

https://dspace.org/

Molecular epidemiology and cellular immunology of respiratory syncytial virus in South Africa.

Venter., Maria.

2003-03-31
https://hdl.handle.net/10539/26606

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

MOLECULAR EPIDEMIOLOGY AND CELLULAR IMMUNOLOGY OF RESPIRATORY SYNCYTIAL VIRUS IN SOUTH AFRICA

Maria Venter

A Thesis Submitted To The Faculty Of Health Science
University Of The Witwatersrand, Johannesburg
For The Degree Of Doctor Of Philosophy

Johannesburg 2003

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other university.

.....

Maria Venter

On this 31 day of March 2003

For Hannes

For supporting and encouraging me in all big challenges I undertake and always believing that I'll succeed.

"Nothing in life is to be feared. It is only to be understood".

Marie Curie (1867 - 1934) Polish-French chemist

PUBLICATIONS

1. M. Venter, S. A. Madhi, C. T. Tiemessen, B.D. Schoub Genetic diversity and Molecular epidemiology of Respiratory Syncytial Virus over four consecutive seasons in South Africa: Identification of new subgroup A and B genotypes. *Journal of General Virology*, (2001), 82, 2117-2124
2. S.A. Madhi; M. Venter, A. Madhi, K. Peterse, K. Klugman. Differing manifestations of respiratory syncytial virus-associated severe lower respiratory tract infections in human immunodeficiency virus type 1-infected and uninfected children. *The Pediatric Infectious Disease Journal (2001),* 20:2, 164-170.
3. M. Venter, M. Collinson, B.D. Schoub. Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: comparison of viruses and genotypes causing different disease manifestations. *Journal of Medical Virology,* Vol. 2002 Nov;68(3):452-61.
4. S. A. Madhi; M. Venter, R. Alexandra, H. Lewis, Y. Kara, M. Greef, C. Larsen. Respiratory syncytial virus associated illness in high risk children and national characterisation of the virus genotype in South Africa. *Journal of Clinical Virology*, In Press.
5. M. Venter; M. Rock; A. Puren; C.T. Tiemessen; J.E. Crowe Jr. Respiratory syncytial virus nucleoprotein-specific cytotoxic T-cell epitopes in a South African population of diverse HLA types are conserved in circulating field strains, *Journal of Virology*, In Press.

PRESENTATIONS

1. M. Venter, S.H. Madhi, B.D.Schoub. Molecular epidemiology of RSV in South Africa. Oral presentation, WHO meeting on respiratory syncytial virus epidemiology in developing countries, Johannesburg, November 2000.
2. M. Venter, S.H. Madhi, B.D. Schoub. Genetic diversity and Molecular epidemiology of Respiratory Syncytial Virus over four consecutive seasons in South Africa: Identification of new subgroup A and B genotypes. Poster presentation, International RSV meeting, RSV after 45 years, Segovia Spain, September 2001.
3. M. Venter, M. Collinson, S.H. Madhi, B.D. Schoub. Molecular epidemiology of RSV in South Africa. Oral presentation, WHO meeting on respiratory syncytial virus epidemiology in developing countries, Madrid, Spain, October 2001.
4. M. Venter, C.T. Tiemessen, J.E. Crowe Jr. Identification of novel Respiratory Syncytial virus specific memory CD8+ cytotoxic T-lymphocyte epitopes in humans. Oral presentation, XIIth International Congress of Virology, Paris, July 2002.

Acknowledgements

The experimental work was performed at the Molecular Virology Unit at the National Institute for Communicable Diseases, (former National Institute for Virology), Johannesburg, South Africa. Part of Chapter 5 was carried out during my visit to Vanderbilt University Medical School, Nashville Tennessee, USA as part of a collaborative study with the Pediatric Infectious Disease Department.

My Sincere thanks to:

My supervisors, Dr. Caroline T. Tiemessen and Prof Barry Schoub at the National Institute for Communicable diseases for their constant support and encouragement and for revision of this thesis.

Dr. James E. Crowe Jr. at Vanderbilt University Medical School in Nashville Tennessee for an excellent collaboration and mentorship through out my RSV studies.

Dr. Shabir Madhi at Chris Hani Baragwanath hospital for provision of most of the specimens used in this study. The staff of the Agincourt Health and Population Unit, University of the Witwatersrand and pediatricians that took part in the Abbott RSV study, for specimen collection. My Colleagues at the NICD for their interest and support, especially Stephina Nyoka and Sharon Shalekoff for helping me with the flow cytometry work, Dr. Clive Gray for technical advice on CTL immunology; Dr. Adrian Puren and Maria Paximadis for HLA typing, Antoinette Grobbelaar for showing me how to use PAUP; and finally Amelia Buys, Michelle Morgan, Areeva Oliver and Lorraine Cranston for diagnostic work as well as the medical staff at the NICD and blood donors that took part in the CTL study.

Dr. Peter Wright and Dr. Susan Robertson of the RSV-WHO study group for their interest in this study and for helping me to initiate the RSV Research program in South Africa through the opportunity to visit Vanderbilt University.

And last but not least, Hannes for his patience, constant support and for always believing in me and my parents for inspiring me to study, teaching me to trust in God and providing me with all the opportunities and encouragement I could ever need throughout my life. ... Baie dankie!

ABSTRACT

Respiratory syncytial virus (RSV) is the major viral cause of severe lower respiratory tract infection in children and there is currently no commercial vaccine available. Research about RSV epidemiology and protective immunology is crucial for the development of effective vaccines and vaccination programs. No information exists about the molecular epidemiology and RSV-specific cytotoxic T-cell (CTL) response in South Africa, which is thought to be essential for the disease-free control of RSV.

The molecular epidemiology of RSV in South Africa was firstly characterised over four consecutive seasons (1997-2000) at a single hospital, serving Soweto. Both subgroups A and B were shown to be important in epidemics. Sequence- and phylogenetic analysis of part of the G-protein identified 5 subgroup A and 4 subgroup B genotypes. Of these, one subgroup A and three subgroup B genotypes did not cluster with the genotypes previously identified in the rest of the world and were named SAA1, SAB1, SAB2 and SAB3. Different circulation patterns were identified for subgroups A and B. Subgroup A was more diverse and dominant genotypes were replaced by new genotypes in consecutive seasons, although a single genotype remained dominant for two seasons. The same two subgroup B genotypes remained dominant in all four seasons. Positive selection was identified in both subgroups, suggesting that immune-driven selection may contribute to reestablishment of annual epidemics. RSV specimens from infants attending rural community clinics were next compared to specimens from hospitalised children. Evidence was found that the same viruses could cause mild upper respiratory tract infections, lower respiratory tract infections and severe RSV in young infants suggesting no correlation between the infecting strain and disease pathogenesis. Comparisons to isolates from Mozambique and five communities from the rest of South Africa suggested that cross-border transfer and wide distribution of strains might occur in a single epidemic season. The genotype that was dominant in Johannesburg in 2000 was isolated throughout the rest of the country, suggesting that successful genotypes may have the ability to spread nationwide.

To characterise the RSV-specific CTL response in South Africans, IFN-gamma ELIspot was used to screen 37 healthy adults with diverse HLA backgrounds for

memory CTL responses to peptides covering the complete N-protein of RSV. Responses of more than 40 spot forming units per million (SFU/million) cells were detectable in 21 individuals. The dominant responses were further characterized and two 14-mer peptides identified with cytolytic activity restricted to HLA B*08, A*02 and B*15. Finer mapping identified a novel 9-mer B*08-restricted, RSV-specific, CD8+ CTL epitope. These HLA types are common in the South African population and will be useful for subsequent studies of CTL responses to RSV in infants. This is the first time that a dominant CTL response has been characterised within a complete RSV protein, and is only the fourth human CTL epitope to be identified in RSV, and the first to these particular HLA types. The identified CTL epitope was shown to be completely conserved in the nucleoproteins of all genotypes identified in South Africa to date. No indications of positive selection could be found in the selected epitopes or in the N-protein, suggesting that regular updating would not be required if this epitope is used to monitor the CTL response in RSV vaccine trials.

Data collected in this study may contribute to the development of effective vaccines and vaccination programs for the control of RSV in South Africa and other countries.

TABLE OF CONTENTS

CHAPTER 1 Literature Survey 1

1.1 Introduction 1

1.2 The Virus 3

1.2.1 Classification 3

1.2.2 Viral structure and Replication 4

1.3 The epidemiology 10

1.3.1 Molecular epidemiology 12

1.3.2 Immune response 18

1.3.3 Vaccine development 38

1.4 Study Objectives 41

CHAPTER 2 Genetic diversity and molecular epidemiology of RSV over four consecutive seasons in Soweto, South Africa: Identification of new subgroup A and B genotypes 46

2.1 Introduction 46

2.2 Methods 49

2.2.1 Patient specimens and viruses 49

2.2.2 RNA extraction 49

2.2.3 Subgroup identification: RT-PCR and multiplex nested PCR 50

2.2.4 Analysis of the PCR products 52

2.2.5 Nucleotide sequencing 53

2.2.6 Phylogenetic analysis 53

2.2.7 Genotype assignment 54

2.3 Results: 56

2.3.1 Nested PCR for subgroup identification 56

2.3.2 RSV subgroup distribution over four years at Chris Hani Baragwanath hospital 56

2.3.3 Phylogenetic analysis and genotype assignment of South African RSV clinical specimens 58

2.3.4 Amino acid sequence comparison 65

2.3.5 Other estimates of selection 67

2.3.6 Genotype distribution pattern over four years in a single community in South Africa 70

2.4 Discussion 72

CHAPTER 3 Molecular epidemiological analysis of community circulating RSV in rural South Africa: comparison of viruses and genotypes causing different disease manifestations 76

3.1 Introduction 76

3.2 Materials and Methods 79

3.2.1 Patients, specimens and viruses 79

3.2.2 RNA extraction 80

3.2.3 Subgroup identification: RT-PCR, multiplex nested PCR, and Nucleotide sequencing 80

3.2.4 Phylogenetic analysis 80

3.3 Results 81

3.3.1 Subtyping and phylogenetic analysis of the Agincourt specimens 81

3.3.2 Amino acid analysis 84

3.4 Discussion 88

CHAPTER 4 Distribution of RSV genotypes across Southern Africa during a single epidemic season 93

4.1 Introduction 93

4.2 Methods 97

4.2.1 Patients and viruses 97

4.2.2 Specimen collection and RSV diagnosis in study group 1 98

4.2.3 Subtyping, genotyping and phylogenetic analysis 98

4.2.4 Calculation of positive selection: 99

4.3 Results 100

4.3.1 RSV genotype circulation throughout South Africa during one season 100

4.3.2 Evolvement in RSV genotypes: Comparison of all clinical specimens from South Africa from study groups 1, 2 and 3 102

4.4 Discussion 107

CHAPTER 5 Identification and characterisation of dominant human RSV cytotoxic T-cell epitopes in a population of diverse HLA types 110

5.1 Introduction 110

5.2 Methods 112

5.2.1 Subjects 112

5.2.2 Media and Components 112

5.2.3 HLA typing. 112
5.2.4 Synthetic peptides. 112
5.2.5 ELIspot assay 113
5.2.6 Intracellular cytokine staining 113
5.2.7 Peptide-specific cell lines 114
5.2.8 Establishment of EBV-transformed B-cell lines. 114
5.2.9 Cytotoxicity Assay 115
5.2.10 Cell subset depletion studies. 116
5.3 Results 117
5.3.1 Identification of an RSV-specific CTL response by IFN-gamma ELIspot and ICC 117
5.3.2 Characterisation of RSV N-protein specific CTL epitopes by chromium release assay 122
5.3.3 Mapping of optimal epitopes using 9- and 10-mer truncated peptides 126
5.3.4 Confirmation of MHC class I phenotype 127
5.4 Discussion 128
CHAPTER 6 Conservation of Nucleoprotein-Specific Cytotoxic T-cell Epitopes of Human Respiratory Syncytial Virus 133
6.1 Introduction 133
6.2 Materials and Methods 135
6.2.1 Viruses and RNA 135
6.2.2 RT-PCR 135
6.2.3 Sequencing and sequence analysis 136
6.3 Results 137
6.3.1 RT-PCR 137
6.3.2 Amino acid sequence comparison 137
6.4 Discussion 141
CHAPTER 7 Concluding remarks 145
REFERENCES 154
APPENDIX A Fisher's exact test for positive selection in Subgroup A and B RSV clinical specimens described in Chapter 4. 189
APPENDIX B South African RSV sequences submitted to the genbank. 192
APPENDIX C Ethics clearance 194

LIST OF FIGURES

Figure 1.1: Map of the negative sense genomic RNA of RSV strain A2 Reproduced from (Collins *et al.*, 1999).....5

Figure 1.2: A three dimensional representation of the RSV virion. 5

Figure 1.3: Variable and constant regions of the G-protein of the A2 isolate of RSV. 7

Figure 1.4: Genes activated in early phase RSV infection of respiratory epithelial cells. 20

Figure 1.5: Immune determinants in RSV disease.....23

Figure 1.6: CD8+ CTL antigen recognition and effector function. 27

Figure 1.7: MHC class I (left) and II (right) molecules consisting of the peptide binding cleft and the membrane spanning domains (A). The empty binding cleft shown from above (B.) and bound to peptide (C.) The MHC class I molecule binds peptides by their ends (left) while the MHC class II molecule binds peptides by interactions along the length of the binding groove. (Adapted from Janeway *et al.*1999).....30

Figure 1.8 a & b: The frequency of the major HLA type A and B alleles in selected populations. 33

Figure 2.1 a) RSV Subtyping by multiplex RT-PCR and b.) RSV subtyping by multiplex nested PCR. 57

Figure 2.2: RSV subgroup circulation at Chris Hani Baragwanath Hospital during 1997-2000.....59

Figure 2.3: Maximum Parsimony trees of South African RSV subgroup A and B nucleotide sequences. 64

Figure 2.4: Amino acid alignment of the second variable region of the G-protein for subgroup A (A) and subgroup B (B) South African specimens and reference sequences of the different genotypes. 66

Figure 2.5: Identification of genotype-specific mutations and mutations that may influence the secondary structure of the G-protein. 68

Figure 2.6: Distribution of subgroup A and B genotypes over four seasons (1997-2000) at Chris Hani Baragwanath hospital. 71

Figure 3.1: Phylogenetic trees of South African RSV Subtype A and B Specimens. 83

Figure 3.2: Comparison of the G-protein amino acid sequences of community and hospital specimens of RSV subgroup A (A.) and subgroup B (B.) genotypes....85

Figure 4.1: The map of South Africa (A.) indicates the sites included in this study with stars. The climatic regions of the world according to Koeppen's classification system are shown in (B.) with an enlargement of Southern Africa shown in (C.). 96

Figure 4.2. Phylogenetic analysis of South African subgroup A and B RSV, isolated from different geographic locations.....104

Figure 5.1: Reaction profiles of subjects D16, D5, and D10 to individual N-protein 14mer peptides as measured in IFN-gamma.....120

Figure 5.2: The ELIspot screening assay used to test PBMCs obtained from thirty-seven healthy adult subjects for IFN-gamma production against a panel of 127 individual peptides spanning the RSV N protein.. 121

Figure 5.3: Analysis of peptide-reactive $CD8^+$ T cells from subject D16. PBMC were incubated in the presence or absence of indicated peptides for 6 hr in an IFN-γ/ICS assay. The data are presented as the percentage of $CD3^+/CD8^+$ lymphocytes that were $CD69^+$ and scored positive for IFN-γ production.....121

Figure 5.4: Recognition of viral epitopes by peptide-specific effector cells generated from subject D16 (HLA-type indicated) in standard ^{51}Cr-release assays. (see Figure 5.4 on previous page).....125

Figure 6.1: RT-PCR of RSV subgroup A and B nucleoprotein genes. 138

Figure 6.2: Amino acid alignment of the N protein sequences of human RSV subgroup A and B genotypes with bovine RSV, ovine RSV, pneumovirus of mice, and avian and human metapneumoviruses (MPV) relative to the A2 prototype. (See figure on previous page).....140

Figure 6.3:. Phylogenetic relationship of the *Pneumovirinae*. 144

LIST OF TABLES:

Table 2.1: RSV-specific primers for RT-PCR and nested PCR.....51

Table 2.2: Average genetic P-distances between and within subgroup A and B genotypes. 61

Table 2.3: The ratio of non-synonymous substitutions per non-synonymous site (Ka) to synonymous substitutions per synonymous site (Ks) calculated according to the Modified Nei-Gojobori method (Kumar *et al.*, 2001).....69

Table 3.1: Distribution of age, disease severity (upper respiratory tract infection (URTI); Lower respiratory tract infection (LRTI); severe LRTI and gender of children infected by the individual RSV strains. a.) Individual cases; b.) Summary for identified genotypes 87

Table 4.1: RSV-positive specimens from children in study group 1 with lower respiratory tract infections enrolled from April 2000 through to September 2001 in different demographic regions of South Africa (Madhi *et al.* 2002). 101

Table 4.2: Respiratory syncytial virus genotypes isolated in distinct geographical locations across South Africa during the year 2000.....101

Table 5.1: HLA types and ethnic groups of subjects selected after IFN-gamma ELIspot screen for which EBV cell lines were established and additional cell lines used for determining HLA restrictions 119

Table 5.2: RSV N-protein specific IFN-gamma production and cytolytic activity to selected peptides.....123

Table 6.1: Distance matrix of the average between group amino acid P-distances of the members of the Pneumovirinae. 138

SELECTED ABBREVIATIONS USED IN THE TEXT

ALRI	Acute lower respiratory tract infections
BCL	B-lymphocyte cell line
CD4/8/3	Clusters of differentiation
CTL	Cytotoxic T-cell
dNTP	Deoxynucleoside triphosphate
EBV	Epstein Bar Virus
ELIspot	Enzyme linked immunospot
FACS	Fluorescence activated cell sorter
FCS	Fetal calf serum
GM-CSF,	Granulocyte-macrophage colony-stimulating factor
H-2	Histocompatibility antigens of mice
HIV	Human immunodeficiency virus
HLA	Human Leukocyte Antigen
ICAM-1	Intracellular adhesion molecule 1
ICC	Intracellular cytokine
IFA	Immunofluorescence assay
IFN	Interferon
Ig	Immunoglobulin
IL-6, IL-8, IL-11	Interleukin –6,-8,-11
iNOS	Inducible Nitric Oxide synthase
LRTI	Lower respiratory tract infection
MHC	Major histocompatibility complex
MIP1α	macrophage inflammatory protein 1 alpha
mM	Millimolar
mRNA	Messenger Ribonucleic acid
n/c	Not calculated
ng	Nanogram
NICD	National Institute for Communicable diseases
NIV	National Institute for Virology
NK-cells	Natural killer cells

ORF	Open reading frame
PCR	Polymerase chain reaction
PFU	Plaque forming unit
RFLP	Restriction fragment length polymorphism
RNase	ribonuclease
RSV	Respiratory syncytial virus
RT-PCR	Reverse transcriptase-polymerase chain reaction
TAP	Transporters associated with antigen processing
Th-1	Type 1 T-helper cells
TNF	Tumour necrosis factor
U	Unit
µ	Micro
WHO	World Health Organisation

CHAPTER 1

Literature Survey

1.1 INTRODUCTION

Respiratory syncytial virus (RSV), a member of the Pneumovirus genus in the family Paramyxovirus, is the major cause of serious lower respiratory tract infection in infants and young children in both developed and developing countries (Selwyn, 1990). RSV affects approximately 90% of infants and young children by age 2 years, with peak rates occurring in infants of 6 weeks to 6 months, but particularly those of less than 3 months of age (Simoes, 1999). Infection results in lower respiratory tract infections in 40% of primary infections and is responsible for hospitalisation in 0.1-2% of infants less than one year of age (MRC Report, 1978). A three-year long study carried out at the Chris Hani Baragwanath hospital, has also confirmed RSV to be the most important single etiological agent of acute respiratory infection in South Africa (Madhi *et al.*, 2000). RSV has further been implicated as a serious pathogen in certain adult populations especially immunocompromised patients like bone marrow transplant recipients, and the elderly (Englund *et al.*, 1988; Falsey *et al.*, 1992) however, it also contributes to community-acquired pneumonia in adults of all ages (La Montagne, 1997).

RSV was first identified in 1956 as the agent causing chimpanzee coryza (Morris Jr *et al.*, 1956) and was named respiratory syncytial virus after its subsequent isolation from children with pulmonary disease in Baltimore, USA (Chanock, 1957). Although the World Health Organisation (WHO) has identified RSV as one of highest priority organisms for development of a preventative vaccine (Crowe, 1995), there is presently still no commercial vaccine available (Crowe, 2001). Several challenges exist to RSV vaccine development:

In the first place, infection may occur in the presence of pre-existing immunity, e.g. infection occurs in young infants in the presence of maternally-derived antibodies and reinfection is the norm throughout life (Henderson *et al.*, 1979).

Although reinfection may occur by repeat exposure to the same viral isolate (Hall *et al.*, 1991) which suggests that antigenic variation is not strictly required to allow reinfection, evidence suggests that antigenic variation may play a role in the ability of RSV to escape the immune response and establish infection (reviewed by Sullender, 2000).

Secondly, infants need to be immunised in the first month of life at a time of immunological immaturity and interference of maternal antibodies (Wright *et al.*, 2000).

Finally, one of the major setbacks to RSV vaccine development was the unfortunate incident whereby a formalin-inactivated vaccine tested in the 1960's not only failed to protect, but also resulted in enhanced disease among the recipients upon reinfection (Kim *et al.*, 1969; Kapikian *et al.*, 1969). In contrast children receiving live RSV intramuscularly or intranasally did not experience vaccine-enhanced disease (Belshe *et al.*, 1982; Wright *et al.*, 1976). Following this incident, research was focussed extensively on determining the immunological determinants of vaccine induced disease as well as immunity needed for protection. These studies suggested that protective immunity is associated with a type 1 T-helper cell response with the induction of CD8+ cytotoxic T-cells (CTL), while a type 2 T-helper cell response is associated with disease (reviewed by Graham, 1996). Although CTL activity has been shown to be important for viral clearance, it has also been shown to be capable of causing immunopathology when passively transferred in large amounts into the lungs of mice (Zinkernagel & Hengartner, 1994). This emphasises the fine balance that exists between protection and disease-enhancement, and the importance of a thorough surveillance of the immune response during testing of experimental vaccines.

In preparation for RSV vaccine programs, background information about RSV epidemiology is needed. The WHO has initiated intensive RSV epidemiology programs throughout the world to assess RSV-associated disease burden (WHO Department of Vaccines and Biologicals (V&B), 2001). In conjunction with this the circulating strains and evolution of RSV need to be determined to identify relevant isolates for inclusion in vaccines and to aid in identification of relevant vaccination strategies. Since very little data are currently available about RSV in South Africa,

for this investigation, studies have been identified that will provide relevant information about RSV in our country that may contribute in the development and testing of relevant vaccines.

Firstly since no information is yet available about South African RSV strains, the epidemiological characteristics and evolution of circulating RSV genotypes need to be determined in order to design appropriate vaccines and vaccination strategies. The distribution of different strains across the country during an epidemic and the relationship between the infecting strains and disease pathogenesis caused by them also needs to be determined. Secondly, in order to be able to control for the cytotoxic T-cell (CTL) response evoked by experimental vaccines, CTL epitopes common to South African human leukocyte antigen (HLA) types need to be determined in order to produce reagents that can be used in CTL assays in infants. Having access to assays for evaluating quantitative aspects of epitope specific CTL responses will facilitate RSV vaccine development and improve our understanding of RSV induced disease pathogenesis. Finally, the amount of conservation in important immunogenic regions, especially in the dominant CTL epitopes, among the different strains identified in South Africa needs to be characterised to determine if immune selection will drive the need for regular updates in future vaccine programs.

The following literature review will give an overview of RSV virology, focussing on the molecular virology, -epidemiology and immunology, especially the importance of CD8+ cytotoxic T-cells in protection and disease. A short description of the current status of vaccine development will also be given.

1.2 THE VIRUS

1.2.1 Classification

Human RSV is an enveloped virus with a non-segmented negative sense single-stranded RNA genome (Collins *et al.*, 1996b). It belongs to the Pneumovirus genus in the subfamily *Pneumovirinae* in the family *Paramyxoviridae*. Other members of this genus include bovine, ovine and caprine RSV, and pneumovirus of mice, while the subfamily also includes the metapneumovirus genus. Other

members of the *Paramyxoviridae* are the *Morbillivirus* genus (Measles virus); *Respirovirus* (Sendai virus and human parainfluenza virus types 2 and 3) and *Rubellavirus* genus (Mumps virus and human parainfluenza type 2) (Collins *et al.*, 1996b, Collins *et al.*, 1999, Pringle & Easton, 1997). Two antigenically distinct RSV subgroups have been identified and designated subgroup A and B (Anderson *et al.*, 1985).

1.2.2 Viral structure and Replication

The RSV genome consists of 15,222 nucleotides for the prototype strain A2 and encodes 10 subgenomic mRNAs that are capped and polyadenylated. Eleven different proteins are encoded, each from a different transcript, except for the M2-1 and M2-2 proteins that are translated from two overlapping reading frames (ORF) on the M2 RNA (Collins *et al.*, 1996a) (Figure 1.1). The virion consists of a nucleocapsid within a lipid envelope and contains eight of these proteins (Collins *et al.*, 1999), while two are non-structural (NS1 and NS2) and only found in infected cells (Huang *et al.*, 1985). The location of M2-2 is not yet determined. The nucleocapsid consists of the nucleocapsid (N) protein that confers nuclease resistance, the phosphoprotein (P), the major polymerase subunit (L), and the transcription elongation factor (M2-1) (Figure 1.2). It is unclear if the M2-2 protein is a virion component, however it is found much less abundantly than the other proteins intracellularly, either because of instability or inefficient translation of its overlapped ORF (Collins *et al.*, 1999). The envelope is derived from the host cell plasma membrane and contains protruding glycoprotein spikes consisting of three transmembrane homo-oligomer viral envelope proteins, the attachment protein G, the fusion protein F and the small hydrophobic protein SH (Collins *et al.*, 1996b).

RSV replication is restricted to the surface epithelium of the respiratory tract where it binds to an unidentified cellular receptor. Fusion occurs at the cell surface and allows the nucleocapsid to enter the cytoplasm. Transcription and replication of viral RNA occurs in the cytoplasm. The mRNA is transcribed from the 3' end by a sequential stop-start mechanism and is directed by the gene start (GS) and the gene end (GE) transcription signals flanking the viral genes.

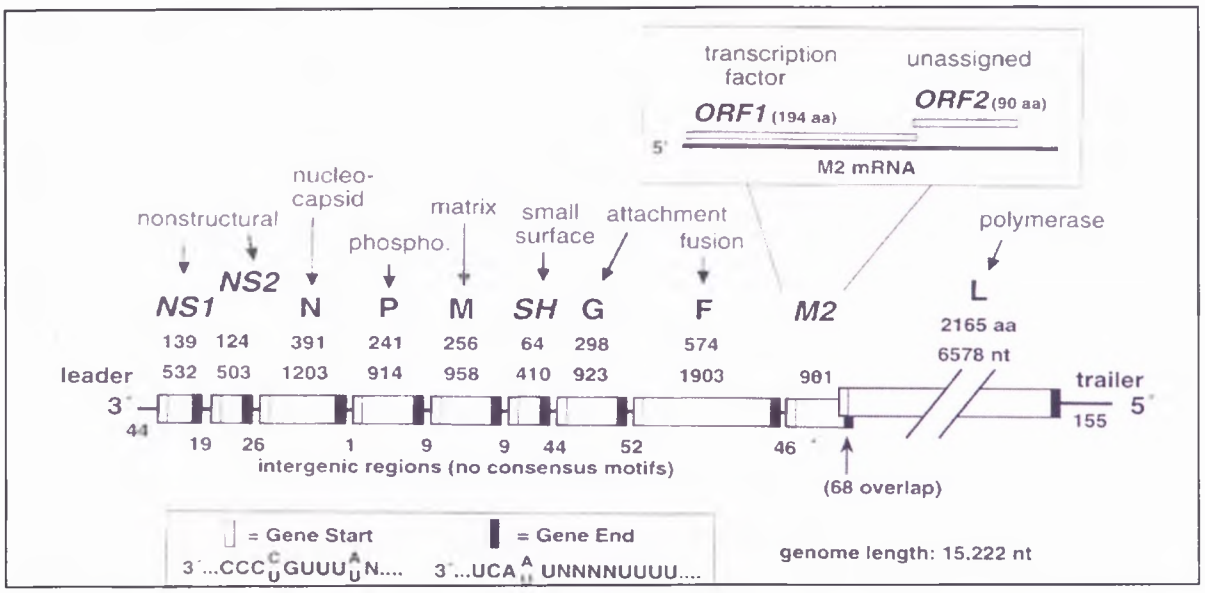


Figure 1.1: Map of the negative sense RNA genome of RSV strain A2. The encoded mRNA's are indicated by the rectangles. The shaded and filled regions at the ends indicate the gene start and gene end signals. The length of each mRNA is indicated in number of nucleotides immediately above the box, with the corresponding amino acid length of the encoded protein above that. The two open reading frames of M2 are indicated in the expanded box at the top right. The extragenic leader, intergenic and trailer regions are indicated by thin lines and the nucleotide lengths shown underneath. (not drawn to scale) Reproduced from Collins *et al.* (1999).

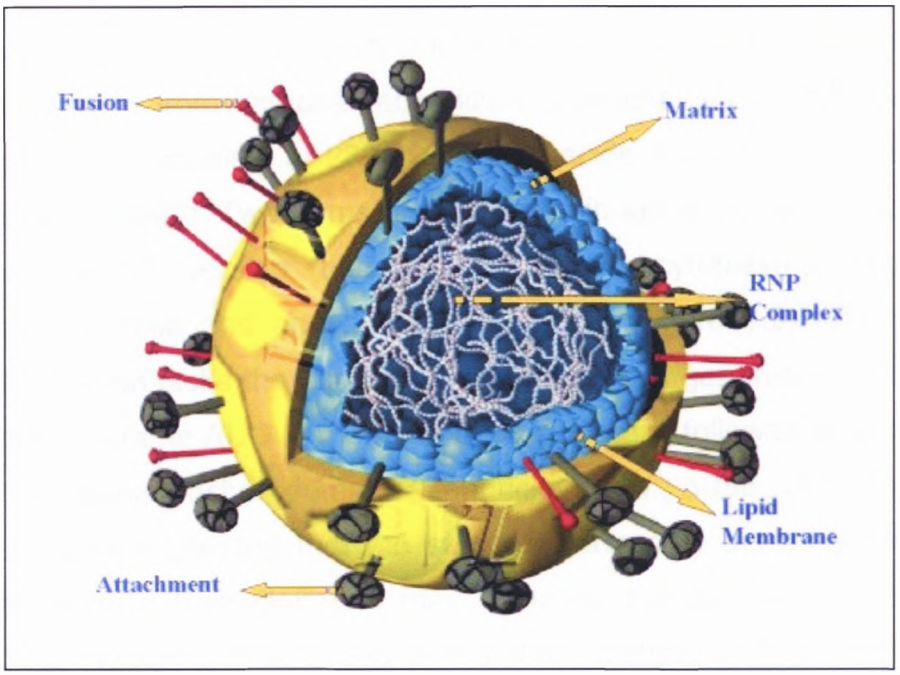


Figure 1.2: A three-dimensional representation of the RSV virion. The **envelope** consists of the lipid membrane with the attachment protein (G), the fusion protein (F) and the SH protein; the **matrix** consists of the Matrix protein (M); and the RNA-nucleoprotein complex (RNP) consisting of the **nucleocapsid** containing the Nucleoprotein (N); the Phosphoprotein (P), the Large polymerase subunit (L) and the M2-1 protein associated with the RNA genome. Reproduced from: (http://www.bio.warwick.ac.uk/easton/IMAGES/Diagrams/3dvirus.jpg).

During replication, a full-length positive sense copy, the antigenome, is synthesised. This acts as a template for the progeny genome in the nucleocapsid. The virions form by budding at the cell surface. (Reviewed by Collins *et al.*, 1999). The following section will describe the function of the individual proteins, emphasising their relevance to the immune response and to this study.

1.2.2.1 The G-protein

The G-protein is thought to be the attachment protein of RSV, but unlike other *paramyxoviridae* lacks haemaglutinin or neuraminidase activity (reviewed by Sullender, 2000). The cellular receptor that it binds to, has however, not yet been identified (Collins *et al.*, 1999). The G-protein is heavily glycosylated by post-translational addition of N- and O-linked sugars to the more than 70 potential glycosylation sites provided by a 30.6% serine and threonine content (Wertz *et al.*, 1985). This glycosylation pattern is unusual for paramyxoviruses but is characteristic of mucinous proteins secreted by respiratory epithelium and is thought to have a role in viral infectivity and antigenicity, possibly by shielding the viral-encoded protein from immune recognition (Lambert, 1988; Palomo *et al.*, 1991). The G-protein consists of 289-299 amino acids depending on the strain (reviewed in Cane, 2001). Two forms of the G-protein are expressed. The first is a membrane bound version that consists of a cytoplasmic region, a transmembrane domain and an extracellular region (Figure 1.3). The second secreted soluble form lacks the former two regions and is generated by initiation at an alternative in-frame AUG in the hydrophobic domain, followed by proteolytic removal of the signal/anchor domain (Hendricks *et al.*, 1988, Roberts *et al.*, 1994). It has been suggested that the secreted form may act as a decoy for antibody-mediated neutralisation and may influence the immune response by potentiating the induction of IL-5 and eosinophilia (Johnson *et al.*, 1998). The ectodomain contains two hyper-variable regions with substantial sequence variation between the subgroups, separated by a highly conserved central region, containing four cysteine residues that are conserved in all isolates. The central region, which is slightly hydrophobic, has been proposed as a putative receptor binding site (Johnson *et al.*, 1987b).

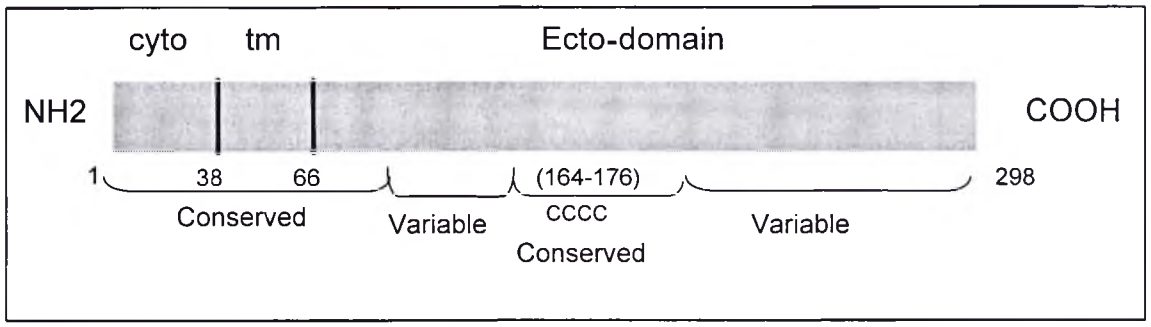


Figure 1.3: Variable and constant regions of the G-protein of the A2 isolate of RSV. The cytoplasmic tail (cyto), transmembrane domain (TM), and ectodomain are indicated as well as the variable and conserved regions. The position of the 13 amino acid conserved region (164-176) and four conserved cysteine residues are shown. (Adapted from Sullender 2000)

The G-protein is a target for the neutralising antibody response, and immunisation with G partially protects animals from RSV challenge, however protection is highly group-specific (Johnson *et al.*, 1987a; Johnson *et al.*, 1987b). This is the RSV protein that is the most variable both between and within the two subgroups, with 47% amino acid differences between subgroup A and B (Johnson *et al.*, 1987b). Most strain-specific variable epitopes are found in the second hyper-variable domain while the conserved- and group- specific epitopes have been mapped to the central conserved domain (Melero *et al.*, 1997). It has been shown that the G-protein is not essential for cell entry although the virus with G deleted is highly attenuated, suggesting a second attachment activity (Teng *et al.*, 2001). Furthermore, recent studies have shown that the central conserved cysteine noose region is not required for efficient viral infection *in vitro* or *in vivo* or effective neutralisation in mice, suggesting that although the G-protein is necessary for efficient virus replication *in vivo*, this activity does not require this conserved domain (Teng & Collins, 2002). The extent of variability found in the G-protein both between and within the two subgroups, as well as its role in neutralising immunity and pathogenicity has made the G-protein a popular target for molecular epidemiological studies (reviewed in Cane 2001 and Sullender, 2000).

1.2.2.2 The F-protein

The fusion protein (F) is responsible for syncytia formation by promoting fusion of viral and cell membranes. It allows entry of the viral replication complex after

attachment and is expressed on the surface of infected cells to permit viral spread without exposure to the extra-cellular environment (Walsh *et al.*, 1986). The F-protein is synthesised as a 70 kDa precursor (F_0) that is cleaved by a furin-type intracellular protease into two disulphide linked subunits (F_2 and F_1). It is thought that the subsequent exposure of a hydrophobic domain on F_1 may be important in membrane fusion, cell tropism and pathogenicity (Walsh *et al.*, 1986). The F-protein is considered to be the major neutralisation and protective antigen of RSV (Collins *et al.*, 1999). It is highly conserved with only 11% amino acid difference between subgroup A and B (Johnson & Collins, 1988). Several neutralisation epitopes are found on the F-protein of which two are related with fusion activity and highly conserved between the two subgroups (Beeler & van Wyke Coelingh, 1989). The F-protein is an important target for the CTL response both in mice and humans (Cannon & Bangham, 1989; Nicholas *et al.*, 1990; reviewed by Walsh & Graham, 1999). CTL epitopes on the F-protein have been mapped both in mice (Jiang *et al.*, 2002) and humans (Brandenburg *et al.*, 2000a).

1.2.2.3 The SH protein

The small hydrophobic protein is a transmembrane surface protein of unknown function. Recombinant RSV with SH deleted is fully viable in culture (Bukreyev *et al.*, 1997). It does not contain any neutralisation epitopes (Walsh & Graham, 1999) but has been shown to induce CTLs in humans (Cherrie *et al.*, 1992a).

1.2.2.4 The Matrix protein

The matrix protein (M), is an internal protein and has been shown to be important for virion assembly (Teng & Collins, 1998). Little is known about the functions of the matrix (M) protein of RSV, however, in other negative-strand RNA viruses, the M protein inhibits the viral polymerase prior to packaging and facilitates virion assembly. Recent investigations into the role of the M protein has shown that in RSV infected cells the M protein co-purified with viral ribonucleoproteins (RNPs), and that the transcriptase activity of purified RNPs was enhanced by treatment with antibodies to the M protein in a dose-dependent manner, suggesting that the

M protein is associated with RSV nucleocapsids and that it can inhibit virus transcription (Ghildyal *et al.*, 2002).

