in Hp^1 individuals is due to anomalous inheritance (Koda *et al*, 1998). As no distinction was made between anhaptoglobinemia and hypohaptoglobinemia in this study, it cannot be said which of these two conditions was associated with the polymorphisms. Nonetheless, these results show a fourth set of genetic polymorphism that is associated with the Hp0 allele. The association of Hp0 with different haptoglobin genetic polymorphisms suggests that this phenotype is caused by more than one polymorphism. These polymorphisms could be acting singly or in combination, with an exception of the Hp^{del} which is not found in Africans.

The frequency of Hp0 phenotype in the entire study population was 0.30. These findings are in accordance with the previous studies in which the frequency of the Hp0 phenotype in Sub-Saharan Africa was reported to range between 10 and 40% (Constants, 1981). Since in the Gauteng region of South Africa there is no malaria endemic, the Hp0 phenotype found in this population would be due to genetic influences, and/or other factors.

In South African blacks, the frequency of Hp^1 allele ranged between 0.43 and 0.59, which is within the reported range in Africans (Giblett, 1989; Nurse *et al*, 1985). The Zulu group sample size was large enough to get a close estimation of the frequency of the allele in the group; the other groups were present in small numbers in the study population. The allele frequencies were close in Zulu, Xhosa and Tswana groups, Xhosa group frequency being the median point. Pedi and Tsonga groups showed very close frequencies. The Sotho group (Southern Sotho) showed the lowest frequency, lower than the previously reported one (Nurse *et al,* 1985). The expectations were that the frequencies in the groups under the Nguni group (Zulu and Xhosa groups), and the groups under the Sotho-speakers (Pedi, Tswana and Sotho groups) would show frequencies be very similar to each other (Lane *et al*, 2001). Intercultural mating, which was common in our study population, and the small numbers of other ethnic groups would explain the shift from the expected groupings. Nevertheless, the frequency distribution in this study does not differ significantly from a previous study where Zulu, Xhosa, Swazi, Tswana and Sotho groups showed very similar frequencies (Nurse *et al*, 1985). Although there are differences in frequencies among South African blacks, different languages, cultures and geographic locations, they all originate from the Bantu expansion and share a common ancestral population (Lane *et al*, 2001) which explains the small difference in allele frequencies.

The very strong linkage disequilibrium between the promoter region site -61 and the coding region was not unexpected. Site -61 is one of the IL-6 response elements. The strong association is due to the fact that this site correlates with the

coding region. Sites -55/-61, -61/-104 and -55/-104 showed strong to very strong linkage disequilibrium. These site are close to each other and they demonstrate the $D_{-55,-61} > D_{-61,-104} > D_{-55,-104}$ order, which indicates that the closer the loci, the greater the linkage disequilibrium (Lewontin, 1964). It is not clear why -61/-295 showed such strong linkage disequilibrium. The association is stronger than that of -55/-104 and -61/-104 but the sites are farther apart, and site -295 is not a functional site.

Among the four polymorphisms of the promoter region, -61A/C showed marginal association with susceptibility to HIV infection (P=0.045). The frequency of the C allele was higher in the general population than in the HIV positive population. As mentioned previously, position -61 is one of the response elements for the inducer of the protein expression. The A to C base substitution at this site reduces the activity of the promoter (Maeda, 1991). Haptoglobin protein is indirectly involved in the entry of the HIV into the host cell during infection (El Ghamati, 1996) or cell to cell transmission (Quaye *et al*, 2000). The effect of the reduced protein expression on the infection by HIV has been reported (Quaye *et al*, 2000) where the Hp0 phenotype was more common in HIV negative individuals. A previous study reported the association of Hp0 and the -61C allele (Teye *et al*, 2003). Although in this study there was no significant association between the -61C allele and Hp0, the decreased protein expression in -61C individuals would explain the association between the reduced risk of HIV infection and the -61C allele. We also observed a high frequency of Hp0 phenotype in the general population, 0.43

compared with 0.25 in the HIV positive population, though the difference was not statistically significant.

The results of this study showed no association between the haptoglobin genotypes and susceptibility to HIV infection. The Hp^1 and Hp^2 allele frequencies did not differ significantly between the HIV positive and the general populations. This implied that the Hp^1 and Hp^2 alleles have the same chances of HIV infection. These findings are in agreement with the study done in Ghana (Quaye *et al*, 2000).

There was no association between susceptibility to infection TB and any of the polymorphisms. Previous studies reported that polymorphism in the coding region of the haptoglobin gene is associated with susceptibility TB infection (Fedoseeva *et al*, 1993) and outcome after treatment initiation (Kasvove *et al*, 2000). In this study, the effect of the genetic polymorphism on severity of the disease and treatment outcome remained uninvestigated. It could happen that the polymorphisms affect the outcome after treatment initiation or severity of the disease in our population. Thus, the lack of significant association between haptoglobin polymorphism and TB infection in this study does not mean the complete lack of association between TB and haptoglobin polymorphism.

Even though the total sample size was large, only a few samples had usable clinical data for studying the impact of the haptoglobin polymorphism on the rate of the disease progression. In addition, the clinical data we had was self reported, which made it unreliable to a certain extent. That made it difficult to conclude

with certainty that there was a complete lack of significant correlation between other polymorphisms and the rate of progression.

There was no statistically significant association between any of the polymorphisms, including the coding region polymorphism which has been previously reported to be associated with the rate of the disease progression in Europeans (Delanghe *et al*, 1996) and Ghanaians (Quaye *et al*, 2000).

Because $CD4^+$ cell depletion is one of the indicators of the rate of HIV/AIDS progression (Carré *et al*, 1998), the association between polymorphism and the rate of $CD4^+$ cell depletion was examined. Although there was no significant association between the polymorphisms and the rate of $CD4^+$ cell depletion, two polymorphisms are noteworthy: -104A/T and the coding region. A high frequency of -104AT genotype (0.72) was observed in the group with faster rate of $CD4^+$ cell depletion, and the higher frequency of -104AA genotype (0.38) in the group with slower rate of cell depletion. It must be pointed out that the frequency of AT in the whole study population was 0.46, meaning there was no excess of this genotype, and the AA genotype was found in lower frequency in the whole population (0.19). The -104 site is within the IL-6 response segment, between positions -186 and +30 (Oliviero and Cortese, 1989). It is possible that variation at this site could have an effect on the level of protein expression and the disease, though no conclusion could be drawn at this stage. The effect of this polymorphism on the activity of the promoter needs to be studied, and its possible effects on the rate of $CD4^+$ cell destruction.