1.2.2.5 The M2 protein

As mentioned above, two proteins are transcribed from the M2 RNA. The major product is the M2-1 protein which is encoded by the 5' proximal ORF and promotes transcription elongation (Collins *et al.*, 1996a). The M2-1 protein has been shown to be a transcriptional processivity and antitermination factor. It has a Cys3His1 zinc binding motif which is essential for function (Hardy & Wertz, 2000), requires phosphorylation for efficient function, and has been shown to interact with the RSV nucleocapsid (N) protein (Cartee & Wertz, 2001). This is the major target for CTL in the mouse model (Kulkarni *et al.*, 1995; Openshaw *et al.*, 1990)(Nicholas *et al.*, 1990) and has also been shown to induce CTL in humans (Cherrie *et al.*, 1992a). The dominant murine H2-K^d CTL epitope on this protein has been defined at amino acid positions 82-90 (Kulkarni *et al.*, 1995). Different groups have shown that immunisation of mice with vaccinia virus expressing the M2 protein provides transient protection to RSV challenge (Connors *et al.*, 1991; Kulkarni *et al.*,1993). The M2-2 protein is the second product encoded by the M2 mRNA and thought to be a negative regulatory factor for replication and transcription (Bermingham & Collins, 1999).

1.2.2.6 Proteins of the nucleocapsid complex

The nucleoprotein (N) the phosphorylated (P) protein and the Large (L) polymerase protein make up the replication machinery and are bound tightly to the viral RNA in the nucleocapsid complex (Walsh & Graham, 1999). The **N-protein** is the major nucleocapsid protein. Baculovirus expression of the N-protein in insect cells results in nucleocapsid-like structures (Meric *et al.*, 1994), suggesting that the N-protein has an important function in maintaining the nucleocapsid structure (Schlender *et al.*, 2000). It has been shown to bind tightly to the genomic and antigenomic RNA to form the ribonucleoprotein template for RNA synthesis (Collins *et al.*, 1999). The N-protein is the most conserved of all RSV proteins between the two subgroups with only 4% amino acid differences (Johnson *et al.*, 1998). Comparisons of the amino acid sequences of the human

RSV nucleoprotein with the nucleocapsid proteins of other pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses revealed regions of conserved structure that most likely have the common functions of RNA binding and protein/protein interactions in the virus nucleocapsid (Barr *et al.*, 1991). The RSV N-protein is a target for the CTL response in mice, and has been shown to be the major target for memory CTL response in humans (Bangham *et al.*, 1986; Cherrie *et al.*, 1992a).

The **Phosphoprotein (P)** is thought to associate with free N and L proteins to keep them soluble prior to their assembly into the nucleocapsid and to act as a polymerase co-factor. It has not been implicated in the immune response to RSV. The **L-protein** is the major polymerase subunit. The N, P and L proteins are necessary and sufficient for RNA replication and a poorly processive form of transcription (Grosfeld *et al.*, 1995; Yu *et al.*, 1995; Collins *et al.*, 1999).

1.2.2.7 The non structural proteins NS1 and NS2

Two non-structural proteins NS1 and NS2 are found in RSV infected cells, of which the function is not yet clearly defined. NS1 has been shown to mediate inhibition of transcription and RNA replication, suggesting a role in negative regulation (Atreya *et al.*, 1998). Studies in bovine RSV (bRSV) suggested that NS1 and NS2 cooperatively antagonise the alpha/beta interferon-induced antiviral response (Schlender *et al.*, 2000). These proteins have been shown to block activation of interferon regulatory factor 3, suggesting a role in dysregulation of the immune response to help the virus evade the host defence mechanism against the viral infection and contribute to the pathogenesis of bRSV (Schlender *et al.*, 2000). A role in host range determination has also been described for NS1 and NS2 (Bossert & Conzelmann, 2002). Neither of the non-structural proteins have been shown to contain neutralising antibody or CTL epitopes involved in the RSV-specific immune response (Collins *et al.*, 1999).

1.3 THE EPIDEMIOLOGY

RSV is a seasonal virus, and causes annual epidemics in the cold season in temperate climates, while tropical climates experience outbreaks during the rainy

season when temperatures fall (Weber *et al.*, 1998). Studies in Johannesburg, South Africa, which has a temperate climate have suggested that RSV epidemics occur mostly in autumn and winter months when the rainfall is at its lowest (Joosting *et al.*, 1979; Madhi *et al.*, 2001). RSV seasonality has not yet been determined for the winter rainfall, tropical and subtropical regions of South Africa, but is expected to correspond with other tropical and temperate regions.

RSV is a major cause of morbidity in infants. Infection is uncommon in the first month of life when infants are protected by maternal antibodies, although infection can occur when maternal antibody titres are low (Glezen *et al.*, 1981). After the neonatal period, virtually all (70%) infants are infected in the first year of life with the remainder being infected before the age of 2 years (Glezen *et al.*, 1986). Although acute lower respiratory tract infections (ALRI) are common during the first year (22.4%-40%) most children will experience only one episode of ALRI and reinfection illnesses are generally mild (Brandt *et al.*, 1973; Glezen *et al.*, 1986). Hospitalisation rates varies from 1%-2% in under one year olds in the first world (Brandt *et al.*, 1973; Glezen *et al.*, 1986; Martin *et al.*, 1978). The groups most at risk are infants of 1-3 months old, those with a history of premature birth, bronchopulmonary dysplasia, and congenital heart disease (Simoes, 1999), while adults most affected by severe RSV include immunocompromised patients (Englund *et al.*, 1988) and the elderly (Falsey *et al.*, 1992). Although traditionally considered to be a pediatric pathogen, RSV is a contributor to community-acquired pneumonia in adults of all ages and a major cause of winter morbidity (La Montagne, 1997). RSV is also a major nosocomial hazard in pediatric wards, as well as in transplant units where it is associated with high mortality rates (Hall, 2000).

Less information is available about RSV in developing countries, however RSV is suggested to be the major viral cause of viral ALRI in children followed by influenza, parainfluenza and adenovirus (Weber *et al.*, 1998). Although the risk factors in developing countries are not yet clearly defined, crowding, indoor smoke pollution and malnutrition may contribute to the risk of developing more severe disease (Simoes, 1999). RSV-associated ALRI is slightly more common in boys than in girls in both developing and first world countries (Glezen *et al.*, 1986;

Weber *et al.*, 1998). Limited studies about the effect of HIV/AIDS have been published, but a study conducted in South Africa suggests that the estimated burden of RSV associated severe LRTI is higher in HIV seropositive children and that HIV seropositive children with severe LRTI have a higher mortality rate (7%) relative to HIV seronegative children (0%). Furthermore, RSV isolation was not limited to season in HIV seropositive children (Madhi *et al.*, 2000). Further studies suggested that HIV infected children remained predisposed to developing RSV-associated severe LRTI beyond the first six months of life and are more likely to develop pneumonia and have greater evidence of bacterial co-infection and greater mortality than HIV uninfected children (Madhi *et al.*, 2001).

1.3.1 Molecular epidemiology

One possible explanation for reinfections caused by RSV is the substantial variability that exists in RSV strains. Although RSV does not have the capacity to undergo reassortment of genome segments, it has a highly mutable genome due to an RNA polymerase that lacks proofreading functions (Collins *et al.* 1996b). This results in the establishment of quasispecies that exist in equilibrium around theoretical consensus sequences that are shaped by selective pressure of the environment, providing greater adaptability (reviewed by Sullender, 2000). Two distinct RSV subgroups, A and B, have been identified by serological methods, and within these, various antigenic variants or subtypes of RSV (Anderson *et al.*, 1985). These subtypes have also been shown to be distinct at the genetic level. Sequencing studies of the nucleoprotein (N), the phosphoprotein (P), the small hydrophobic (SH) protein and the attachment protein (G) were used to confirm the existence of two distinct groups, but also numerous variants or lineages within each group (Johnson *et al.*, 1987a; Cane & Pringle, 1992; Cane & Pringle, 1991; Cane & Pringle, 1995c; Cristina *et al.*, 1990; Storch *et al.*, 1991; Sullender *et al.*, 1991; Cane *et al.*, 1992; Cane *et al.*, 1994; Garcia *et al.*, 1994; Sanz *et al.*, 1994).

1.3.1.1 Methods used for studying the genetic variability of RSV

Different strategies for examining the variability of RSV for investigation of the molecular epidemiology of RSV have been pursued. These include reactions with monoclonal antibodies against G (Anderson *et al.,* 1991; Hall *et al.,* 1990), an

RNase A mismatch cleavage method (Storch *et al.*,1991; Christina *et al.*,1991), and analysis of individual genes of the virus by RT-PCR in combination with either restriction mapping or nucleotide sequencing (reviewed by Cane & Pringle; 1995b; Cane, 2001; Sullender, 2000). Since RSV grows poorly in tissue culture and the virus is difficult to purify, RT-PCR analysis of clinical specimens has become increasingly popular because it allows rapid analysis of large numbers of samples (reviewed by Cane & Pringle, 1995b).

As mentioned before, the nucleoprotein (N) has been shown to be the most conserved between the subtypes showing 96% amino acid identity between sub-groups A and B, while the G-protein is the most variable with 53% amino acid identity. In addition the G-protein has been found to be highly variable between isolates of the same subgroup with subgroup A isolates being up to 20% variable in amino acid sequence and subgroup B isolates up to 9% variable. (Reviewed by Cane and Pringle, 1995). Variability of the G-protein gene has been shown to be primarily located in the two hyper-variable regions of the extracellular domain (Johnson *et al.,* 1987; Cane and Pringle, 1991). Group-specific oligonucleotides have been identified on the G-protein that can be used in PCR for differentiation between subgroup A and B isolates (Sullender *et al.,* 1991; Sullender *et al.,* 1993). Peret and colleagues have shown that sequence analysis of the second hyper-variable region, consisting of 270 nucleotides, provides a reliable proxy for the entire G gene variability (Peret *et al.,* 1998). This provides a means for rapid sequencing and phylogenetic comparison of a large number of isolates. The advantage of sequencing is that genetic variation at all positions in the gene fragment under investigation can be identified while only a small number of sites are used with restriction mapping.

A study conducted over 5 years in a single community in the USA identified 5 subgroup A clades (genotypes) and 22 subtypes (based on more than 96% nucleotide similarity), and 4 distinct genotypes and 6 subtypes among group B isolates (Peret *et al.,* 1998). Further genotypes were identified in a study of 5 different communities across North America suggesting that the genotype system may resolve further as more sequences become available. Currently this system that includes most isolates from the developed world up to the year 2000,

consists of seven subgroup A genotypes (GA1-7) and four subgroup B genotypes (GB1-4) (Peret *et al.*, 2000). Cane has used restriction mapping patterns of the N, and sequencing of part of the SH and G gene to divide RSV isolates into a distinct number of genotypes (reviewed by Cane and Pringle 1995). There is no standardised method for genotyping at this stage, although phylogenetic analysis of the second variable region of the G-protein has become most popular, making the method published by Peret *et al.* useful for comparison of strains identified in different studies (Cane, 2001; Peret *et al.*, 2000; Peret *et al.*, 1998; Sullender, 2000).

1.3.1.2 Genetic variation and molecular epidemiological analysis

When studying the prevalence of different genotypes over 11 years in epidemics in Birmingham in the UK, it was shown that each epidemic was characterised by a number of different genotypes and could therefore be considered an aggregate of concurrent separate epidemics. The prevalence of the genotypes varied each year and the dominant genotype in one year declined in the next with some genotypes reappearing or disappearing in subsequent years. Minor changes were identified in the reappearing genotypes, although it was not clear if this is a prerequisite for its reestablishment (Cane *et al.*, 1994). Similar observations were made in studies over subsequent years in the USA (Peret *et al.*, 2000) and in Korea (Choi & Lee, 2000). It has been shown that both antigenic groups can co-circulate in one epidemic, or that one may predominate. However different genotypes seem to be predominant from year to year. This antigenic variability may contribute to the ability of RSV to establish annual epidemics (Cane and Pringle 1995; Peret *et al.*, 1998; Cane *et al.*, 1999).

Anderson has shown in a study in the USA that RSV outbreaks can vary in the pattern of group and subgroup isolations between communities during the same year and between years for the same community, indicating that RSV outbreaks are community- and possibly regional-based, but not national (Anderson *et al.* 1990). Genetic analysis of isolates in different communities in North America confirmed this finding. Peret *et al.* found co-circulation of 5-7 genotypes with one or two dominant strains in each community, but that the patterns of genotype

occurrence were distinct between communities and less variation existed between members of the same strain within rather than between communities (Peret *et al.*, 2000).

Although high levels of variability may exist even within a single local epidemic, studies have shown that very similar viruses circulate world-wide (Cane *et al.*, 1994; Sanz *et al.*, 1994). Studies in the Gambia have however found numerous isolates that did not correlate with any of the previously identified genotypes (Cane *et al.*, 1999). Furthermore, a study in Mozambique (Roca *et al.*, 2001) and the work done in this study in South Africa (Venter *et al.*, 2001) identified further strains that did not correlate with any of the genotypes identified in the rest of the world. It has been suggested that RSV epidemics in the Gambia are effectively a mixture of viruses circulating the developed world, entering the country via the coastal and urban districts and viruses derived from more remote interior regions of Africa (Cane *et al.*, 1999). The same may hold true for Mozambique and South Africa.

In contrast to the above-mentioned studies, a study in Cuba revealed isolates that remained much more genetically stable over subsequent years. Isolates from Cuba from 1994 and 1995 had only 5 nucleotide differences from the Long strain, which was isolated in 1956 in the USA, and comparisons to viruses from later epidemics revealed the same viruses with identical G-protein sequences to that of the 1994 and 1995 seasons. The authors suggested that differences in seasonality or restrictions in travel might have favoured viral stability in Cuba. RSV infections are distributed throughout the year in Cuba; there seem to be a lack of displacement of the dominant genotype and in accumulation of changes in the G-protein (Valdes *et al.*, 1998).

1.3.1.3 Evolution of the RSV G-protein

Several investigations have found signs of positive selection in the G-protein, suggesting possible immune evasion by RSV (Cane & Pringle, 1995c; Garcia *et al.*, 1994; Woelk & Holmes, 2001). The progressive accumulation of amino acid changes in the attachment protein of subgroup A RSV isolated over 38 years

occurred at a rate of 0.25% per year. The reactivity of isolates with monoclonal antibodies showed that the antigenic profiles correlated with the genetic variation, suggesting antigenic drift due to immune selection (Cane & Pringle, 1995c). Evidence that homologous strains of RSV provide protection against reinfection with the same strains but that mice are only partially protected against heterologous strains, suggests that genetic variation helps the virus to evade the immune response and to reestablish infections in the presence of immunity (unpublished data, P. Young, International RSV congress 2001, Segovia, Spain). Strain- and genotype-specific antibodies and epitopes have been identified which may suggest that protection against one genotype may not necessary provide cross protection against another (Jones *et al.*, 2002). Specific protective epitopes (protectotopes) have been identified on the G-protein that appear to be conserved in subgroup A in humans (Power *et al.*, 2001); however the G-protein has been shown to only provide subgroup-specific protection (Walsh *et al.*, 1987). Monoclonal antibodies have been used to show the existence of epitopes that are 1.) Conserved between subgroup A and B, 2.) Strain-specific variable epitopes, and 3.) Group-specific epitopes specific to either subgroup A or B. Most of the conserved and group-specific epitopes have been mapped to the conserved central region, while the strain-specific epitopes are located at the C-terminal variable region (Martinez *et al.*, 1997; Melero *et al.*, 1997). Recognition of many of the variable epitopes is dependent on cell type specific glycosylation. Human polymorphism for the expression of certain carbohydrate epitopes may thus add to the antigenic variation in that the same hRSV strain infecting different individuals may have different carbohydrate side-chains which will affect the expression of certain epitopes (Garcia-Beato *et al.*, 1996). Evidence also suggests that glycosylation may contribute to RSV strain variation (Palomo *et al.*, 2000).

Analysis of the type and position of mutations occurring on the G-protein gene revealed that synonymous nucleotide changes are distributed uniformly along the G-protein while nonsynonymous mutations accumulate preferentially in the two variable regions of the G-protein providing further confirmation of positive immune driven selection (Cane *et al.*, 1991; Garcia *et al.*, 1994). The C-terminal antigenic relevant region of the G-protein can accommodate multiple sequence

changes. Mutations that result in single amino acid substitutions, frame-shift mutations and premature stop codons have been identified suggesting that RSV may overcome antibody pressure raised during natural infection by generating variants with changes in the key epitopes of the G-protein (Melero *et al.*, 1997).

1.3.1.4 Relationship between strains and virulence

Different groups have attempted to correlate the severity of disease with the infecting strains or genotypes with contrasting findings. Walsh *et al.* found in a study over four years that subgroup A infections were more severe, and infants infected with subgroup A required ventilatory support much more often than subgroup B infections which rarely required ventilatory support (Walsh *et al.*, 1997). Fletcher *et al.* identified a specific subtype A genotype (SHL2) that resulted in more severe disease (Fletcher *et al.*, 1997). In contrast, a study in Copenhagen suggested that subgroup B infections produced more severe disease in this location than did subtype A infections, and that this difference was age-related. This group identified a specific type B genotype that produced more severe disease than the type A genotype and correlated the Interleukin-6 (IL-6): tumour necrosis factor alpha (TNF-alpha) ratio in samples of nasopharyngeal secretions to the severity of the disease, with a high ratio being related to a low severity (Hornsleth *et al.*, 1998). Other studies found no correlation between disease severity and genotype (Brandenburg *et al.*, 2000b; Smyth *et al.*, 2002). Smyth *et al.* found a strong, dose-related association between IL-8 mRNA produced in the airways and disease severity, and no association with virus genotype and suggested that clinical manifestations of RSV bronchiolitis are determined by local immunological responses to infection, rather than by characteristics of the infecting virus (Smyth *et al.*, 2002).

It has been suggested that a reason for these conflicting data could be the fact that only hospitalised infants that are already severely ill are considered and that the severity of disease may be affected by the levels of maternal strain-specific antibodies, which will change from year to year (Cane, 2001). Recently studies in inbred mice have shown that susceptibility to RSV infection was attributable to genetic differences between the mouse strains with specific strains being more

prone to severe disease. This finding suggests that genetic factors may also influence RSV susceptibility and disease severity (Stark *et al.*, 2002).

Although the antigenic variation seen in RSV may have an important role in reestablishing infections, this appears to be driven by the immune response and several studies suggest that an individual's immune response may have larger influences on susceptibility, and severity of disease. The following section will give a description of RSV immunity, concentrating on the immune response needed for protection.

1.3.2 Immune response

1.3.2.1 RSV immunity

RSV replication is restricted to the surface epithelium of the respiratory tract by the immune response of the host, and disease may range from rhinitis to severe bronchiolitis and pneumonia (reviewed in Collins *et al.*, 1999; Dudas & Karron, 1998).

Innate immunity: The portal of infection is the upper airway epithelium. Here, as a first line of defence, the host's immune response releases proinflammatory cytokines and initiates the nonadaptive immune response. *In vitro* studies have shown that following RSV infection, pulmonary epithelial cell lines release IL-6, IL-8, IL-11, granulocyte-macrophage colony-stimulating factor (GM-CSF), and soluble TNF receptor and TNFalpha (Arnold *et al.*, 1995; Elias *et al.*, 1994). This has been confirmed *in vivo* although, TNFalpha is detectable only at low levels in nasal washes of infected children (Matsuda *et al.*, 1995; Nakayama *et al.*, 1993). The major role of the epithelium-derived proinflammatory cytokines is thought to be on downstream immunological events, since the expression of the cell surface adhesion molecules, major histocompatibility complex (MHC) antigens I and II, and the differentiation and activation of effectors of the adaptive immune response are affected by them (reviewed in Walsh & Graham, 1999). Cytokines (IL-1beta), IL-6, IL-11, TNFalpha) and chemokines (IL-8, RANTES, macrophage inflammatory protein-1 alpha (Mip-1 alpha)) that are produced by RSV infected epithelial cells and airway macrophages modulate airway inflammation and attract and activate different kinds of inflammatory cells such as neutrophils and

eosinophils (reviewed by Tsutsumi *et al.*, 2001). Eosinophilia is associated with severe RSV disease (Ingram *et al.*, 1995). Intracellular adhesion molecule-1 (ICAM-1) is also expressed in RSV infected epithelial cells and is thought to be regulated by IL-1 alpha (Patel *et al.*, 1995). This may result in increased adhesion of neutrophils and possibly eosinophils to RSV infected cells (Stark *et al.*, 1996). Upregulation of MHC class I molecules has been documented in RSV infected epithelium which may contribute to the elimination of RSV infected cells by RSV-specific CD8+ CTL (Garofalo *et al.*, 1996). The upregulation of certain apoptotic proteins (ICE; Fas; IEX-IL; and iNOS), by the transcription factors IRF-1 and NF-kappa beta in RSV infected cells has also been documented (Domachowske *et al.*, 2000; O'Donnell D *et al.*, 1999; Takeuchi *et al.*, 1998; Tsutsumi *et al.*, 1999). Figure 1.4 illustrates the activation of genes and proteins expressed during early phase RSV infection.

It is thought that early activation of Natural Killer (NK) cells is important in a protective response and that they are an important source of IFN-gamma (Hussell & Openshaw, 1998). Macrophages have a role as antigen presenting cells and are a source of cytokines, which may influence the outcome of RSV-induced disease (Walsh & Graham, 1999). RSV infected human alveolar macrophages produce IL-6, TNF-alpha, IL-8 and IL-10 the latter of which has an important regulatory role in cytokine production as well as MHC-class II expression and T-helper 1 (Th1) CD4 lymphocyte differentiation (Panuska *et al.*, 1995; Panuska *et al.*, 1990). Macrophages found in the bronchoalveolar lavage fluid from children also express IL-1beta and TNF-alpha, while those from RSV infected adult transplant patients expressed viral proteins, class II MHC molecules and IL-1beta (Panuska *et al.*, 1992; reviewed in Walsh & Graham, 1999).

Adaptive immunity: RSV immunity is largely mediated by virus-specific immune responses but protection is incomplete and infection occurs throughout life (reviewed by Dudas & Karron, 1998, Sullender, 2000). Viraemia is not produced by RSV infection and serum-neutralising antibodies do therefore not prevent infection. However, high titres of serum-neutralising antibodies do protect against lower respiratory tract infection as shown in the cotton rat model (Prince *et al.*, 1985).

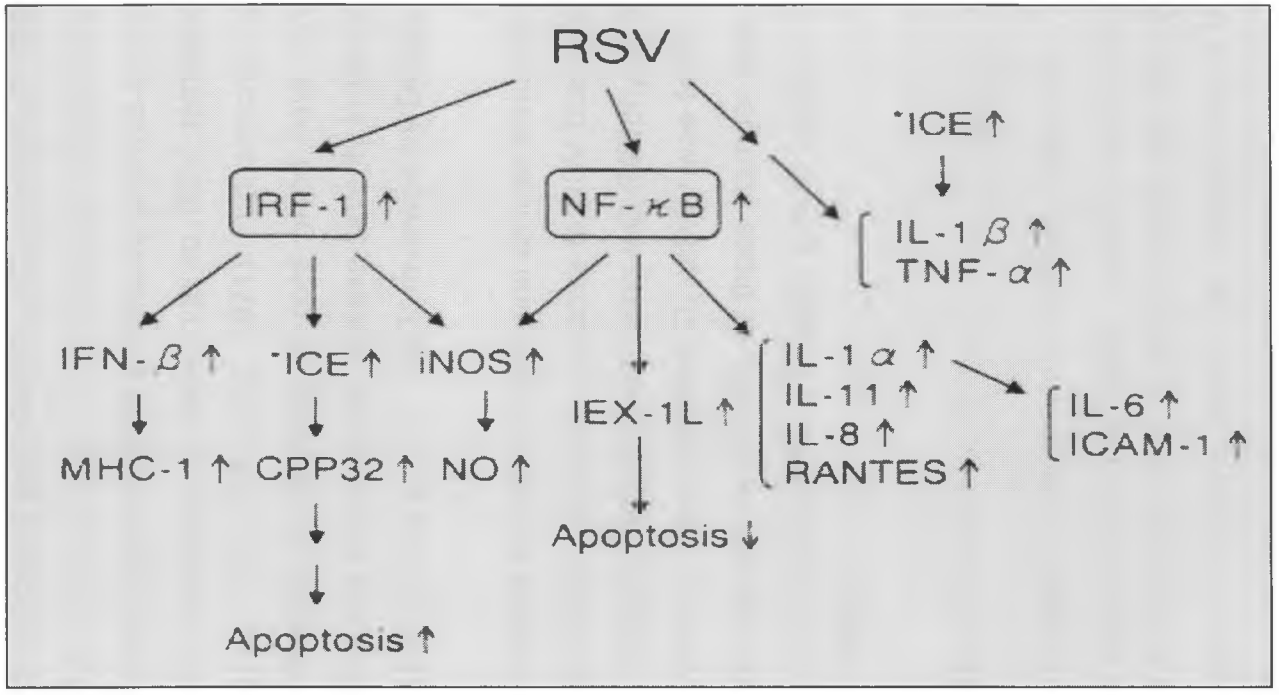


Figure 1.4: Genes activated in early phase RSV infection of respiratory epithelial cells. Adapted from Tsutsumi *et al.* (2001).

Serum RSV specific IgG antibodies (acquired after infection or maternally-transferred), provide protection against lower respiratory tract infection (Crowe, 2001), while secretory antibodies (IgA) are thought to protect against upper respiratory tract infection (Wright *et al.*, 2000). MHC class I CTL are thought to mediate the resolution of acute infection (Cannon *et al.*, 1988), and animals or patients that lack CTLs exhibited prolonged viral shedding (Cannon *et al.*, 1987; King *et al.*, 1993). Although studies in mice suggest antibodies are not necessary for viral clearance, the CTL response appears to be short-lived and it is thought that antibodies are the principle mediators of resistance to RSV reinfection (Cannon *et al.*, 1987; Connors *et al.*, 1991; Mills *et al.*, 1971). Furthermore, high levels of maternally-derived antibodies in maternal or cord blood have been shown to correlate with the incidence of infection and disease severity in the first 6 months of life (Glezen *et al.*, 1981; Holberg *et al.*, 1991; reviewed by Dudas & Karron, 1998).

Age appears to have an important effect on both humoral and cellular immunity to RSV. Infants of 6-8 months elicit a poorer immune response to RSV than older infants and often fail to develop secretory IgG or neutralising IgG (Murphy *et al.*, 1986a; Murphy *et al.*, 1986b). Similarly, RSV-specific CTL activity was found in 65% of infants aged 6-24 months but only in 35-38% of those younger than 5 months, which may in part account for more severe disease encountered in the younger age group (Chiba *et al.*, 1989).

1.3.2.2 Immune determinants of disease

Although a major part of RSV disease is due to virus-mediated destruction of epithelium and therefore also the extent of viral replication, certain components of RSV disease following natural infection may also be due to the host's immune response (Collins *et al.*, 1999; Walsh & Graham, 1999). The potential role of the immune response in RSV disease was first revealed when a formalin-inactivated alum precipitated RSV vaccine (FI-RSV) administrated to children not only failed to protect against infection, but caused enhanced disease symptoms, with 80% of the vaccinees being admitted to hospital (Kim *et al.*, 1969). Children that received this vaccine were wheezing, which is a feature of RSV-induced bronchiolitis and they had elevated numbers of eosinophils in their blood (Chin *et al.*, 1969; Kim *et*

al., 1969). Much of the research in subsequent years have focussed on elucidating the cause of this vaccine-enhanced disease especially since vaccination with a live RSV vaccine, either intramuscularly or intranasally did not result in enhanced disease (Belshe *et al.*, 1982; Wright *et al.*, 1976).

It is thought that the same immunological determinants that induce asthma in a small number of children following primary RSV infection may have induced the vaccine-enhanced disease. Both of these syndromes seem to be mediated by the activation of IL-4 and IL-5, suggesting that the activation of type 2 T-helper cells may be associated with the vaccine-enhanced disease (Graham, 1996) (summarised in Figure 1.5). Evidence in mice suggests that, a type 2 T-helper cell response, with induction of IL-4 as the dominant cytokine, is associated with disease pathogenesis while activation of a type 1 T-helper cell response with IFN-gamma as the dominant cytokine and the activation of CD8+ T-cells are associated with protection against RSV (Graham *et al.*, 1993). This model suggests that certain genetic factors like atopy, prematurity, or immunisation with the formalin-inactivated vaccine (FI-RSV) as well as subunit vaccines (F- and G-protein), result in the activation of a type 2 T-helper cell response following RSV infection with induction of IL-4 and IL-5 as the major cytokines. IL-4 promotes the production of IgE, which in turn triggers mast cell degranulation while IL-5 induces recruitment of eosinophilia. Both of these are associated with wheezing in RSV infected children. The Th2 responses are furthermore auto-stimulatory and inhibit the Th1 responses, thereby becoming the dominant response on all subsequent exposure to RSV antigen (reviewed by Graham, 1996). Cotton rats developed similar lung pathology following immunisation with either the G-protein or F-protein (Connors *et al.*, 1992; Murphy *et al.*, 1990). Subsequent studies have also suggested roles for IL-10 and IL-13 in enhanced disease pathogenesis (Johnson & Graham, 1999; Waris *et al.*, 1996). On the other side of this model, a type 1 T-helper cell response is elicited following natural RSV infection or immunisation with live RSV and is associated with the production of IFN-gamma as the major cytokine, a shift towards IgG2a antibody isotype and earlier activation of CD8+ cytotoxic T-cell responses. IFN-gamma activation results in upregulation of antigen presentation, increased cytolytic activity and antiviral cytokines and finally viral clearance (Graham, 1996).

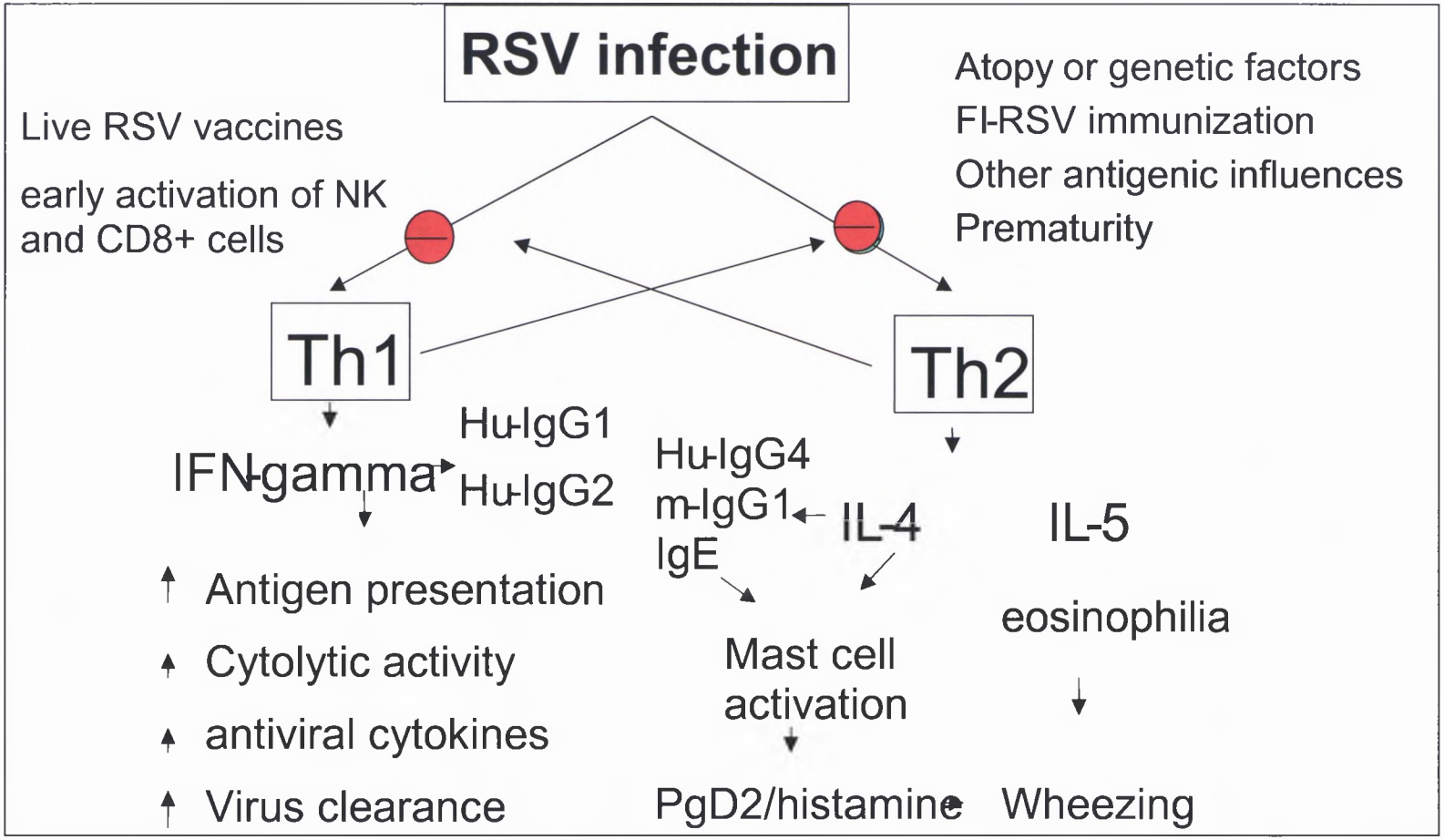


Figure 1.5: Immune determinants in RSV disease.
The Features of the RSV induced illness suggest that, a type 2 T-helper cell pattern of cytokine production may be an important component of the pathogenesis while activation of CD8+ CTL and a type 1 T-helper cell pattern of cytokines may be associated with protection (Adapted from Graham, 1996). Factors that could possibly promote the Th2 response are listed at the right, while factors that may promote a Th1 response are listed in the left corner of the figure. The Th2 response is auto stimulatory and inhibits the Th1 response and *vice versa* as indicated by the red circles on each leg. Cytokine inductions associated with either Th1 or Th2 type responses, as well as the outcome of each type of T cell response are listed. (See text for details)

Studies in mice have also shown that increased levels of NK cell cytotoxicity and IFN-gamma production correlated with increased levels of CD8+ T-cells and that mice with few NK cells subsequently had low levels of CD8+ T-cells and developed lung eosinophilia (Hussell & Openshaw, 1998). Evidence suggests that older persons (> 65 years) produce significantly less IFN-gamma in response to RSV than younger subjects (< 39 years) suggesting that aging may be associated with a defect in the cellular response to RSV and may be related to the increased morbidity observed in elderly persons (Looney *et al.*, 2002).

Although the type 1 T-helper cell response and the induction of CD8+ CTL appear to be essential for protection and correlate with recovery (Anderson *et al.*, 1990; Graham *et al.*, 1991b) some reports of immune pathology induced by CD8+ CTL have also been made. In mice passive transfer of large numbers of CD8+ CTL to the lungs is associated with enhanced pathology (Askonas & Taylor, 1987; Cannon *et al.*, 1988a). Others have found CTL-mediated clearance in mice without enhanced disease (Munoz *et al.*, 1991). Schwarze *et al.* have shown that although IL-5 and eosinophils are essential for the development of airway hyper-responsiveness and exacerbated asthma following acute RSV infection, influx of eosinophils cannot occur in CD8+ T-cell deficient mice, (Schwarze *et al.*, 1999a; Schwarze *et al.*, 1999b; Schwarze *et al.*, 1999c).

Conversely, evidence suggested that abnormal T-cell regulatory mechanisms might be related to a hyperactive IgE response. It has been shown that a reduced number of CD8+ cells, which may include cytotoxic or suppressor T-cells, are present during convalescence in those who wheeze with RSV compared with children with upper respiratory illness alone. RSV-IgE in nasopharyngeal secretions was inversely related to the number of CD8+ T-cells in peripheral blood suggesting an IgE suppressive role for some of the CD8+ cells (Welliver *et al.*, 1984). Furthermore, CD8+ CTL and NK cells are a source of IFN-gamma which inhibits CD4+ T-cells from producing IL-4 and IL-5, and thus also from stimulating the production of IgE and eosinophilia (Openshaw, 1995; reviewed in Openshaw *et al.*, 2002).

It is thought to be unlikely that CD8+ T-cells were involved in the FI-RSV vaccine induced disease since inactivated antigens are in general unable to induce CD8+ MHC class I restricted CTL, because replication is in general required for entry into the MHC class I antigen processing pathway (Walsh & Graham, 1999).
Still, this suggests a fine line between protection and immune pathology and emphasises the importance of monitoring the CTL responses in future experimental vaccine trials in humans.

The following sections (1.3.2.3-1.3.2.4) will focus on the CTL response. To clarify the terminology used in the rest of this thesis some basic background about CTL immunology will firstly be given to provide a better understanding about

i. The CD8+ CTL function,
ii. The function of the MHC molecules, and
iii. HLA types in South Africa.

The role of CTLs in the protective immune response to RSV will then be discussed.

1.3.2.3 CTL immunology

1.3.2.3.1 Cytotoxic T-cell function

As described above, evidence suggests that an RSV-specific type 1 T-helper cell response and CD8+ cytotoxic T-cells (CTL) are critical for the disease-free control of RSV infection in the murine model (Cannon *et al.*, 1988a). CTLs control acute viral infection directly by destroying virus-producing cells and possibly by releasing cytokines with antiviral activity such as IFN-gamma and TNF-alpha (reviewed by Walsh & Graham, 1999).

Figure 1.6 illustrates the process of antigen recognition and killing of the infected cell by the CTL: After viral infection of the target cell, virus replication and protein synthesis occur in the cytosol. Some of the viral proteins are degraded in the cytosol by a multicatalytic protease complex called the proteasome into small peptide fragments, which are then transported into the lumen of the endoplasmic reticulum by the TAP (Transporters associated with antigen processing) transporter. Here the peptide binds to the partially folded MHC class I molecule.

The fully folded peptide MHC class I-complex is then transported through the Golgi complex to the cell surface (Shepherd *et al.*, 1993, Song & Harding, 1996). The initial interaction of the CD8 T-cell with the target cell is made by non-specific adhesion molecules (LFA-1 expressed on the T-cell and ICAM-1 expressed on the target cell). This allows the T-cell to remain in contact with the target cell and to scan the surface of the target cell for the presence of specific peptide:MHC complexes. The peptide-MHC class I molecule complex on the cell surface is then recognised by CD8+ cytotoxic T-cells through binding to the T-cell receptor and the CD8+ co-receptor. A B7 co-stimulatory molecule is expressed by the antigen-presenting cell, which associates with the CD28 receptor on the CTL, thereby stimulating expansion of the T-cell and induces production of T-cell growth factor IL2. B-7-CD-28 interaction also stimulates the expression of a regulatory receptor, CTLA-4, which limits the proliferative response of activated T-cells to antigen and B-7 on the surface of antigen-presenting cells and limits the amount of IL-2 produced (Azuma *et al.*, 1992a; Rudd, 1996; Azuma *et al.*, 1992b). The activated CTL then produces effector molecules consisting of cytotoxins (perforin, granzymes and Fas ligand) and cytokines (IFN-gamma, TNF-alpha and TNF-beta). The cytotoxins kill the virus infected cell by different non-specific mechanisms: **Perforin** creates pores in the target cell's plasma-membrane (Liu *et al.*, 1995) after which osmotic lysis and **granzyme**-induced apoptosis of the target cell takes place (Heusel *et al.*, 1994), while the membrane bound **Fas-ligand** (CD95L) triggers the death of Fas (CD95)-expressing cells by apoptosis (Squier & Cohen, 1994; Suda *et al.*, 1993). IFN-gamma functions by direct inhibition of viral replication, by inducing increased expression of MHC class I and activating macrophages by recruiting them to the site of infection as effector and antigen presenting cells. IFN-gamma also stimulates IgG2a synthesis, inhibits Th2 cell growth, and activates NK cells. TNF-alpha and –beta can synergize with IFN-gamma in macrophage activation and in killing of certain target cells by interacting with TNFR I (TNF Receptor family) (Arai *et al.*, 1990; Ramshaw *et al.*, 1992). CTL effector cell function is highly specific inducing apoptosis of the viral infected cells, but sparing adjacent normal cells (reviewed in Janeway *et al.*, 1999).

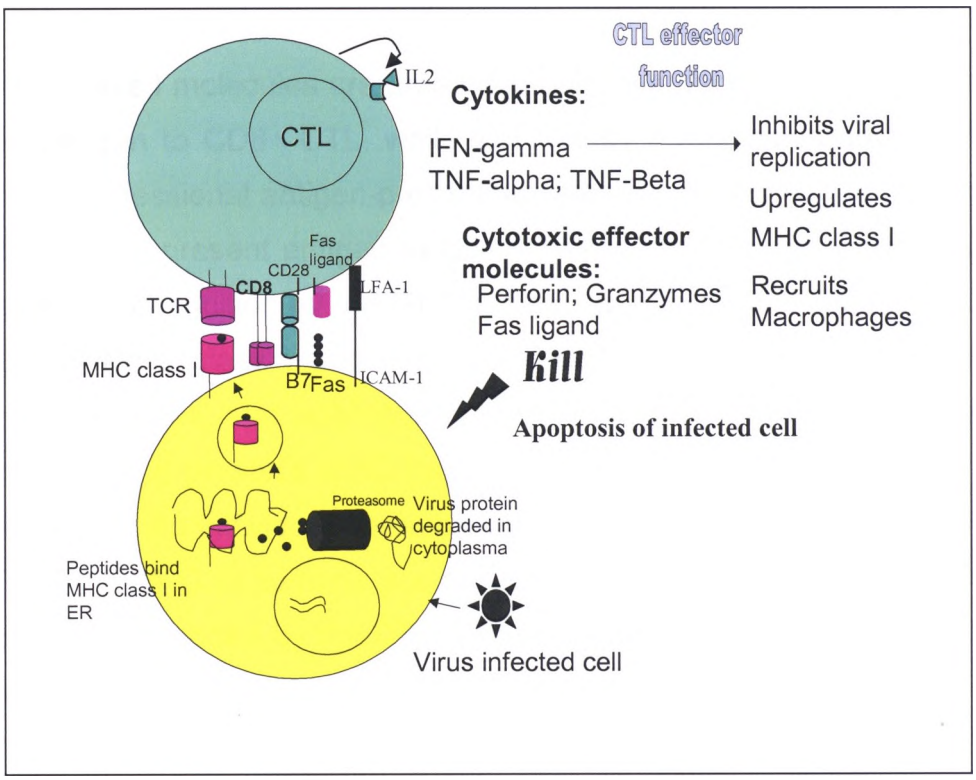


Figure 1.6: CD8+ CTL antigen recognition and effector function. See description in the text.

1.3.2.4 The MHC molecules.

The MHC class I molecules are presented on the surface of all somatic cells and present antigen to CD8+ CTL, while MHC class II molecules are found on the surface of professional antigen-presenting cells like macrophages, dendritic cells and B-cells and present antigen to CD4+ T-cells. In order to prevent pathogens from evading recognition by the MHC molecules, these molecules are *polygenic*, i.e. several MHC class I and II genes encode proteins with different peptide binding specificities, and *polymorphic*, with each gene having multiple alleles in the population. A person's MHC haplotype is determined by the combination of HLA alleles on an individual chromosome. The class I genes are called HLA-A, HLA-B, and HLA-C in humans and H2-K, -D, and -L in mice. The MHC class II molecules are encoded by the DR; -DP; and -DQ genes in humans (reviewed in Janeway *et al.*, 1999).

Both these molecules are cell surface glycoproteins although they have different subunits and peptide binding properties. The MHC class I molecule consists of an α-chain encoded by the MHC genes non-covalently associated to a β_2-microglobulin. The α-chain spans the membrane. It consists of 3 subunits (α1-3), one (α3) that is associated with the β_2-microglobulin to form a immunoglobulin-like structure, while the other two (α1 and α2) are paired to generate a cleft on the surface of the molecule that is the site for peptide binding (Smith *et al.*, 1996). The MHC class II molecule consists of a non-covalently associated α and β chain folded much like the MHC class I molecule, with the major difference in the peptide binding cleft, which is more open (Fremont *et al.*, 1996)(Figure 1.7 a).The protein products of the MHC genes are highly polymorphic especially in the peptide binding grooves which determine the peptide binding properties of the different MHC molecules (Babbitt *et al.*, 1985). An individual's T-cells recognise a foreign virus's T-cell epitopes as a peptide bound by a specific allelic variant MHC molecule that is encoded by HLA genes of his/her own haplotype and will not recognise the same peptides bound to foreign MHC molecules. The antigen recognised by the T-cell is thus MHC-restricted. (Zinkernagel & Doherty, 1974).

MHC class I molecules bind small 8-10 mer peptides, which have been processed intracellularly after infection by viruses or intracellular parasites. MHC class II molecules usually bind longer peptides of 10-20 residues that are produced after degradation of extracellular pathogens by acid proteases in endocytotic vesicles by professional antigen-presenting cells following endocytosis. Peptides are bound to the MHC molecules in an elongated conformation, however in the case of the MHC class I molecule both ends of the peptide are tightly bound at either end of the peptide binding cleft by the amino and carboxyl termini, while the MHC class II molecule does not bind the ends of the peptide which are extended beyond the cleft. The peptides bind the MHC molecules through structurally related anchor residues that differ for different MHC class I alleles both in position and identity, but are similar for all peptides binding the same MHC molecule. Most peptides binding MHC class I molecules have an anchor residue at the carboxy terminus that is usually hydrophobic or basic. The amino acids in the anchor residue that bind a specific MHC molecule do not need to be identical, but are always related amino acids. (Babbitt *et al.*, 1985, Bouvier & Mwiley, 1994; Southwood *et al.*, 1998; Rammensee, 1995; reviewed in Janeway et al 1999). (Figure 1.7 b).

1.3.2.4.1 HLA class I diversity: HLA types in South Africa.

MHC polymorphisms: The human MHC genes are highly polymorphic and many alleles have been identified for each gene locus. More than 200 alleles of some of the MHC class I and II loci have been identified that are present at a relatively high frequency in the population. Furthermore, because of this high allele frequency in the population, most individuals are heterozygous with both alleles being expressed in a co-dominant fashion, providing six MHC class I molecules expressed per cell (Marsh, 1998). The MHC polymorphism in the population is thought to have evolved as a strategy to overcome the evasive strategies of pathogens. Reports suggest that HLA class I homozygosity may result in accelerated disease progression in HIV-type I infection (Tang *et al.*, 1999).

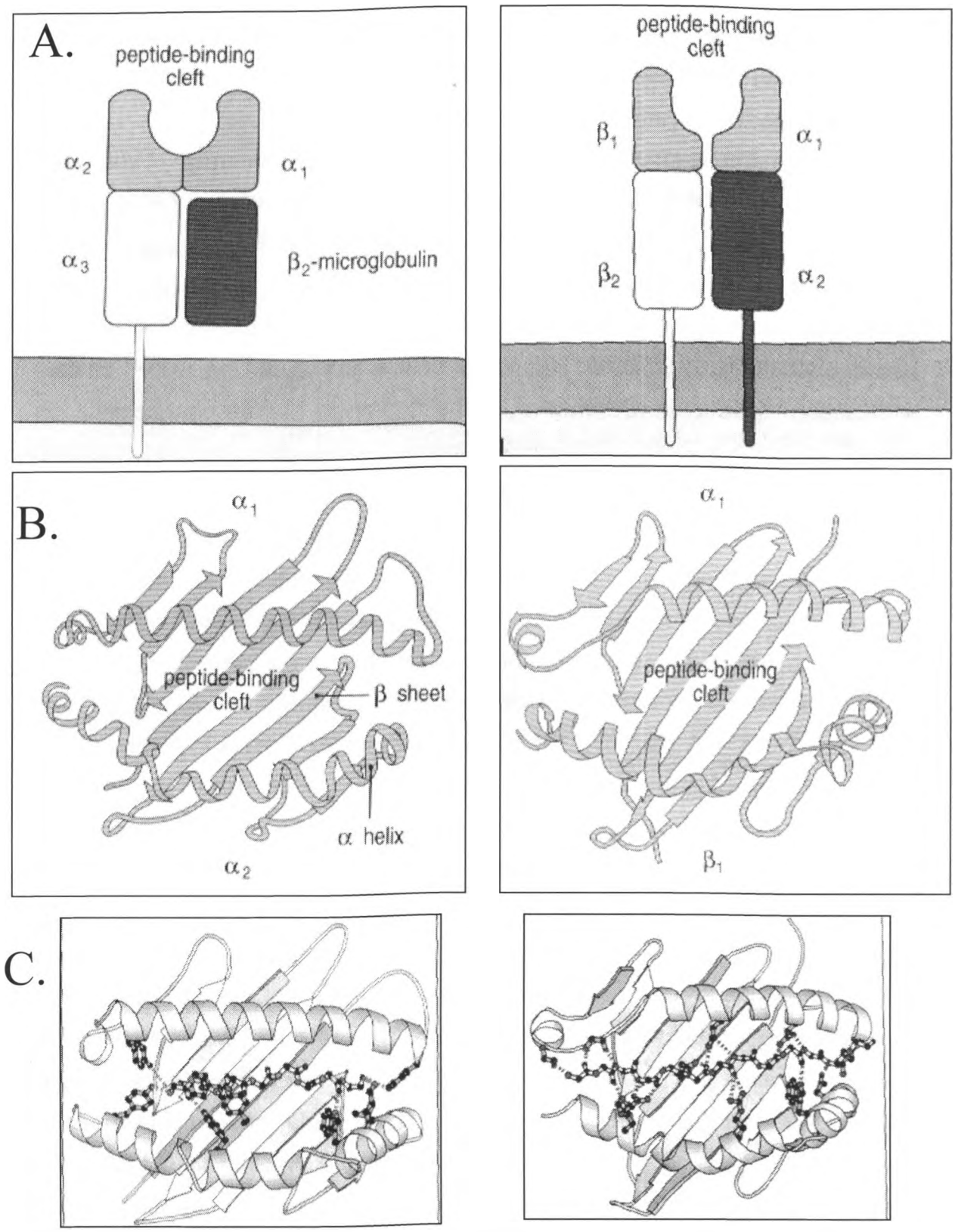


Figure 1.7: MHC class I (left) and II (right) molecules consisting of the peptide binding cleft and the membrane spanning domains (A). The empty binding cleft shown from above (B.) and bound to peptide (C.) The MHC class I molecule binds peptides by their ends (left) while the MHC class II molecule binds peptides by interactions along the length of the binding groove. Adapted from Janeway *et al.*(1999).

Various instances of CTL evasion have been documented where the pathogen mutates certain immunodominant CTL epitopes, or the flanking regions to prevent these peptides from binding the MHC class I molecules (Goulder *et al.*, 2001b; Potts & Slev, 1995; Yellen-Shaw *et al.*, 1997). This is more difficult if there are many MHC molecules present in the population, suggesting an evolutionary advantage for polymorphisms in the MHC molecules. On the other hand, positive selection of certain HLA types has also been documented, for instance HLA-B53, which is common in West Africa but rare in other populations, is thought to have been selected because of its association with recovery from potentially lethal malaria (Hill *et al.*, 1992). The allele frequencies can differ dramatically between populations, and are still being defined. An overview of the allele frequencies in different population groups can be found in Geer *et al.* (1998).

HLA types in South Africa:

South Africa has a highly diverse population as can be appreciated from its 11 official languages of which nine are from African origin (Venda, Sesotho, Sepedi, Ndebele, Setswana, siSwati, Xhosa, Zulu and Tsonga) and two from European origin (Afrikaans and English). The population consists of approximately 43.4 million people of which 75.2% are Black, 13.6% Caucasian, 8.6 % Coloured, and 2.6% Indian according to census estimations of 2000 and 1996 (http://geography.about.com/library/cia/blcsouthafrica.htm). The African groups originated from the Khoisan and the southern migration of the Niger-Congo linguistics groups during the great Bantu migration in the last 3 centuries BC. South African Caucasians are of mainly Dutch (first settlers arrived in 1652), French and German (17th century) and British (19th century) origin. A third major population group are the Cape Coloureds who are an anthropologically distinct mixed population of recent origin. This group possesses western European, southern-African and south-east Asian genes, but in significantly different proportions than the Caucasians. The southern African component of the Cape Coloured ancestry was derived mainly from the Khoikhoi and to a lesser extent from Negro slaves originating mostly from Madagascar. Apart from these three groups, the San (Bushmen) and recent immigrants of south east Asian origin also make up a small proportion of the South African population (Bowyer, 2002; Hammond *et al.*, 1997).

The HLA types in South African ethnic groups are not yet clearly defined, however two reports of comprehensive comparisons between different ethnic groups of southern Africa to ethnic groups from the rest of Africa, or to other parts of the world have been documented (Bowyer, 2002; Hammond *et al.*, 1997). In general the highest frequency allele amongst the different South African population groups for HLA-A was A2 (10-25%). However, although HLA*201 has a high frequency in Caucasians it is relatively rare in the South African Zulu, where A*207 is more common (Clayton & Lonjou, 1997; Hammond *et al.*, 1997). A30 is also very common in the Black and Coloured populations (10-25%) (Hammond *et al.*, 1997), but rare in the South African Caucasians. The Cape Coloureds had different allele frequencies to most of the African groups. The South African Black populations had high frequencies of B12, B7, and B8. B70 had high frequencies by serological testing, but proved to be B*1503(B72) by DNA testing. B13, B15, B44, and B8 were common in the South African Caucasians. The Cw7 (12-35%), Cw4 (7-24%), and Cw6 (3-20%) were the most common alleles of the Cw locus (du Toit *et al.*, 1988; Hammond *et al.*, 1997). Figures 1.8 a. and b. show the frequency for selected South African HLA type A and B alleles in comparison with populations in the rest of the world (adapted from Bowyer, 2002).

Despite the substantial amount of polymorphism that exists in the human population, cross-reactivity exists between certain HLA class I molecules and it has been suggested that the majority of HLA-A and B-alleles can be grouped into a relatively small number of supertypes as defined by their broad peptide binding specificity. Cross- binding peptides are frequently recognised by specific T-cells in the course of natural disease processes and have implications for the design of broadly reactive vaccines (Bertoni *et al.*, 1998; Sette & Sidney, 1998; Sette & Sidney, 1999b).

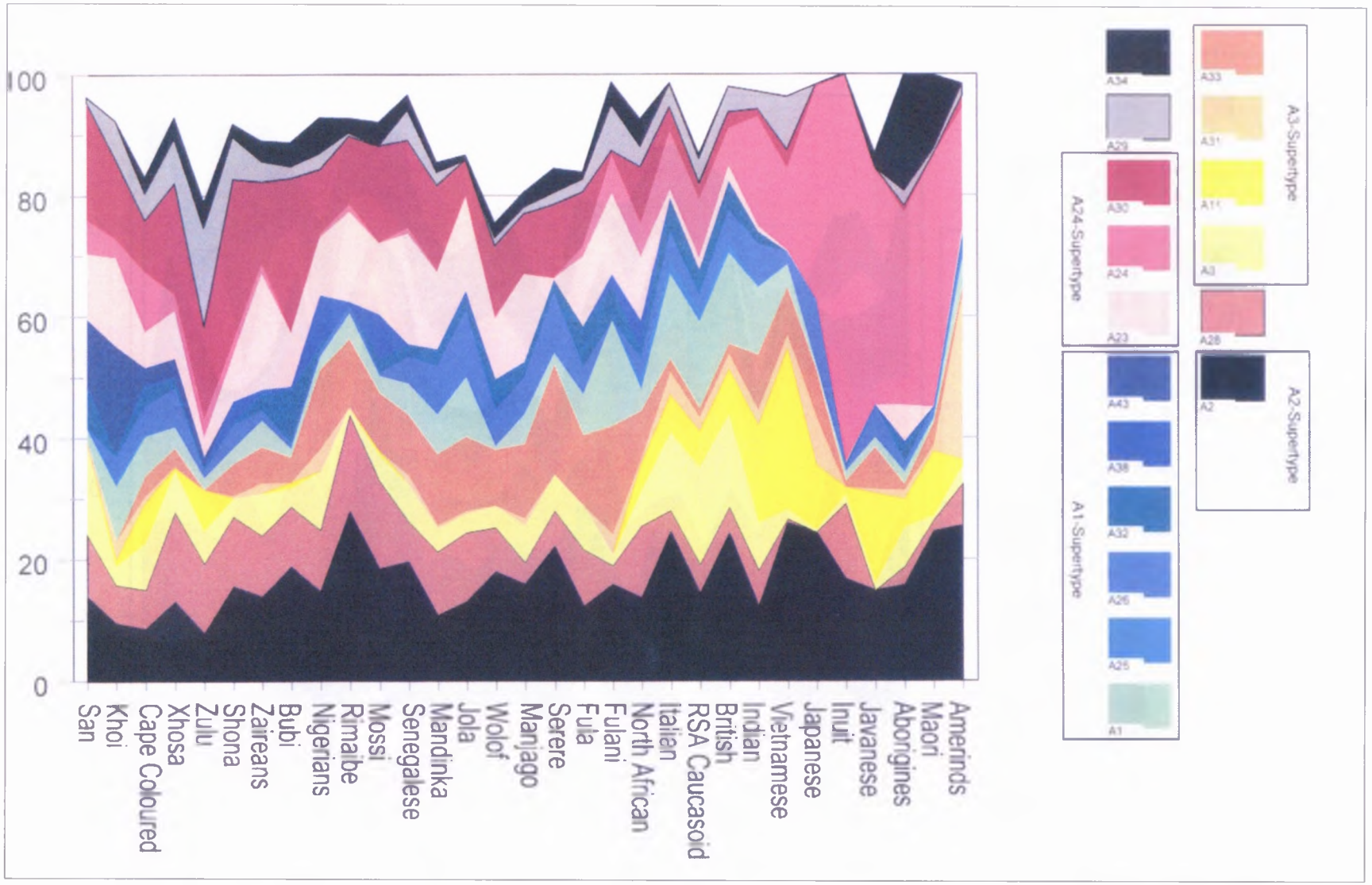


Figure 1.8.a. HLA –A antigen percentage frequencies in selected populations. The colour code and supertype grouping of individual alleles are shown at the right. Adapted from Bowyer (2002).

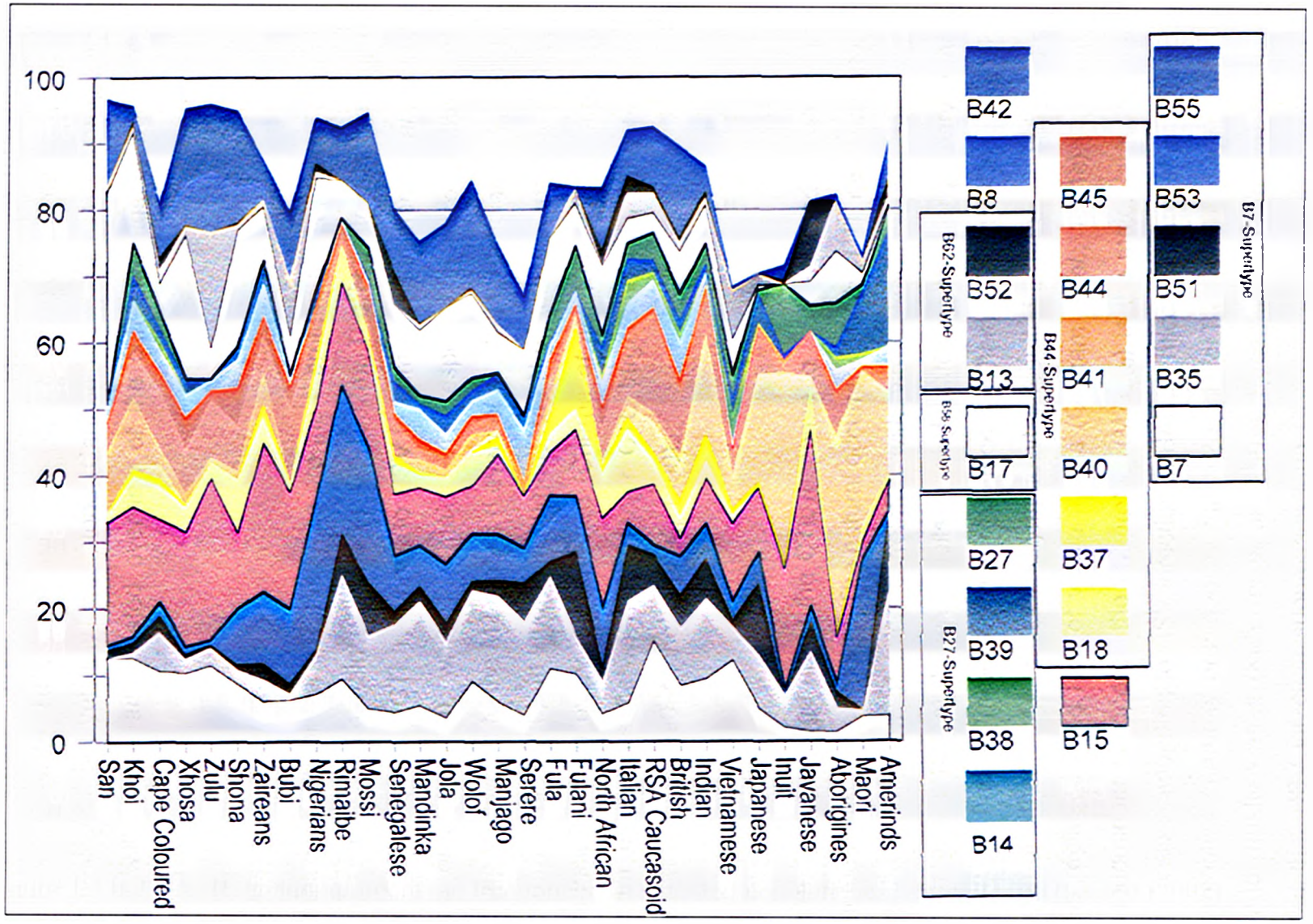


Figure 1.8.b. HLA –B antigen percentage frequencies in selected populations. Adapted from Bowyer (2002).

1.3.2.5 The RSV-specific CTL response.

As for many other viruses, CTLs are likely to be the principle mediators of resolution of acute RSV infection, although their role in protection and pathogenesis is not yet fully understood (Cannon *et al.*, 1988a). The following section will provide an updated overview of the current status of RSV CTL research.

Target proteins

The major targets for the CTL response are the F-protein, the N-protein, the M2 protein and the SH protein (Cherrie *et al.*, 1992a). The N-protein is the major target for the memory CTL response to RSV in humans (Bangham *et al.*, 1986; Cherrie *et al.*, 1992a), while the major CTL response in mice of the H-2^d haplotype is directed to the M2 protein (Levely *et al.*, 1991; Nicholas *et al.*, 1991; Openshaw *et al.*, 1990). In studies in humans, the N-protein was the only protein that induced CTL in every donor tested (Cherrie *et al.*, 1992a). This suggests important differences between the major CTL targets in mice and humans (Openshaw *et al.*, 1990; Nicholas *et al.*, 1990).

Studies in mice

Much of our knowledge about the CTL response to RSV is based on research conducted in the mouse model. The natural occurrence of the CTL response after experimental infection of mice correlated with virus clearance and recovery (Graham *et al.*, 1991b). When B-cells were depleted in mice in primary RSV infection, this had no impact on viral clearance (Graham *et al.*, 1991a), however depletion of either CD4+ or CD8+ T-lymphocytes resulted in an extended period of virus replication and lack of visible illness (Graham *et al.*, 1991c). A single immunodominant epitope on the M2-protein ($M2_{82-90}$) is the major target of the CD8+ T-cells responding to RSV infection in BALB/c mice (Kulkarni *et al.*, 1995). The protective immunity provided by the M2-specific T-cells is however short-lived and rapidly wanes (Kulkarni *et al.*, 1993). Recently, Braciale and Chang reported that RSV infection suppresses lung CD8+ T-cell effector activity and peripheral CD8+ T-cell memory in the respiratory tract of experimentally infected mice through interference with T-cell receptor mediated signalling (Chang &

Braciale, 2002). They found that only 50% of M2-specific CTL present in RSV infected lungs produced IFN-gamma following RSV infection, while there was no deficiency in spleen CTL. CTL directed to a suboptimal RSV F-protein specific CTL epitope had a comparable defect in cytokine synthesis suggesting that this is not only a function of the M2 epitope. These IFN-gamma defective CD8+ CTLs were unable to upregulate perforin expression in the lung and down modulate cell surface TCRs in response to antigen and were defective in *ex vivo* cytolytic activity. This suppressive effect on CD8+ CTL memory in the lungs may explain the limited duration of protective immunity to RSV infection. If RSV has a similar inhibitory effect on CD4+ T-cells, RSV infection may impact on the strength and durability of both cellular and humoral immunity. (Chang & Braciale, 2002).

Aung *et al.* have shown that RSV clearance in perforin-knockout mice, although slightly delayed relative to wild-type mice, occurred by day 10 through Fas ligand-mediated lysis of target cells. They suggested that perforin was not necessary for clearance of primary RSV infection and that the FasL pathway may compensate for the absence of perforin although the alternative mechanism of CTL killing is less efficient and may lead to enhanced disease (Aung *et al.*, 2001). The authors did however point out the fact that NK cells, which possess the same cytolytic machinery as CD8+ CTL, may have been affected by the perforin deficiency. This may have important implications for initiation of viral clearance and promotion of the adaptive response. The possibility that the delayed illness and exaggerated response seen in the perforin-knockout mice could be related to the altered NK cell response cannot be ruled out.

Evidence suggests that priming with CD8+ T-cells before vaccination with the G-protein can modulate the magnitude of pulmonary eosinophilia and Th2 T-cell response that occur during subsequent RSV challenge (Sullender *et al.*, 1991). Although CTL induced by a M2 CTL epitope afforded protection in mice, it also resulted in enhanced weight loss (Anderson *et al.*, 1991). RSV replication was terminated in the lungs of persistently infected mice by infusion of low numbers of RSV-specific cytotoxic T cells without inducing illness (Cannon *et al.*, 1987), yet transferring a larger number of RSV-specific cytotoxic T cells resulted in haemorrhagic pneumonitis (Cannon *et al.*, 1988a). Other groups have found

CTL-mediated RSV clearance in mice without enhanced illness (Munoz *et al.*, 1991).

RSV-specific CTL in humans

RSV-specific CD8+ cytotoxic T cells have been derived from adult peripheral blood mononuclear cells (Bangham *et al.*, 1986) and from children following acute RSV infection (Isaacs *et al.*, 1987; Chiba *et al.*, 1989), and is thought to be crucial for clearing RSV infection since children lacking cell mediated immunity shed virus for months in contrast to healthy children who stop shedding virus 7-21 days after infection (Fishaut *et al.*, 1980). RSV-specific memory CTLs have been documented in both mice (Bangham *et al.*, 1985) and in humans (Bangham & McMichael, 1986; Bangham *et al.*, 1986). Mbawuike *et al* detected RSV-specific CTL activity in 80% of primary infected infants, and although the response was short-lived, a memory response was induced in subsequent infections that correlated with a lower incidence of lower respiratory tract infections in subsequent RSV epidemics. This group also found a positive correlation between CD8+ CTL levels and IFN-gamma and an inverse correlation with IL-4 implicating a contribution of CD8+ CTL and IFN-gamma in control of RSV disease in infants and children (Mbawuike *et al.*, 2001).

Studies in humans have been limited by technical difficulties imposed by the traditional techniques used to enumerate CTL responses. These include a small sample volume from acutely infected children or low frequency of RSV-specific memory CTL response in adults (Goulder *et al.*, 2000b). The availability of more sensitive technologies has now made it possible to study the CTL response in humans. These include the use of tetrameric MHC class I molecules for binding specific epitopes (Altman *et al.*, 1996) and peptide epitope stimulation of T-cells followed by measurement of IFN-gamma secretion by ELIspot or intracellular cytokine staining (ICC) using flow cytometry (Murali-Krishna *et al.*, 1998). In order to be able to use these techniques in studies in infants, RSV specific CTL epitopes need to be identified that can be used as peptide reagents in either of these assays. Identification of RSV-specific CTL epitopes is also needed to study the role of CTL in the pathogenesis and protection of RSV infections.

Currently, only 3 CTL epitopes have been defined for RSV. These include one HLA B7 epitope to the N-protein (Goulder *et al.*, 2000b), and two epitopes on the F-protein restricted to HLA B57 and HLA C12, respectively (Brandenburg *et al.*, 2000a). Brandenburg *et al.* identified these epitopes from T-cell clones that were generated from infants that had just recovered from RSV infection. The authors could show that all of the T-cell clones were CD8+ CTL producing IFN-gamma. Although this confirmed that a CTL response occurs in young infants, and the identified epitopes were relevant in eliciting CTL during acute RSV infection, methods that generate T-cell clones are stochastic in nature and not designed to measure the frequency of the protein-specific CTL. The epitope identified by Goulder *et al.* was in healthy adults, and demonstrated that detection of memory CTLs at low frequencies is possible using IFN-gamma ELIspot techniques. However the epitope was identified by screening donors with peptides to a limited region of the N-protein and only in 6 subjects of who 4 shared the restricting B*07 allele. Neither of these studies aimed to identify the dominant CTL response to any RSV protein in a given population, although they contributed to elucidate the total RSV-specific CTL response in humans. Further epitope mapping studies will be necessary to identify the dominant RSV-specific response and to address the role of RSV-specific CTL in controlling disease. Although it may prove to be a much more daunting task to define the most relevant responses in humans, with the substantial HLA diversity found in an outbred population relative to inbred mice, studies with HIV-1 have indicated that screening of 30-50 previously infected individuals with peptide libraries easily and rapidly defines the epitopes for the most common HLA types (Altfeld *et al.*, 2000; Goulder *et al.*, 2001a).

Knowledge about the role of CTL in protection and disease in humans will contribute to the development and testing of effective vaccines to RSV.

1.3.3 Vaccine development

Despite 30 years of research there is still no licensed RSV vaccine available. In addition to the obstacles listed in section 1.1, the lack of a proper animal model also hinders the testing of experimental vaccines. Although the mouse model helped to highlight the immunological determinants of disease and protection, mice are not fully susceptible to RSV replication (Graham *et al.*, 1988; Prince *et*

al., 1979). The cotton rat proved to be more susceptible (Prince *et al.*, 1978), but the lack of immunological reagents proved to be problematic, and a high body temperature (39°C) may result in an over-estimation of attenuation of temperature sensitive vaccines (Crowe, 2001). The use of non-human primates provided critical information in preclinical trials (Karron *et al.*, 1997b), but even these animals are not as permissive as humans, so that eventually only studies in humans can provide definite information about safety, genetic stability, immunogenicity and protective efficacy (Crowe, 2001). From the immunological studies described above, it can be concluded that the ideal protective and safe RSV vaccine would be one that induces both neutralising antibodies and appropriate levels of CTLs in a Th1 type cytokine milieu, and that the induction of mucosal immunity would be important. This type of immune response was traditionally achieved using live attenuated vaccines but can today also be accomplished with a variety of novel vectors and gene delivery vehicles.

The nasal administration of live-attenuated vaccines that had temperature sensitive (ts) or cold adapted (ca) mutations, achieved many of these immunological requirements, however, although these vaccines were initially promising, they were always over- or under-attenuated in the target population (Karron *et al.*, 1997b; Wright *et al.*, 2000; Wright *et al.*, 1976). Despite these setbacks, recent advances in the field of genetic engineering, improved knowledge about the protein function, the pathogenesis and immunology of RSV infection, have provided new strategies and hope that a successful vaccine can be developed. The reverse genetics system developed for RSV by Collins *et al.* now provides the opportunity to introduce site-directed mutations and even delete complete genes (Collins *et al.*, 1995). Novel strategies of immunomodulation with co-expression of certain cytokines by RSV (Bukreyev *et al.*, 2000), replacement of genes of bovine RSV with human RSV to produce a much-attenuated virus (Buchholz *et al.*, 2000), or production of chimeric viruses whereby subgroup A G- and F-proteins are replaced by those of subgroup B to produce an attenuated subgroup B vaccine (Whitehead *et al.*, 1999), are only some of the vaccines currently under investigation. Deletion of genes encoding proteins involved in pathogenesis like NS1, or the G-protein provides much-attenuated viruses that

will not undergo reversion and may produce very safe vaccines (Karron *et al.* 1997a; Teng *et al.*, 2000).

Subunit vaccines are also still under investigation, even though it is unlikely that these will be used in very young infants. Although similarly altered immunity to that induced by the FI-vaccine has been observed for F and G subunit vaccines (Connors *et al.*, 1992; Murphy *et al.*, 1990), it still holds promise for maternal immunisation or as a booster in previously infected high-risk older children or the elderly to prevent lower respiratory tract disease (Groothuis *et al.*, 1998; Falsey & Walsh, 1997). Currently, the only commercial prophylaxis for RSV is the use of the humanized, mouse monoclonal antibody targeted to the F-protein, palivizumab (Synagis®), which has been shown to be safe and effective in reducing the incidence of RSV hospitalisation by 55% in premature infants (reviewed in Groothuis & Nishida, 2002). However, the cost involved is substantial (approximately $1216.58/100mg wholesale price) and monthly administration is needed to maintain protection (Meissner *et al.*, 1999).

Although sterilising immunity is probably not an achievable goal for RSV vaccines, immunisation of high-risk groups, including all normal infants in the first month of life, premature infants, and patients with chronic lung disease, bronchopulmonary dysplasia and the elderly will reduce the incidence of lower respiratory tract infections and hospitalisation. Therefore, research that will support the development of a vaccine to RSV remains a very important goal.

1.4 STUDY OBJECTIVES

Aims: Firstly, to characterise the molecular epidemiology of RSV in South Africa by determining the virus genotypes, analysing strain circulation, antigenic selection and determining if there is an association between pathogenesis and infecting strains; and secondly to map dominant CTL epitopes on the N-protein of the RSV used by South Africans and determine if immune selection occurs in RSV-specific CTL epitopes.

To achieve these aims the study was divided into two major sections and the following objectives set:

A. THE MOLECULAR EPIDEMIOLOGY OF RSV IN SOUTH AFRICA

1. To establish a means of typing RSV into the two subgroups, A and B for use in South Africa. (Chapter 2)
2. To determine the epidemiological pattern of strain circulation of the RSV subtypes over four consecutive seasons in one specific community in South Africa. (Chapter 2)
3. To identify the RSV genotypes present in South Africa. (Chapters 2-4)
4. To determine if there is an association between the infecting strains and pathogenesis. (Chapter 3)
5. To analyse the distribution pattern of strains in one season in different climatic regions across South Africa. (Chapter 4)

B. DEFINING EPITOPE SPECIFICITY AND HLA RESTRICTION PATTERNS OF RESPIRATORY SYNCYTIAL VIRUS (RSV)-INDUCED CD8+ T-CELLS IN SOUTH AFRICAN ADULTS

1. To identify RSV memory CTL responses specific to the nucleoprotein in South African adults. (Chapter 5)
2. To map the position of specific dominant T-cell epitopes on the N-protein using overlapping 14-mer peptides that cover the complete N-protein. (Chapter 5)

3. To determine the phenotype of peptide specific responses as CD8+/ CD4+. (Chapter 5)
4. To characterise the cytotoxicity of CD8+ CTLs elicited by the identified epitopes and to determine the HLA restriction of the epitopes. (Chapter 5)
5. To map the exact position of selected epitopes on the RSV nucleoprotein by using truncated 9-mer peptides. (Chapter 5)
6. To determine if the population selects for variability in the identified CTL epitopes by sequencing the N protein genes of the different RSV genotypes identified in South Africa. (Chapter 6)

The WHO has initiated extensive RSV epidemiological programs throughout the world to assess the RSV-associated disease burden. In conjunction with this effort, the circulating strains and evolution of RSV need to be determined to identify relevant isolates for inclusion in vaccines and to aid in the identification of relevant vaccination strategies. The need for a vaccine and the obstacles to its development also emphasise the importance of a thorough understanding of the immune response in RSV-associated disease and protection. RSV has the ability to infect infants in the presence of maternal antibodies and to cause repeat infections throughout life. This may be due either to an inadequate immune response or to the variability of the virus. Since very little data are currently available about RSV in South Africa, studies were designed that will provide background information about RSV in our country needed for the development and testing of relevant vaccines.

Therefore, the purpose of this project was firstly the identification and molecular characterisation of RSV genotypes in South Africa (Chapters 2-4). The studies undertaken will define the molecular epidemiological characteristics, determine the circulating genotypes in South Africa, determine the evolutionary pattern of RSV over several years in one community, establish the circulation pattern of RSV across South Africa, and determine if there is a relationship between pathogenicity and the infecting RSV genotype in epidemics.

Until now no data was available about the molecular epidemiology of RSV in South Africa. Only two other reports exist about RSV molecular epidemiology in

Africa, one from Mozambique (Roca *et al.*, 2001) and one from Gambia (Cane *et al.*, 1999). Both of these studies suggest the existence of certain genotypes in Africa that have not been identified elsewhere. No data are available about the molecular epidemiology over consecutive seasons from sub-Saharan Africa, which has one of the worst HIV epidemics in the world.

To achieve these objectives, specimens were obtained from three different study groups:

1. Children hospitalised for ALRI at Chris Hani Baragwanath hospital from 1997-2000. RSV positive specimens: n = 1187
2. Children visiting community clinics in Agincourt in the Limpopo Province from 1998-2000 as part of a WHO RSV initiative. RSV positive specimens: n =96 (1998 =3; 1999=7; 2000=86)
3. Children with ALRI visiting 5 private hospitals in different parts of South Africa, namely Johannesburg, Pretoria, Bloemfontein, Cape Town and Durban, as part of an RSV prevalence study, conducted during 2000 and sponsored by Abbott Laboratories. RSV positive specimens: n = 50

For study group 1, specimens were selected systematically to include every month of each year in order to obtain a true representation of genotypes during this time. The quality of specimens has in part influenced the number of samples used for molecular analysis. Specimens of study group number 1 were of high quality since these specimens were collected at Chris Hani Baragwanath hospital in Soweto and sent for diagnosis to the National Institute for Communicable Diseases in Sandringham, (both sites which are in the vicinity of Johannesburg) on the same day, where an aliquot was frozen for molecular analysis. These specimens were used for establishment of the molecular techniques (RT-PCR and nested PCR for subtyping and sequencing), but were of high enough quality to be subtyped by RT-PCR alone. For study group number 1, 30-50 specimens were subtyped from each year (250 in total) to determine the dominant subgroup in each year while approximately 30 specimens from each year were sequenced to identify the circulating genotypes (103 in total). Study group 2 consisted of community specimens collected at the WHO study site in Agincourt in the

Limpopo Province of South Africa. This was a remote rural site without electricity, which made specimen preservation more difficult. Since the specimen quality was in general quite poor, a nested PCR was developed for subtyping and to increase the amount of PCR product for subsequent sequencing. Patient records were analysed to determine the disease severity and the genotypes identified in the community compared to the hospital specimens to determine if there was any correlation between disease severity and infecting strains. Study group 3 consisted of specimens collected at hospitals in different regions in South Africa during 2000-2001. Specimens were frozen in liquid nitrogen and sent to the National Institute for Communicable Diseases in one batch for diagnosis and an aliquot used for molecular analysis. The quality of these specimens was good, but only a small number were collected by some of the sites. Study groups 2 and 3 were used to determine the distribution of genotypes across South Africa during one year. All available clinical specimens from study group 2 and 3 have been analysed by PCR and sequencing.