In the coding region polymorphism, an excess of the Hp2-1 genotype was observed in the whole study population and all the other categories, but not in the group with faster rate of $CD4^+$ cell depletion (and the general population). High frequencies of Hp2-2 genotype and Hp2-2 protein phenotype were observed in the group with the faster rate of $CD4^+$ cell depletion. The association between $CD4^+$ cell depletion and Hp2-2 protein phenotype has been reported (Quaye *et al*, 2000). Since $CD4^+$ cell destruction is one of the indicators of the rate of HIV/AIDS progression (Carré *et al*, 1998), the high frequency of Hp2-2 in the group with faster rate of $CD4^+$ cell depletion is in agreement with the previous study of Delanghe *et al* (1996), where the association of this protein phenotype and the faster rate of HIV/AIDS progression was reported, but the difference was not statistically significant.

Another frequency difference which was not statistically significant was in the Hp0 phenotype. The frequency of this phenotype was higher in the group with a slower rate of $CD4^+$ cell depletion. This suggested a possibility of association between the Hp0 phenotype with the slower rate of the disease progression, as has been reported previously (Quaye *et al*, 2000).

We looked at the haplotypes for three polymorphic sites (-55, -61 and -104) and their effect on TB and HIV infection, and the rate of HIV/AIDS disease progression. The reason for taking haplotypes for three polymorphic sites is that our sample size was small; taking haplotypes for all five sites would result in a large number of haplotypes with few samples of each. The polymorphic sites

chosen are close to each other and they show strong to very strong linkage disequilibrium and they are found within the IL-6 response segment. There was no association between any of the haplotypes and either susceptibility or the rate of disease progression. The lack of correlation between the haplotype for these polymorphic sites and the disease status was not unexpected as there was no significant association between -55 and -104 polymorphic sites and the disease status, and the association with -61 was marginal.

The hemoglobin clearance efficiency differs between haptoglobin types. The poor prognosis in Hp2-2 individuals with HIV was explained by retention of iron in the plasma by Hp2-2 phenotype that leads to oxidative stress, resulting in high rate of viral replication (Delanghe *et al*, 1996). In this study no association was found between the haptoglobin genotypes/phenotypes that have been previously reported to be associated with TB and the rate of HIV/AIDS progression. There are two possible explanations for this. One, that is more likely, could be that our sample size was small. Alternatively, it could be explained by the findings of Kasvosve *et al* (2002) that iron metabolism is not influenced by haptoglobin phenotypic variation in African blacks. This could also explain the lack of association between Hp phenotypes and TB infection in the Zimbabweans (Kasvove *et al*, 2000). There are contradicting results regarding susceptibility to malaria and Hp phenotypes (Trape and Fribourg-Blanc, 1988; Joshi *et al*, 1987; Elagib *et al*, 1998; Aucan *et al*, 2002) which could point out population diversity or differences in the functioning of haptoglobin.

South Africa has the highest number of people living with HIV/AIDS than any country, and the prevalence is still increasing. That poses a challenge to researchers to find a cure for the millions of people already infected with the virus, and a vaccine for those who are uninfected. Studying the influence of genetics on susceptibility to HIV infection and the rate of the disease progression is another way that could lead to treatment and prevention of AIDS. A number of genes have been reported to be associated with HIV infection and disease progression; most of these studies have been done in non-Africans. These associations may not be a general picture for all the populations. As genetic diversity is high in Africans, searching for additional polymorphisms found in Africans, and determining whether these polymorphisms have an impact on susceptibility to HIV infection and the rate of the disease progression may lead to the solving of the HIV/AIDS endemic.

The findings of this study present a need to further investigate the impact of haptoglobin genetic variation on susceptibility to HIV infection and the rate of HIV/AIDS disease progression in black South Africans. The study should be done on a long term established cohort with large sample size.

The novel association of Hp0 phenotype and -55GG and -104AA genotypes in South Africans also needs further investigation. It should be investigated whether the -55GG and -104AA genotypes are associated with anhaptoglobinemia or hypohaptoglobinemia, and if so how these genotypes affect the level of haptoglobin expression.

CONCLUSIONS

Until now, the effect of polymorphism in the promoter region of the haptoglobin gene on HIV infection and the rate of progression of AIDS has not been reported. We have found a significant association between the -61C allele and low risk of HIV infection. This association could be related to the low levels of haptoglobin protein expression caused by the A-C base substitution at position -61. Since we have looked at small sample sizes, further investigation on larger population size with reliable clinical data need to be done to confirm this.

The significant association of the Hp0 phenotype with -55GG and -104AA genotypes suggests that there is a relationship between these genotypes, singly or in combination, and Hp0 phenotype, not that they are exactly the cause of the Hp0 phenotype because the genotypes were also found in the other phenotypes. The association of Hp1-1 genotype with Hp0 could be due to anomalous inheritance, or related to the level of protein expression in Hp1-1 individuals.

We have also observed frequency differences which were not statistically significant, but noteworthy: a high frequency of -104AA genotype and Hp0 phenotype in individuals with a slower rate of $CD4^+$ cell depletion; high frequency of -104AT genotype and Hp2-2 phenotype in individuals with a faster rate of $CD4^+$ cell depletion. The observed associations, both the statistically significant and the one that were not statistically significant, portray a need to do further investigations on the effect of the polymorphisms in the noncoding region of the haptoglobin gene on HIV infection and the rate of disease progression. Since few samples had usable data in this study, further investigation needs to be done on a large established cohort with more reliable clinical data, including the rate of viral

load increase, would likely lead to a concrete conclusion. If significant associations are found, they could be used in drug development and treatment trials.