The purpose of the second part of this study (Chapter 5) was to map RSV-specific CTL epitopes and to define their MHC class I restriction elements in South Africans. In order to develop an effective vaccine, the appropriate protective immune response to RSV needs to be identified. Although serological assays for RSV are available, no assays for measuring cell-mediated immunity exist. A type 1 T-helper (Th1) cell response with the induction of CD8+ cytotoxic T-cells is thought to be essential for the clearance of RSV from the lungs, although some evidence also exists that CTLs may enhance disease. In order to study the role of the CTL response in humans, CD8+ T cell epitopes need to be defined because they constitute the reagents needed for detecting the CTL specific responses rapidly and effectively in children. Having access to assays for evaluating quantitative aspects of epitope-specific CTL responses will facilitate RSV vaccine development and improve our understanding of RSV-induced disease pathogenesis.

Only 3 CTL epitopes restricted to human HLA types have been identified to date, none of these studies have been aimed at determining the dominant responses to any of the RSV proteins in a population, and none are to HLA types that are

common in the South African population. The second part of this thesis is focussed on mapping RSV CTL epitopes and their MHC class I restriction elements for South Africa. Thirty-seven South Africans from various ethnic origins were screened with overlapping peptides covering the complete RSV N-protein by IFN-gamma ELIspot. The N-protein was selected because this is the most conserved protein both within and between the RSV subgroups and has been reported to be the major target for the CTL response in humans. Since all individuals are infected by RSV during the course of their lives, any adult can be used to study RSV specific CTL epitopes and their MHC class I restriction elements. T-cell epitopes were tested for cytotoxicity by Chromium (Cr^{51}) release assays and the phenotype determined by intracellular cytokine assay or by using CD4+ or CD8+ sorted cells in IFN-gamma ELIspots. The HLA allelic expression of individuals with a positive reaction to at least one epitope was determined by molecular techniques. The HLA restriction of epitopes was determined using target cells that were mismatched in all but the HLA allele in question.

In the final chapter (Chapter 6) the data accumulated about the molecular epidemiology of RSV in South Africa was used to determine if immune selection occurs in the population against defined CTL epitopes. Such selection could help the virus evade the immune response and so establish a repeat infection. Immune evasion does occur in chronic infections like HIV and EBV, and recently apparent immune evasion in CTL epitopes of influenza virus was also found (Goulder *et al.* 2001;Boon *et al.*, 2002). However, it is hypothesised that immune selection does not occur in the CTL epitopes in the highly conserved N protein of RSV and that this will therefore not need regular updates or influence vaccine development. In order to prove this, the N-protein of RSV clinical specimens that represent all genotypes identified in South Africa in this study were sequenced and compared by phylogenetic and amino acid analysis.

CHAPTER 2

Genetic diversity and Molecular epidemiology of RSV over four consecutive seasons in Soweto, South Africa: Identification of new subgroup A and B genotypes

2.1 INTRODUCTION

Little is known about RSV epidemics in developing countries, however it has been estimated that a third of all deaths in children under five years of age are due to acute lower respiratory tract infections (ALRI), with RSV being responsible for 27-96% of hospitalised cases caused by a viral infection (reviewed by Weber *et al.*, 1998).

The existence of distinct lineages within the RSV subgroups has been demonstrated at both antigenic (Hendry *et al.*, 1988; Mufson *et al.*, 1988; Storch *et al.*, 1991) and nucleotide levels (reviewed by Cane & Pringle, 1995b; Sullender, 2000). Nucleotide sequencing has revealed the G-protein to be the most divergent, both between and within the two subgroups (Johnson *et al.*, 1987b). The G-protein, which is one of the targets for the neutralising and protective immune response (Johnson *et al.*, 1987a), appears to accumulate amino acid changes with time, suggesting evolution under selective pressure (Cane & Pringle, 1995a). It has been reported that antigenic differences between the RSV strains influence the immune response in children, which is predominately group-specific and reveals poor cross-reactivity (Hendry *et al.*, 1988; Cane *et al.*, 1996; Sullender *et al.*, 1990b). These properties make the G-

protein an ideal target for epidemiological studies. The second hyper-variable region, which makes up the C-terminal of the G-protein, has been reported to provide a reliable proxy for the entire G-gene variability (Peret *et al.*, 1998) and has subsequently been used in phylogenetic analyses for molecular epidemiological studies (Peret *et al.*, 2000; Peret *et al.*, 1998).

RSV epidemiology is community-based in nature, but transmission of viruses between communities may occur (Anderson *et al.*, 1991; Peret *et al.*, 2000). It is thought that several strains are introduced or circulate endemically in communities each year and that local factors, possibly immunity induced by exposure to strains from previous years, determine which strains predominate during an RSV season (Peret *et al.*, 2000). Although RSV can be variable within a single local epidemic, very similar viruses circulate worldwide, indicating temporal but little geographical clustering (Cane & Pringle, 1992; Garcia *et al.*, 1994).

Although extensive analysis of RSV G-protein variability has been performed, most studies have been based on subgroup A isolates (Cane *et al.*, 1991; Cane & Pringle, 1995a; Cane, 1997; Cane *et al.*, 1999; Garcia *et al.*, 1994), with little information available on subgroup B isolates (Sullender *et al.*, 1990a; Sullender *et al.*, 1991; Coggins *et al.*, 1998; Martinez *et al.*, 1999). Few studies have reported on the molecular epidemiology of both subgroups (Coggins *et al.*, 1998; Peret *et al.*, 2000; Choi & Lee, 2000; Peret *et al.*, 1998). Our knowledge of the molecular epidemiology of RSV has so far been based mainly on studies conducted in the developed world with few reports from Africa (Cane *et al.*, 1999; Roca *et al.*, 2001). No data are available about the molecular epidemiology over consecutive seasons from sub-Saharan Africa, which has one of the worst HIV epidemics in the world (Lepage *et al.*, 1998). Recent reports suggest that the perennial isolation of RSV in HIV infected children may be changing the epidemiology of RSV illness in South Africa (Madhi *et al.*, 2000).

This chapter describes the first molecular epidemiological analysis of RSV over consecutive seasons in a single community in sub-Saharan Africa. The

evolutionary pattern of RSV was examined over four consecutive epidemics in hospitalised children in an academic hospital (Chris Hani Baragwanath hospital) serving Soweto in South Africa. Soweto is located in the Gauteng province of South Africa and has a temperate climate with warm summers and mild winters. The rainy season occurs in the summer months. Techniques for RNA isolation, subtyping and phylogenetic analysis that are used in Chapters 2-4 were established using these specimens. A published RT-PCR was adapted for use in clinical specimens, and a nested PCR was developed for subtyping and amplification of clinical specimens that contained low levels of RNA. The genotyping method published by Peret *et al.* (Peret *et al.*, 2000; Peret *et al.*, 1998) was extended to include all South African genotypes. The co-circulation pattern and genetic variability of both subgroup A and B clinical specimens were determined, and the relatedness of South African RSV specimens with genotypes identified in the the rest of the world was established.

2.2 METHODS

2.2.1 Patient specimens and viruses

This study was conducted on RSV infected children between the ages of 0-60 months admitted with ALRI to Chris Hani-Baragwanath hospital, Soweto, South Africa. This hospital serves a population of approximately 1.2 million Black urban South Africans, including 45,000 children under the age of two years (Daponte, 1995). During the first year only children that fulfilled the criterion of severe ALRI were recruited, while in the following three years all children hospitalised with ALRI were enrolled. RSV-positive nasal pharyngeal aspirates (NPA) from infants collected over a period of four years (1997-2000), were frozen at –80°C following diagnosis by immunofluorescence assay (IFA). Specimens were selected from every month that RSV was isolated (January to November, 1997-2000) thereby giving a representative proportion of the RSV specimens found in each month. A total of 225 specimens out of 1187 RSV-positive specimens (18.9%) over the four year period were selected for subtyping. HIV status was not a selection criterion for inclusion in this study.

2.2.2 RNA extraction

Total RNA was isolated from frozen NPA. NPA were thawed on ice, and immediately supplemented with 20 Units of RNase inhibitor (Roche molecular biochemicals, Mannheim, Germany). Different RNA extraction methods were tested and the results compared by assessing the success of RT-PCR. These methods included the use of Trizol LS reagent (GIBCOBRL, Life technologies, Inchinnan, Scotland); RNeasy™ spin columns (QIAGEN, Hilden, Germany); Viral RNA isolation kits (Roche molecular biochemicals, Mannheim, Germany) and the High Pure™ RNA kit (Roche molecular biochemicals, Mannheim, Germany). Of these methods, the Viral RNA isolation kit and the High Pure™ RNA isolation kit (Roche) gave the most consistent results, but when using the High Pure™ RNA RNA isolation kit, the following modifications had to be made: 200 µl of the clinical specimen was added directly to 400µl of the lysis buffer instead of producing a cell pellet first.

In the case of very viscous specimens the lysates were homogenised using QIAshredder™ homogeniser columns (QIAGEN, Hilden, Germany). Total RNA was extracted directly from 200 µl of clinical specimen according to the manufacturer's instructions.

2.2.3 Subgroup identification: RT-PCR and multiplex nested PCR

A multiplex nested RT-PCR was developed for subgroup identification of clinical specimens that combines an RT-PCR method for distinguishing between group A and B (Sullender *et al.*, 1993), with an additional nesting step using subgroup-specific primers.

RT-PCR

The original cDNA synthesis and PCR method described by Sullender *et al.* was modified to a one step RT-PCR reaction using the Titan™ One Tube RT-PCR system (Roche molecular biochemicals). Primers were as published in Sullender *et al.* (1993) (Table 2.1). 10 µl of RNA was added to 10 µl standard 5X reaction buffer, 10 mM of each dNTP, 20 pmol of each primer (G32B, G267A, F164A), 5 mM DTT-solution, 10U RNase inhibitor, and 1 µl the Titan ™ enzyme mix and the volume was made up to 50 µl with deionised distilled water. The reaction was performed in a GeneAmp PCR system 9600 (Perkin-Elmer Applied biosystems, Foster City, USA) thermocycler according to the following program: 30°C for 30 minutes, 94°C for 2 minutes, (94° for 30 seconds, 55°C for 30 seconds, 68°C for 1minute) for 35 cycles, 68°C for 7 minutes. Negative water controls were carried through the whole procedure.

Nested PCR

Primers for the nested PCR were specific for the 5' ends of either the subgroup A or B RT-PCR products (Table 2.1). The subgroup B nesting primer (GB52) has been published (Christensen *et al.*, 1999). The subgroup A nesting primers (G314A and G283A) were designed for this study to be specific for subgroup A. G314A gave more specific results and was used standardly.

Table 2.1: RSV-specific primers for RT-PCR and nested PCR

Primer	Specificity	Orientation	Fragment size	Sequence	Author
F164: RT-PCR and cDNA	Subgroup A & B F-protein, nt 164 Reverse	3'		GTTATGACACTGGTATACCAACC	(Sullender *et al.*, 1993)
G267A RT-PCR	Subgroup A, G-protein, nt 267 Forward	5'	900 bp	GATGCAACAAGCCAGATCAAG	(Sullender *et al.*, 1993)
G32B RT-PCR	Subgroup B, G-protein, nt 32 Forward	5'	1100 bp	GCAACCATGTCCAAACACAAG	(Sullender *et al.*, 1993)
F1 Nested PCR	Subgroup A and B, F-protein, nt1, Reverse	3'		CAACTCCATTGTTATTTGCC	(Peret *et al.*, 1998)
GB52B Nested PCR	Subgroup B, G-protein nt, 52	5'	946 bp	AATCAACGCACTGCCAGKACTC (where K=G/T)	(Christensen *et al.*, 1999)
G314A	Subgroup A, G-protein, nt 314	5'	697 bp	GAATCCCCAGCTTGGAATCA	Venter *et al.* 2002
G283A*	Subgroup A, G-protein, nt 283	5'	728 bp	CAAGAACACAACCCCAACAT	(Venter *et al.*, 2002)

*Primer G283A was only used in cases where no product could be detected with G314A.

In cases where no product was detected with G314A, G283A was used to try to increase the number of specimens that could be subtyped. Although both are subgroup A-specific, G283A would in some specimens give an extra non specific band, however; all specimens with double bands were shown to be subtype A by sequencing. The nesting 3' primer (F1) (Peret *et al.*, 1998) is specific for the detection of both subgroup A and B viruses.

The nested PCR was conducted in a 100 µl reaction. 1 µl of the RT-PCR product was mixed with 0.2 mM of each deoxinucleotide (dTTP, dATP, dGTP, dCTP) (Roche molecular biochemicals, Mannheim, Germany), 30 pmol of each primer

(G52B, G283A, F1), 2.6U Expand high fidelity PCR enzyme mix (Roche molecular biochemicals, Mannheim, Germany), 10X Expand HF buffer with 15 mM $MgCl_2$.

The reaction was carried out using the following program: 94°C for 2 minutes; (94°C for 30 seconds; 58°C for 1 minute; 72°C for 1 minute) for 30 cycles; and finally 72°C for 7 minutes. This reaction was used for most nested PCR reactions used in chapters 2-4.

At the end of the project further optimisation showed that non-specific bands could be reduced if a hotstart PCR is performed with the same primers. For this the HotstarTaq DNA polymerase system (QIAGEN, Hilden, Germany) was used. HotStarTaq DNA Polymerase is a thermostable, modified form of the recombinant Taq DNA polymerase and is only activated by a 15 min, 95°C incubation step. HotStarTaq DNA Polymerase provides high PCR specificity and often increases the yield of the specific PCR product. The following reaction conditions were used:

1 µl of RT-PCR product was added to a reaction mix consisting of 5 µl of 10X PCR buffer (containing 15 mM $MgCl_2$), 200 µM dNTPs, 0.3 µM of each primer (F1, G52B, and G314A), 1.25U HotstarTaq DNA polymerase and made up to a final volume of 50 µl with sterile water. The reaction was carried out using the following cycling conditions: 95°C, 15 minutes; (94°C, 1 minute; 59°C, 1 minute; 72°C, 1 minute) x 35; 72°C, 10 minutes.

2.2.4 Analysis of the PCR products

PCR products were analysed on a 2% agarose gel using a 100 bp ladder as molecular weight marker (DNA molecular marker XIV, Roche diagnostics, Mannheim, Germany). The gels were visualized and documented on a UVP transilluminator using Grab IT annotating Grabber 2.51 software (UVP (Inc), USA).

2.2.5 Nucleotide sequencing

PCR products were purified using the Qiaquick 8 PCR purification kit (QIAGEN, Hilden, Germany). Primers used for sequencing were as follows:

Subgroup A 5'primer: G598A (5' GGAAAGAAAACCACCACCAA)
3' primer: FG-A (GATTCTGTTTGATTTGGTCAT)
Subgroup B 5': G604B (5'AAACCAACCATCAAA CCCACA3');
3'primer: FG-B CTTGCTTGATAGATCACGGTT

Primers were designed using DNASIS for windows version 2.5. (Hitachi Software engineering Co, Ltd) and multiple alignments of several published RSV G-protein gene sequences.

Sequencing products were purified with Autoseq GE-50 columns (AEC-Amersham, Sandton, South Africa). Nucleotide sequencing was carried out on both strands using ABI Prism big dye terminator cycle sequencing on an ABI 377 sequencer model; or with the ABI Prism^{R} Big Dye^{TM} Terminator Cycle sequencing reaction kit v2.0 on an ABI 3100 sequencer (Perkin-Elmer Applied biosystems, Foster City, USA).

Nucleotide sequence editing was performed with DNASIS for windows version 2.5 (Hitachi Software engineering Co, Ltd.) or with Sequencher™ version 4.0.5.

2.2.6 Phylogenetic analysis

A region spanning 270 nucleotides, representing the second variable region of the G-protein gene was used for the phylogenetic analysis (Peret *et al.*, 1998). This corresponds to nucleotides 649-918 of prototype strain A2 for subgroup A and nucleotides 652-921 of prototype strain 18537 for subgroup B G-protein genes (Johnson *et al.*, 1987). Nucleotide sequences of subgroup A and B viruses were aligned separately with Clustal X 1.64b (Thompson *et al.*, 1997) using the multiple alignment option. Phylogenetic analyses were carried out with the maximum parsimony method (MP), using PAUP version 4.0b4a (Swoford DL, 1993). A heuristic search was done for the best tree. Statistical evaluations of the Maximum parsimony trees were carried out by bootstrap evaluation under MP-assumption. Sequences were added randomly and one tree was held at each

step (100 replicates and 100 bootstrap replicates). Bootstrap values were obtained from a 50% majority rule consensus tree.
TreeView (win32) version 1.6.1 (Page, 1996) or PAUP version 4.0b4a was used for reconstructing trees. Trees drawn with PAUP were midpoint rooted with minimum F-value optimisation.

Trees were also computed using Maximum Parsimony, Neighbour-joining, and Fitch-Margoliash methods of Phylip Version 3.5c (Felsenstein, 1993), quartet maximum likelihood reconstruction using Puzzle v.4.0.2. (Strimmer and von Haeseler, 1996); and Neighbour-Joining and Minimum evolution methods of MEGA version 2.1 (Kumar *et al.*, 2000) (Results not shown). Comparison of phylograms obtained by Maximum Parsimony and Neighbour joining methods suggested that similar trees were obtained with either method when tested by 100 bootstrap evaluations. Due to time and computer capacity requirements of Maximum Parsimony methods, and since trees of similar quality could be obtained under above-mentioned conditions, trees were constructed by Neighbour-joining methods using Mega 2.1 in later chapters (Chapters 3 and 4).

The Pairwise distance between and within the genotypes of the RSV subgroups were determined as the proportion of differences (P-distance), i.e. the number of pairwise nucleotide or amino acid differences divided by the total number of nucleotides or amino acids in the sequenced region, with MEGA version 2.1 (Kumar *et al.*, 2000). The ratio of synonymous substitutions per synonymous site (Ks) to non-synonymous substitutions per non-synonymous site (Ka) was calculated according to the Modified Nei-Gojobori method using Mega version 2.1 (Kumar *et al.*, 2001). Ka/Ks were calculated as the average ratio for South African sequences in subgroup A and Subgroup B genotypes respectively, and relative to the prototype strains A2 and B1. If Ka/Ks >> 1, strong evidence exists that selection has acted to change the protein (Hurst, 2002; Nei & Kumar, 2000)

2.2.7 Genotype assignment

Subgroup A and B clinical specimens were genotyped by phylogenetic comparisons of sequences that had previously been assigned to specific genotypes (Peret *et al.*, 2000; Peret *et al.*, 1998). Sequences were arbitrarily

considered a genotype if they clustered together with bootstrap values of 70-100%. After analysis of the data the criteria of a genotype were further refined to those clinical specimens that had a P-distance of less than 0.07 from all other members in the same phylogenetic cluster. Subgenotypes were assigned within the genotypes, to clusters of sequences with P-distances of less than 0.04.

2.3 RESULTS:

2.3.1 RT- PCR and nested PCR for subgroup identification

The modified one step RT-PCR was effective for subtyping RSV and resulted in PCR products of 900 and 1100 bp PCR products that could easily be distinguished on a 2% agarose gel, as indicated by Sullender *et al.* (1993). However, although a single PCR was sufficient for cultured isolates and for certain high quality clinical specimens (Figure 2.1.a), the RNA levels were too low in many of the nasopharyngeal aspirates, especially in low quality specimens that were not transported or stored correctly. This made it difficult to amplify the G-protein gene product to high enough levels for subtyping or sequencing. Subsequently, a multiplex nested PCR was developed that could be used for subtyping RSV clinical specimens directly and to improve the yields of the RT-PCR. The nested PCR drastically improved the sensitivity of the RT-PCR. It was used successfully for subtyping clinical specimens and resulted in products of 728 bp for subgroup A and 946 bp for subgroup B that could without difficulty be distinguished by agarose gel electrophoresis (Figure 2.1 b). In a test of sensitivity, out of 35 clinical specimens that were RSV-positive by IFA but negative by RT-PCR, 32 could be shown to be positive by the nested PCR. Negative controls that were included throughout this experiment remained negative. To verify the specificity of the nested primers, a total of 52 of the nested PCR products were sequenced and compared to known subgroup A or B isolates. All of these were shown to have been subgrouped correctly. The nested PCR improved the quality and yield of DNA for use in sequencing.

2.3.2 RSV subgroup distribution over four years at Chris Hani Baragwanath hospital

A total of 225 specimens (18.9%) were selected systematically from every month for which RSV was isolated (January to November, 1997-2000) for subtyping, 44 from 1997 (42%); 69 from 1998 (22%); 55 from 1999 (15%) and 57 from 2000 (14%). Of the 225 clinical specimens, 144 belonged to subgroup A and 81 to subgroup B.

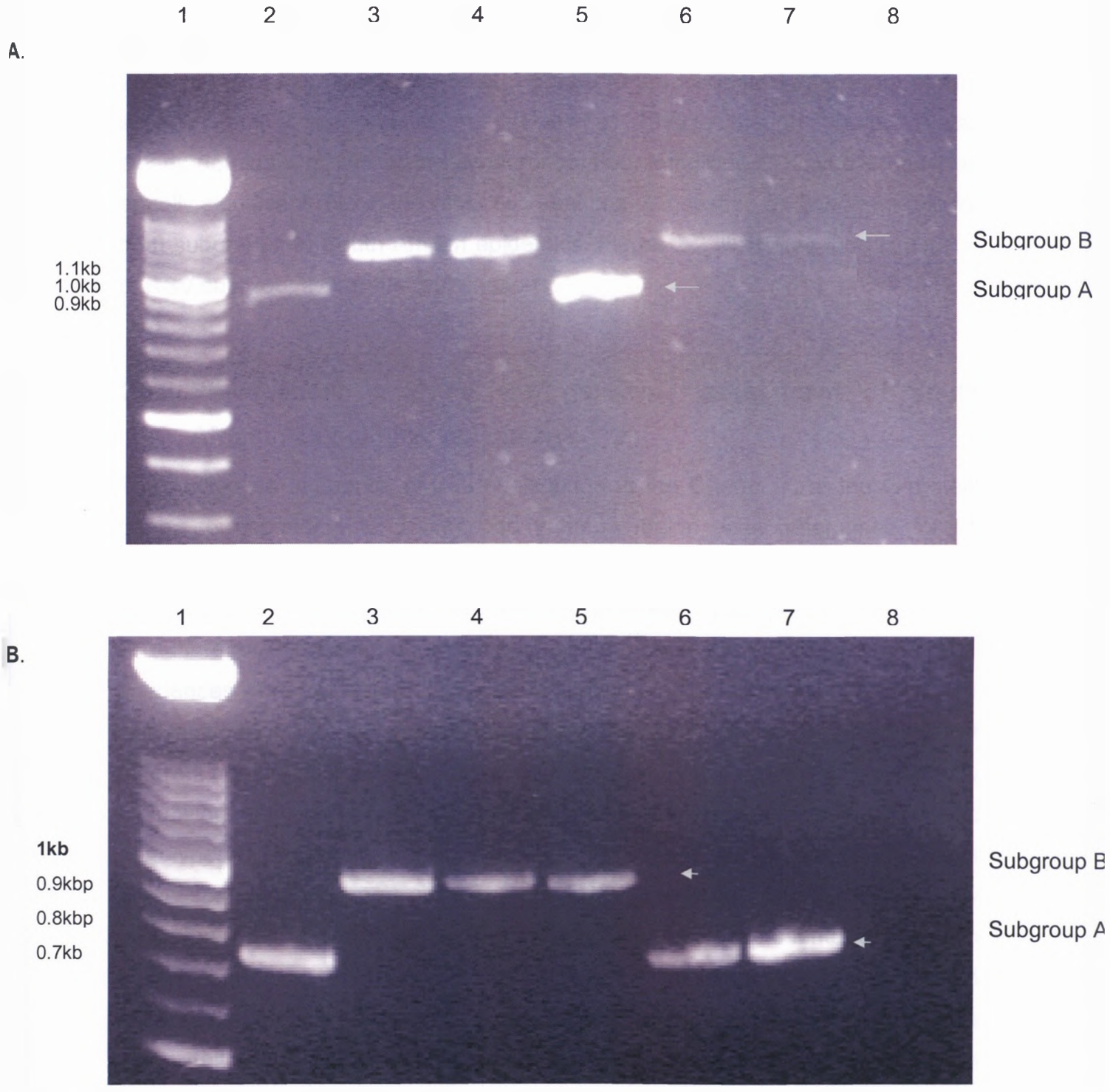


Figure 2.1: a.) RSV subtyping by multiplex RT-PCR. Lane 1: Molecular weight size marker (100 bp ladder (Roche, Manheim Germany)); Lanes 2-7 RT-PCR products of RSV-positive nasopharyngeal aspirates; Lane 8: Negative control.

b.) RSV subtyping by multiplex nested PCR. Lane 1: Molecular weight size marker; Lanes 2-7 multiplex nested PCR of RSV-positive nasopharyngeal aspirates that could not be detected by RT-PCR alone; Lane 8: Negative control.

During 1997 subgroup A and B co-circulated approximately equally at Chris Hani Baragwanath hospital, while subgroup A predominated during 1999 (74%) and 2000 (89%). Subgroup A has been reported to predominate in three out of four seasons in other geographical areas (Hall *et al.*, 1990). Although subgroup A was clearly dominant in two of the seasons studied, subgroup B played at least an equally significant role in the other two seasons, suggesting an important role for both subgroups in South African epidemics. Figure 2.2 illustrates the distribution of RSV subgroups over the four seasons.

2.3.3 Phylogenetic analysis and genotype assignment of South African RSV clinical specimens

The nucleotide sequence of 270 nucleotides at the C-terminal of the G-protein gene, corresponding to the second variable region, was determined for 65 subgroup A and 38 subgroup B specimens, spanning each of the four RSV seasons. Unique sequences were selected from each year (45 subgroup A and 35 subgroup B specimens) for phylogenetic analysis (Figure 2.3). Where identical sequences were found in more than one season, one representative of each year was included in the phylogenetic tree in order to demonstrate the distribution of different genotypes over the study period. Subgroups were genotyped separately by phylogenetic comparisons with sequences previously assigned to specific genotypes (Peret *et al.*, 1998; Peret *et al.*, 2000) and by comparison to various other sequences from around the world (Figure 2.3). Authenticity of new genotypes was tested by performing blast searches with representatives of each cluster formed by South African sequences and including the closest related sequences identified on the Genbank in the phylogenetic analysis. Sequences were arbitrarily considered a genotype if they clustered together with bootstrap values of 70-100%. After analysis of the data the criteria for a genotype was further refined to specimens with a pairwise-distance (P-distance) of less than 0.07 to all other members in the same phylogenetic cluster.

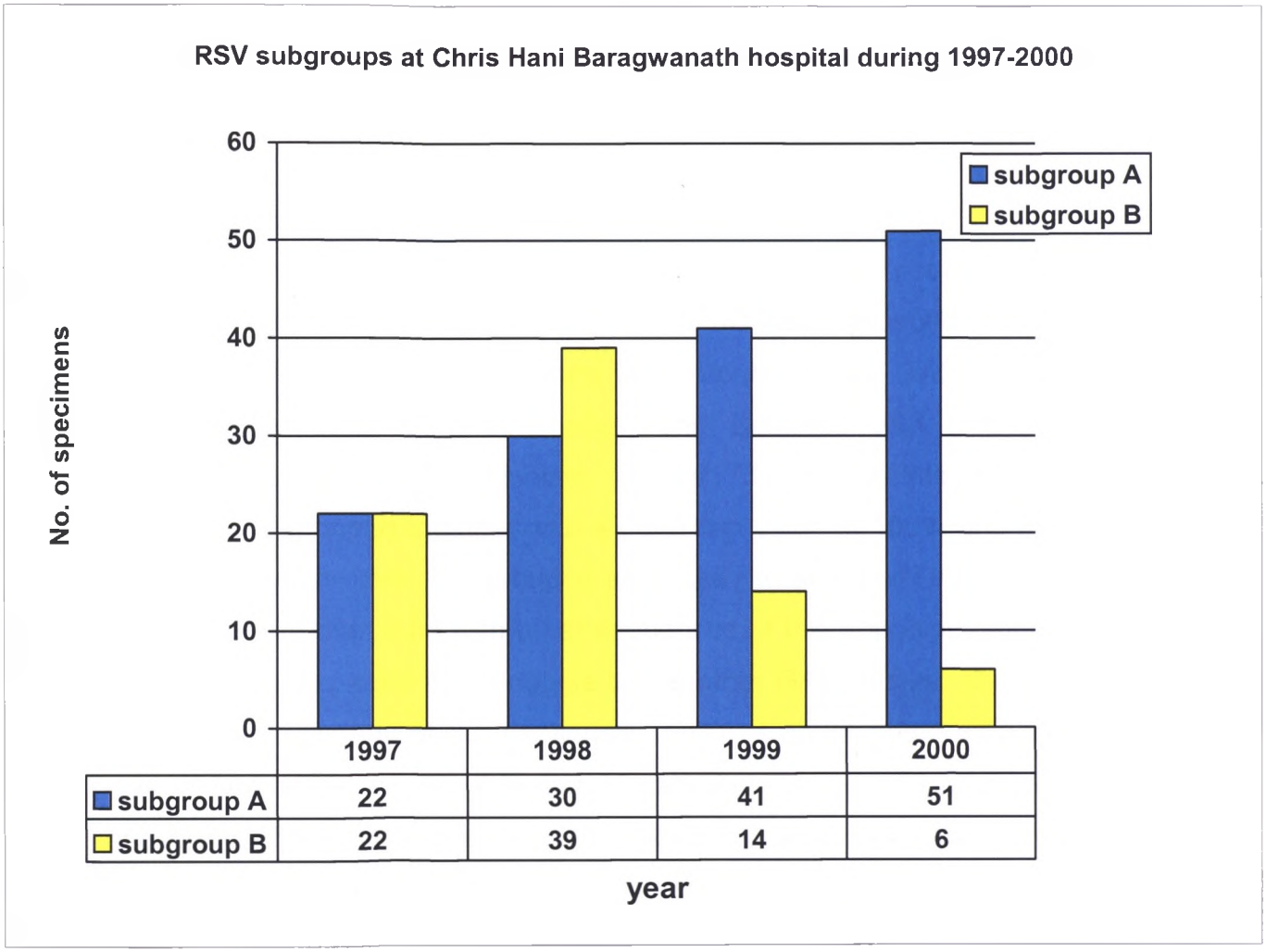


Figure 2.2: RSV subgroup circulation at Chris Hani Baragwanath hospital during 1997-2000. The numbers of isolates that were subgrouped as either A or B in each season are indicated in the Table.

*Chi square statistics determined over all four years and between individual years: (1997-2000: $P<0.0001$); (1997 vs 1998: $P=0.6$); (1998 vs 1999: $P=0.001$); (1999 vs 2000: $P=0.069$).

One sequence of each of two subgroup A and the two subgroup B clusters of sequences identified in Mozambique in 1999 were also included in the analysis (Roca *et al.*, 2001), since Mozambique is the closest geographical region to our study site where sequencing data are available.

Subgroup A:

South African subgroup A specimens clustered into 4 groups with bootstrap values of 70-100% (Figure 2.3.a). Three of these grouped with previously assigned genotypes, GA5, GA7 and GA2. South African GA2 viruses grouped separately to other GA2 sequences, and clustered with isolate WV12342 (Sullender *et al.*, 1998). Isolate SA98V173 and isolate Moz/9/99 from Mozambique grouped together with a bootstrap value of 100% and a P-distance of 0.007 to each other, but grouped separately to all other GA2 viruses with P-values of more than 0.04 with other sequences of this genotype, suggesting that it may form a separate subgenotype to the other GA2 viruses. The fourth group was distinct from any of the previously identified genotypes, and was therefore named SAA1 (South Africa A1). The Mozambique sequences clustered with GA2 and GA5. The P-distances determined between and within the genotypes are shown in Table 2.2.A. The average inter-genotypic P-distances between the South African subgroup A specimens ranged from 0.076 (SAA1 and GA7) to 0.098 (GA2 and GA5). P-distances between individual sequences within each genotype ranged from 0 to 0.059. The closest related sequence on the Genbank to the SAA1 sequences was isolate Sel91242 from Korea (1993) (Choi & Lee, 2000) which has P-values ranging between 0.026-0.059 relative to SAA1 specimens, however significant bootstrap statistics could not be obtained when it was included in the phylogenetic analysis. Three subgenotypes exist within GA5 with distances greater than 0.04 between individual branches. The South African GA5 specimens had P-values of more than 0.04 relative to the reference sequences except for SA97D804 and CN3114, which were identical.

Table 2.2: Average genetic P-distances between and within subgroup A and B genotypes. The average P-distance (The proportion of differences; i.e. the number of pair-wise nucleotide or amino acid differences divided by the total number of nucleotides or amino acids in the sequenced region, calculated with MEGA2 (Kumar *et al.*, 2001) for subgroup A (A.) and subgroup B (B.) calculated between the individual genotypes is shown below the diagonal in each table, and the P-distance within each genotype in bold on the diagonal. The same calculations are also shown if only South African sequences are taken into consideration.

A.

	GA1	GA2	GA3	GA4	GA5	GA6	GA7	SAA1
GA1	**0.039**							
GA2	0.144	**0.025**						
GA3	0.155	0.055	**0.033**					
GA4	0.181	0.119	0.112	**N.C.**				
GA5	0.167	0.090	0.085	0.121	**0.029**			
GA6	0.143	0.058	0.050	0.106	0.071	**0.019**		
GA7	0.146	0.068	0.060	0.105	0.080	0.053	**0.021**	
SAA1	0.146	0.070	0.070	0.107	0.077	0.052	0.067	**0.018**

South African clinical specimens:

	GA2	GA5	GA7	SAA1
GA2	**0.015**			
GA5	0.098	**0.030**		
GA7	0.081	0.090	**0.033**	
SAA1	0.074	0.081	0.076	**0.017**

B.

	GB1	GB2	GB3	GB4	SAB1	SAB2	SAB3	SWED&USA
GB1	**0.022**							
GB2	0.046	**n/c**						
GB3	0.056	0.043	**0.030**					
GB4	0.104	0.086	0.101	**0.027**				
SAB1	0.072	0.059	0.069	0.112	**0.053**			
SAB2	0.091	0.078	0.077	0.110	0.107	**0.037**		
SAB3	0.062	0.052	0.040	0.102	0.076	0.069	**0.010**	
SWED&USA	0.109	0.096	0.113	0.162	0.122	0.153	0.121	**0.020**

South African clinical specimens:

	GB3	GB4	SAB1	SAB2	SAB3
GB3	**0.038**				
GB4	0.116	**0.03**			
SAB1	0.073	0.119	**0.053**		
SAB2	0.083	0.132	0.119	**0.037**	
SAB3	0.046	0.111	0.076	0.069	**0.01**

Subgroup B:

South African subgroup B RSV specimens clustered into five groups (Figure 2.3.b), two of these with previously identified genotypes GB3 and GB4. However, as predicted by Peret *et al.* (2000) the current classification scheme would have to be expanded to accommodate all the South African clinical specimens. One group of sequences clearly grouped with a previously assigned genotype GB4. However, although some South African sequences clustered with the GB3 reference sequences, these could not be supported by significant bootstrap statistics. GB3 RSV specimens formed various subgenotypes with inter genotypic P-distances of up to 0.058, which may give rise to separate genotypes in future studies. The remaining three South African clusters were distinct to the previously assigned genotypes and were named SAB1, SAB2 and SAB3. Mozambican specimens grouped with GB3 and SAB2. The average inter-genotypic P-distances between South African subgroup B specimens ranged from 0.046 (GB3 and SAB3) to 0.132 (GB4 and SAB2) (Table 2.2.B). The SAB2 sequences (SA99V800 and SA99V1325) grouped with isolate MOZ198/99 (Mozambique, 1999) (Roca *et al.*, 2001) and Sel93366 (Korea, 1993) but not with any of the genotypes identified by Peret *et al.* (2000). SAB3 was most closely related to GB3 with P-distances ranging from 0.021-0.062 between individual sequences, but did not cluster with significant bootstrap values to any of the GB3 sequences. The closest related sequence to the SAB3 sequences on the Genbank was Mad11-92 from Madrid (Martinez *et al.*, 1999), which has P-distances of 0.021-0.03 with SAB3, however it clustered separately to the SAB3 specimens when included in the phylogenetic analysis. SAB2 has an average intergenotypic P-distance of 0.069 with SAB3. Distances between individual sequences in these groups of up to 0.145 and insignificant bootstrap statistics placed them in different genotypes. No sequences could be identified on the Genbank that clustered with the SAB1 specimens.

These results suggest that SAA1, SAB1, SAB3 and SAB2 are new genotypes of which the former three have so far only been identified in South Africa.

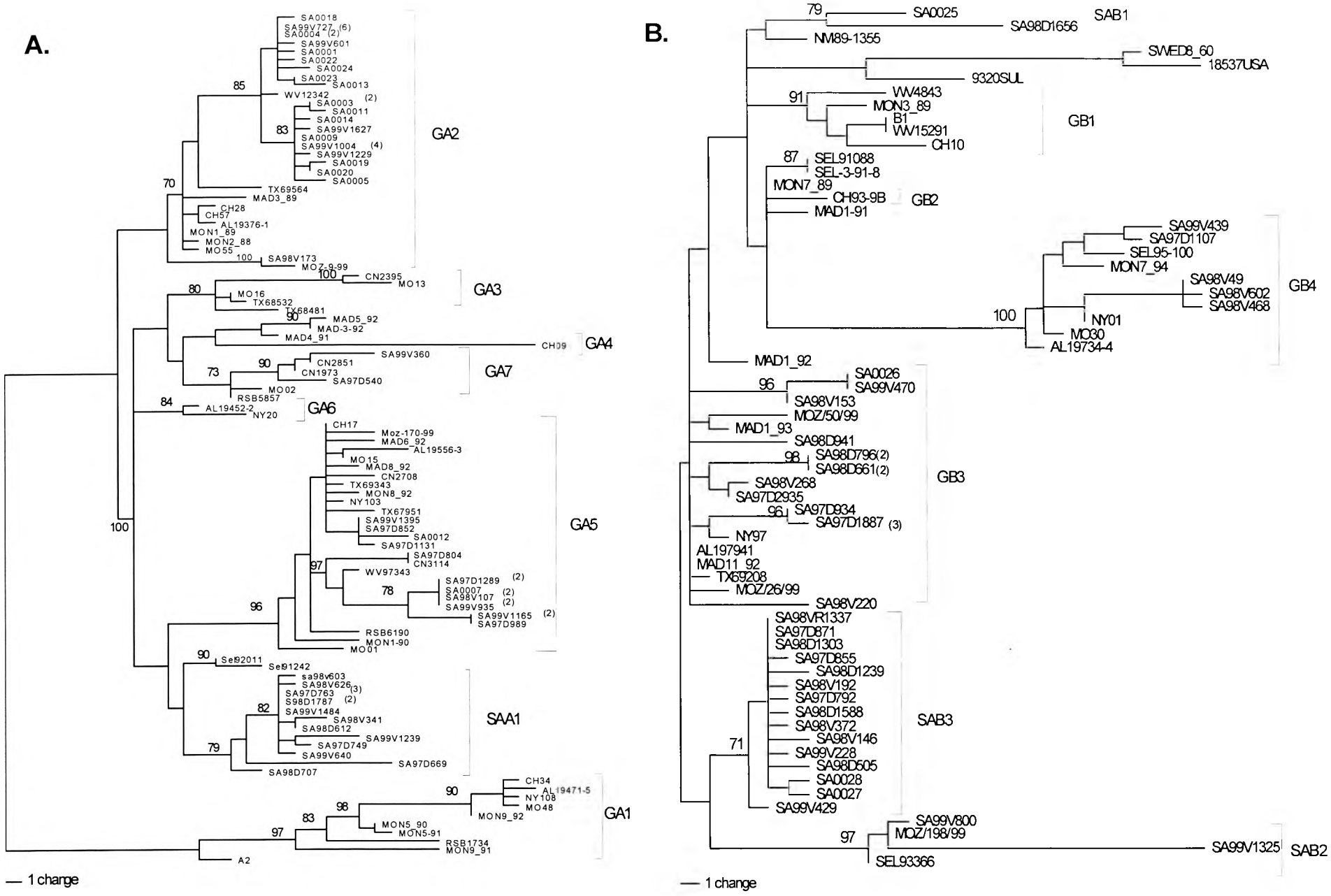


Figure 2.3: : Maximum Parsimony trees of South African RSV subgroup A and B nucleotide sequences.
See legend on next page.