APPENDIX 1

Table E.1. Raw data – Sample genotypes and phenotypes

Sample	Group	-55	-61	-104	-295	Hp Genotype	Phenotype
206	W	AG	A	A	T	Hp2-1	Hp0
207	W	A	A	AT	CT	Hp1-1	Hp0
208	W	AG	A	AT	CT	Hp2-1	Hp0
209	W	AG	A	AT	CT	Hp1-1	Hp0
210	W	A	A	T	CT	Hp2-2	Hp2-2
211	W	A	AC	T	CT	Hp2-1	Hp2-1
212	W	A	A	T	CT	Hp2-1	Hp2-1
213	W	AG	A	AT	T	Hp2-1	Hp2-1
214	W	G	A	A	T	Hp1-1	Hp0
215	W	A	A	T	CT	Hp2-1	Hp2-1
216	W	AG	A	T	CT	Hp2-1	Hp2-1
217	W	A	A	T	T	Hp2-2	Hp2-2
218	W	AG	AC	T	CT	Hp2-1	Hp0
219	W	G	A	A	CT	Hp1-1	Hp0
220	W	AG	A	T	CT	Hp2-1	Hp2-1
221	W	AG	A	AT	T	Hp2-1	Hp2-1
222	W	AG	A	T	T	Hp2-2	Hp2-2
223	W	G	A	AT	CT	Hp1-1	Hp0
224	W	A	A	T	CT	Hp2-2	Hp2-2
225	W	G	A	AT	CT	Hp1-1	Hp0
226	W	A	A	T	CT	Hp2-1	Hp2-1
227	W	AG	AC	AT	T	Hp2-1	Hp2-1
228	W	G	A	T	T	Hp1-1	Hp0
229	W	AG	A	A	T	Hp2-1	Hp2-1
230	W	AG	A	AT	T	Hp2-1	Hp2-1
231	W	AG	A	AT	T	Hp1-1	Hp1-1
232	W	AG	A	AT	T	Hp2-1	Hp0
233	W	G	A	AT	T	Hp2-1	Hp2-1
234	W	AG	AC	A	T	Hp2-1	Hp0
235	W	AG	A	A	T	Hp1-1	Hp0
236	W	A	A	T	CT	Hp1-1	Hp0
237	W	AG	AC	AT	T	Hp2-1	Hp0
238	W	AG	A	AT	T	Hp1-1	Hp0
239	W	AG	AC	AT	CT	Hp2-2	Hp2-2
240	W	AG	A	AT	CT	Hp2-2	Hp2-2
241	W	AG	A	T	T	Hp2-1	Hp2-1
242	W	AG	AC	AT	CT	Hp2-2	Hp0
243	W	AG	A	AT	T	Hp1-1	Hp1-1
244	W	G	A	AT	T	Hp2-2	Hp0
245	W	AG	A	AT	CT	Hp2-1	Hp2-1
A001	W	G	AC	T	T	Hp2-2	
A002	W	G	A	T	T	Hp2-1	

Sample genotypes and phenotypes (continued)

Sample	Group	-55	-61	-104	-295	Hp Genotype	Phenotype
A003	W	AG	A	AT	T	Hp2-1	Hp2-1
A004	W	A	AC	T	T	Hp2-2	Hp2-2
A005	W	AG	AC	AT	T	Hp2-1	Hp2-1
A006	W	A	AC	T	T	Hp2-2	Hp2-2
A007	W	A	AC	T	T	Hp2-2	Hp2-2
A008	W	AG	A	T	T	Hp2-2	Hp2-2
A009	W	AG	AC	AT	T	Hp2-1	Hp2-1
A010	W	A	AC	T	T	Hp2-2	Hp2-2
101	P	G	A	T	T	Hp2-2	Hp2-2
102	P	G	A	AT	TC	Hp2-1	Hp2-1
103	P	G	A	A	T	Hp1-1	Hp0
104	P	G	A	AT	T	Hp2-1	Hp0
105	P	G	A	AT	TC	Hp1-1	Hp0
106	P	G	A	A	T	Hp1-1	Hp0
107	P	A	A	T	T	Hp2-1	Hp2-1
108	P	G	A	AT	T	Hp2-1	Hp2-1
109	P	A	A	T	T	Hp2-1	Hp2-1
110	P	G	A	A	T	Hp1-1	Hp0
111	P	G	A	AT	T	Hp2-1	Hp0
112	P	AG	A	AT	T	Hp2-1	Hp2-1
113	P	A	A	T	T	Hp2-1	Hp0
114	P	AG	A	AT	T	Hp2-1	Hp0
115	P	AG	A	AT	T	Hp2-1	Hp2-1
116	P	A	AC	T	TC	Hp2-1	Hp2-1
117	P	AG	AC	T	T	Hp1-1	Hp0
118	P	AG	A	AT	TC	Hp1-1	Hp1-1
119	P	A	A	T	T	Hp2-1	Hp2-1
120	P	A	A	T	T	Hp2-2	Hp2-2
121	P	AG	A	A	TC	Hp1-1	Hp1-1
122	P	AG	A	AT	TC	Hp2-1	Hp2-1
123	P	A	A	T	T	Hp2-2	Hp2-2
124	P	G	A	A	TC	Hp2-1	Hp0
125	P	AG	A	AT	TC	Hp2-1	Hp2-1
126	P	AG	A	AT	T	Hp2-1	Hp2-1
127	P	AG	AC	AT	TC	Hp2-1	Hp0
128	P	A	AC	T	T	Hp2-1	Hp2-1
129	P	AG	A	AT	T	Hp2-1	Hp2-1
130	P	AG	A	AT	T	Hp2-1	Hp2-1
131	P	AG	A	AT	T	Hp2-1	Hp2-1
132	P	A	AC	T	T	Hp2-1	Hp2-1
133	P	AG	A	AT	TC	Hp2-1	Hp0
135	P	G	A	A	T	Hp2-1	Hp0
136	P	A	A	T	T	Hp2-1	Hp0

Sample genotypes and phenotypes (continued)

Sample	Group	-55	-61	-104	-295	Hp Genotype	Phenotype
137	P	AG	A	AT	T	Hp2-1	Hp2-1
138	P	AG	A	AT	T	Hp2-1	Hp2-1
139	P	A	AC	T	T	Hp2-2	Hp2-2
140	P	A	AC	T	T	Hp2-2	Hp2-2
141	P	A	A	T	T	Hp2-1	Hp2-1
142	P	A	C	T	T	Hp2-2	Hp0
143	P	AG	A	AT	TC	Hp2-1	Hp2-1
144	P	AG	A	AT	TC	Hp2-1	Hp2-1
145	P	G	A	A	T	Hp1-1	Hp1-1
146	P	G	A	A	C	Hp2-1	Hp0
147	P	A	A	AT	T	Hp1-1	Hp1-1
148	P	AG	A	AT	T	Hp2-2	Hp2-2
149	P	G	A	A	TC	Hp2-1	Hp2-1
150	P	AG	A	AT	TC	Hp2-1	Hp2-1
151	P	AG	A	AT	T	Hp2-1	Hp0
152	P	G	A	AT	TC	Hp2-1	Hp2-1
153	P	A	AC	AT	T	Hp2-2	Hp2-2
154	P	G	A	A	TC	Hp2-1	Hp0
155	P	AG	A	AT	TC	Hp2-1	Hp2-1
156	P	A	A	AT	T	Hp2-1	Hp2-1
157	P	G	A	AT	T	Hp2-1	Hp2-1
158	P	AG	A	AT	T	Hp2-1	Hp2-1
159	P	AG	A	AT	T	Hp2-2	Hp2-2
160	P	G	A	A	TC	Hp2-1	Hp2-1
161	P	AG	AC	AT	T	Hp2-1	Hp2-1
162	P	AG	A	T	T	Hp2-1	Hp2-1
163	P	G	A	A	TC	Hp2-1	Hp0
164	P	AG	A	AT	TC	Hp2-1	Hp2-1
166	P	AG	A	AT	T	Hp2-1	Hp2-1
167	P	A	A	AT	T	Hp2-2	Hp2-2
168	P	A	A	AT	T	Hp2-1	Hp0
169	P	A	AC	T	T	Hp2-2	Hp2-2
170	P	AG	A	A	T	Hp2-1	Hp2-1
171	P	AG	A	AT	TC	Hp2-1	Hp2-1
172	P	AG	A	AT	T	Hp2-1	Hp0
173	P	AG	A	T	TC	Hp2-1	Hp2-1
175	P	AG	A	T	T	Hp2-1	Hp2-1
176	P	AG	A	AT	T	Hp2-1	Hp0
177	P	AG	A	T	T	Hp2-1	Hp2-1
178	P	AG	A	A	C	Hp1-1	Hp1-1
179	P	AG	A	AT	TC	Hp2-2	Hp2-2
180	P	A	A	AT	T	Hp2-2	Hp2-2
181	P	G	A	A	T	Hp2-1	Hp2-1