Figure 2.3: Maximum Parsimony trees of South African RSV subgroup A and B nucleotide sequences.

The phylograms were the best trees under Kishino-Hasegawa test and were drawn with PAUP version 4.0b4a (Swoford DL, 1993). The trees were midpoint rooted using minimum F-value optimisation and drawn to scale with the bar indicating one change. Bootstrap probabilities obtained from a 50% majority rule consensus tree are shown at the branch nodes. Only values of >70 are indicated. For bootstrap evaluation under MP assumption, sequences were added randomly and one tree was held at each step (100 replicates and 100 bootstrap replicates) applying the tree bisection-reconnection branch-swapping algorithm. Where relevant, the number of identical sequences from one year is indicated in brackets. South African specimens (1997-2000) are indicated by SA97-SA00. Reference sequences for each genotype assigned by Peret *et al.* (1998; 2000) (GA1-GA7) and (GB1-4) were obtained from the Genbank: from the USA (NY, New York; Al, Alabama; MO, Missouri; Tx, Texas; CN, Canada) (Peret *et al.*, 2000); CH (Rochester, New York) (Peret *et al.*, 1998). Additional sequences from around the world were included in the comparison, by selecting representatives of distinct clusters found in previous studies and selecting isolates from the Genbank that gave the best hits in blast searches with each of the South African clusters: WV (West Virginia) (Sullender *et al.*, 1991); Birmingham, UK (RSB) (Cane *et al.*, 1994); Montivideo, Uruguay (MON) (Garcia *et al.*, 1994; Martinez et al. 1999); Madrid, Spain (MAD (Garcia *et al.*, 1994; Martinez *et al.*, 1999); (Sel) Seoul, Korea (Choi and Lee, 2000), Mozambique (MOZ) (Roca *et al.*, 2001). Prototype strains for subgroup A: Strain A2 (Australia) (Wertz *et al.*, 1985); subgroup B: Swed8-60 (Sweden) (Sullender *et al.*, 1991); USA (18537) (Johnson *et al.*, 1987); B1 (Karron *et al.*, 1997). The genotype assignment is indicated at the right by brackets.

South African RSV sequences also clustered with 3 previously assigned subgroup A and two subgroup B genotypes confirming the notion that RSV variants can spread worldwide but also that RSV evolution is dynamic with new variants appearing in individual communities (Cane & Pringle, 1992; Garcia *et al.*, 1994).

Comparison of the Mozambican specimens to South African specimens also suggest transfer between communities as all genotypes identified in Mozambique occurred in the same year in South Africa although at low frequencies. Neither of the prototype strains A2 nor B1, currently used for the production of experimental vaccines, clustered with any of the South African RSV specimens.

2.3.4 Amino acid sequence comparison

Sequence divergence was higher at the amino acid level than at nucleotide level, confirming previous findings (Cane & Pringle, 1995b; Martinez *et al.*, 1999; Sullender *et al.*, 1991). The predicted amino acid sequences of clinical specimens in the different genotypes were compared to prototype strains A2 and B1 for subgroup A and B, respectively, suggesting significant differences between the genotypes found in South Africa and the prototype strains (Figure 2.4). The translated amino acid sequences of **subgroup A** specimens had an overall P-distance of 0.113. Average intra-genotype variability within the South African specimens ranged from 0.017 for GA2 to 0.065 for GA5, while the average inter-genotype variability was 10 times as high ranging from 0.12 (GA2 and SAA1) to 0.21 (GA2 and GA5) and (GA5 and GA7). The average distance between A2 prototype strain and South African specimens ranged from 0.189 for SAA1 specimens to 0.269 for GA5 specimens. The overall distance between the amino acid sequences of South African **subgroup B** specimens was 0.101. Intra-genotype P-distances ranged from 0.053 for (GB4) to 0.114 for SAB1. The average P-distance between genotypes ranged from 0.133 (GB3 and SAB1) and 0.229 between GB4 and SAB2. The average p-values between South African specimens and prototype sequence B1 ranged from 0.125, between B1 and GB3 specimens, and 0.207 between B1 and GB4 specimens.

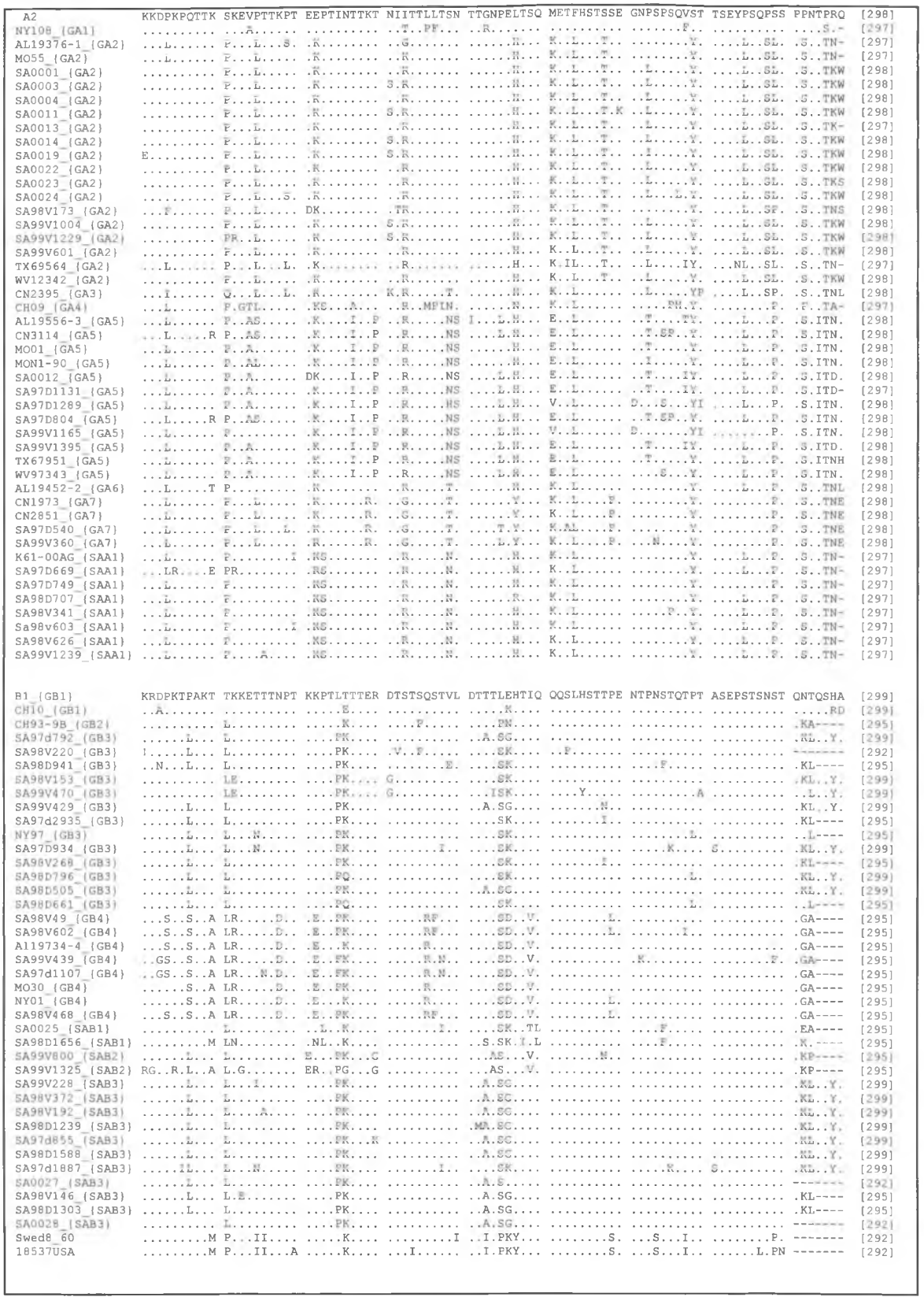


Figure 2.4: Amino acid alignment of the second variable region of the G-protein for subgroup A (A.) and subgroup B (B.) South African isolates and reference sequences of the different genotypes. Alignments are shown relative to prototype sequences A2 and B1, respectively.

Genotype-specific amino acid substitutions could be identified, a phenomenon also observed when the genotypes were first identified (Peret *et al.*, 1998). Figure 2.5 illustrates the genotype-specific mutations by comparing the different genotypes to a consensus sequence of all specimens. Certain amino acid substitutions that may result in changes in protein secondary structure between the genotypes are visible when comparing the profile of hydrophobic (red), hydrophilic (blue), and special amino acids (Cysteine, Glycine and Proline) (yellow), which are involved in the secondary protein structure.

The G-protein of South African specimens (relative to prototype strains A2 and B1) consisted of 297- (SAA1), 298- (GA7) or both (GA2 and GA5) amino acids for subgroup A, and 295- (GB4, SAB1 and SAB2) or 292-, 295-, and 299- (GB3 and SAB3) amino acids for subgroup B. Subgroup A specimens used the UAG stop codon in either nucleotide position 259 or 262. Subgroup B specimens used either the UAA or the UAG stop codons in positions 241, 250, or 262. Changes in stop codon usage are thought to be associated with antigenic variation in RSV escape mutants that recognise strain-specific epitopes (Melero *et al.*, 1997).

2.3.5 Other estimates of selection

The mean nucleotide P-distance was 0.06 and 0.07 for subgroup A and B, respectively, while the mean amino acid P-distance was 0.113 and 0.10 for South African subgroup A and B specimens. A respective transition to transversion ratio of 5.16 and 6.9 was calculated for subgroup A and B specimens. The mean proportion of nucleotide substitutions that resulted in amino acid changes among South African specimens was 70.8% for subgroup A and 67% for subgroup B. When comparing the South African specimens in Subgroup A to the prototype strain A2 and the South African subgroup B specimens to prototype B1, a mean Ka/Ks ratio of 1.5 and 1.14 were calculated for subgroup A and B respectively, where Ka is the non-synonymous substitutions per non-synonymous sites and Ks is the synonymous substitutions per synonymous sites. When the individual genotypes were compared to each other and to the prototype strains A2 (subgroup A) and B1 (subgroup B) several Ka/Ks ratios of < 1 could be identified, which suggests positive selection (Table 2.3). Substitutions were the only type of mutations detected.

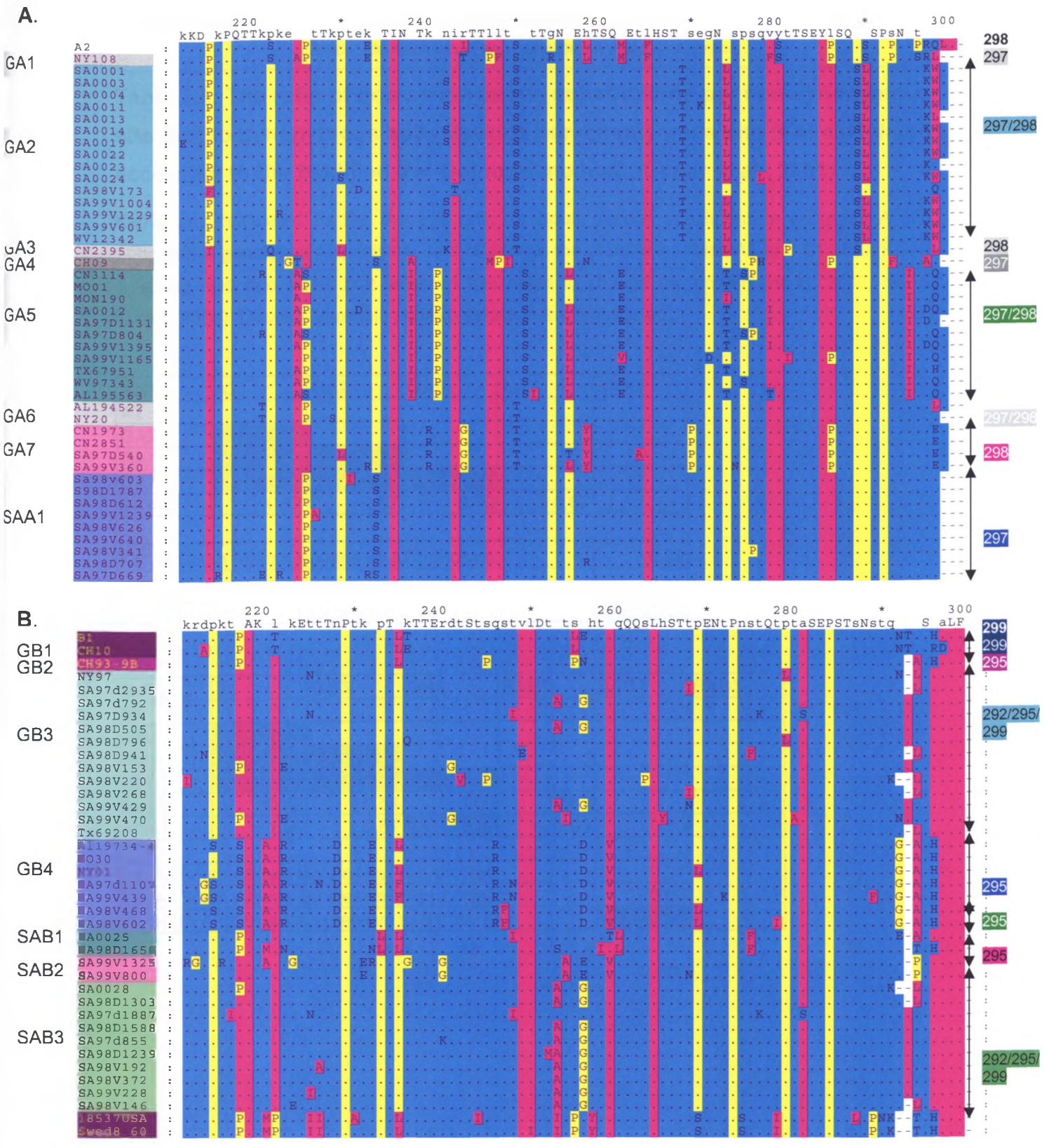


Figure 2.5: Identification of genotype-specific mutations and mutations that may influence the secondary structure of the G-protein. The amino acid alignments of subgroup A (A.) and B (B.) genotypes are compared to consensus sequences shown in the top rows to identify genotype-specific mutations. Amino acid changes that may result in changes in the secondary structure were defined by changes in hydrophobic (red); hydrophilic (blue) or special amino acids involved in secondary folding Cysteine, Glycine and Proline (yellow) (Lodish *et al.*, 1995); sequence profile. The different amino acid sequences lengths in each genotype are shown at the right of the sequence.

Table 2.3: The ratio of non-synonymous substitutions per non-synonymous site (Ka) to synonymous substitutions per synonymous site (Ks) calculated according to the Modified Nei-Gojobori method (Kumar *et al.*, 2001).

Ka/Ks was calculated as the average ratio for South African specimens in subgroup A and Subgroup B genotypes respectively, and relative to the prototype strains A2 and B1. If Ka/Ks >> 1, strong evidence exists that selection has acted to change the protein (Hurst, 2002; Nei & Kumar, 2000). Cases of possible positive selection are highlighted.

A.

Ka/Ks	A2	GA2	GA5	GA7	SAA1
A2					
GA2	1.57				
GA5	1.56	2.20			
GA7	1.44	1.80	2.20		
SAA1	0.77	0.67	0.70	0.67	
Ks: All South African isolates vs A2 prototype:					0.08
Ka: All South African isolates vs A2 prototype:					0.12
All South African isolates relative to A2: Ka/Ks					1.50

B.

Ka/Ks	B1	GB3	GB4	SAB1	SAB2	SAB3
B1						
GB3	1.14					
GB4	1.50	1.18				
SAB1	1.33	1.14	1.30			
SAB2	0.82	0.54	1.00	0.83		
SAB3	1.00	0.67	1.09	1.00	0.55	
Ks: All South African isolates vs B1 prototype:					0.07	
Ka: All South African isolates vs B1 prototype:					0.08	
All South African isolates relative to B1: Ka/Ks					1.14	

2.3.6 Genotype distribution pattern over four years in a single community in South Africa

Comparison of the genotypes circulating within each season revealed a rather complex pattern. Different genotypes co-circulated each year, with certain genotypes becoming dominant and then declining before being replaced with a different dominant genotype (Figure 2.6.). Some genotypes were found throughout the 4 years of study (GA5, GB3 and SAB3), while others were only found at a low frequency in some seasons (GA7, SAB1 and SAB2). Different circulation patterns were identified for subgroup A and B genotypes. More diversity was found among subgroup A specimens during the 4 year study period than in subgroup B, supporting findings by Choi & Lee (2000). Subgroup A showed a gradual build-up and then replacement of dominant genotypes, e.g. GA5 (1997) was replaced by SAA1 (1998), which was then replaced with GA2 (1999). GA2 predominated for more than one season increasing from 42% (1999) to 78% (2000). Subgroup B was more consistent, and two subgroup B genotypes (GB3 and SAB3) remained co-dominant during all four years, constituting 78% of subgroup B specimens over the study period. Previous studies on subgroup A isolates have reported that a new genotype becomes dominant each year (Peret *et al.*, 2000; Cane *et al.*, 1994). Although this was true for the first 3 years of our study, the same genotype, GA2, remained dominant during the last two years (1999 and 2000). This possibly suggests that successful variants may be able to persist and remain dominant for more than one season.

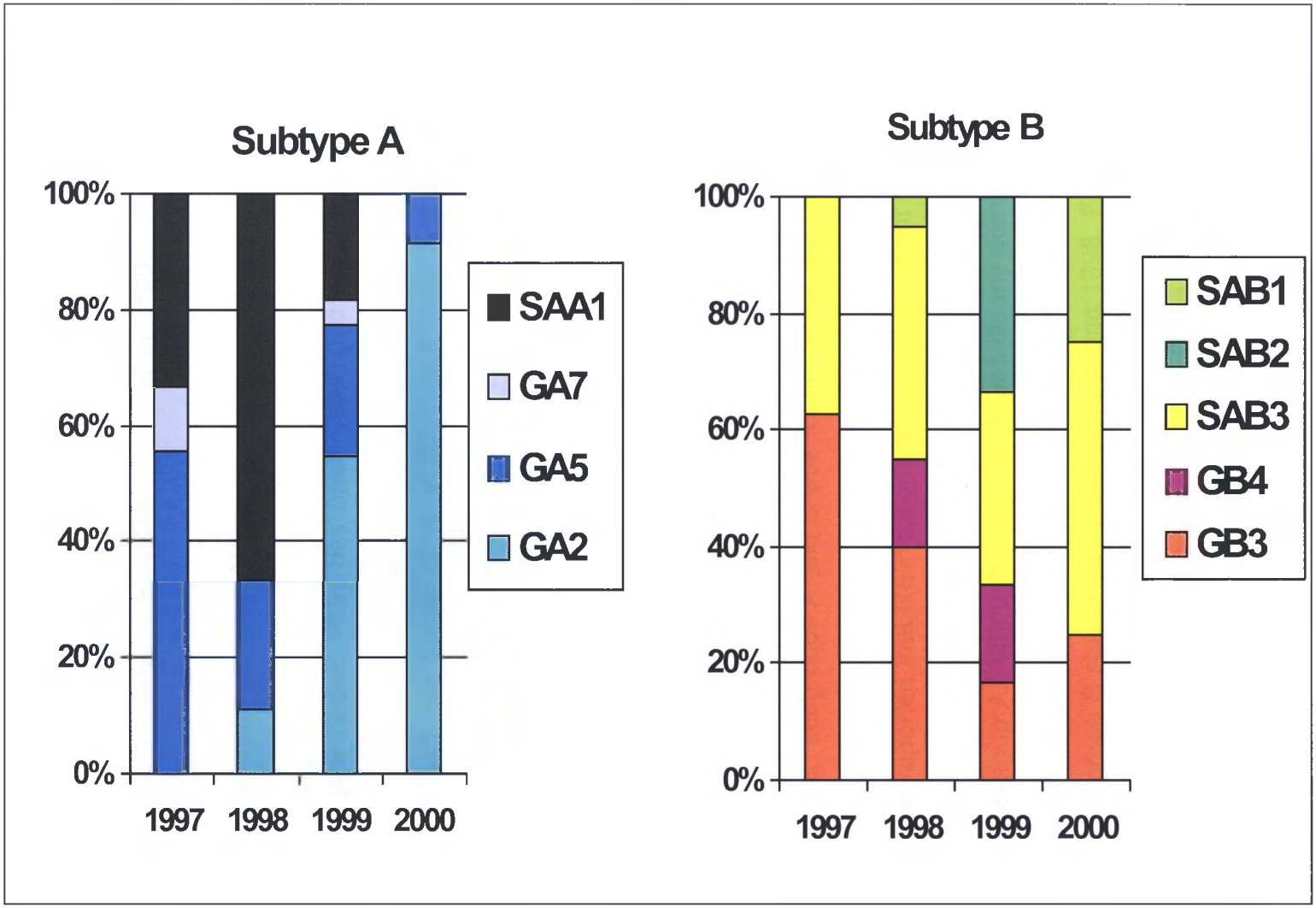


Figure 2.6: Genotype distribution over four seasons (1997-2000) at Chris Hani Baragwanath hospital.

2.4 Discussion

It is thought that the antigenic and genetic variation found both between and within the two RSV subgroups might contribute to its ability to cause repeat infections throughout life. Limited data are available about the molecular epidemiology of RSV in developing countries (Cane *et al.*, 1999; Weber *et al.*, 1998). Until now no information was available about the evolution of RSV over consecutive epidemics in sub-Saharan Africa.

In this study a multiplex nested PCR has been developed for subgrouping of clinical specimens of RSV into subgroups A and B since nasopharyngeal aspirates often contained low levels of RSV RNA and could not always be detected by a single RT-PCR. The nested PCR allowed rapid typing of clinical specimens without the need for culturing of the virus. Sufficient good quality DNA could be amplified to use for nucleotide sequencing. Since the PCR products include the second variable region of the G-protein, successfully amplified subgroup A or B genes can be sequenced directly avoiding the potential inclusion of mutations that could result during viral culture.

Co-circulation of the two subgroups A and B was found, which is consistent with previous reports (Hendry *et al.*, 1986; Cane *et al.*, 1992; Choi & Lee, 2000). However, It has been reported that RSV subgroup A is predominant in three out of four seasons in other geographical areas (Hall *et al.*, 1990), suggesting a more important role for subgroup A. Although subgroup A was clearly dominant in two of the seasons studied, subgroup A and B co-circulated equally during 1997, and a slight excess of subgroup B specimens were found during 1998. Furthermore, subsequent studies not included in this analysis showed that during 2001 there was a switch back to subgroup B with 28 out of 30 typed specimens from Chris Hani Baragwanath hospital belonging to subgroup B. This suggests an important role for both subgroups A and B in South African epidemics.

RSV outbreaks can vary in the pattern of group and subgroup isolations between communities during the same year and between years for the same community,

indicating that RSV outbreaks are community- and possibly region-based, but not national (Anderson *et al.*, 1991). Few studies have so far analysed the genetic variability of RSV specimens of both subgroups during consecutive epidemics in the same region (Choi & Lee, 2000; Peret *et al.*, 1998). It has been reported that multiple RSV lineages circulate in single epidemics and that different lineages predominate in each epidemic (Cane *et al.*, 1994; Peret *et al.*, 1998; Choi & Lee, 2000; Coggins *et al.*, 1998).

In order to place this study in context with previous studies, a published genotyping method (Peret *et al.*, 2000; Peret *et al.*, 1998) was used to assign the South African specimens to genotypes. Closer analysis of phylogenetic clusters formed by the South African specimens and comparison to representative sequences of the published genotypes revealed that sequences in each individual genotype in both subgroups A and B had P-distances of less than 0.07. Therefore, the criteria for the assignment of new genotypes were set to include relative P-distances of less than 0.07 and bootstrap statistics of greater than 70% to other members of the cluster. A new subgroup A genotype (SAA1), and two new subgroup B genotypes, (SAB1 and SAB2) were identified this way. SAB2 grouped with a recently identified cluster of specimens from Mozambique (Roca *et al.*, 2001). However, no published sequences could be identified that clustered with SAA1 and SAB2, suggesting these are new genotypes that have so far only been identified in South Africa. The use of reference sequences was found useful for assigning the South African specimens to genotypes, but it is suggested that a standardised method for genotyping RSV would allow more effective comparison of epidemics occurring worldwide.

South African RSV specimens also clustered with 3 previously assigned subgroup A and two subgroup B genotypes. This confirms the notion that RSV variants can spread worldwide, but also that RSV evolution is dynamic and that new variants may appear in individual communities (Cane & Pringle, 1992; Garcia *et al.*, 1994).

Sufficient subgroup A and B clinical specimens were obtained in order to be able to assess the molecular epidemiology of both subgroups. Multiple genotypes co-circulated for both subgroups, however subgroups A and B revealed different circulation patterns. Subgroup A genotypes were more variable between seasons than subgroup B. Different subgroup A genotypes became dominant, increasing in numbers before declining and eventually disappearing, showing a clear displacement of the dominant genotype, while two subgroup B genotypes remained co-dominant for all four years. This is consistent with a recent study (Choi & Lee, 2000) conducted over 9 years in Korea. It has been suggested that this higher variability may be advantageous to subgroup A and could contribute to its higher worldwide predominance (Coggins *et al.*, 1998; Hall *et al.*, 1990). Previous studies on subgroup A isolates have reported that a new genotype becomes dominant each year (Cane *et al.*, 1994; Peret *et al.*, 2000). Although this was true for the first 3 years of our study, the same genotype, GA2, remained dominant during the last two years (1999 and 2000). GA2 showed a gradual increase in occurrence from 1998 to 2000 until 78% of all clinical specimens belonged to this genotype. This suggests that successful variants may be able to persist and remain dominant for more than one season. Increased levels of herd immunity may cause the displacement of genotypes over consecutive epidemics (Choi & Lee, 2000). This argument is strengthened by subsequent typing data from our laboratory that showed a return to subgroup B predominance at Chris Hani Baragwanath hospital during 2001. Most other studies had only limited data about subgroup B strains, making it difficult to identify epidemiological patterns among subgroup B specimens.

Sequence analysis revealed a high percentage of nucleotide changes that result in amino acid changes and a high ratio of transitions to transversion and non-synonymous substitutions relative to synonymous substitutions. These are all indicators of positive selection, which may be a result of immunological pressure. It has furthermore been shown that changes in stop codon usage are associated with important antigenic variation found in RSV escape mutants selected with monoclonal antibodies that recognise strain-specific epitopes (Melero *et al.*, 1997). Two different protein lengths were identified among group A and three

among group B South African specimens. Genotype-specific amino acid mutations were also identified that may result in changes in the secondary structure. Collectively this is suggestive of immune evasion by RSV, possibly accomplished through epitope changes. Evidence for progressive accumulation of amino acid changes on the G-protein and for antigenic drift due to immune pressure have been reported. Cane and Pringle have shown, using monoclonal antibodies, that the changes in nucleotide and amino acids of the G-protein are reflected by the antigenic profiles of specimens which suggests antigenic drift over time (Cane & Pringle, 1995c). Although the significance of this for the development of effective vaccines remains to be determined, immune driven selection may necessitate the use of more recently isolated strains from different areas for vaccine production. Regular updating of vaccine strains may also be required. However, amino acids substitution occurred sporadically and in specific regions, which appear to be "hotspots" where mutations are accepted. This is consistent with previous studies and may have implications for synthetic vaccine design (Peret *et al.*, 1998; Roca *et al.*, 2001).

In summary, this study has described a method for amplifying clinical specimens for use in phylogenetic analysis, and has reported on the first molecular epidemiological analysis of RSV over consecutive seasons in southern Africa. Three new genotypes were identified in addition to known genotypes identified in the the rest of the world. The high percentages of nucleotide changes that resulted in amino acid changes suggest that a strong selective pressure is evident in both RSV subgroups. Evidence of different circulation patterns between subgroup A and B viruses over consecutive seasons in the same community was found. Subgroup A revealed more variability and a displacement of genotypes while subgroup B remained more consistent. These findings provide additional information regarding RSV evolution and strain circulation worldwide and may contribute to future vaccine development.

CHAPTER 3

Molecular epidemiological analysis of community circulating RSV in rural South Africa: comparison of viruses and genotypes causing different disease manifestations

3.1 INTRODUCTION

To determine the burden of RSV disease in developing countries, the WHO has initiated epidemiological studies in rural communities in Nigeria, South Africa, Mozambique and Indonesia (WHO Department of Vaccines and Biologicals (V&B), 2001), and compiled a generic protocol to examine the incidence of lower respiratory infection due to RSV in children less than five years of age (Wright & Cutts, 2000). This will provide valuable baseline information for future immunisation programs. In conjunction with this effort, molecular epidemiological data about strains circulating developing countries would be important for vaccine development and to ensure relevant strains are used in this process. Most molecular epidemiological studies have until now been focussed on specimens from hospitalised infants in industrialised countries. Limited information is available about the molecular epidemiology of RSV in developing countries (Cane *et al.*, 1999; Roca *et al.*, 2001). In Chapter 2 the first study about RSV molecular epidemiology over consecutive seasons in a single hospital in South Africa is described (Venter *et al.*, 2001). However, limited data have so far been accumulated about community circulating RSV, especially from Africa.

It is still unclear if any correlation exists between community circulating RSV that results in mild infections and RSV strains that can cause severe disease that

require hospitalisation. Several studies have attempted to correlate differences in clinical severity of infections with the different groups or genotypes of RSV (Cane, 2001; Walsh *et al.*, 1997). Some studies showed no difference between subgroup A and B, while others reported that either subgroup A or B caused more severe disease (Brandenburg *et al.*, 2000b; Fletcher *et al.*, 1997; Hornsleth *et al.*, 1998; Walsh *et al.*, 1997). Certain studies have suggested that specific genotypes may result in more severe disease (Fletcher *et al.*, 1997; Hornsleth *et al.*, 1998). However, most of these studies focussed on hospitalised infants, which is problematic when defining clinical severity, since only severely ill babies are considered while those with mild disease are not examined because they are unlikely to be hospitalised (Cane, 2001).

In this chapter, a molecular epidemiological analysis of community-acquired RSV clinical specimens obtained from infants and children attending rural primary healthcare clinics in the Agincourt district of the Limpopo Province of South Africa is described. The Agincourt site was one of the WHO RSV surveillance sites and had a remote rural location (600 km north-east of Johannesburg) with limited infrastructure and only 13 % of the population has an adequate water supply. The study site was located on the border of Mozambique only 150 km from the WHO study site in Manhiça, Mozambique. Although these two sites are located in close proximity to each other, their climates are quite different, which makes them interesting to compare. The rainy season in Manhiça coincides with the warm season (mean temperature 30°C) and lasts from October through to May with an average of 1,000 mm rain. The cool dry season is from June through September. Agincourt is much drier than Manhiça, with a semi-arid climate and is located at a higher altitude. The cool dry season is from April to September, while summer is from October to March with temperatures ranging from 16-28 °C. The rainy season is from November to March (rainfall ranges from 500 mm in the east to 1,200 mm in the west). In Agincourt cases of RSV occurred mainly in the dry cool season, (February- August 1999; March-September 2000) while in Mozambique RSV occurred mostly in the rainy season (February-May 1999; January-April 2000) (Report of a workshop on the epidemiology of respiratory syncytial virus (RSV) in developing countries, Johannesburg 2000 (WHO Department of Vaccines and Biologicals (V&B), 2001; unpublished data from the WHO RSV

study group). Because of the remote location and limited infrastructure of the Agincourt site the specimens were not of ideal quality for use in RT-PCR and the aim of the WHO study was therefore only to determine the incidence of RSV by rapid ELISA. However, using the multiplex nested PCR described in Chapter 2 a selection of these specimens could also be analysed at the molecular level as will be described in this chapter. These specimens were sequenced and used in phylogenetic analysis to compare to genotypes previously found in severely sick hospitalised infants. The distribution of genotypes between Mozambique, Agincourt and Johannesburg were also compared during one epidemiological season to determine if cross-border transfer occurs. Information about the correlation between RSV genotypes found in hospitalised infants and community-circulating RSV will be beneficial to future immunisation programs, and will also help to answer questions about RSV pathogenicity.

3.2 MATERIALS AND METHODS

3.2.1 Patients, specimens and viruses

The study site is described in detail in Tollman *et al.* (1999). Nasopharyngeal aspirates were collected from November 1998-December 2000 according to WHO recommendations (Wright & Cutts, 2000) from children aged 0-59 months attending primary health care clinics in the Agincourt district of Bushbuckridge region, Limpopo Province, South Africa, located 500 km north-east of Johannesburg on the border with Mozambique. This is a remote area of South Africa and considered rural in terms of lack of infrastructure and being a substantial distance from any tar road (Tollman *et al.*, 1999). The HIV sero-prevalence in the Bushbuckridge area is estimated to be 19.4% as indicated by the 1999 South African National antenatal survey (Department of Health Report, 2000).

For the purpose of this study, the WHO criteria were followed to define RSV disease severity (Wright & Cutts, 2000). In short, the disease was classified as upper respiratory tract infection if the child had a cough or difficulty breathing but with no chest in-drawing or fast breathing; lower respiratory tract infection was defined if the child had fast breathing or chest in-drawing or stridor or wheezing or apnoea. Severe pneumonia was identified if the lower respiratory tract infection symptoms included chest in-drawing or stridor.

Nasopharyngeal aspirates were collected by nurses at the clinics, and stored in a clinic-based domestic fridge until collection twice a week by the RSV study coordinator. Specimens were transported on ice to the central RSV collection point in Agincourt. During 1999 the specimens were tested on site using the rapid ELISA test (ABBOTT Laboratories, North Chicago, USA), and the RSV-positives sent to the National Institute for Communicable Diseases in Johannesburg for confirmatory testing. In 2000 and 2001 the Agincourt site was provided with a liquid nitrogen tank. Specimens were snap frozen and twice a month transported on dry ice to the National Institute for Communicable Diseases where the RSV rapid ELISA (ABBOTT Laboratories, North Chicago, USA) was carried out.

Specimens were again snap-frozen and stored at -80°C for later use in molecular analysis.

3.2.2 RNA extraction

RNA isolation from nasopharyngeal aspirates was carried out as described in Chapter 2 using the Viral RNA™ kit (Roche molecular biochemicals, Mannheim, Germany) according to the manufacturer's recommendations.

3.2.3 Subgroup identification: RT-PCR, multiplex nested PCR, and Nucleotide sequencing

Subgroup identification, RT-PCR, nested PCR and nucleotide sequencing have been described in Chapter 2.

3.2.4 Phylogenetic analysis

Phylogenetic analyses were carried out as described in Chapter 2, using MEGA version (2.3) (Kumar *et al.*, 2000). Subgroup A and B clinical specimens were genotyped by phylogenetic comparisons to sequences that had previously been assigned to specific genotypes (Peret *et al.*, 2000; Venter, *et al.*, 2001; Peret *et al.*, 1998). Amino acid analysis was conducted with MEGA version (2.3) (Kumar *et al.*, 2000) and Genedoc version 2.6001 (www.psc.edu/biomed/genedoc). The positions of N-glycosylation sites were predicted with Prosite, (http://ca.expasy.org/cgi-bin/scanprosite). The position of mucin type O-glycosylated residues was predicted with NetOglyc 5.0 (Gupta *et al.*, 1999). Alignments were done with Clustal X 1.64b and shading with Genedoc 2.6001. (www.psc.edu/biomed/genedoc).

3.3 RESULTS

3.3.1 Subtyping and phylogenetic analysis of the Agincourt specimens

A total of 26 RSV-positive specimens from Agincourt (December 1998-December 2001) could be amplified by nested RT-PCR (17 subtype A and 9 subtype B). These included 12 subtype A and 4 subtype B from 2000, 4 subtype A and 3 subtype B from 1999 and 1 subtype A and 2 subtype B from 1998. These specimens were sequenced and compared by phylogenetic analysis to RSV clinical specimens from infants of 0-60 months hospitalised for acute lower respiratory tract infections at Chris Hani Baragwanath hospital in Soweto, South Africa, from 1997-2000 (Chapter 2) (Venter *et al.*, 2001); and RSV isolated in 1999 at the WHO study site in Manhiça in Mozambique approximately 150 km from the Agincourt site (Roca *et al.*, 2001) (Figure 3.1). Of the community-acquired subtype A RSV specimens, 13 clustered with GA2 (1998, 1999 and 2000) and 3 with SAA1, with P-distances ranging from 0.00-0.04 compared to Johannesburg specimens from the same genotypes. GA2 was found at low levels in 1998 in Johannesburg, but constituted 42% of specimens in 1999 and 78% in 2000. SAA1 was not found in Johannesburg in 2000. RSV isolated in Mozambique clustered with GA5 and with isolate SA98V173, which grouped separately from all other GA2 specimens. Moz-9-99 and SA98V173 only contained 2 nucleotide differences. Of the community-acquired subgroup B specimens, 2 clustered with SAB3 (1998 and 1999) and 6 with GB3 (1999 and 2000) with P-distances of 0.04-0.07 compared to Johannesburg specimens in the same genotype. Some GB3 Agincourt specimens from the same year clustered together, however all specimens from different years clustered with different subgenotypes. Mozambican specimens from 1999 clustered with GB3 and SAB2. SAB2 specimens SA99V800 and Moz/198/99 were identical suggesting a cross-border transfer.

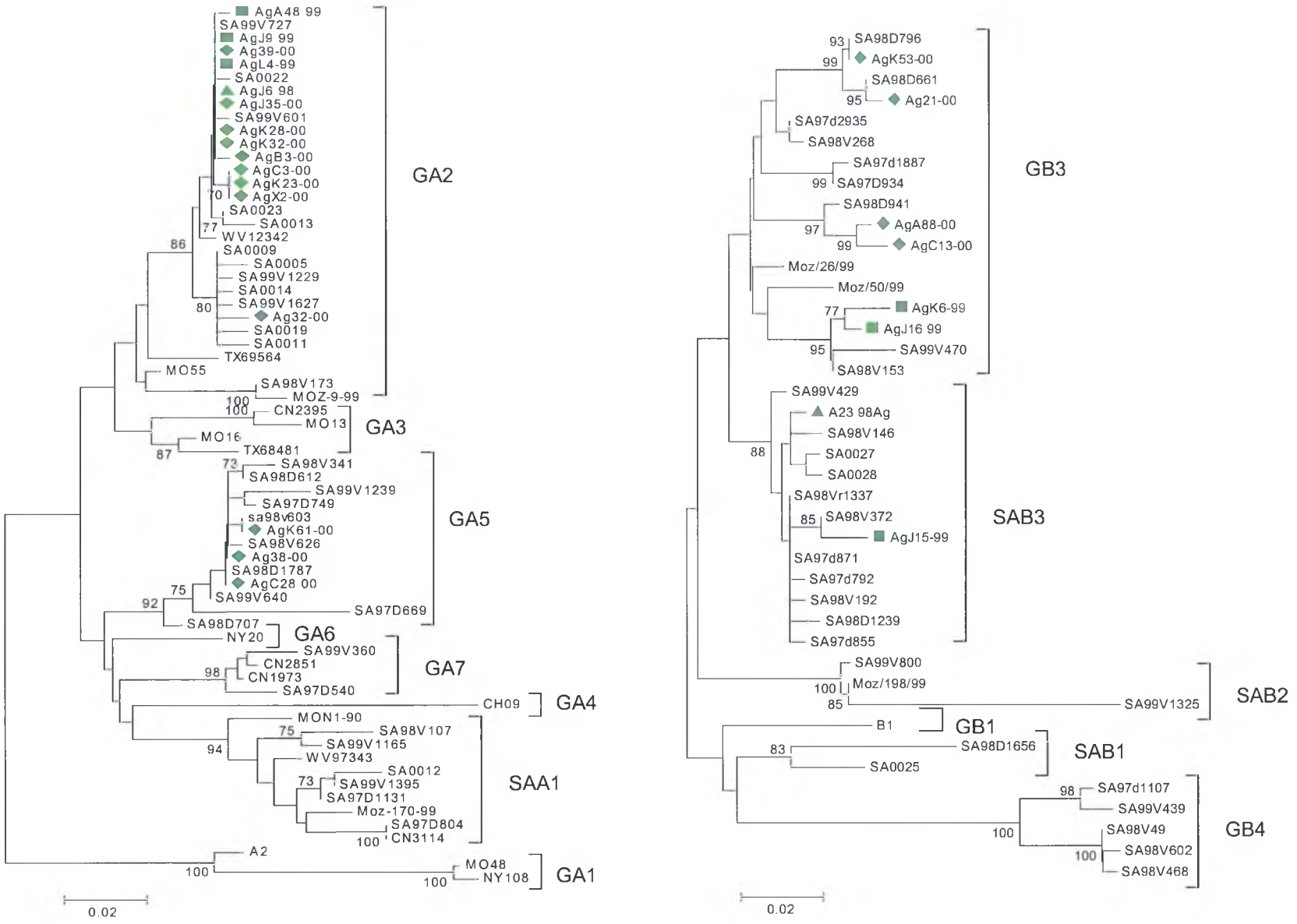


Figure 3.1: Phylogenetic analysis of South African community and hospital Subgroup A and B RSV isolates from Agincourt and Johannesburg. (see legend on following page)

Figure 3.1: Phylogenetic trees of South African RSV Subtype A and B Specimens.

The Neighbour-Joining Trees were computed with Mega version 2.1 using the Nucleotide Kimura 2-parameter. Trees were midpoint rooted and drawn to scale with the bar indicating 0.02 substitutions per site. Bootstrap probabilities for 100 iterations are shown at the branch nodes. Only values of 70-100 are indicated. Clincal specimens from the Agincourt region are indicated by Ag98-Ag00 (1998-2000) and marked with a green bullet, using a ♦ for 2000; ■ for 1999; ▲ for 1998. Clincal specimens from Soweto, South Africa are indicated by SA97-SA00 (Venter *et al.*, 2001). Reference sequences for each genotype assigned by Peret (GA1-GA7) and (GB1-4) are included. Reference sequences were obtained from Genbank: from the USA (NY, New York; Al, Alabama; MO, Missouri; Tx, Texas; CN, Canada (Peret *et al.*, 2000); CH (Rochester, New York) (Peret *et al.*, 1998); WV (West Virginia) (Sullender *et al.*, 1991); Uruguay (MON) (Garcia *et al.*, 1994); Spain (MAD) (Garcia *et al.*, 1994); prototype strains for type A: strain A2 (Australia) (Wertz *et al.*, 1985); type B: Swed8-60 (Sweden) (Sullender *et al.*, 1991); USA (18537) (Johnson *et al.*, 1987a); B1 (Karron *et al.*, 1997b). Clincal specimens from each cluster identified in Mozambique (MOZ) in 1999 are included (Roca *et al.*, 2001). The genotype assignment is indicated at the right by brackets.

3.3.2 Amino acid analysis

The mean amino acid P-distance was 0.05 (subgroup A) and 0.08 (subgroup B) between the Agincourt specimens and 0.1 and 0.11 between all South African subgroup A and B specimens, respectively. The amino acid alignments of subgroup A and B Agincourt specimens with South African and Mozambican specimens are shown in Figure 3.2. Four putative N-glycosylation sites were identified for subgroup A and B, respectively. The positions of three of these were conserved for all subgroup A sequences, although the asparagine position of the second site was moved one position for GA5 and SAA1. Two of the sites were conserved in all but two subgroup B specimens. The positions of the motifs K-P-X, and T-T-K-X identified by Cane (1991) are indicated. These motifs were found to be repetitive throughout the G-protein, and although the function is unknown they may be associated with the extensive O-glycosylation of the G-protein (Cane, 1991). For subgroup A, two of these sites (site 4 and 5) were genotype-specific, and were only found in GA2 and GA5, respectively. Four of these motifs were found in subgroup B, but none were conserved in all sequences. The positions of mucin type O-glycosylated serine and threonine residues were predicted with NetOglyc 5.0 (Gupta *et al.*, 1999). A high serine and threonine content resulted in extensive O-glycosylation. Although certain sites were isolate-specific, genotype-specific patterns could be identified and some sites that were conserved in all specimens. The positions of O-glycosylated sites that were conserved in all specimens are indicated below the alignment in Figure 2. In Subgroup A, of 32 putative O-glycosylation sites, 5 were conserved in all sequences, and 11 were genotype-specific. In subgroup B, 30 putative O-glycosylation sites were identified, of which 8 were conserved in all sequences and 9 in specific genotypes. Out of 22 threonine residues in subgroup A, 15 were conserved in all sequences, while 10 out of 28 subgroup B threonine residues were conserved. No significant difference could be found between community specimens and hospital specimens from the same genotype, except for one SAB3 community isolate (AgJ15-99) which had a premature stop codon in position 262. When comparing the differences in Agincourt sequences in the same genotype isolated in different years, an observation was made that the amino acid changes often resulted in changes in the secondary structure, e.g. in

genotype GA2, Agincourt specimens of 1999 all had the following pattern: 227(L); 242(T); 275(L), while specimens from 2000 of genotype GA2 had the following changes to the pattern: 227(P); 242(T); 275 (L) or 227(L);242 (A); 275(L); or 227(P); 242(T); 275(L);or 227(L); 242(T); 275(P).

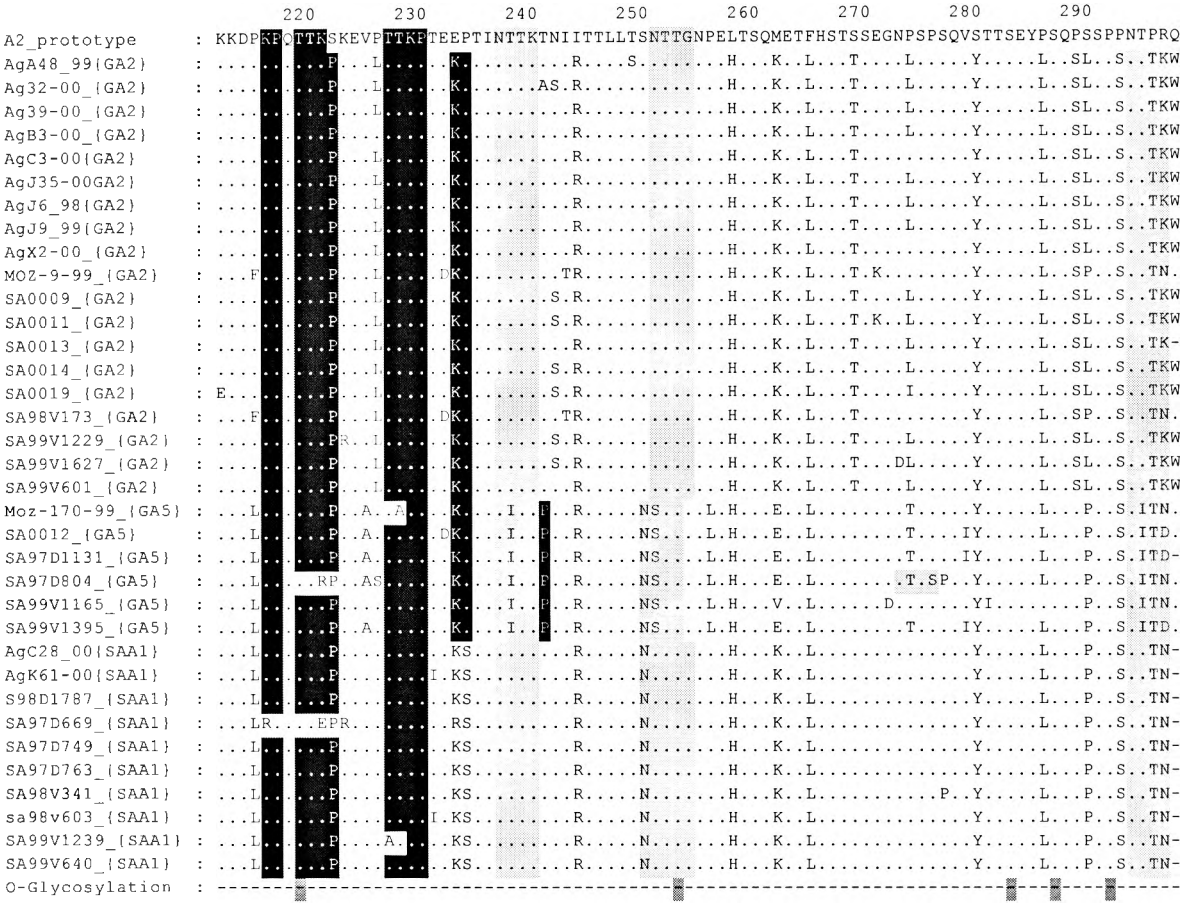
Disease severity and infecting genotype

The breakdown of age group, disease severity and gender of the infected infants and the genotype of the infecting strain identified in the rural community is shown in Table 3.1. Both subgroup A and B resulted in lower respiratory tract infections (LRTI). Different GA2 specimens that had identical G-protein sequences resulted in upper respiratory tract infections (URTI), lower respiratory tract infections and severe RSV. Severity of disease was defined according to the criteria stipulated by the WHO generic protocol (Wright *et al.*, 2000). Certain GA2 and GB3 specimens from hospitalised infants in Soweto, which had caused severe RSV, had identical sequences to specimens found in the community that only resulted in an upper respiratory tract infection. Within the sample that was genotyped, more male patients were affected than females within both RSV subgroups.

Figure 3.2: Comparison of the G-protein amino acid sequences of community and hospital specimens of RSV subgroup A (A.) and subgroup B (B.) genotypes

Alignments are shown relative to prototype sequences A2 and B1, respectively. The positions of N-glycosylation sites are indicated by grey blocks for subgroup A and B respectively (predicted by Prosite, http://ca.expasy.org/cgi-bin/scanprosite). The position of the conserved motifs K-P-X, and T-T-K-X identified by Cane (1991) are indicated in black blocks. The position of mucin type O-glycosylated serine and threonine residues predicted with NetOglyc 5.0 (Gupta *et al.*, 1999) that were conserved in all sequences are indicated below the alignment by dark grey blocks. Alignments were done with Clustal X 1.64b and shading with Genedoc 2.6001 (www.psc.edu/biomed/genedoc). Identical sequences are indicated by dots (.); truncations relative to the prototype sequences are indicated by stripes (-). (See Figure 3.2 on following page)

A.



B.

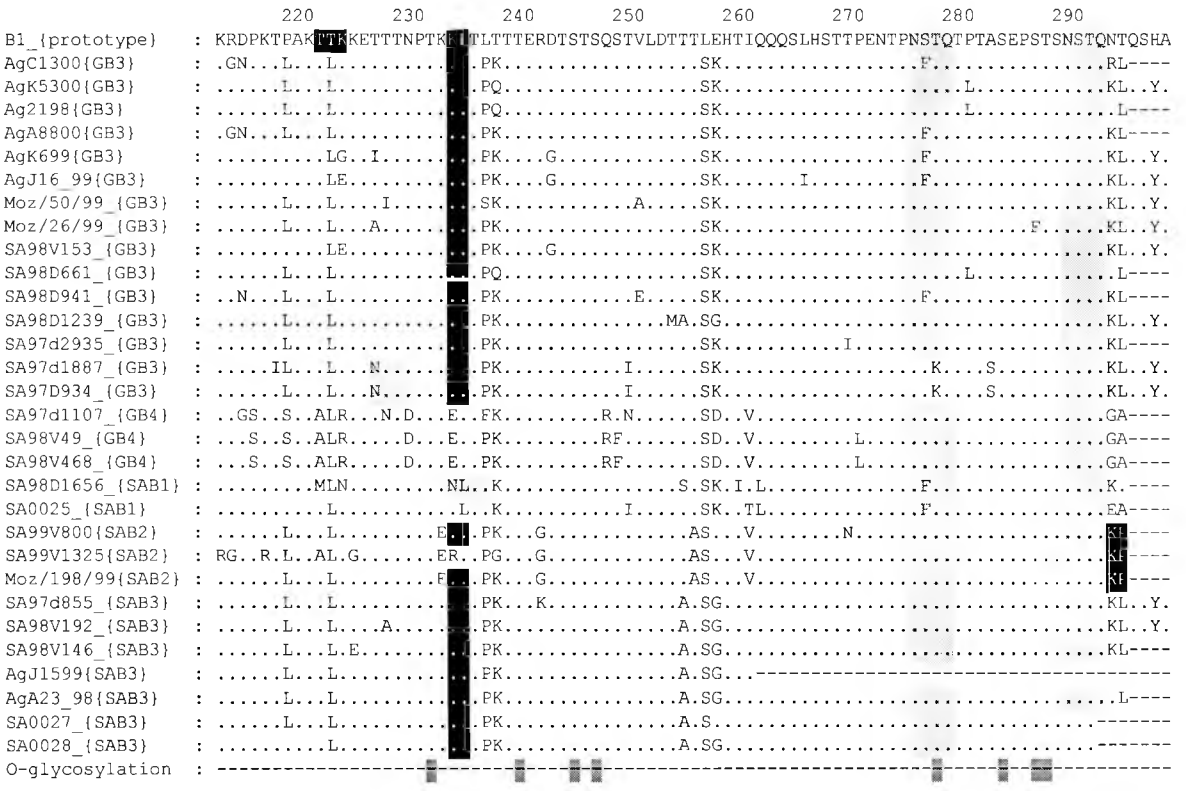


Figure 3.2 See legend on previous page.

Table 3.1: Distribution of age, disease severity (upper respiratory tract infection (URTI); Lower respiratory tract infection (LRTI); severe LRTI and gender of children infected by the individual RSV strains. a.) Individual cases; b.) Summary for identified genotypes

Table 3.1.a: Individual cases:

Date collected	Specimen No.	Subtype	Genotype	Gender	Age (Months)	Disease manifestation
7-Nov-98	AgA21/98	B	GB3	M	17	LRTI
15-Nov-98	AgA23/98	B	SAB3	M	20	LRTI
24-Dec-98	AgJ06/98	A	GA2	M	10	Severe LRTI
11-Jan-99	AgL04/99	A	GA2	M	3	LRTI
19-Feb-99	AgJ09/99	A	GA2	M	4	LRTI
16-Mar-99	AgJ/15/99	B	SAB3	M	32	LRTI
17-Mar-99	AgJ16/99	B	GB3	F	11	LRTI
20-Apr-99	AgA039/99	A	GA2	F	7	LRTI
13-May-99	AgK6/99	B	GB3	M	6	URTI
27-Jul-99	AgA048/99	A	GA2	M	2	LRTI
31-Mar-00	AgL03/00	A	GA2	F	34	LRTI
31-Mar-00	AgL32/00	A	GA2	M	15	LRTI
4-Apr-00	AgK23/00	A	GA2	M	26	URTI
11-Apr-00	AgX2/00	A	GA2	F	10	LRTI
20-Apr-00	AgK28/00	A	GA2	M	5	URTI
26-Apr-00	AgL35/00	A	Undetermined*	F	36	LRTI
3-May-00	AgK32/00	A	GA2	F	17	URTI
6-Jun-00	AgC13/00	B	GB3	M	49	LRTI
7-Jun-00	AgJ/5/00	A	GA2	M	28	URTI
5-Jul-00	AgA88/00	B	GB3	M	22	URTI
12-Jul-00	AgB3/00	A	GA2	M	2	LRTI
28-Jul-00	AgC28/00	A	SAA1	F	30	URTI
3-Aug-00	AgC38/00	A	SAA1	M	36	URTI
16-Aug-00	AgC36/00	B	Undetermined*	M	23	URTI
23-Aug-00	AgK53/00	B	GB3	M	33	URTI
5-Sep-00	AgK61/00	A	GA5	F	9	URTI

Table 3.1 b: Summary for identified genotypes:

Genotype:		GA2	SAA1	GB3	SAB3	Total
Age group:	**0-2 months**	2	0	0	0	2
	3-12 months	6	1	2	0	9
	12-36 months	5	2	4	2	13
Disease severity:	**LRTI**	8	0	3	2	13
	URTI	4	3	3	0	10
	Severe LRTI	1	0	0	0	1
Male		9	1	5	2	17
Female		4	1	1		6
Number of cases:		13	3	6	2	24*

***The PCR products of one subtype A and one subtype B isolate could not be sequenced because of insufficient quality or yield.**

3.4 DISCUSSION

This chapter describes the genetic characterisation of RSV clinical specimens collected at rural community clinics in the Limpopo Province of South Africa. During 1998-1999 specimens were tested on site by rapid ELISA, kept at 4°C and then sent to Johannesburg, while in 2000 specimens were frozen in liquid nitrogen before being sent to Johannesburg for diagnosis. The rapid ELISA performed well on both frozen and fresh specimens. Because of the remote location of these clinics and the lack of infrastructure, it was not possible to culture the viruses. Using the subtype-specific multiplex nested RT-PCR described in Chapter 2, molecular analysis of the clinical specimens could be facilitated. Twenty-six specimens from the rural site could be successfully amplified using the nested-PCR. The use of the liquid nitrogen tank for specimen storage improved the number of specimens that could be confirmed antigenically, but several days elapsed before they could be frozen which may have resulted in RNA degradation, and possibly explains why more specimens from the rural site could not be amplified. This again emphasises the importance of specimen collection when molecular analysis is to be performed. RSV is extremely unstable, heat-labile and sensitive to freeze-thawing and is therefore difficult to culture (Wright & Cutts, 2000). This may contribute to the fact that most studies are conducted on specimens collected at clinical care sites that are close to well-established laboratories or from hospitalised infants.

Agincourt subgroup A specimens all grouped with genotypes GA2 and SAA1. All the Agincourt SAA1 specimens were isolated in 2000 while all the Soweto specimens in this genotype were isolated during 1997, 1998 and 1999 epidemics, but none from 2000. Amino acid changes that may alter the hydrophilic profile and the antigenic structure were observed in Agincourt specimens in GA2 that had been isolated in different seasons, although one isolate that was identical between seasons was also found. Subgroup B Agincourt specimens in different seasons all clustered with different subgenotypes. This may be due to selective pressure over consecutive seasons. It has been suggested that herd immunity to a particular predominant strain develops sufficiently during an epidemic to restrict circulation of that strain in subsequent epidemics, or that high levels of maternal immunity in a population after an epidemic of a particular genotype results in

reduction of the severity of disease caused by that genotype in infants born after that epidemic, resulting in a decline in the prevalence of the genotype (Cane, 2001). The influence of HIV on this has not yet been established.

When comparing the Agincourt specimens to specimens from Chris Hani Baragwanath hospital in Soweto, specimens from Soweto in GA2 (1999); SAA1 (1998); GB3 (1998) were identified that were identical to specimens found in subsequent years in Agincourt. Specimens from Mozambique (Roca *et al.*, 2001) were also identified that clustered with Soweto specimens, but not with any other sequences in the Genbank (SA98V173, outlier to GA2 and the SAB2 isolate, SA99V800) (Venter *et al.*, 2001). Furthermore, although the Agincourt and the Mozambican study sites were only 150 km from each other, the RSV epidemic season was entirely different in these sites, as discussed in the introduction to this chapter. Of interest, the Agincourt site and the Johannesburg study sites (some 500 km apart) showed similarities in their RSV epidemic seasons. It is suggested that these patterns may represent human transport of virus (RSV study group, WHO unpublished data). Comparison of the Mozambican specimens identified in 1999 with South African specimens also suggests transfer between communities as all genotypes identified in Mozambique occurred in the same year in South Africa, although at low levels. An isolate belonging to a newly identified Mozambican subgroup B cluster was identical to one of the SAB2 strains isolated in the same year in South Africa. It is postulated that this variant might have spread from Mozambique to South Africa or vice versa, around May 1999 when SA99V800 was isolated, but that additional drift had occurred during the same epidemic in South Africa to give rise to a new subgenotype by September when SA99V1325 was isolated. Soweto is located next to Johannesburg, in the gold mining district of South Africa, and attracts migrant labour from both Agincourt and Mozambique. Agincourt is situated in close proximity to the Mozambican study site, and a quarter of the population are in fact Mozambican refugees (Tollman *et al.*, 1999).

Previous reports suggested that predominant strains and overall patterns of circulating genotypes were distinct between geographical diverse locations in North America during one epidemic season and the authors didn't find wide

regional spread of one predominant genotype (Peret *et al.*, 2000). The regular migration of members of the population between Johannesburg and Agincourt may thus facilitate transfer of genotypes and influence the epidemiology of RSV in our study site.

There was no significant difference between the G-protein sequences of viruses circulating in the community and specimens from hospitalised infants, except for one SAB3 isolate that had a premature stop codon in position 262. Martinez *et al.* have shown that strain-specific epitopes are located at the C-terminal region of the G-protein and that escape mutants that had premature stop codons in this area, could be selected with monoclonal antibodies (Martinez *et al.*, 1997). Although RSV specimens of especially subtype B, that differ in the length of their G- proteins have been described, to my knowledge this is the first time that a wild-type virus has been found with such a severely truncated G-protein. The truncated isolate (AgJ15/99) caused lower respiratory tract infection in a 32-month old male, and no evidence could be found that this mutation resulted in an attenuated phenotype. In general RSV subtype B isolates have G-proteins of three different lengths (292, 295 and 299 amino acids) (Martinez *et al.*, 1999). Martinez *et al.* (1999) have reported a correlation between the time of isolation and protein length. Viruses with G-proteins of 292 amino acids represented archaic strains isolated in the 1960's, viruses isolated in the 1980's encoded proteins of 299 amino acids, while viruses isolated in the 1990's mostly had sequences of 295 amino acids (Martinez *et al.*, 1999). Only South African specimens from GB4, SAB1 and SAB2 contained 295 amino acids. Specimens in GB3 and SAB3 from Agincourt, Soweto and Mozambique had predominately 299 amino acids. Two specimens from Soweto 2000 only had 292 amino acids. However, except for the protein length, the South African specimens were not closely related to other archaic sequences, suggesting independent evolution to the strains documented by Martinez *et al.* which were mostly from the Uruguay and Spain.

Except for this single truncated isolate, no significant difference in the amino acid sequence, or N- and O-glycosylation sites between the community and hospital specimens could be found. The N-glycosylation sites were mostly conserved in

the subgroups, while O-glycosylation sites were conserved in some positions but also revealed more genotype- or isolate-specific patterns. It has been reported that carbohydrates in the C-terminal third of the G-protein influence the expression of certain epitopes by either masking or contributing to antibody recognition (Palomo *et al.*, 2000). Glycosylated regions of the C-terminal third of the G-protein have also been shown to be partially protease-resistant and to be glycosylated in a cell type-specific manner (Garcia-Beato & Melero, 2000). This suggests that there may be differences in the glycosylation of the G-protein in different human hosts. Further variation could thus be introduced through O-glycosylation that may be influenced by the genetics of the individual hosts that cannot be observed through sequence analysis alone. The G-protein is one of the targets for the neutralising and protective immune response (Johnson *et al.*, 1987a). It is the most variable both between and within the two subgroups and appears to accumulate amino acid changes with time suggesting evolution under selective pressure (Cane & Pringle, 1995a). Indications of positive selection were found in both subgroups in this study (see Chapter 2). Therefore, although attenuating point mutations in other proteins cannot be ruled out, it would be expected that if any major differences exist between strains the G-protein should reflect this.

Several studies have attempted to correlate differences in clinical severity of infections with the different groups or genotypes of RSV (Cane, 2001; Walsh *et al.*, 1997). Some studies showed no difference between subgroup A and B, while others reported that infection with subtype A results in more severe disease. However, most studies have so far been focussed on hospitalised infants, which is problematic when defining clinical severity, since only severely ill babies are considered while those with mild disease are not examined because they are unlikely to be hospitalised (Cane, 2001). Little data are available about community circulating viruses, especially in developing countries. Within the limitations imposed by the sampling data, this study suggested that the same virus might in fact cause mild upper respiratory tract infections, lower respiratory tract infections and severe RSV. The infants from the rural community were in general also exposed to the same risk factors such as open fires and crowding. This may suggest that other factors such as genetics and the individual's own

immune response may have a larger influence on disease severity than the infecting strain of RSV.

In summary, RSV specimens from infants attending rural community clinics in the Limpopo Province of South Africa were compared to specimens from hospitalised infants in Soweto and Mozambique. Wide distribution of specific strains and evidence of identical viruses resulting in both severe RSV and mild upper respiratory tract infections have been found.

CHAPTER 4

Distribution of RSV genotypes across Southern Africa during a single epidemic season

4.1 INTRODUCTION

As explained in previous chapters, RSV epidemiology is thought to be a community-based phenomenon, and it has been suggested that although transmission of viruses between communities may occur, the pattern of genotypes are distinct between them (Peret *et al.*, 2000). Analysis of RSV isolated at the same time from different sites around the world suggested that although RSV can be variable within a single local epidemic, very similar viruses circulate worldwide, indicating temporal but little geographical clustering (Cane & Pringle, 1992). The importance of strain variations to RSV epidemiology and prophylaxis is not yet clearly defined, however, evidence of positive selection that may assist the virus in escaping pre-existing immunity suggests that the variation is likely to be important in RSV infection and disease (Cane & Pringle, 1995c; Sullender *et al.*, 1991; Garcia *et al.*, 1994). Most of these data are however based on studies conducted in the first world. No data are as yet available about the circulation of RSV strains across South Africa within one epidemic year.

Findings in the previous two chapters already suggested that there might be differences between epidemics in South Africa and regions in the northern hemisphere. In the first instance, one subgroup A and three subgroup B genotypes have been identified that have only been found in southern Africa to date. Secondly, one genotype (GA2) showed predominance over two consecutive seasons in Soweto, and instead of showing heterogenity in genotypes with co-circulation of genotypes, specimens of 2000 were almost exclusively genotype GA2. Finally, specific viruses were identified with identical

sequences between Agincourt and Johannesburg, and Mozambique and Johannesburg, suggesting wide distribution of strains.

In the previous Chapter, the differences in RSV epidemic seasons between a site in Mozambique and a site located only 150 km from there in South Africa was described. South Africa is divided into quite diverse climatic regions (Figure 4.1) and covers an area of approximately 1,219,912 sq km. The central regions where Johannesburg, Pretoria and Bloemfontein are located have dry winters and rainy summer seasons, while the climatic conditions in Cape Town in the south are more temperate with a rainy winter season and a dry summer season. On the east coast, Durban has a tropical-like climate with perennial rainfall patterns. Agincourt, which was described in Chapter 3, is also a summer rainfall district; however it is drier than the central regions and has warmer winters. The labour force frequently travels between rural areas and the cities, especially Johannesburg, which is the financial capital of the country. In Chapter 3 it was suggested that the regular migration of members of the population between Johannesburg and Agincourt might facilitate transfer of genotypes and influence the epidemiology of RSV between these two sites. The extent of genotype distribution in one season between the different climatic regions has, however, not been determined.

Although the overall mortality related to RSV associated LRTI is low in developed countries (1-2%) (Wang *et al.*, 1995) and in South Africa (Madhi *et al.*, 2000), there are children who remain at high risk of severe morbidity as well as mortality (Wang *et al.*, 1995). Both of these studies showed that these children account for a disproportionately large number (20-25%) of those children who subsequently required hospitalisation for RSV-associated LRTI. Children with underlying risk factors in a developing community in South Africa accounted for 80% of those who require mechanical ventilation and 75% of all deaths due to RSV-associated LRTI (Madhi *et al.*, 2000). Recent advances in immunoprophylaxis, particularly the licensing of Palivizumab (Synagis®) in preventing RSV-associated hospitalisations among some high risk groups of children, have stimulated research in further characterising the burden of RSV disease in high-risk infants.

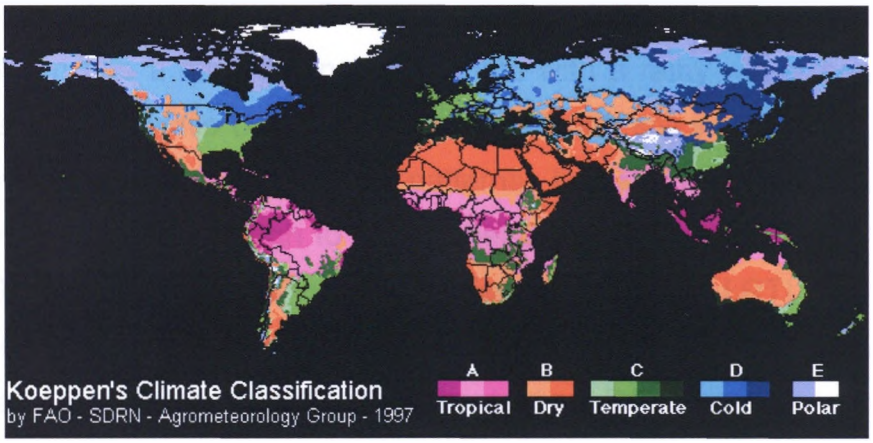
Inclusion of this group in the molecular epidemiological analysis of RSV may help determine the extent and impact of RSV strain distribution during an epidemic throughout the sector of the population most affected by RSV.

In this chapter RSV-positive specimens were taken from high-risk infants attending hospitals in Johannesburg, Pretoria, Bloemfontein, Durban and Cape Town during 2000 and 2001. These specimens were subtyped, sequenced and compared phylogenetically to specimens from hospitalised infants from Soweto and community specimens in Agincourt to determine the distribution pattern of RSV during one epidemic season in different climatic regions of South Africa. These specimens were collected as part of an impact study to define the frequency of underlying risk factors predisposing to severe RSV-associated LRTI in children aged <24 months. The genotyping method described in Chapter 2 was also evaluated for its usefulness in typing of RSV specimens from across South Africa. Knowledge of the distribution of RSV genotypes across South Africa during one epidemic season, definition of epidemic seasons in the different climatic regions, and the burden of RSV in high-risk infants will be beneficial to the design of future vaccination strategies and the implementation of effective prophylaxis schedules.

A.



B.



C.



Figure 4.1: The map of South Africa (A.) indicates the sites included in this study with stars. The climatic regions of the world according to Koeppen's classification system are shown in (B.) with an enlargement of Southern Africa shown in (C.). Maps were obtained from http://www.expedia.co.uk/ and http://weather.about.com/gi/dynamic/offsite.htm?site=http://www.fao.org/WAICENT/FAOINFO/SUSTDEV/EIdirect/climate/EIsp0054.htm

4.2 METHODS

4.2.1 Patients and viruses

Study group 1:

Study group 1 consisted of high-risk infants and children less than 24 months of age presenting to the emergency ward with symptoms consistent with RSV disease during 2000 and 2001 at private hospitals in Johannesburg, Pretoria, Durban, Bloemfontein and Cape Town (sponsored by Abbott laboratories as part of an RSV impact study). RSV-positive specimens that were available for subtyping: N=50.

A prospective study was performed in four geographically distinct regions involving five paediatric practices mostly in the private sector. The sites were spread across four provinces in South Africa with different climatic conditions (Figure 4.1). Johannesburg and Bloemfontein have dry winter and rainy summer seasons, Cape Town has a more temperate climate with a rainy winter season and a dry summer season while Durban, in which the fourth site was located has a tropical-like climate with perennial rainfall patterns. The study was conducted prospectively from April 2000 through to September 2001. During period 1 (April to December 2000), all outpatient as well as inpatient children with LRTI who met the pre-determined criteria were eligible for inclusion into the study if they were <24 months of age and were diagnosed as having a LRTI based on clinical evaluation by the attending paediatrician. Children were categorised as being at high risk for severe RSV-associated disease if they met any of the following criteria:

i). Gestational age of ≤ 35 weeks at birth and <6 months of chronological age;

ii). < Two years of age with chronic lung disease that required any medical therapy in the previous six months;

iii). Children under two years of age with a cyanotic congenital heart disease or an acyanotic congenital heart disease that was associated with a clinically significant shunt;

iv). Children with any known immunosupression.

During period 2 (January to September 2001), recruitment was only restricted to those children that met the criteria of being at high risk for severe disease.

Study group 2:

Infants aged 0-60 months from Chris Hani Baragwanath hospital, Soweto (1997-2000), hospitalised for lower respiratory tract infections that varied from mild to severe ICU cases. (Number subtyped: N=225; described in Chapter 2).

Study group 3:

Infants aged 0-60 months attending rural community clinics in the Agincourt region of Bushbuckridge on the border of Mozambique in the Limpopo Province of South Africa (WHO study), from November 1998-May 2001.(Number subtyped, N=26; described in Chapter 3)

4.2.2 Specimen collection and RSV diagnosis in study group 1

Nasopharyngeal aspirates were obtained as described in Madhi *et al.*, (2000). The aspirates were mixed with viral transport medium, and an aliquot used immediately for diagnostic testing on site with the Abbott TESTPACK® RSV (ABBOTT Laboratories, North Chicago, USA) while a second aliquot was snap-frozen within 24 hours of collection in liquid nitrogen. The latter was transported on dry ice to the National Institute for Communicable diseases, and used for molecular analysis.

4.2.3 Subtyping, genotyping and phylogenetic analysis

RNA was isolated from nasopharyngeal aspirates as described in Chapter 2. Specimens were subtyped by subtype-specific RT-PCR (Sullender *et al.*, 1993; Madhi *et al.*, 2001) followed by multiplex nested PCR (Venter *et al.*, 2001), sequenced and genotyped as described in Chapter 2. Phylogenetic analyses were carried out with MEGA version (2.1) (Kumar *et al.*, 2000). The sequenced specimens (study group1) were compared in phylogenetic trees to specimens found at Chris Hani Baragwanath hospital in Soweto during 1997-2000 (study group 2) (Venter *et al.*, 2001) and at community clinics in the Agincourt region of the Limpopo Province of South Africa (study group 3) (Venter *et al.*, 2002).

4.2.4 Calculation of positive selection:

Cases of positive selection were identified with Fisher's Exact Test for positive selection, using the modified Nei-Gojobori methods for direct computation of the numbers of synonymous and nonsynonymous differences, and the number of synonymous and nonsynonymous sites. Fisher's Exact Test for positive selection is conducted to examine the null hypothesis of the neutral evolution. The probability must be <0.05 for hypothesis rejection at 5% level, and thereby accepting that positive selection has occurred. Calculations were carried out with Mega version 2.1 (Kumar *et al.*, 2001).

4.3 RESULTS

Study group 1 consisted of children recruited as part of an RSV impact study conducted in collaboration with Abbott laboratories. Fifty-three (24%) of 220 children in study group 1 enrolled during period 1 had risk factors for severe RSV disease; while 38 high-risk children were enrolled during 2001. During 2000 RSV was isolated from 53 children of whom 16 (30%) were high risk and 37 (22%) were not considered to be high risk for severe RSV disease (P = 0.31) (Table 4.1). The results suggested that high-risk children were more likely to require intensive unit care (25% vs. 2.7%, P = 0.02) and to be hospitalised for a longer duration than non high-risk infants. Overall (2000-2001), RSV was isolated from 34 (37.4%) of the 91 high-risk infants enrolled (Madhi *et al.*, 2002).

4.3.1 RSV genotype circulation throughout South Africa during one season

A total of 40/53 RSV positive specimens obtained in the year 2000 from the 5 different study sites in study group 1 were subtyped. Other specimens were lost to subtyping during the transport of specimens to the NICD. Subtyping of the 40 specimens obtained during the year 2000 showed 37 (92.5%) as belonging to subtype A and 3 (7.5%) as belonging to subtype B. For the year 2001 only 9 specimens were available for subtyping (5 subtype A; 4 subtype B). GA2 was the dominant genotype in the year 2000 in all the study sites, with 25/33 specimens in study group 1, 22/28 specimens from study group 2 and 8/14 specimens from study group 3 clustering with GA2 (Table 4.2). Certain GA2 specimens with identical G-protein sequences were identified between the different communities (Figure 4.2.A.). Identical specimens and nearly identical GA2 specimens were isolated in 2000 in Johannesburg, Agincourt, Bloemfontein, Pretoria, Cape Town and Durban in the private sector hospitals, the community clinics and the government academic hospital. Other subtype A genotypes identified in study group 1, although at much lower levels, include SAA1 (1/33), and GA5 (4/33).

Table 4.1: RSV-positive specimens from children in study group 1 with lower respiratory tract infections enrolled from April 2000 through to September 2001 in different demographic regions of South Africa (Madhi *et al.* 2002).

(Period 1)	High- risk	Otherwise healthy	p value
RSV infection by region	n=53	n=167	
Johannesburg	4/17 (23.5) †	16/ 88 (18.2)	0.74
Cape Town	4/14 (28.6)	16/39 (41.0)	0.61
Durban	4/12 (33.3)	2/25 (8.0)	0.07
Bloemfontein	4/10 (40.0)	3/15 (20.0)	0.39
Total	16/53 (30.2)	37/167 (22.2)	0.31

Table 4.2: Respiratory syncytial virus genotypes isolated in distinct geographical locations across South Africa during the year 2000

Genotype:	**Study group 1:** Bloemfon-tein	Cape Town	Johannes-burg	Durban	**Study group 1: Total**	**Study group 2:** Baragwanath hospital	**Study group 3:** Agincourt
Year: 2000							
GA2	4	8	11	2	**25**	**22**	**8**
SAA1			1		**1**		**3**
GA5	1	1	1	1	**4**	**2**	
GB3						**1**	**3**
SAB1	1				**1**	**1**	
SAB3		2			**2**	**2**	
Total	**6**	**11**	**13**	**3**	**33**	**28**	**14**
Year: 2001						ND	ND
GA2	2				**2**		
GA5		2	1		**3**		
SAB3		1	2		**3**		
GB3		1			**1**		
Total	**2**	**4**	**3**	**0**	**9**		

ND: Not determined.