Sample genotypes and phenotypes (continued)

Sample	Group	-55	-61	-104	-295	Hp Genotype	Phenotype
182	P	A	A	T	T	Hp2-1	Hp2-1
183	P	AG	A	T	TC	Hp1-1	Hp1-1
184	P	G	A	A	TC	Hp2-1	Hp2-1
185	P	G	A	A	C	Hp1-1	Hp1-1
186	P	A	AC	T	C	Hp2-2	Hp2-2
187	P	A	A	T	T	Hp2-1	Hp2-1
188	P	G	A	A	T	Hp2-1	Hp0
189	P	AG	A	AT	T	Hp1-1	Hp1-1
190	P	A	A	T	TC	Hp2-1	Hp2-1
191	P	G	A	AT	C	Hp1-1	Hp1-1
192	P	A	A	T	TC	Hp2-1	Hp2-1
193	P	A	A	T	T	Hp2-1	Hp2-1
194	P	AG	A	AT	TC	Hp2-1	Hp2-1
195	P	AG	A	AT	C	Hp2-1	Hp2-1
196	P	AG	A	A	TC	Hp2-1	Hp2-1
197	P	AG	A	AT	T	Hp2-1	Hp2-1
198	P	AG	A	AT	TC	Hp2-1	Hp2-1
199	P	AG	A	AT	TC	Hp2-1	Hp2-1
200	P	AG	A	AT	T	Hp2-1	Hp2-1
201	P	G	A	AT	TC	Hp2-2	Hp2-2
203	P	A	A	T	TC	Hp2-2	Hp2-2
204	P	AG	AC	AT	T	Hp2-1	Hp2-1
205	P	A	AC	AT	T	Hp2-1	Hp2-1
220+	P	AG	A	A	T	Hp2-1	Hp0
525 171	CG	AG	A	AT	TC	Hp2-1	
525 327	CG	A	A	T	T	Hp2-2	
525 301	CG	AG	A	T	T	Hp2-1	
525 343	CG	AG	A	AT	T	Hp1-1	
525 316	CG	AG	A	A	T	Hp2-1	
525 298	CG	A	A	T	T	Hp2-2	
525 160	CG	A	A	A	T	Hp1-1	
525 028	CG	G	A	A	T	Hp1-1	
536 031	CG	AG	AC	AT	T	Hp2-1	
536 121	CG	A	A	T	T	Hp2-1	
536 107	CG	AG	AC	A	T	Hp2-1	
536 015	CG	A	A	T	T	Hp2-1	
536 149	CG	G	A	A	C	Hp2-1	
536 173	CG	AG	AC	AT	T	Hp2-1	
541 144	CG	A	A	AT	C	Hp2-1	
541 085	CG	A	A	T	T	Hp2-2	
541 049	CG	A	A	T	T	Hp2-1	
541 062	CG	AG	A	A	T	Hp1-1	

Sample genotypes and phenotypes (continued)

Sample	Group	-55	-61	-104	-295	Hp Genotype
541 036	CG	AG	AC	T	T	Hp2-2
541 073	CG	A	A	T	T	Hp2-1
541 115	CG	A	A	T	TC	Hp2-1
541 131	CG	AG	A	A	TC	Hp2-1
541 178	CG	A	AC	T	T	Hp2-2
541 098	CG	A	A	T	T	Hp2-2
541 180	CG	AG	A	AT	TC	Hp1-1
541 193	CG	AG	A	A	TC	Hp1-1
541 228	CG	A	AC	T	T	Hp2-2
541 234	CG	AG	A	A	T	Hp1-1
541 242	CG	AG	A	T	T	Hp2-1
541 256	CG	G	A	T	T	Hp1-1
541 353	CG	A	AC	T	T	Hp1-1
615 015	CG	A	A	T	T	Hp2-1
615 093	CG	AG	AC	A	T	Hp2-1
615 080	CG	AG	A	AT	T	Hp1-1
615 078	CG	AG	A	A	T	Hp1-1
615 067	CG	AG	A	A	T	Hp1-1
615 059	CG	A	A	AT	T	Hp2-2
615 044	CG	AG	A	AT	TC	Hp1-1
615 031	CG	AG	A	AT	T	Hp2-1
615 026	CG	AG	A	AT	TC	Hp1-1
615 107	CG	A	A	T	TC	Hp2-2
615 110	CG	A	A	AT	T	Hp2-2
615 121	CG	AG	A	AT	T	Hp2-1
615 136	CG	AG	A	AT	T	Hp2-1
615 325	CG	A	A	AT	T	Hp2-2
615 332	CG	AG	A	AT	TC	Hp2-1
615 340	CG	AG	A	T	T	Hp2-1
615 377	CG	A	A	T	T	Hp2-2
615 381	CG	AG	A	T	TC	Hp1-1
615 394	CG	AG	A	AT	TC	Hp2-1
615 406	CG	AG	A	AT	TC	Hp1-1
615 366	CG	AG	A	A	TC	Hp1-1
616 017	CG	AG	A	AT	T	Hp2-2
616 042	CG	AG	A	T	TC	Hp1-1
616 091	CG	AG	A	A	T	Hp2-1
616 358	CG	A	A	AT	T	Hp2-2
616 445	CG	G	A	AT	T	Hp2-1
616 453	CG	A	A	T	T	Hp1-1
616 457	CG	A	A	A	T	Hp1-1
616 472	CG	A	AC	AT	TC	Hp2-2
616 486	CG	AG	A	AT	TC	Hp2-1
616 499	CG	AG	A	A	T	Hp2-1
616 503	CG	AG	A	T	T	Hp1-1

P = Samples collected from HIV positive subjects

W = With unknown HIV status

CG = Samples provided by Dr Clive Gray (also from HIV positive subjects)

REFERENCES

An P, Martin MP, Nelson GW *et al* (2000), **Influence of CCR5 promoter haplotypes on AIDS progression in African-Americans,** *AIDS,* 14:2117-2122.