GA5 specimens that were identical were identified from Bloemfontein and Durban. Only 3 subtype B specimens were identified in the year 2000, two from Cape Town (SAB3) and one from Bloemfontein (SAB1). From Agincourt 57% of specimens belonged to GA2, while 3/14 (21.4%) clustered with GB3, and 3/14 with SAA1. The largest sample size was taken from Chris Hani Baragwanath hospital where 78% of specimens belonged to GA2, 7.1% to GA5, 4.5% to GB3, 4.5% to SAB1 and 7.1% to SAB3 (Figure 4.2.B.).

Study group 1 specimens from 2001 grouped with GA2, GA5, GB3 and SAB3, with most clustering with GA5 and SAB3, however the sample size was too small to make a firm conclusion about the dominant genotypes in 2001 (Figure 4.2.A and B.). As mentioned in Chapter 2, subsequent typing data from Chris Hani Baragwanath hospital suggested that subtype B was the dominant subtype in 2001. Indications of positive selection were found between the GA5 2001 specimens Ab3059Ct01 and Ab5076Pt01 relative to GA5 specimens from 2000, however isolate Ab3061Ct01 had an identical sequence to a Mozambican isolate from 1999. This suggests that although positive selection may contribute to establishment of new infections in the population, virus strains from previous seasons from other regions may also be a source of infections in the immunologically-naïve population.

4.3.2 Evoultion in RSV genotypes: Comparison of all specimens from South Africa from study groups 1, 2 and 3

Positive selection could be identified between various specimens between different seasons using Fisher's exact test for positive selection but also between specimens from different sites in the same season in both subgroup A and B (Appendix A). For subgroup A the genotyping method introduced in Chapter 2 still holds, with the only outliers being SA98V173 and Moz-9-99. The South African GA2 specimens all grouped separately from the published GA2 reference sequences (Peret *et al.*, 2000), but still grouped together (Figure 4.2.A.). Subgroup B genotyping is however more cumbersome, and the current system will most probably evolve even further. After addition of the Bloemfontein isolate

AB17BL00, SAB1 specimens SA98D1656 and SA0025 did not cluster together anymore. Closer analysis revealed that isolate SA98D1656 had a P-distance of 0.06 with SA0025, but of P=0.1 with AB17B00, and that SA0025 and AB17BL00 were more closely related with a P-distance of P=0.05, even though bootstrap statistics of only 53 could be obtained. This suggests that this genotype has evolved and needs to be adjusted. For now, SA0025 and AB17BL00 have been typed as SAB1 and SA98D1656 as SAB4 because the latter does not group with any other sequences. SAB3 specimens from 2001 (AB5078P01 and AB3064C01) have P-distances of P=0.05 to each other and signs of positive selection relative to other members of the genotype, however significant bootstrap statistics could still be obtained for this genotype. GB3 specimens AgK699 and Ag2100 have P-distances of 0.07 to each other, and do not cluster together with significant bootstrap values, which suggest that they should be in different genotypes. When SAB2 and SAB3 specimens are however removed from the equation, the GB3 specimens do cluster together with significant bootstrap values, although indications of positive selection and P-values of up to 0.07 between specimens in this genotype suggest that it will probably evolve further. The 2001 GB3 isolate AB3062C01 and SAB3 isolate AB5078P01 have a P-distance of P=0.08, which suggests that further drift occurred between specimens in these two genotypes, reconfirming that they should be genotyped separately. Therefore, the subgroup B genotyping method was useful for characterisation of the South African specimens, but an additional genotype, SAB4, needs to be added to accommodate all sequences.

Figure 4.2. Phylogenetic analysis of South African subgroup A and B RSV, isolated from different geographic locations

(Figure 4.2.A.: Subgroup A; Figure 4.2.B.: Subgroup B follows on the next two pages.) The Neighbour-joining trees were drawn to scale with the bars indicating 0.02 nucleotide substitutions for subgroup A and B, respectively. Bootstrap statistics for 100 re-sampling iterations are shown at the node. Only values of more than 70% are indicated. The phylograms were midpoint-rooted. RSV specimens sequenced as part of this study are indicated by an Ab prefix which indicates the study name (Abbott). The city of isolation is indicated by a suffix, as follows: Cape Town (Ab-00/01-Ct); Bloemfontein (Ab-00/01Bl); Pretoria (Ab-00/01Pt); Johannesburg (Ab-00/01-J) and Durban (Ab-00/01-D). RSV specimens from Agincourt in the Limpopo Province of South Africa are indicated by Ag-00 while specimens from hospitalised infants from Soweto (1997-2000) are indicated as SA-97-00. Colour bullets are used to indicate the different sites as follows: Bloemfontein (yellow); Pretoria (purple); Cape Town (blue), Johannesburg/Soweto (red), Agincourt (green), Durban (turquoise). Different epidemic seasons are indicated as follow: upside down triangle=1997; triangle= 1998; square=1999; diamond= 2000; circle =2001. Reference sequences of each genotype identified by Peret *et al.*, (2000) were obtained from the Genbank and included in the analysis: from the USA (NY, New York; Al, Alabama; MO, Missouri; Tx, Texas; CN, Canada); Sequences were also included from Uruguay (MON); Spain (MAD) (Garcia *et al.*, 1994; Martinez *et al.*, 1999; and Mozambique (MOZ) (Roca *et al.*, 2001) prototype strains for type A: strain A2 (Australia) (Wertz *et al.*, 1985); type B: Swed8-60 (Sweden) (Sullender *et al.*, 1991); USA (18537) (Johnson *et al.*, 1987). The genotype assignment is indicated at the right in brackets.

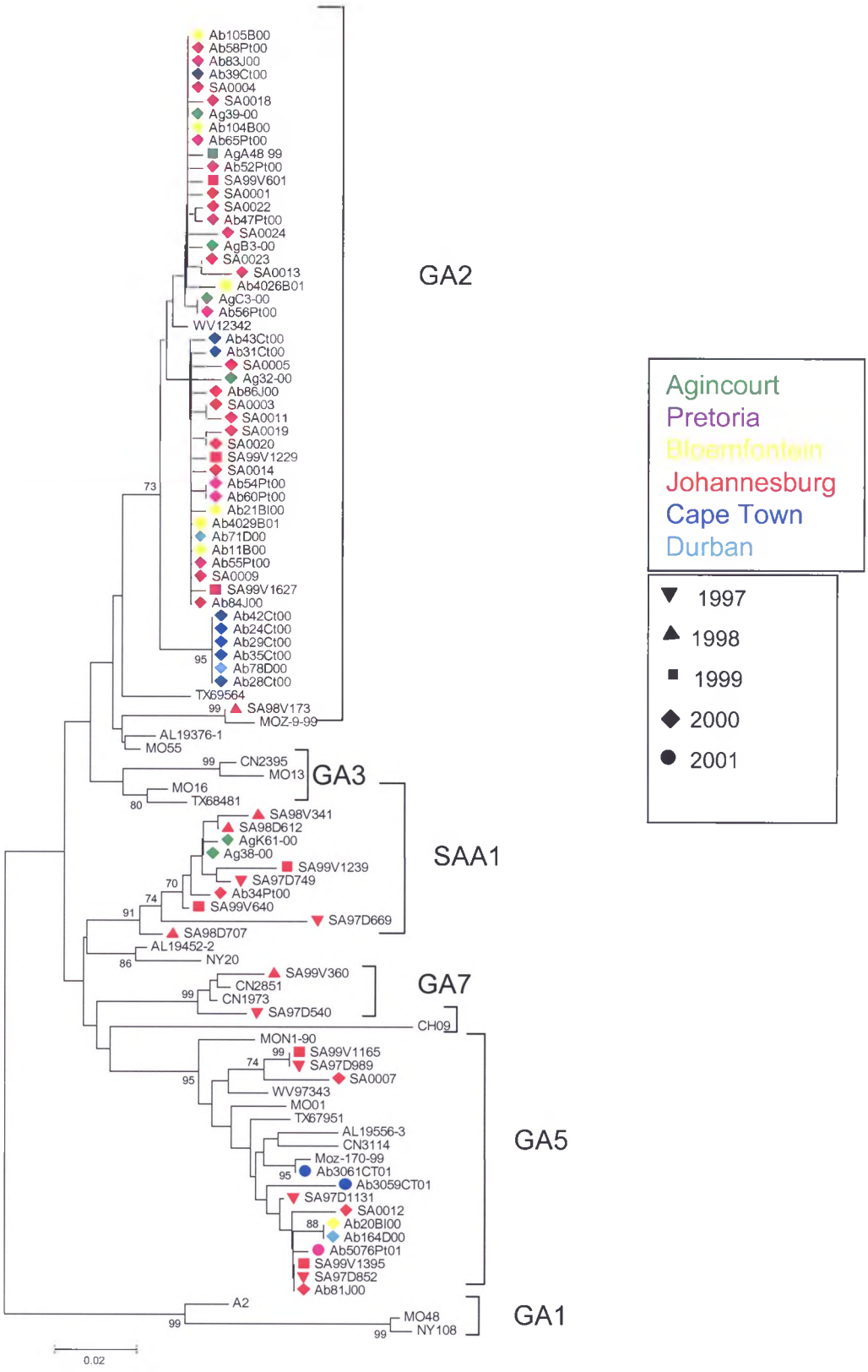


Figure 4.2.A. Subgroup A isolates from different geographic sites across South Africa

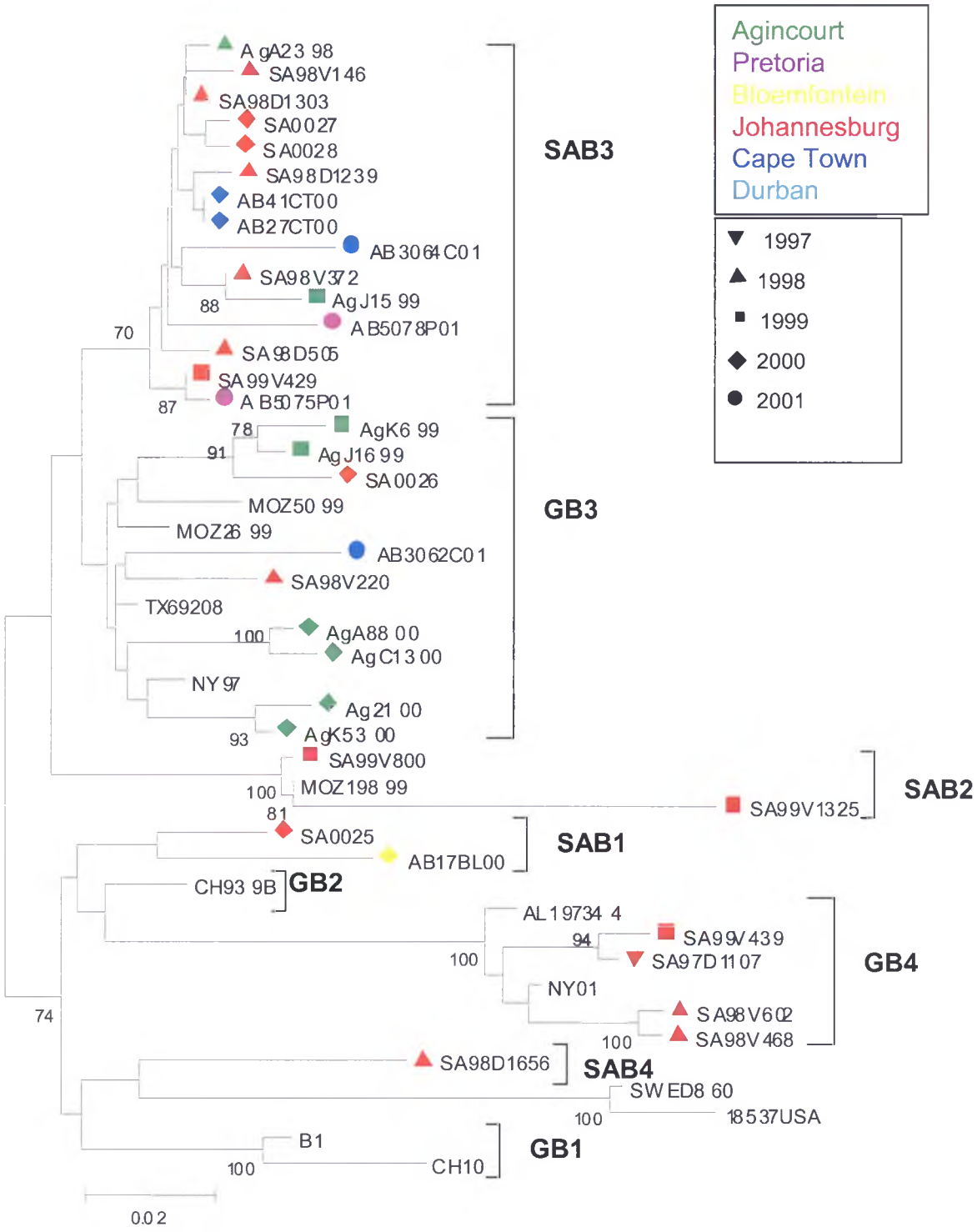


Figure 4.2.B. Subgroup B isolates from different geographic sites across South Africa

4.4 Discussion

A study conducted between five communities in the USA, suggested that RSV epidemiology is community-based, not regional or national (Peret *et al.*, 2000). The authors found that the predominant strains and overall patterns of circulating genotypes were distinct between communities and that wide regional spread of one predominant outbreak genotype did not occur, although transmission of viruses may occur. Peret *et al.*, (2000) also suggested that local factors, e.g. previous strain-specific immunity to the virus, and possibly the relative fitness of the circulating virus, dictate which strains will predominate in a given season. In Cuba on the other hand, little variation in the G-protein genes occurred in isolates from 1994-1995 relative to archaic strains of 1956. There also appears to be a lack of accumulation of amino acid changes and replacement of the dominant genotypes each year and RSV infections are distributed throughout the year. It has been suggested that this genetic stability may be due to a restriction on travel within and outside the island (Valdes *et al.*, 1998). In South Africa, genotypes were identified that have not been found elsewhere, although some of the same genotypes that circulated the rest of the world were also identified. In the study described in Chapter 2 that was conducted at Chris Hani Baragwanath Hospital in Soweto over four consecutive seasons (1997-2000), switching of the predominant genotype between individual seasons, was found with genotypes appearing and disappearing from one year to the next. However in 1999 GA2 became the dominant genotype, with 42% of specimens from Chris Hani-Baragwanath Hospital falling into this group, and increased even further in 2000 when 78% of all specimens belonged to this genotype. In this Chapter a study is described where specimens were taken in different climatic and geographic regions of South Africa during one epidemic. Although the sample from each community was too small to clearly define the predominant strain for each of the five sites, GA2 was isolated most frequently in all study sites and GA2 specimens were identified that were identical between Cape Town, Johannesburg, Durban and Bloemfontein. GA2 specimens were shown to be the dominant genotype at Chris Hani Baragwanath hospital in Soweto in 2000 and were also isolated in 2000 in Agincourt in the Limpopo Province of South Africa, approximately 500 km

north-east of Johannesburg. This data suggest wide distribution of this genotype throughout South Africa during 2000.

When considering other studies, it appears to be unusual for one genotype to predominate for more than one season. Different groups found that the predominant genotype differed in each epidemic (Choi & Lee, 2000; Peret *et al.*, 1998; Cane, 2001), with the exception of Cuba (Valdes *et al.*, 1998). The data described in this chapter also indicate that genotype GA2 had spread across the whole country during 2000, and identical sequences were found in specimens from study sites located 500-1400 km apart. This may suggest that highly effective and fit strains may have the ability to spread nation-wide.

Although South Africa is only about twice the size of Texas, 3 different climatic regions are represented here. The sample sizes were too small to make clear-cut conclusions about the seasonality of RSV in the different sites, however previous studies suggested that there might be year-to-year variation of the RSV season even within a specific region (Madhi *et al.*, 2001). The RSV season in general starts in March possibly through to August in Cape Town (Hussey *et al.*, 2000) and Johannesburg (Madhi *et al.*, 2001), but may start earlier in Durban where isolation appears to be more perennial with the peak possibly being late summer (December through to February) (Dr P. Jena- personal communication). Analysis of the data from the current study suggested that the peak RSV activity in the year 2000 was between April through to August in most regions, excluding Durban where isolation of RSV started to wane after April (Madhi *et al.* in press). This suggests that RSV is isolated in the rainy summer months in Durban which has a subtropical climate and in autumn to winter in the temperate regions of Cape Town which has a winter rainfall, as well as Johannesburg, Bloemfontein and Pretoria which have summer rainfall. In Chapter 3 the differences in RSV seasonality between the Agincourt site and the WHO Mozambique site (150 km from each other) were also described. This suggests that RSV can be isolated in either summer or winter in South Africa depending on the geographic region. Travelling between the different climatic regions may influence the seasonality of RSV infections and reestablishment of genotypes between sites. The high HIV

incidence rates in South Africa may also play a role in harbouring and prolonged shedding of RSV genotypes although this has not yet been proven. Further work over consecutive seasons in different locations would have to be carried out to determine how often genotypes circulate throughout the country or remain dominant in more than one season. More restricted genotype circulation of RSV in individual communities in South Africa may also occur in more remote locations where less travelling to the major cities takes place.

Overall analysis of all clinical specimens studied in Chapters 2-3 suggests that the genotyping method is effective, but that genotypes will probably evolve in future studies. This is consistent with the findings of positive selection between viruses of different seasons, between viruses of different sites as well as between specimens in the same site and season. Amino acid changes in the G-protein have been shown to accumulate over time, suggesting that the designation of specimens into genotypes cannot be static since new genotypes will steadily emerge while others may disappear entirely (Cane & Pringle, 1995c). It is thought that the immune response is the selective pressure for the emergence of new variants of RSV and that this could either be at an individual level whereby variability contributes to the ability of the virus to re-infect, or at a community level through restriction of circulation of the virus (Cane, 2001).

To conclude, this study has shown that for the year 2000 the dominant genotype found in Johannesburg could be isolated in different climatic regions throughout South Africa in the same season, suggesting possible wide distribution of effective strains. The same genotypes also affected high-risk infants attending the private sector hospitals (in Johannesburg, Pretoria, Bloemfontein, Durban and Cape Town), government hospitals in Soweto and health care clinics in the Limpopo Province.

CHAPTER 5

Identification and characterisation of dominant human RSV cytotoxic T-cell epitopes in a population of diverse HLA types

5.1 INTRODUCTION

Evidence suggests that an RSV-specific type 1 T-helper cell response and the induction of CD8+ cytotoxic T-cells (CTL) are critical for the disease-free control of RSV infection in the murine model (Cannon *et al.*, 1988a; Graham *et al.* 1993; Graham *et al.* 1991). It is thought that activation of Th-2 cells is the key factor in disease pathogenesis while a Th1 response with induction of CD8+ T-cells is needed for a protective immune response (Reviewed in Graham, 1996). However, reports of CTL-induced immunopathology in mice during the transfer of large numbers of CTLs have also been made, suggesting that a fine balance exists between protective and disease-enhancing effects (Cannon *et al.*, 1988a; Cannon *et al.*, 1987). In the past, most studies of the CTL response to RSV have been focussed on mice (reviewed in Walsh & Graham, 1999). CTLs have been documented in humans following RSV infection (Bangham & Askonas, 1986; Chiba *et al.*, 1989) and are expected to play a crucial role in viral clearance, however their role in protection and immune pathology remains unclear. RSV proteins that induce CTL activity include the F-protein, the M2-protein, N-protein, and the SH-protein (reviewed by Walsh, 1999; Walsh & Graham, 1999). The N-protein, which is the most conserved between the two RSV subgroups, is the major target for the memory CTL response in humans (Bangham & Askonas, 1986; Cherrie *et al.*, 1992a) while the M2 protein induces the dominant response in mice of the H-2^d haplotype (Openshaw *et al.*, 1990). Studies in humans have been limited by technical difficulties imposed by the traditional techniques used to

enumerate CTL responses. This includes a small sample volume from acutely infected children or low frequency of RSV-specific memory CTL response in adults (Goulder *et al.*, 2000b). The availability of more sensitive technologies has now made it possible to study the CTL response in humans. These include the use of tetrameric MHC class I molecules for binding specific epitopes (Altman *et al.*, 1996) and peptide epitope stimulation of T-cells followed by measurement of interferon-gamma (IFN-gamma) secretion (Murali-Krishna *et al.*, 1998) by ELIspot or intracellular cytokine (ICC) staining using flow cytometry. In order to be able to use these techniques in studies in infants, RSV-specific CTL epitopes need to be identified that can be used as peptide reagents in either of these assays.

Recent studies have shown that memory responses in adults can be used to enumerate RSV epitopes using IFN-gamma ELIspot (Goulder *et al.*, 2000b). The attraction of this strategy is the availability of higher volumes of blood from adults that allow further characterisation of the identified epitopes using chromium release assays to confirm specific cytotoxicity. Furthermore epitopes likely to be identified were probably involved in clearance of RSV infection in the past. Currently, only 3 CTL epitopes have been defined for RSV. These include one HLA B7 epitope to the N-protein (Goulder *et al.*, 2000b), and two epitopes on the F-protein restricted to HLA B57 and HLA C12, respectively (Brandenburg *et al.*, 2000a).

To date no study has taken all possible epitopes on any RSV protein into consideration during the mapping process, or looked at the dominant RSV-specific CTL epitope response to a specific protein in a diverse population. In this study South African adults were screened by IFN-gamma ELIspot using overlapping peptides covering the complete N-protein of RSV. The dominant responses were further characterised and the epitopes with the strongest cytolytic activity mapped with truncated peptides. South Africa has a very diverse population consisting of Southern African Black ethnic groups, Caucasian of Dutch, French, German and British origin, Cape Coloureds and Indians, with a wide diversity of HLA types (Goulder *et al.*, 2000a; Hammond *et al.*, 1997). In order to further characterise the CTL response in children, dominant epitopes that are restricted to HLA types of the study population have to be defined. This will contribute to vaccine development and provide a better understanding of immune induced disease pathogenesis of RSV.

5.2 METHODS

5.2.1 Subjects

Thirty-seven healthy South African adult subjects (18-50 years) representing a wide range of South African ethnic groups were recruited from Johannesburg, South Africa. These included South-Sotho, North-Sotho, Tswana, Zulu, Swazi, Ndebele, Xhosa, Cape Coloureds, Caucasian and Indian individuals (Table 5.1). This work was approved by the Committee for Research in Human Subjects (Medical), University of the Witwatersrand, Johannesburg, South Africa.

5.2.2 MEDIA AND COMPONENTS

For the ELIspot, ICC assay and Cytotoxicity experiments R10 media was used containing RPMI 1640 (Sigma-Aldrich, St. Louis, MO), 10% Fetal calf serum (Highveld Biologicals, Johannesburg, South Africa), 10 mM HEPES buffer (Roche diagnostics, Mannheim, Germany) with 2 mM glutamine (Sigma-Aldrich, St. Louis, MO), 0.13 mg/ml Penicillin (Sigma-Aldrich, St. Louis, MO), and 0.27mg/ml Streptomycin (Laboretoria, Pretoria, South Africa).

5.2.3 HLA typing.

HLA typing was performed by the HLA Laboratory (National Institute for Communicable Diseases, Sandringham, South Africa) using low-resolution sequence-specific primer PCR.

5.2.4 Synthetic peptides.

The panel of 127 peptides used in ELIspot screening experiments was 14 amino acids (aa) in length overlapping adjacent peptides by 11 aa. The overlapping peptide sets correspond to the N-protein of subgroup A2 RSV, and were purchased from Mimotopes Pty. Ltd. (Clayton Victoria, Australia). Selected 9- and 10-mer peptides used in cytotoxicity and ICC assays were synthesized by the Natural and Medical Sciences Institute (University of Tuebingen, Germany). These peptides were HPLC purified to more than 90% purity.

5.2.5 ELIspot assay

Prior to the addition of cells, 96-well plates (Millipore (Immobilon-P, Sterile, high protein binding for ELIspot assays, clear), Millipore, Bedford, MA) were coated with anti-IFN-gamma mAb 1-DIK (5 µg/ml); Mabtech, Stockholm, Sweden) at 4 °C overnight. After washing the plates three times with PBS, each plate was blocked by addition of RPMI 1640 containing 10% FCS (R10) for at least 1 hour at room temperature. Peptides were added directly to wells in a volume of 50 µl followed by addition of freshly isolated PBMCs at 200,000 cells/well in 50 µl R10 media. The final concentration of peptides in the screening assay was 10 µM. The plates were incubated for 18-20 hours overnight at 37°C in 5% CO_2. The plates were washed, labelled with biotinylated anti-IFN-gamma mAb 7-B6-1 (2 µg/ml; Mabtech) and incubated for 3 hours at room temperature. After additional washes, streptavidin-conjugated horseradish peroxidase (Pharmingen, BD Bioscience, Europe) was added for 1 hour at room temperature. The plates were washed and IFN-gamma producing cells detected as dark spots after a 40 min colour reaction using NovaRed substrate (Vector, Burlingame, Ca). IFN-gamma producing cells were counted using either a KS-ELIspot reader, (Zeiss Göttingen, Germany) or a Zoom Stereo microscope, (Olympus, Hamburg, Germany). The number of peptide-specific IFN-gamma specific T-cells was calculated by subtracting the negative control value, and the results were expressed as the number of spot-forming cells per million PBMC (SFC/10^6).

5.2.6 Intracellular cytokine staining

ICC assays were performed by incubating freshly isolated PBMC (2 x 10^6) with 10 µM peptide and co-stimulatory antibodies anti-CD28 and anti-CD49d mAb (1 µg/ml each; Becton Dickinson, San Jose, CA) at 37 °C in 5% CO_2 for 2 hours. According to the manufacturers, CD49d (Anti-VLA-α4) can block or enhance fibronectin-stimulated T-lymphocyte proliferation, while the combination of CD28 and CD49d provides optimal co-stimulatory signals for antigen specific cytokine production. Brefeldin-A (10 µg/ml; Sigma-Aldrich, St. Louis, MO) was added to inhibit protein transport in order to increase the intracellular staining signal by enhancing the accumulation of intracellular cytokines. The cells were incubated for an additional 4

hours, and then placed at 4 °C overnight prior to staining. The next day, EDTA (2 mM final) was added to each sample for 10 min at room temperature and vortexed before transfer to polystyrene tubes containing 2 ml cold wash buffer (1% BSA, 0.1% sodium azide in PBS). After an additional wash, the cells were lysed, using BD-FACS lysing solution and then permeabilized using BD-FACS Permeabilization solutions (Becton Dickinson) prior to staining. Cells were stained using peridinin chlorophyll protein (PerCP)-, allophycocyanin (APC)-, fluorescein isothiocyanate (FITC)- and phycoerythrin (PE)- fluorescent conjugated antibodies: anti CD3-APC, anti-CD8-Per-CP,anti-CD69-PE and anti-IFN-gamma–FITC (Becton Dickinson). After a final wash, cells were resuspended in PBS containing 1% paraformaldehyde. Six-parameter flow cytometric analysis was performed using a FACSort flow cytometer (Becton Dickinson). Data acquisition was done using CellQuest software measuring PerCP, APC, and FITC and PE as fluorescent parameters and acquiring 50,000 lymphocytes using a forward-scatter side-scatter gate. Data were analysed with Paint-A-Gate software gating small viable lymphocytes on forward scatter side scatter plots, CD3+/CD8+, CD8/IFN-gamma and CD69+/IFN-gamma+ lymphocytes. To confirm the CD8+ response, CD8- cell populations were also analysed.

5.2.7 Peptide-specific cell lines

Peptide-stimulated lymphocyte cultures were established from healthy adult subjects as previously described (Goulder *et al.*, 2000a). Briefly, freshly isolated PBMC at a high cell density (20-30x10^6) in 100 μl R10 were pulsed for 1 hour at a peptide concentration of 100 μM and then diluted up to 2 x 10^6 cells/ml in R10 containing 25 ng/ml recombinant human IL7 (rIL7; Endogen, Woburn, MA). Cells were plated out into 2 ml wells of 24-well tissue culture plates and culture at 37°C in 5% CO_2. Medium changes were performed on day 3 and every 4 days thereafter with R10 containing 10 U/ml rIL2 (Roche Diagnostics, Mannheim, Germany) replacing the rIL7. Peptide-specific CTL lines were tested in ^{51}Cr release cytotoxicity assays after 10 days of stimulation.

5.2.8 Establishment of EBV-transformed B-cell lines

Blood was collected from volunteers and peripheral blood mononuclear cells (PBMC) isolated by Ficoll-Hypaque density gradient centrifugation. PBMCs were used to

establish EBV-transformed B cell lines (Sugden & Mark, 1977, Tosato *et al.*, 1982). In Brief, 1-10 x 10^6 PBMCs were resuspended in 5 ml of RPMI containing 20% heat inactivated Fetal calf serum (FCS), 0.01% of 200 mM L-glutamine, 0.01% of 100 mM Na-Pyruvate, 0.01% 10000 U/ml Penicillin and 10,000 µg/ml streptomycin. 5 ml filtered EBV supernatant prepared from the EBV-cell line B95-8 marmoset cells and 10 µl of cyclosporin A (1 mg/ml) was added and the cells incubated at 37°C with 5% CO_2. Once transformation was apparent EBV-transformed BLCLs were expanded and maintained in the above medium.

5.2.9 Cytotoxicity Assay

Cytotoxicity assays were performed using peptide-specific cell lines as effector cells and autologous EBV-transformed B cell lines (BCL) or BCL presenting appropriate HLA-class I molecules as target cells. A peptide that contains a known A*02 restricted CTL epitope of Influenza (Flue MP.58) was used as a positive control in donors that had the A*02 allele, while the HIV RT.476 peptide, which contains a HIV CTL epitope, was used as a negative control only in donors that were known to be HIV negative and with their permission. Briefly, BCLs were labelled with 100 µCi $Na_2{}^{51}CrO_4$ (Amersham Pharmacia Biotech, Buckinghamshire, England) for 1 hour, washed three times, and pulsed with 10 µM peptide for an additional 1 hour at 37 °C in 5% CO_2. Target cells were washed three times and then incubated with effector cells at indicated effector to target (E/T) ratios at 37°C and 5% CO_2 for 4 hours. Cellular release of ^{51}Cr into the supernatant was measured by using a Top Count Microplate scintillation counter (Packard Instrument Company, Meridien, CT), and the percent specific lysis was calculated by the following formula: percent lysis = [(experimental release – spontaneous release)/(maximum release – spontaneous release)] X 100. Results are reported as the mean of either duplicate or triplicate values, depending on the amount of available effector cells.

In some experiments, BCL target cells were infected with recombinant vaccinia virus expressing the RSV N-protein (obtained from Dr. J. Beeler, FDA, USA), the HIV Nef protein (vT197 (96ZM651 Nef, Subtype C)) or wild-type vaccinia virus as negative controls, (Vaccinia NYCBH) (These reagents were obtained through the AIDS research and reference reagent program, Division of AIDS, NIAID, NIH: VT197 from Therion Biologics corporation) at a multiplicity of infection of 5 for 1 hour at 37°C in 5% CO_2. Excess vaccinia virus was removed by washing with R10 and cells were

placed in 10 ml R10 overnight in a 25 cm^2 area tissue culture flask. The following day, a chromium release assay was performed as described. Cold unlabelled competitor target cells were added to the assay plates at a cold: hot target cell ratio of 40:1 when autologous target cells were used.

5.2.10 Cell subset depletion studies

Cells were sorted using magnetic Dynabeads® coated with anti-CD4 or anti-CD8 monoclonal antibodies according to the manufacturer's recommendations (Dynal Biotech, Oslo Norway). Briefly, PBMCs or peptide-specific cell lines were labelled with pre-washed T-cell subset-specific Dynabeads at a bead to cell ratio of 5:1. Following 15 minutes of incubation at 4 °C, cells were washed with RPMI containing 1% FCS before the beads were adhered to a magnet for 2 minutes. The supernatant was collected; the cells washed and then concentrated by centrifugation for their subsequent use in chromium release or ELIspot assays. Cells sorted with anti-IgG or anti-CD19 beads were used to control for the amount of cells used in further experiments. The depletion efficiency was measured by staining with fluorescently-labelled CD4 or CD8-specific antibodies, followed by flow cytometric analysis.

5.3 RESULTS

5.3.1 Identification of an RSV-specific CTL response by IFN-gamma ELIspot and ICC

Thirty-seven healthy South African adults from various different ethnic groups were screened by IFN-gamma ELIspot for reactivity to 127 overlapping 14-mer peptides that cover the complete nucleoprotein (N) of RSV. Of these 21 blood donors had a peptide-specific memory response of higher than 40 sfu/million, while 17 reacted to at least one peptide with more than 50 sfu/million cells. Each subject had a unique profile of reactivity to the N-protein peptides as shown in Figure 5.1. Figure 5.2 summarises the total number of donors that reacted to each of the individual peptides in IFN-gamma ELIspot. Blood was collected from donors that reacted favourably for establishment of EBV transformed B-cell lines and HLA typing (Table 5.1). The 14-mer peptides that had reactivity with most donors in ELIspot were peptides N-16, N-78, N-84, N-85, N-86 and N-100, which were selected for further characterisation. Two spot phenotypes were observed, fine small spots (F) usually present in a large number of cells, and dark big spots usually present in a smaller amount of cells. The ELIspot results (in SFU/million) of donors reacting to each of the selected peptides are summarised in Table 5.2

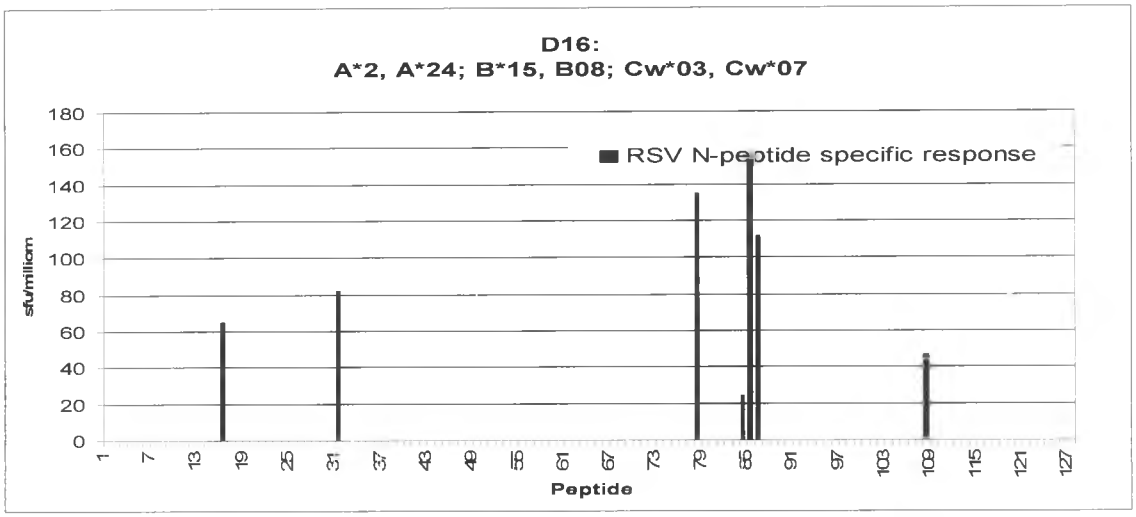
Seventeen reacting donors for whom EBV cell lines could be established were HLA typed. Donors that had the strongest reactions to the selected peptides in ELIspot were used in subsequent studies. Low frequencies of IFN-gamma release could be detected in some donors after stimulation with each of the peptides N-16, N-78, N-84, N-85, N-86 and N-100, in each case due to a $CD8^+$ response (data not shown). The frequency of $CD8^+$ RSV N-peptide-specific $CD8^+$ T-cells induced by the selected peptides was low, ranging from 0.02-0.07% of $CD8^+$ T lymphocytes. The low precursor frequency of these cells was not surprising given that these responses were elicited in healthy donors who were not known to be recently reinfected. The ICS procedure confirmed the ELIspot results and suggested peptides that induced a $CD8^+$ response that could be used for further analyses. The N-84/N-85/N-86 cluster of nested peptides that was detected in a significant group of donors by ELIspot (Figure 5.1) also induced intracellular interferon gamma signal as detected by the

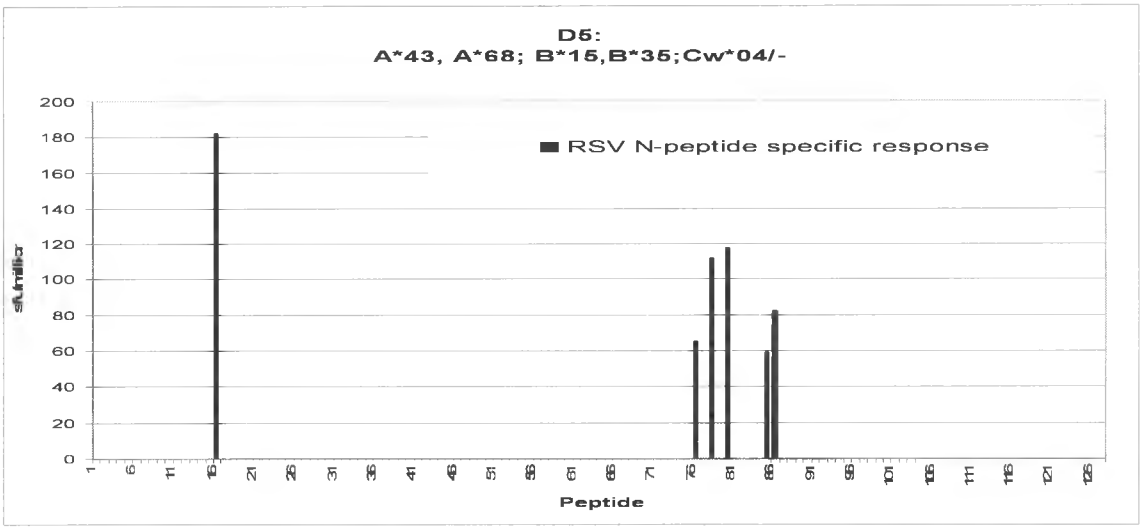
ICS assay. These peptides elicited the strongest and most reproducible response by ICS and therefore these peptides were further characterised in the most detail. Representative data for ICS analysis using these peptides is shown in Figure 5.3 for Donor 16 (HLA A2/24, B8/15, Cw3/7). Flow cytometric detection was used to enumerate the percentage of gated CD8+ T cells that exhibited a high level of the surface activation marker CD69 that also had a significant increase of intracellular gamma interferon. The donor shown had CD8+ T cells that recognised each of the three N-84/N-85/N-86 peptides, while a negative control peptide did not induce detectable interferon gamma. This finding (and assays with other donors, data not shown) suggested that these peptides contained a CD8+ restricted CTL epitope or epitopes. Gating of the CD4-negative populations suggested that N-86 induced both CD4+ and CD8+ responses which may indicate that it also contain a CD4+ restricted epitope. Subsequent staining of Donor 16's CD4+ cells suggested IFN-gamma was also induced in 0.05% of CD4+ cells following stimulation with N-86 however this response was not pursued any further. N-84 and N-85 did not induce significant IFN-gamma release in CD4+ cells (0.01 and 0.00 respectively) (results not shown). The ICC procedure was useful for assistance in confirmation of the ELIspot results and in selecting peptides that induced a CD8+ response that could be used for further analyses.