Arbour NC, Lorenz E, Schutte BC *et al* (2000), **TLR mutations are associated with endotoxin hyporesponsiveness in humans**, *Nat Genet*, 25: 187-191.

Asakawa J, Kodaira M, Nakamura N *et al* (1999**), Chimerism in humans after intragenic recombination at the haptoglobin locus during early embryogenesis**, *Proc Natl Acad Sci*, 96: 10314-10319.

Asleh R, Marsh S, Shilkrut M *et al* (2003), **Genetically Determined Heterogeneity in Hemoglobin Scavenging and Susceptibility to Diabetic Cardiovascular Disease***, Circulation Research*, 92: 1193-1200.

Aucan C, Walley AJ, Greenwood BM and VS Hill (2002), **Haptoglobin genotypes are not associated with resistance to severe malaria in The Gambia**, *Transactions of the royal Society of tropical medicine and hygiene*, 96: 327-328.

Barnicot NA, Garlick JP, Singer R and Weiner JS (1959), **Haptoglobin and transferrin variations in the Bushmen and some other South African people**, *Nature*, 184: 2042.

Barre-Sinoussi F, Chermann JC, Rey F *et al* (1983), **Isolation of T-lymphotropic retrovirus form a patient at risk for acquired immunodeficiency syndrome (AIDS),** *Science*. 20:868-871.

Bednarik DP and Folks TM (1992), **Mechanism of HIV-1 latency**, *AIDS,* 6:3-16.

Bensi G, Raugei G, Klefenz H and Cortese R (1985), **Structure and expression of the human haptoglobin locus**, *EMBO J*, 4: 119-126.

Black JA and Dixon GH (1968), **Amino acid sequence of alpha chains of human haptoglobins,** *Can J Biochem*, 48: 123-132.

Boelaert JR (1996), **Iron and infection**, *Acta Clin Belg*, 51: 213-221.

Borghans JAM, Beltman BJ and De Boer RJ (2004), **MHC polymorphism under host-pathogen coevolution,** *Immunogenetics*, 55:732-739.

Bottini E, Palmarino R, Lucarelli P, Lista F and Bottini N (2001), **ACP1 and human adaptability: Association with past malarial morbidity in the Sardinian population,** *American Journal of Human Biology*, 13: 753 – 760.

Bowman BH (1993), Haptoglobin, In: Bowman BH, editor, **Hepatic plasma proteins**, San Deigo, Academic Press, 159-167.

Carré N, Boufassa F, Hubert JB *et al* (1998), **Predictive value of viral load and other markers for progression to clinical AIDS after CD4+ cell count falls below 200/µL,** *International Journal of Epidemiology*, 27: 897-903.

Carrington M, Dean M, Martin MP and O'Brien SJ (1999), **Genetics of HIV infection: chemokine receptor CCR5 polymorphism and its consequences**, *Hum Mol Genet*, 8:1939-1945.

Carrington M, Nelson G and O'Brien SJ (2001), **Considering genetic profiles in functional studies of immune responsiveness to HIV-1**, *Immunology Letters*, 79: 131 -140.

Constants J, Viau M, Gouaillard C *et al* (1981), **Haptoglobin polymorphism among Subsaharian and West African groups. Haptoglobin phenotype**

determination by radioimmuno-electrophoresis on Hp0 samples, *Am J Hum Genet,* 33: 606-616.

Crow JF (1986), **Basic concepts in population, quantitative and evolutionary genetics,** W.H. Freeman, New York.

Dean M, Dean M, Carriton M, Winkler C (1996), **Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene, Hemophilia growth and development study, multicenter hemophilia cohort study, San Francisco City Cohort, alive study,** *Science,* 273:1856-1862.

Delanghe JR, Langlois MR, Boelaert JR *et al* (1998), **Haptoglobin polymorphisms, iron metabolism and mortality in HIV infection**, *AIDS,* 12: 1027-1032.

Dobryszycka W (1997), **Biological functions of haptoglobin – New pieces to an old puzzle,** *Eur J Clin Chem Biochem,* 35: 647-654.

Dorrington R, Bouren D, Bradshaw D, Laubscher R and Timaeus IM (2001), **The Impact of HIV/AIDS on adult mortality in South Africa: Technical report, Burden of Disease Research Unit,** Medical Research Council**,**

Eisaev B (1995), **Results of the treatment of patients with recurrence of pulmonary tuberculosis with different types of haptoglobin,** *Probl Tuberk*, 6: 20-22.

El Ghmati SM, van Hoeyveld EM, van Strijp JAG *et al* (1996), **Identification of haptoglobin as an alternative ligand for CD11b/CD18**, *J Immunol*, 2542-2552.

Elagib AA, Kider AO, Akerstrom B and Elbashir MI (1998), **Association of the haptoglobin phenotype (1-1) with falciparum malaria in Sudan,** *Trans R Soc Trop Med Hyg,* 92: 309-311.

Ezekowits RA, Kuhlman M, Groopman JE and Byrn RA (1989), **A human serum mannose-binding inhibits *in vitro* infection by human immunodefiency virus,** *J Exp Med*, 169: 185-196.

Farzadegan H, Henrard DR, Kleeberger CA, *et al* (1996), **Virologic and serologic markers of rapid progression to AIDS after HIV-1 seroconversion**, *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology,* 13:448-455.

Fedoseeva SV, Iusopova MM, Chukanova VP and Pospelov LE (1993), **Course of infiltrating pulmonary tuberculosis depending on the patient's genotype,** *Probl Tuberk*, 2:8-10.

Gallo RC, Salahuddin SZ, Popovic M, et al (1984**), Frequent detection and isolation of cytopathic retrovirus (HTLV-III) from patients with AIDS and at risk for AIDS.** *Science*, 224:500-503.

Galvani AP and Slatkin M (2003), **Evaluating plague and smallpox as historical selective pressures for the CCR5-Δ32 HIV-resistance allele,** *Proc Natl Acad Sci,* 100: 15276-15279.

Gao F, Bailes E, Robertson DL *et a*l (1999), **Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*,** *Nature,* 397:436-441.

Garred P, Madsen HO, Balslev U *et al* (1997), **Susceptibility to HIV infection and progression of AIDS in relation to variant alleles of mannose-binding lectin**, *Lancet*, 349: 236-240.

Gaynor R (1992), **Cellular transcription factors involved in the regulation of HIV-1 gene expression,** *AIDS,* 6:347-363.

Georges-Courbot MC, Lu CY, Makuwa M *et al* (1998), **Natural infection of a household pet red-capped mangabey (*Cercocebus torquatus torquatus*) with a new Simian Immunodeficiency Virus,** *J Virol*, 72: 600-608.

Giblett ER (1969), **Genetic markers in human blood,** 1st Edition, Blackwell Scientific, Oxford, 69 – 107.

Guiney DG and Kagnoff MF (1997), **Host/pathogen interactions: series introduction,** *J Clin Invest*, 99: 155.

Hancock MW (2002), **Chemokines and transplant immunobiology,** *J Am Soc Nephrol*. 13: 821-824.

Harakeh S, Jariwalla RJ and Pauling L (1990), **Suppression of human immunodeficiency virus replication by ascorbate in chronically and acutely infected cells,** *Proc Natl Acad Sci*, 87: 7245-7249.

Hartl DL and Clark AG (1989), **Principles of population genetics**, 2nd ed, Sinauer Associates, Sunderland, Massachusetts, pp 682.

Haynes BF, Pantaleo G, Fauci AS (1996), **Towards an understanding of correlations of protective immunity to HIV infection**, *Science,* 271: 324-328

Hitzeroth HW and Hummel K (1978), **Serum protein polymorphisms Hp, Tf, Gc, Gm, Inv and Pt in Bantu speaking South African Negroids,** *Anthropologische Anzeiger*, 36: 127-147.

Hopkins DR (2002), **The greatest killer in history: Smallpox,** University of Chicago Press, Chicago.

Isaksson B, Albert J, Chiodi F, Furucrona A, Krook A and Putkonen P (1988), **AIDS two months after primary human immunodeficiency virus infection**, *Journal of Infectious Diseases*, 158: 866-868.

Jenkins T (1972), **Genetic polymorphisms of Man in Southern Africa**, *M.D Thesis.*

Jenkins T and Steinberg AG (1966), **Some serum protein polymorphisms in Kalahari Bushmen and Bantu: Gamma globulins, haptoglobins and transferrins**, *Am J Hum Gen*, 18:399-407.

Jorde LB, Rogers AR, Bamshad M *et al* (1997), **Microsatellite diversity and the demographic history of modern humans,** *Proc Natl Acad Sci*, 94: 3100-3103.

Joshi H, **,** Raghavendra K, Subbarao SK, Sharma VP (1987), **Genetic markers in malaria patients of Delhi,** *Indian J Malariol*, 24: 33-38.

Kaslow RA, Carrington M, Apple R *et al* (1996**), Influence of combinations of human major histocompatibility complex on the course of HIV-1 infection**, *Nat Med*, 2: 405-411.

Kasvosve I, Gordeuk VR, Delanghe JR, *et al* (2002), **Iron status in black persons is not influenced by haptoglobin polymorphism,** *Clin Chem and Lab Med*, 40: 810-813 (Abstract).

Kasvove I, Gomo ZAR, Mvundura E *et al*. (2000), **Haptoglobin polymorphism and mortality in patients with tuberculosis,** *Int J Tuberc Lung Dis*, 4: 771 – 775.

Keyeu D, Zhou K, He F and Shen Y (2003), **LDA – A java-based linkage disequilibrium analyzer,** *Bioinformatics,* 19: 2147-2148.

Kidd JR, Pakstis AJ, Zhao H, *et al* (2000), **Haplotype and linkage disequilibrium at the phenylalamine hydroxylase locus, PAH, in a global representation of populations,** *Am J Hum Genet*, 66: 1882-1889.

Kidd KK, Jenkins T, Morar B, *et al* (1998), **A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus**, *Human Genetics*, 103: 211-227.

Koch W, Latz W, Eichinger M *et al* (2002), **Genotyping of the common haptoglobin Hp1/2 polymorphism based on PCR,** *Clin Chem,* 48: 1377-1382.

Koda Y, Watanabe Y, Soejima M *et al*. (2000), **Simple PCR detection of haptoglobin gene deletion in anhaptoglobinemic patients with antihaptoglobin antibody that causes anaphylactic transfusion reactions,** *Blood*, 95: 1138 – 1143.

Koda Y, Soejima M, Yoshioka N and Kimura H (1998), **The haptoglobin gene deletion responsible for anhaptoglobinemia,** *Am J Hum Genet,* 62: 245 – 252.

Lane AB, Soodyall H, Ratshikhopha ME *et al* (2001), **Genetic structure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies,** *American Journal of Physical Anthropology*, 119: 175-185.

Langlois MR and Delanghe JR (1996)**, Biological and clinical significance of haptoglobin polymorphism in humans**, *Clin Chem,* 42:1589-1600.

Levy JA, Hoffman AD, Kramer SM, et al (1984), **Isolation of lymphocytipathic retroviruses from San Francisco patients with AIDS**, *Science*, 225:840-842.

Lewontin RC (1964), **The interaction of selection and linkage. General considerations; heterotic models,** *Genetics,* 49: 49-67.

Lewontin RC (1988**), On measures of gametic disequilibrium,** *Genetics*, 120: 849-852.

Linke RP (1984), **Typing and subtyping of haptoglobin from native serum disc gel electrophoresis in alkaline buffer: application to routine screening**. *Anal Biochem*, 141: 55-161.

Lipscombe RJ, Sumiya M, Hill AV *et al* (1992), **High frequencies in African and non-African populations of independent mutations in the mannose binding protein gene.** *Hum Mol Genet*, 1: 709-715.

Liu H *et al* (1999) **Polymorphism in RANTES chemokine promoter affects HIV-1 disease progression**, *Proc Natl Acad Sci*, 96: 4581-4585.

Louagie HK, Brouwer TJ, Delanghr JR *et al.* (1996), **Haptoglobin polymorphism and chronic hepatitis C,** *J Hepatol,* 25: 10 – 14.

Madsen HO, Garred P, Thiel S *et al* (1995), **Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein,** *Journal of Immunology,* 155: 3013-3020.

Maeda N (1985), **Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair,** *J Biol Chem,* 260: 6698-6709.

Maeda N (1991), **DNA polymorphisms in the controlling region of the human haptoglobin genes: A molecular explanation for the haptoglobin 2-1 modified phenotype,** *Am J Hum Genet,* 49: 158-166.

Maeda N and Smithies O (1986), **The evolution of multigene families: Human haptoglobin genes,** *Ann Rev Genet*, 20:81-108.

Maeda N, Yang F, Barnett DR, Bowman BH, and Smithies O (1984), **Duplication within the haptoglobin Hp^2 gene**, *Nature*, 309: 131-135.

Manoharan A (1997), **Congenital haptoglobin deficiency,** *Blood*, 90: 1709.

Martin MP, Dean M, Smith MW *et al* (1998), **Genetic acceleration of AIDS progression by a promoter variant of CCR5,** *Science*, 282: 1907-1911.

Martison JJ, Chapman NH, Rees DC, Lui YT and Clegg JB (1997**), Global distribution of the CCR5 gene 32-base pair deletion,** *Nat Genet*, 16: 100-103.

Matheron S, Pueyo S , Damond F *et al* (2003), **Factors Associated with Clinical Progression in HIV-2-Infected Patients Included in the French ANRS Cohort,** *The French ANRS cohort AIDS*, 17:2593-2601.

Matznetter T and Spielmann W (1969), **Blood groups of Mozambique Bantu tribes,** Z. *Morphol Anthropol*, 61:57-61.

McDermid EM and Vos GH (1971a), **Serum protein groups in South African Bantu I: Albumin ceruloplasmin, transferring and haptoglobin,** *SA J of Med Sci*, 36: 7-14.

McDermott DH, Zimmerman PA, Guignard F *et al* (1998), **CCR5 promoter polymorphism and HIV-1 disease progression, Multicenter AIDS Cohort study,** *Lancet,* 352: 866-870.

McMichael A and Klenerman P (2002), **HLA leaves its footprints on HIV**, *Science*, 296: 1410 -1411.

Melamed-Frank M, Lache O, Enav BI, *et al* (2001), **Structure-function analysis of the antioxidant properties of haptoglobin,** *Blood,* 98: 3693 -3698.

Mellors J W, Kingsely LA, Rinaldo CR *et al* (1995), **Quantitation of HIV-1 RNA in plasma predicts outcomes after seroconversion,** *Ann Intern Med*, 122:573-579.

Mellors J W, Kingsely LA, Rinaldo CR Jr *et al* (1996), **Prognosis in HIV-1 infection predicted by the quantity of virus in plasma,** *Science,* 272: 1167-1170.

Michael NL, Louie LG, Rohrbaugh AL *et al* (1997), **The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and progression,** *Nat Med*, 3: 1160-1162.

Minang TJ, Gyan BA, Anchang JK *et al* (2004), **Haptoglobin phenotypes and malaria infection in pregnant women at delivery in western Cameroon,** *Acta Tropica,* 90: 107-114.

Molina JM, Schindler, Ferriani R, *et al* (1990), **Production of cytokines by peripheral blood monocytes/macrophages infected with HIV-1,** *J Infect Dis,* 161: 888-892

Moore CB, John M, James IR *et al* (2002), **Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level,** *Science,* 296: 1439-1443.

Morris L and Williamson C (2001), **Host and viral factors that impact on HIV-1 transmission and disease progression in South Africa**, *S Afr Med J*, 91:212-215.

Murray PJ and Young RA (1992), **Stress and immunological recognition in host-pathogen interactions**, *Journal of Bacteriology,* 174:4193-4196.

Nakayama EE, Hoshino Y, Xin X *et al* (2002), **Polymorphism in the interleukin-4 promoter affects acquisition of human immunodeficiency virus type 1 syncytium-inducing phenotype,** *J Virol,* 74: 5452-5459.

Nei M (1987) **Molecular Evolutionary Genetics,** Columbia University Press, pp.192-193.

Nissapatorn V, Lee C, Fatt QK and Abdullah KA (2003), **AIDS-Related Opportunistic Infections in Hospital Kuala Lumpur,** *Jpn J Infect Dis*, 56: 187-192.

Nurse GT, Elphinstone CD AND Jenkins T, (1974), **Mseleni joint disease: population genetic studies,** *SA J of Sci*, 70:360-365.

Nurse GT, Weiner JS and Jenkins T (1985), **The people of Southern Africa and their affinities,** Oxford University Press, 319.

O'Brien SJ and Dean M (1997), **In search of AIDS-resistance genes**, *Scientific American*, 277:44-51.

O'Brien SJ and Moore JP (2000), **The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS,** *Immunological Reviews*, 177:99-111.

O'Bryan MK, Crima J, Mruk D *et al* (1997) **Haptoglobin is a Sertoli cell product in the rat seminiferous epithelium: its purification and regulation,** *J Androl*, 18, 637–645.

Okayama H, Curiel DT, Brantly ML, Holmes MD and Crystal RG (1989), **Rapid, nonradioactive detection of mutants in the human genome by allele-specific amplification,** *J Lab Clin Med*, 114: 105 – 113.

Oliviero S and Cortese R (1989), **The human haptoglobin gene promoter: Interleukin-6-responsive elements interact with a DNA-binding protein induced by interleukin-6,** *EMBO Journal*, 8:1145-1151.

Olson, G.E., Winfrey, V.P., Matrisian, P.E. *et al*. (1997), **Specific expression of haptoglobin mRNA in implantation-stage rabbit uterine epithelium**. *J. Endocrinol.*,152:69–80.

Plumelle Y (2003), **HIV, 'An evolving species. Roles of cellular activation and co-infections',** *Medical Hypothesis,* 61:136 -157.

Pulgiese A, Gennero L, Pescarmona GP *et al*. (2002), **Serum citrate levels, haptoglobin haplotypes and transferrin receptor (CD71) in patients with HIV-1 infection**, *Infection* , 30: 86 – 89.

Quaye IKE, Brandful J, Ekuban FA, Gyan B and Ankrah N (2000), **Haptoglobin polymorphism in Human Immunodeficiency Virus infection: Hp0 phenotype limits depletion of CD4 cell counts in HIV-1-seropositive individuals,** *The Journal of Infectious Diseases,* 181: 1483 -1485.

Quillent C, Oberlin E, Braun J *et al* (1998), **HIV-1 resistance phenotype conferred by combination of two separate inherited mutations of CCR-5 gene,** *Lancet,* 351:14-18.

Quinn TC, Wawer MJ, Sewankambo N, *et al* (2000), **Viral load and heterosexual transmission of human immunodeficiency virus type 1: Rakai project study group**. *N Engl J Med*, 342:921-929.

Ras GJ, Simson IW, Anderson R *et al* (1983), **Acquired immunodeficiency syndrome: A report of 2 South African cases,** *S Afr Med J*, 64: 104-142.

Rosner B (1990), **Fundamentals of Biostatistics**, 3rd edition, PWS-KENT Publishing Company, USA.

Saah AJ, Hoover DR, Weng S *et al* (1998), **Association of HLA profiles with early plasma viral load, CD4+ cell count and rate of progression to AIDS**

following acute HIV-1 infection: Multi-center AIDS cohort study, *AIDS*, 12: 2107-2113.

Schneider S, Roessli D and Excoffier L (2000), Arelquin ver 2.000: **A software for population genetics data analysis.** Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Schultze HE and Hermans JF (1996), Nature and metabolisms of extracellular proteins, In: Schultze HE and Hermans JF, editors, **Molecular biology of human proteins,** Amsterdam, *Elserivier*, 384-402.

Searl S and Blackwell JM (1998), **Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility**, *J Med Genet*, 36: 295-299.

Sequencher™ 4.0. (1999), Gene Codes Corporation, Inc.

Sharpe-Timms KL, Penney LL, Zimmer RL, Wright JA, Zhang Y and Surewicz K (2002), **Partial purification and amino acid sequence analysis of endometriosis protein-II (ENDO-II) reveals homology with tissue inhibitor of metalloproteinases-1 (TIMP-1*),*** *Journal of Clinical Endocrinology & Metabolism,* 80: 3784-3787.

Shisana O and Simbayi LC (2002), **Nelson Mandela/HSRC study of HIV/AIDS: South African national HIV prevalence, behavioural risks and mass media: household survey,** Human Sciences Research Council, Cape Town.

Smithies O (1955), **Genetic control of some serum proteins in normal humans,** *Nature*, 176: 1265-1266.

Smithies O and Walker NF (1955), **Genetic control of some serum proteins in normal humans**, *Nature*, 176: 1265-1266.

Sokal RR and Rohlf FJ (1981), *Biometry*, 2nd edition, WH Freeman and Company, USA.

Stephens JC, Reich DE, Goldstein DB Stephens (1998)**, Dating the origin of the CCR5-Δ32 AIDS resistance allele by the coalescence of haplotypes**, *Amer J Hum Genet*, 62:1507-1515.

Struyf F, Thoelen I, Charlier N *et al* (2000), **Prevalence of CCR5 and CCR2 HIV-coreceptor gene polymorphisms in Belgium,** *Human Heredity*, 50: 304-307.

Super M, Thiel S, Lu J, Levinsky RJ and Turner MW (1989), **Association of low levels of mannan-binding protein with a common defect in opsonisation,** *Lancet,* 2: 1236-1239.

Swanson P, Devare SG and Hackett J Jr (2003), **Molecular characterization of 39 HIV isolates representing group M (subtypes A-G) and group O: sequence analysis of gag p24, pol integrase and env gp41,** *AIDS Res Hum Retroviruses*, 19: 625-629.

Takemura T and Hayami M (2004), **Phylogenetic analysis of SIV derived from Mandrill and Drill**, *Frontiers in Biosciences*, 9: 513-520.

Teye K, Quaye IKE, Koda Y *et al* (2004), **A novel I247T missense mutation in the haptoglobin 2 β-chain decreases the expression of the protein and is associated with anhaptoglobinemia,** *Human Genetics*, 114: 499-502.

Teye K, Quaye IKE, Koda Y, Tsuneoka M *et al*. (2003), **A-61C and C-101G Hp gene promoter polymorphisms are, respectively, associated with**

ahaptoglobinemia and hypohaptoglobinemia in Ghana, *Clin Genet*, 64: 439 – 443.

Tishkoff SA and Williams SM (2002), **Genetic analysis of African populations: human evolution and complex disease,** *Nat Rev Genet,* 3: 611-621.

Trape JF and Fribourg-Blanc A (1988), **Ahaptoglobinemia in African populations and its relation to malaria endemicity**, *Am J Epidemiol*, 127: 1282-1288.

Tregouet DA, Escolano S, Tiret L, Mallet A and Golmard JL (2004), **A new algorithm for haplotype-based association analysis: the Stochastic-EM algorithm,** *Ann Hum Genet*. 68:165-77.

Turner MW (1996), **Mannose-binding lectin: the pluripotent molecule of the innate immune system,** *Immunol Today*, 17:532-540.

Ullum H, Lepri AC, Hassan A *et al* (1999), **Natural immunity and HIV disease progression**, *AIDS*, 3:557-563.

UNAIDS (2001), **AIDS epidemic update, December.**

UNAIDS (2002), **Report on the global HIV/AIDS epidemic, July.**

UNAIDS (2003), **AIDS epidemic update, December.**

van Vlierberge H, Langlois M and Delanghe J (2004), **Haptoglobin polymorphism and iron homeostasis in health and in disease,** *Clinica Chimica Acta,* 345: 35-42.

Weber J and Alcorn K (2000), **Conference Report: Origins of HIV and the AIDS Epidemic,** *Medscape HIV/AIDS*, 6: 8.

Weinberg (1990), **Cellular Iron Metabolism in Health and Disease,** *Drug Metabolism Reviews,* 22, 531-579.

Weinberg ED (1978), **Iron and infection, Microbiol Rev,** 42: 45-66.

Williamson C and Morris L (2001), **Host and viral factors that impact on HIV-1 transmission and disease progression in South Africa,** *S Afr Med J*, 91:212-215.

Williamson C, Loubser SA, Brice B, Joubert G *et al*. (2000), **Allelic frequencies of host genetic variants influencing susceptibility to HIV-1 infection and disease in South African population,** *AIDS*, 14: 449 -451.

Wuyts B, Hetet G, Grandchamp B and Delanghe JR (2002), **Novel haptoglobin insertion/deletion polymorphism is associated with the lipid profile and C-Reactive Protein (CRP) concentration,** *Clin Chem Lab Med,* 40: 469 – 474.

Yang C, Dash BC, Simon F *et al* (2000), **Detection of diverse variants of human immunodeficiency virus-1 groups M, N and O and simian immunodeficiency viruses from chimpanzees by using generic pol and env primer pairs,** *J Infect Dis,* 181:1791-1795.

Yang F, Brune JB, Baldwin WD, Barnett and Bowman BH (1983), **Identification and characterization of human haptoglobin cDNA,** *Proc Natl Acad Sci*, 80: 5875-5879.

Yip SP, Lovegrove JU, Rana NA, Hopkinson DA and Whitehouse DB (1999), **Mapping recombination hotspots in human phosphoglucomutase (PGM1),** *Human Molecular Genetics*, 8: 1699-1706.