Table 5.1: HLA types and ethnic groups of subjects selected after IFN-gamma ELIspot screen for which EBV cell lines were established and additional cell lines used for determining HLA restrictions

Subject_ID	HLA_A1	HLA_A2	HLA_B1	HLA_B2	HLA_Cw1	HLA_Cw2	Ethnic group (P/M)
D_01	A*30	A*34	B*44	B*57	Cw*04	Cw*18	S.Sotho/S.Sotho
D_03	A*02	A*03	B*07	B*51	Cw*01	Cw*07	Caucasian (RSA)
D_05	A*43	A*68	B*15	B*35	Cw*04		Tswana/Tswana
D_08	A*23	A*68	B*07	B*58	Cw*06	Cw*07	S.Sotho/S.Sotho
D_10	A*02	A*26	B*15	B*40	Cw*02	Cw*04	Cape Coloured/Cape Coloured
D_12	A*31		B*08	B*60	Cw*07	Cw*10	Caucasian/Caucasian (RSA)
D_16	A*02	A*24	B*08	B*15	Cw*03	Cw*07	Caucasian (RSA)
D_17	A*02	A*23	B*15	B*58	Cw*02	Cw*07	Tswana/Tswana
D_18	A*02	A*29	B*18	B*44	Cw*05	Cw*07	Zulu/Swazi
D_23	A*66	A*33	B*42	B*58	Cw*06	Cw*17	Zulu/Xhosa
D_24	A*30	A*68	B*08	B*42	Cw*07	Cw*17	N.Sotho/N.Sotho
D_25	A*02	A*68	B*15	B*49	Cw*03	Cw*07	N.Sotho/N.Sotho
D_28	A*03	A*11	B*07	B*55	Cw*07	Cw*09	Caucasian (RSA)
D_31	A*02	A*25	B*18	B*44	Cw*07	Cw*12	Caucasian (RSA)
D_36	A*23	A*68	B*15	B*15	Cw*03	Cw*16	Zulu/Zulu
D_37	A*30		B*08	B*58	Cw*06	Cw*07	Zulu/Zulu
D_40	A*24	A*31	B*35	B*52	Cw*12		Indian/Indian
CTL_012 [1]	A*02	A*02	B*44	B*40	Cw*05	Cw*10	Caucasian (USA)
HIVNET [2] 6150009-3	A*23	A*74	B*08	B*58	Cw*02	Cw*06	South African Black

[1] and [2] Additional cell lines used for determining HLA restrictions, subjects were however not part of this RSV screen





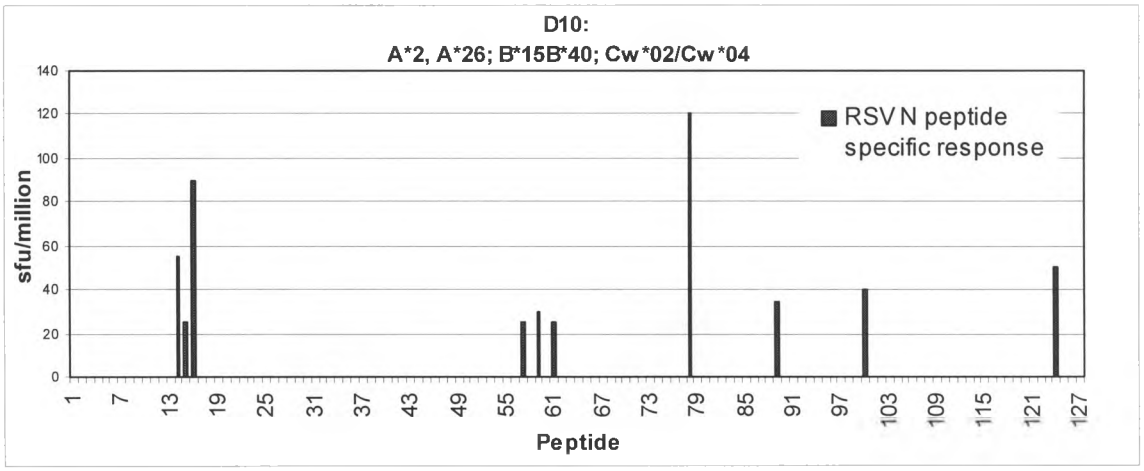


Figure 5.1: Reaction profiles of subjects D16, D5, and D10 to individual N-protein 14-mer peptides as measured in IFN-gamma ELIspot.

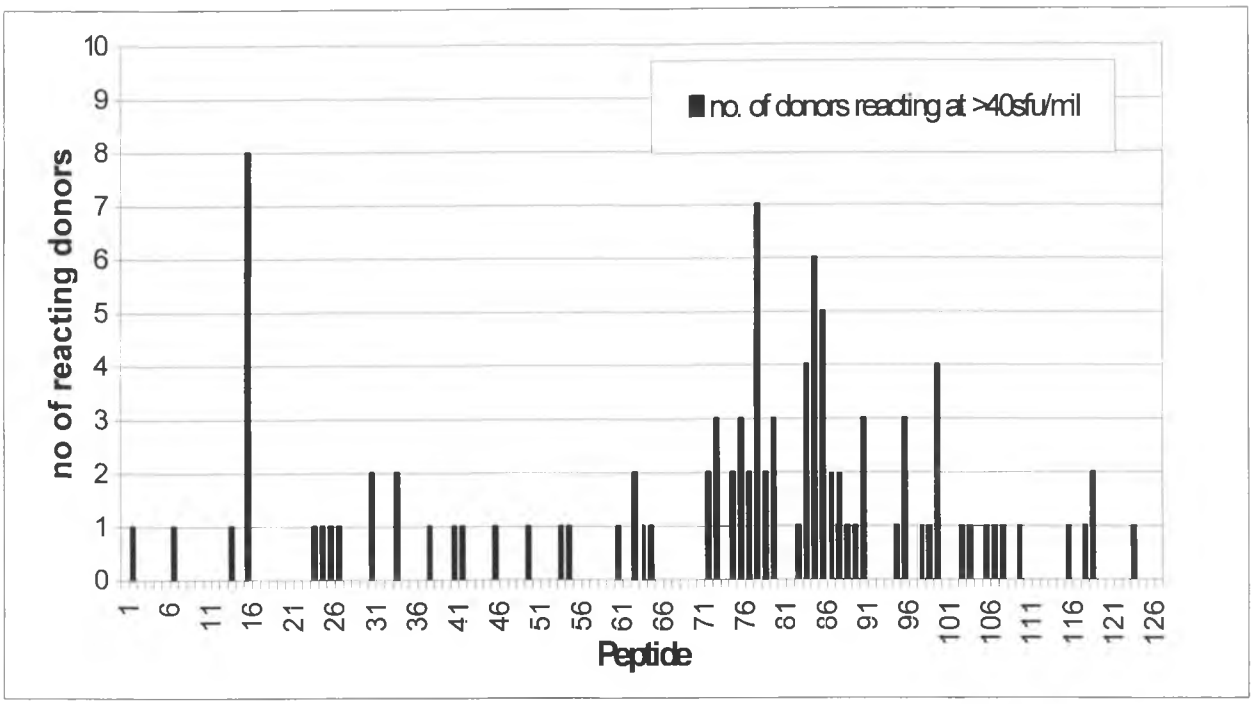


Figure 5-2: The ELIspot screening assay used to test PBMCs obtained from 37 healthy adult subjects for IFN-gamma production against a panel of 127 individual peptides spanning the RSV N-protein. The number of donors reacting to each individual peptide is indicated on the Y-axis and the peptide number on the X-axis.

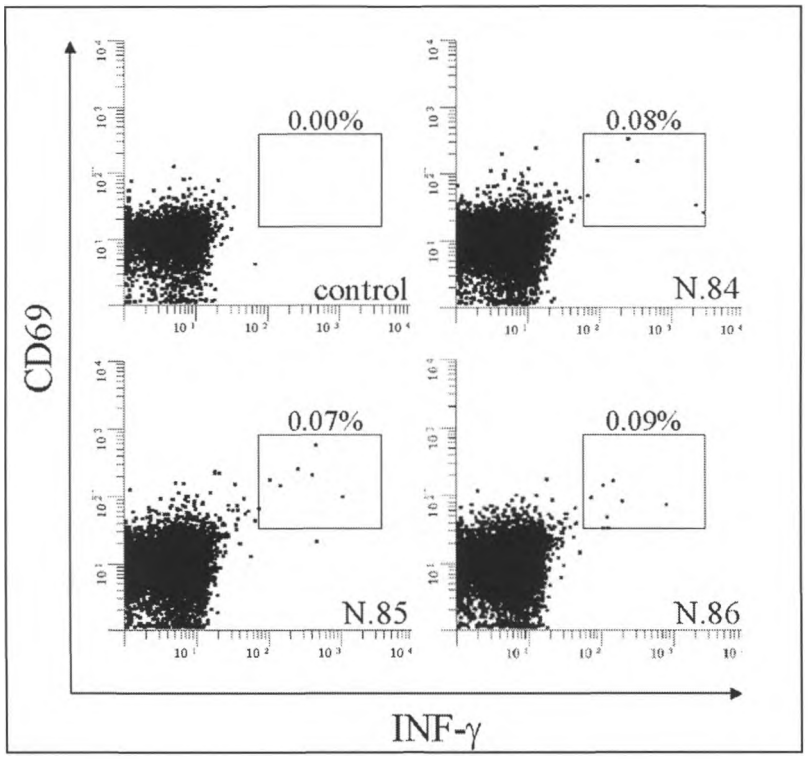


Figure 5.3: Analysis of peptide-reactive CD8+ T cells from subject D16. PBMC were incubated in the presence or absence of indicated peptides for 6 hours in an IFN-gamma/ICC assay. The data are presented as the percentage of CD3+/CD8+ lymphocytes that were CD69+ and scored positive for IFN-gamma production.

5.3.2 Characterisation of RSV CTL epitopes by chromium release assay

Since ELIspot and ICC assays measure the ability of T-cells to produce IFN-gamma but not their ability to kill target cells, the selected peptides were further analysed by chromium release assay for cytolytic ability. Peptide-specific CTL lines were generated using fresh PBMCs and then used as effector cells after 10 days in culture. N-84 and N-85 peptide-specific cell lines induced the highest percentage lysis (>40% at effector: target ratios of 50:1) in autologous target cells of D16 (A*02/ A*24; B*08/ B*15; Cw*03/ Cw*07) that were pulsed with the peptides N-84 and N-85, respectively (Figure 5.4 a.). N-85 also induced superior lysis to the other peptides when tested for D5, D24 and D31 (Results not shown). N-86 induced lyses in autologous cells of D16 and D5. N-16 and N-78 induced only week lysis (<20% at an effector to target ratio of 50:1) when tested for D24 and D16, while N-100 did not induce any lysis in autologous cells of D10 and D31 (summarised in Table 5.2). Peptide-specific effector cells generated against a control peptide did not lyse pulsed target cells. The selected control peptide was an N-protein specific peptide that did not react with the subject's PBMCs in IFN-gamma ELIspot. The HIV RT.476 peptide was only used as negative control for D16 for whom the HIV status was known to be negative and with permission of the donor.

Target cells that were mismatched to all but one HLA allele were used in chromium release assays to define the HLA restriction of epitopes recognised on these peptides. N-84, N-85 and N-86 induced superior lysis to the other peptides, however since N-86 was shown to elicit IFN-gamma secretion in both CD4+ and CD8+ cells further experiments were focussed on the former two peptides. N-84 and N-85 specific lysis were strongly induced in B*08- restricted target cells. A*02- restricted lysis was also induced by N-85 and to a lesser extent by N-84 (Figure 5.4.b). The A*02-restricted lysis was also reflected in A*68 single matched target cells when D24's effector cells were used which could be explained by the fact that A*68 is a member of the A2 supertype. N-84 induced lysis in B*15-restricted target cells, however this response was much weaker than that seen in B*08 (Table 5.2). The highest cytolytic activity induced by all 6 the selected peptides after subtraction of background lyses of unstimulated target cells, the number of sfu/million induced in IFN-gamma ELIspot, and the restrictions determined by lysis in mismatched target cells are summarised in table 5.2.

Table 5.2: RSV N-protein specific IFN-gamma production and cytolytic activity to selected peptides

Peptide:	AA sequence of peptide.	Position on N	subject	SFU/ MIL[a]	%lysis[b] 50:1	%lysis 25:1	%lysis 12.5:1	Restriction
N-16	KLCGMLLITEDANH	46	D16	65F	16.3	9.6	9.1	A*24
			D5	182F	16.5 13.9	16 13.5	14 13.4	B*35 B*15
			D11	65F				
			D7	105F				
			D10	90F				
			D1	176F				
			D24	40F				
			D8	75F				
			D31		12	6	4	B*44
N-78	STRGGSRVEGIFAG	232	D16	135F	12 14.5	10 8.5	9 7.6	autologous A*24
			D5	112F				
			D10	120F				
			D1	59F				
			D24	40		12 12	6 5	A*30 B*08
			D8	35F				
			D25	268F				
N-84	AYGAGQVMLRWGVL	250	D35	250F				
			D37	45				
			D38	105				
			D16	24	13 28 33	14 20 24	18 12 17	A*02 B*15 B*08
			D17	30				
			D24		28	29 20	12 16	A*68 B*08
N-85	AGQVMLRWGVLAKS	253	D16	153	25 15	24.4 19	24 10	A*02 B*08
			D5	59	24	22	16	autologous
			D31	146	27	22 17	17 16	autologous A*02
			D28	90				
			D35	233				
			D37	75				
			D24	30		23 21	15 13	A*68 B*08
N-86	VMLRWGVLAKSVKN	256	D16	112	25	21	20	A*24
			D5	82	23	20.6	20	B*15
			D31	190				
			D12	45				
			D38	56				
N-100	AGFYHILNNPKASL	298	D10	40				
			D31	56	10	7	7	B*44
			D25	43				
			D40	50				
			D18	37f				
N9-254		254	D16		16	13	8	autologous
N9-255	QVMLRWGVL	255	D16	45	85 19 15.3	75 17.6 13.2	52 17 13	B*08 A*02 B*15
N10-256	QVMLRWGVLA	255	D16	40	36	25		B*08

[a] IFN-gamma ELIspot results in spot forming units/million PBMCs, Fine spot phenotype is indicated by a F, the rest were large spots.
[b] %Lysis measured by Cr^{51} release assay after subtraction of unpulsed control target cells.

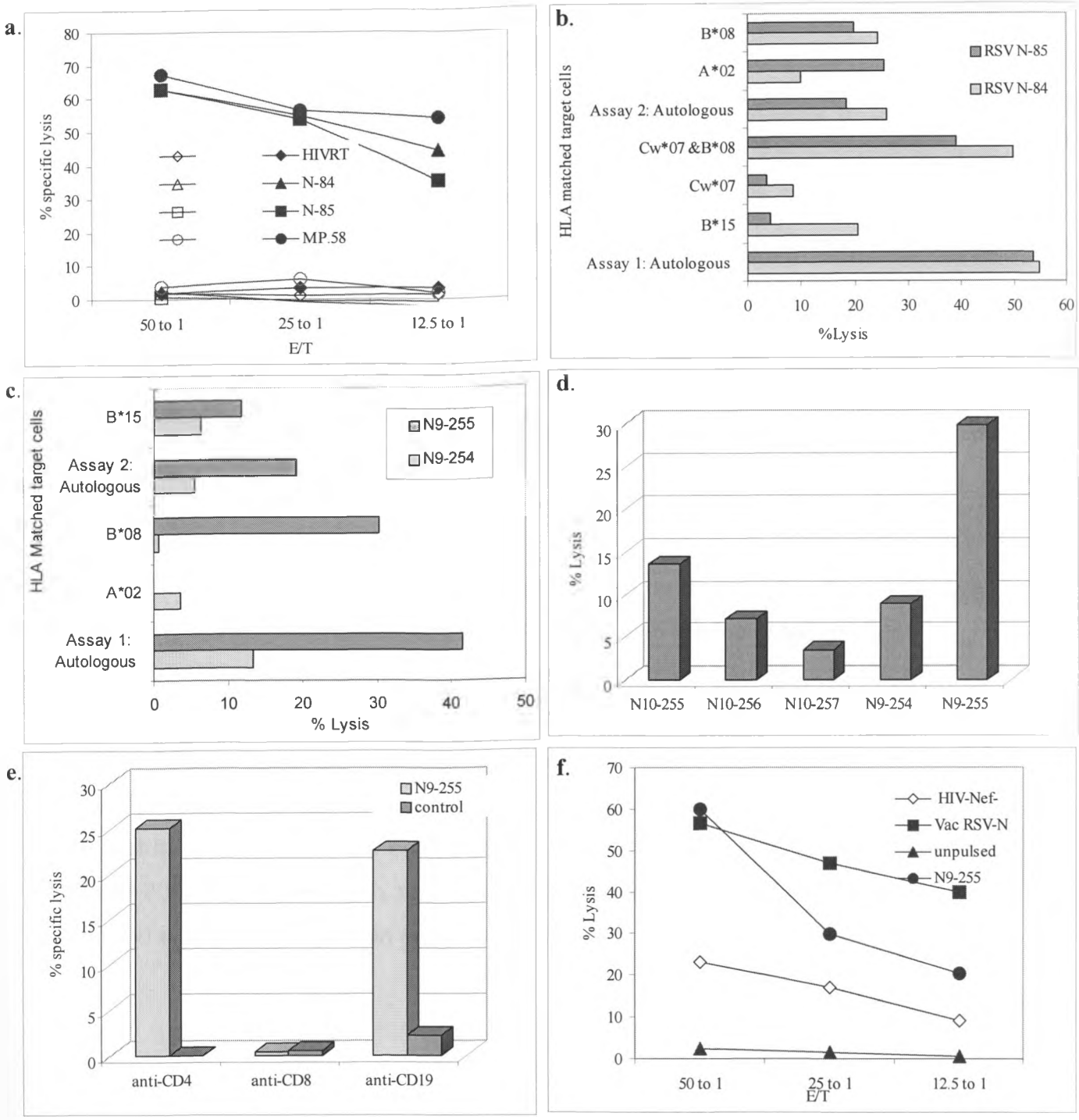


Figure 5.4: Recognition of viral epitopes by peptide-specific effector cells generated from subject D16 (HLA-type indicated) in standard ^{51}Cr-release assays. (See rest of legend description on next page)

Figure 5.4: Recognition of viral epitopes by peptide-specific effector cells generated from subject D16 (HLA-type indicated) in standard ^{51}Cr-release assays. (see Figure 5.4 on previous page)

The data are presented as the percent specific lysis at an effector to target ratio (E:T) of 25:1 (b.-e.) or at the indicated E:T ratios (a. and f.):

(a.) Effector cells were peptide specific cell lines generated with Flu MP.58 (■), HIV RT.476 (♦), RSV N-84 (●), or RSV N-85 (▲) peptides. Target cells were autologous BCL pulsed in the presence (closed symbols) or absence (open symbols) of the indicated peptides.

(b, c.) Effector cells were RSV N-84 or RSV N-85 (b), or RSV N9-254 or RSV N9-255 (c), peptide-specific cell lines. Target cells were autologous or HLA single-matched BCL (corresponding HLA-alleles are indicated) pulsed in the presence of the indicated peptides. Autologous controls for each set of single matched target cells are included (assay 1 and 2).

(d) Peptide-specific effector cells were generated and assayed for cytotoxicity against HLA-B*08-matched BCL target cells pulsed with the indicated peptide.

(e) RSV N9-255 peptide-specific effector cells were depleted of CD4+, CD8+, or CD19+ cells by magnetic beads and assayed for cytotoxicity against HLA-B*08 restricted BCL pulsed in the presence or absence of peptide RSV N9-255.

(f) Effector cells were RSV N9-255 peptide-specific cell lines. Target cells were HLA-B*08-restricted BCL that were pulsed in the presence (●) or absence (■) of peptide RSV N9-255, or infected with recombinant vaccinia virus expressing the HIV nef protein (♦) or the RSV N protein (▲).

5.3.3 Mapping of optimal epitopes using 9- and 10-mer truncated peptides

To identify the exact position of the epitopes within the 14-mer peptides that induced the strongest cytolytic reactions (N-84 and N-85), CTL-epitope prediction algorithms BIMAS and SYFPIETHI were used to aid in the selection of 8-, 9- or 10-mer peptides for finer mapping. A B*08-restricted 9-mer epitope was predicted at position 255 (QVMLRWGVL) in the overlapping region of the two peptides N-84 (amino acids 250-263) and N-85 (amino acids 253-266). This peptide had the highest score of all peptides within the RSV N protein for B*08-predicted CTL epitopes, scoring 80 with the BIMAS algorithm. The two 9-mer peptides located at amino acids 254 and 255 (designated N9-254 and N9-255) were synthesized and used in chromium release assays with HLA-matched or -mismatched target cells, in order to map the optimal epitope within these longer peptides. The N9-255 peptide induced a high level of lysis of B*08 matched cells, whereas the N9-254 peptide did not (Figure 5.4.c and 5.4.d), suggesting that the initial amino acid of the B*08 epitope is at amino acid 255. To further confirm the position of the optimum B*08-restricted epitope, truncated 10-mer peptides covering the remainder of N-85 (N10-255, N10-256, N10-257) were synthesised and tested in chromium release assays (Figure 5.4d). The results showed that the 10-mer peptide starting at position 255 contains a lytic epitope. Nevertheless, the 9-mer peptide starting at that position induced a nearly 2-fold higher % specific lysis, and thus is the optimal B*08-restricted epitope at that position.

The N-85 peptide also appeared to induce A*02-restricted cytolysis in the majority of assays tested with Donor 16 (Figure 5.4b), however the specificity of this epitope could not be mapped with truncated peptides (see Figure 5.4c), perhaps due to variability in reactivity to this peptide over time. This peptide had the highest score of all peptides within the RSV N protein for A*0205-predicted CTL epitopes, scoring 84 with the BIMAS algorithm. The second highest scoring predicted epitope located at position 254 (GQVMLRWGV) is predicted to be restricted to A*0201.

5.3.4 Confirmation of MHC class I phenotype

To determine the phenotype of CTLs induced by the mapped epitope, N9-255 - peptide specific CTL lines were depleted on day 10 of culture using magnetic beads for CD4+, CD8+ and CD19+ cells and then used as effector cells in chromium release assays. Effector cells that were depleted for CD4+ cells lysed target cells at least as effectively as CD19+ depleted cells, however in CD8+ depleted fractions, lyses was reduced to that seen for the negative control (Figure 5.3.e.). The depletion was confirmed by flow cytometry and found to be more than 99% effective (results not shown). This confirms that N9-255 is a MHC class I specific CTL epitope.

The N9.255 epitope is a naturally presented epitope. To confirm that the putative CTL epitope identified with synthetic peptides is presented after processing of the viral protein in cells, target cells were infected with a recombinant vaccinia virus expressing the RSV N protein (Vac-N) and tested by chromium release assay. N9.255 peptide-specific effector cells lysed target cells infected with Vac-N but not target cells infected with wild-type Vaccinia virus or an HIV-Nef recombinant vaccinia virus (Figure 5.4.f). Effector cells stimulated with the N-13 control peptide did not lyse Vac-N infected target cells (data not shown). This finding suggested that the N9.255 B*08-restricted CTL epitope is naturally processed and presented in target cells.

5.4 Discussion

This study demonstrates that a memory CTL response to RSV can be detected in healthy adults using IFN-gamma ELIspot, and describes the characterisation of a dominant CTL epitopic domain recognised by a diverse HLA repertoire. Although the responses were usually present at a low frequency in these healthy adults who were not recently infected, the screening procedure effectively identified potential epitopes that could be further characterised with classical techniques.

Thirty-seven healthy adults of a wide diversity of HLA types were screened for reactivity to overlapping peptides that cover the complete N-protein. Previous studies have indicated that screening of peptide libraries in 30-50 previously infected individuals easily and rapidly defines the epitopes for the most common HLA types (Altfeld *et al.*, 2000; Goulder *et al.*, 2001b). The N-protein was selected because this is the most conserved protein both within and between the RSV subgroups and has been reported to be the major target for the CTL memory response in humans (Cherrie *et al.*, 1992a). Using IFN-gamma ELIspot as a marker for CTL activity, 21 donors were identified with a response of more than 40 sfu/million cells to at least one N-peptide. This was selected as a cut-off point for a positive response, however 17 of these had responses of more than 50 sfu/million cells and a total of 32 had responses of more than 20 sfu/million. In some cases, CTL epitopes may induce SFC frequencies of less than 40 SFC/10^6 cells. For example, the identified B*08 epitope that was located on the 14-mer N-84, induced only 24 SFC/10^6 in ELIspot using PBMCs from Donor 16, however when tested by chromium release assay after 10 days in culture the peptide induced very strong lysis by this same donor's effector cells. This finding suggested that the number of spots observed in ELIspot does not always correlate directly with the strength of cytolytic activity that can be induced by the peptides. Peptides that induced strong IFN-gamma release in a smaller number of cells, observed as a small number of large spots in ELIspot (e.g. peptides N-84 and N-85), appeared to induce stronger specific cytolysis than peptides that induced low levels of IFN-gamma in many cells, (suggested by a high number of small spots in ELIspot, for example peptide N-16).

The screening results suggested that although different individuals may have different patterns of reactivity to individual peptides, a dominant response exists against a small number of peptides, suggesting the existence of epitopic domains with reactivity in different HLA types. This is in keeping with findings for HIV where a small number of highly immunogenic regions on the Gag-protein have been identified that contain two-thirds of the dominant Gag-specific epitopes irrespective of virus clade, ethnicity or age group studied (Goulder *et al.*, 2000a). The strongest epitope identified in the present study was mapped to a 9-mer in position 255. This epitope was located on peptides N-84 and N-85 (amino acids 250-266). Seventeen subjects had a reaction to either one or both of these peptides during the IFN-gamma ELIspot screening, of which 7 (19% of total) had a response of higher than 40 sfu/million. The epitope was shown to induce a strong B*08-restricted response. N-84 also induced B*15 restricted lysis for effector cells of D16 while peptide N-85 also induced A*02 restricted lysis, suggesting that a B*15- and A*02-restricted epitope may also be located on these peptides however these could not be mapped with the selected 9mer peptides. When studying the HLA types of subjects that reacted to either N-84 or N-85 in IFN-gamma ELIspot, the MHC class I alleles B*08, A*02 and B*15 were identified in both Caucasians and Africans (Table 5.1 and 5.2). The A*68 allele, a member of the A*02 supertype was identified in certain African individuals who reacted to peptides N-84 and N-85 but had neither B*08, B*15 nor A*02 alleles. One reacting subject (D28) had none of these HLA types however, this person had a B*07 allele which is part of the B7 supertype to which B*15 also belongs.

Since only low resolution HLA typing was performed, the exact allele specificity of the subjects was not known. The identified B*08-restricted epitope on N9-255 also corresponded to a predicted A*0205-restricted epitope while N9-254 was predicted to be an A*201-restricted epitope by the BIMAS algorithm. Although N9-255 proved to be a potent B*08 restricted epitope, neither of these two peptides could induce the same level of cytolytic activity than peptide N-85 in A*02-restricted target cells, suggesting that another A*02 restricted epitope may be located on peptide N-85. The single matched A*02 B-cell line that was used to determine the restrictions by chromium release assay was from a north American donor, where A*0201 is found predominantly (Goulder *et al.*, 2000a) while the peptide-specific CTL line was

generated from a South African subject. These subjects may have different A*02 subtypes which may result in sub-optimum peptide presentation. A*0201, A*0205 and A*6802 are all members of the A2-supertype (reviewed in Sette & Sidney, 1999a), thus some cross-reactions may however be expected. The BIMAS algorithm did accurately predict the position of the B*08-restricted epitope. This confirms the usefulness of epitope prediction algorithms to help select putative peptides for synthesis, however the experience in this study also suggests that in practice an initial screening process is necessary to determine the study group's individual CTL responses.

The phenotype of the cells activated by the epitope N9-255 was confirmed by ICC and CD4+/CD8+ depletions followed by chromium release assay or ELIspot. No cytolytic activity was induced in chromium release assays when CD8+ cells were depleted and no IFN-gamma stimulation in CD8 negative cells was detected by ICC, confirming that peptide N9-255 is a CD8+ CTL epitope. The 14-mer N-86 induced intracellular IFN-gamma secretion in both CD8+ and CD4+ cells when tested by ICC and may thus also contain a CD4+ epitope. Since N-86 induced IFN-gamma secretion in the detected CD4+ T cells, this suggests a possible type 1 T-helper cell response. MHC class I molecules bind short peptides of 8-10 amino acids (Bouvier & Mwiley, 1994) while peptides binding MHC class II molecules are usually between 10-20 residues long with sizes between 13 and 16 amino acids most frequently observed. The mapped 9-mer epitope N9-255 lacked any CD4+ reactivity suggesting that it should be a safe measure of CD8+ CTL. However, these results show the importance of using peptides of optimum length and confirming both IFN-gamma ELIspot and ICC results by use of cytolytic assays when CTL epitope mapping is performed.

South Africa has a very diverse HLA repertoire. A2 is detected frequently in both the southern African black (16.06%) and Caucasoid populations (29.7%) (A. Puren, HLA laboratory, National Institute for Communicable diseases, unpublished data). The A*02 allelic frequency ranges from 7.8-25% in the different South African ethnic groups (Xhosa, Sotho, Zulu, Shona and Matabele), but the A*201 subtype is rare in Zulus for whom A*207 is detected frequently (Hammond *et al.*, 1997). A*68 is detected in 2.48% of South African Caucasians and in 12.04% of the Black

population. B*08 is also relatively common in both South African Caucasoid (11.62%) and the southern African Black populations (7.44%), while B*15 is detected frequently in both groups with 13.3% and 18.51% in the Caucasian and Black population, respectively (A. Puren, unpublished data). This suggests that the epitope identified in this study will be useful for studies of RSV CTL responses in South Africa.

Although only the B*08 restricted epitope located on the peptides that induced the strongest lysis (N-84 and N-85) were mapped up to 9mer level, chromium release assays also suggested possible B*15 and A*02-restricted epitopes on N-84 and N-85 respectively. The ELIspot, ICC and chromium release assays results also suggested that further epitopes may be located on the 14-mers N-16, N-78 and N-86 that are restricted to further HLA types (Table 5.2). These include potential epitopes restricted to HLA A*24 (found in 1.64% of the South African black population and 8.42% of Caucasians); A*30 (22.26% South African Black-, 8.42% Caucasian population), B*35 (1.91% South African Black-, 8.08% Caucasian population) and B*44 (6.30%South African Black-; 12.63% Caucasian population) (A. Puren, unpublished data). Although these peptides induced only relatively weak lyses by memory CTL in healthy adults they may also be considered when studying the CTL response in recently infected children, in whom the responses may be stronger, especially to define the difference between the CTL responses observed in adults and children. Mapping of these epitopes may also be more effective in recently infected individuals where the frequency of memory cells may be higher.

It is unclear how long CTL memory to RSV may last in humans. This study did indicate that a detectable response might be found to the N-protein in at least 56% of healthy adults in a diverse population. It may be possible that the responses detected in certain individuals by ELIspot may have waned by the time chromium release assays were performed (> 6 months) which may be the reason why a cytolytic response could not be detected to some of the selected peptides that induced a dominant response in the screening process. However, the process described here proved to be an effective method for rapidly defining CTL epitopes to RSV in humans. These epitopes will be useful for further investigations of the CTL response in children and may contribute to vaccine development.

In summary, RSV-specific memory cytotoxic T-cell epitopes were identified in adults using IFN-gamma ELIspot and confirmed by cytolytic assays. Although the frequency of responding cells were in general quite low, these epitopes were most likely involved in clearing natural RSV infection. A population with a diverse range of HLA types was screened with RSV N-peptides and the peptides that elicited a dominant response selected for further characterisation. Finer mapping of the 14-mer peptides that induced the highest cytolytic activity identified a 9-mer MHC class I epitope, N9-255, that was recognised strongly by B*08-restricted target cells. The overlapping 14-mer peptides (N-84 and N-85) containing this epitope also induced lysis in the context of A*02- and to a lesser extent B*15 restricted target cells in some individuals. These alleles are relatively common in the South African population and will be useful for studies in this population. The epitopes were detected in healthy adults, reflecting the response generated in the course of previous natural RSV infection. Identification of these epitopes is a step forward in the process of mapping the total CTL response to the N-protein of RSV and will contribute to characterising the role of RSV-specific CTLs in pediatric infection.

CHAPTER 6

Conservation of Nucleoprotein-Specific Cytotoxic T-cell Epitopes of Human Respiratory Syncytial Virus

6.1 Introduction

In the light of the genetic and antigenic variation detected in respiratory syncytial virus, the identification of a broadly cross-strain protective immune response will be of significant value for vaccine development. This makes identification of CTL epitopes that will provide protection against the different genotypes and for both subgroups an important goal. However, one of the methods used by viruses to evade immune surveillance is the introduction of amino acid mutations within CTL epitopes or in sequences flanking these epitopes (Oldstone, 1997). The epitope sequences themselves are essential for association with the MHC class I molecule and for recognition by the virus specific CTLs, while the flanking sequences are important in cytosolic processing of the proteins to yield the 8-, 9-, or 10-mer peptides that constitute CTL epitopes. Mutations in these regions may be associated with loss of CTL-mediated lysis of virus infected cells (Apolloni *et al.*, 1992; Yellen-Shaw *et al.*, 1997); or may generate peptides that antagonise CTL function (Bertoletti *et al.*, 1994; Klenerman *et al.*, 1994). Previous reports have suggested that the N-protein is the major target of the memory CTL response to RSV in humans (Bangham *et al.*, 1986; Cherrie *et al.*, 1992b). Although it is known that the nucleoprotein is the most conserved of all RSV proteins (Johnson & Collins, 1989), it is not known if the variation that does exist occurs in the regions that contain the CTL epitopes, or if immune pressure may induce positive selection in the CTL epitopes.

Escape mutations in CTL epitopes have been described for several persistent infections like HIV (Barouch *et al.*, 2002; Goulder *et al.*, 2001b; Price *et al.*, 1997) and EBV (Bertoletti *et al.*, 1994). Although CTL escape mutations have been

documented in chronic diseases, it was thought until recently that such mutations would be unlikely to occur in acute viral infections. In influenza virus, CD8+ CTL responses are usually cross-reactive between virus subtypes (Ada & Jones, 1987; McMichael, 1994). Consistent with this is the relative conservation of the dominant CTL target antigens, which are usually internal viral proteins that are expressed early in infected cells before viral release and thus not accessible to antibody selection pressure (Price *et al.*, 1997). However, recently, there have been several reports of CTL escape mutations in influenza virus that abrogate MHC class I presentation and recognition by specific CTLs (Boon *et al.*, 2002; Price *et al.*, 2000; Voeten *et al.*, 2000). These reports suggest that antigenic drift resulting from immune pressure mediated by specific CTLs occurs and can result in escape from CTL directed against influenza (Boon *et al.*, 2002).

Limited data are available about CTL epitopes of RSV in humans, and to date no study has examined sequence variation within human CTL epitopes. To determine if the CTL epitopes recognised by South African subjects are conserved between the virus genotypes and subgroups, viral specimens were selected from all viral genotypes identified in South Africa from 1997-2001, a major part of the N-protein gene was sequenced and compared with historic RSV isolates published in Genbank. To determine the extent of conservation, the N-protein sequences were also compared with those of other members of the Pneumovirinae.

6.2 Materials and Methods

6.2.1 Viruses and RNA

Fourteen RSV specimens that had been genotyped according to their G-proteins in previous studies (Venter *et al.*, 2001; Venter *et al.*, 2002) (Chapters 2 and 3) were selected for analysis in this study. The specimens represented each of the subgroup A and B genotypes described to date in South Africa and isolated from 1997-2001. These included subgroup A genotypes GA2, GA5, GA7, SAA1, and subgroup B genotypes GB4, GB3, SAB1, SAB2, SAB3. RNA isolation was as described in Chapter 2 and 3.

6.2.2 RT-PCR

Primers were designed with Primer 3 software (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi/) using the prototype sequences A2 and B1 for subgroup A or B, respectively. The primers were as follows:

Subgroup A: 5′primer: N73AF: TCCAGCAAATACACCATCCA;
3′primer: N1017AR: CATTATGCCTAGGCCAGCAG;
Subgroup B: 5′primer: N2BF: GGGAAATACAAAGATGGCT;
3′primer: N1031BR: ATGCCTAGACCTGCTGCATT.

The reactions were performed separately for subgroup A and B in a one-step RT-PCR using the Titan™ One Tube RT-PCR system (Roche Molecular Biochemicals) according to the manufacturer's instructions. In brief, 10 μl of RNA was added to 10 μl standard 5X reaction buffer, 10 mM of each dNTP, 20 pmol each of the 5′ and 3′ primers, 5 mM DTT-solution, 10U RNase inhibitor, and 1 μl of the Titan™ enzyme mix and the volume made up to 50 μl with distilled water. The reaction was performed in a GeneAmp PCR system 9600 thermocycler (Perkin-Elmer Applied Biosystems, Foster City, USA) according to the following program: 50 °C for 30 minutes, 94 °C for 2 minutes, (94 °C for 30 seconds, 55 °C for 30 seconds, 68 °C for 1 minute) for 35 cycles, 68 °C for 7 minutes, 50°C for 30 minutes. PCR products were visualised by agarose gel electrophoresis as described in Chapter 2.

6.2.3 Sequencing and sequence analysis

PCR product purification and nucleotide sequence analysis was carried out as described in Chapter 2. In brief, PCR products were purified using the Qiaquick 8 PCR purification kit (QIAGEN, Hilden, Germany). DNA products for sequence analysis were purified with Autoseq GE-50 columns (AEC-Amersham, Sandton, South Africa). Nucleotide sequence analysis was carried out on both DNA strands using ABI Prism big dye terminator cycle sequencing on an ABI 3100 sequencer model according to the manufacturer's specifications (Applied Biosystems, Foster City, USA). Forward and reverse reactions were carried out with the same primers that were used for RT-PCR. An additional primer, N454F [CAGAATACAGGCATGACTCT], was designed to assist in sequencing the 3′ ends where necessary.

Sequences were analysed with Sequencher™ software, version 4.0.5. Sequence alignments were done with Clustal X 1.64b (Thompson *et al.*, 1997). Translations and phylogenetic-analyses were done with MEGA version 2.1 (Kumar *et al.*, 2001). Shading of amino acid alignments were carried out with Genedoc 2.6.002 **(www.psc.edu/biomed/genedoc)**. Sequences used for comparison were obtained from the Genbank: Human RSV prototype subgroup A (A2) (Collins *et al.*, 1985) and subgroup B (B1) (Karron *et al.*, 1997a); human metapneumovirus (van den Hoogen *et al.*, 2002); Avian metapneumovirus (Shin *et al.*, 2002); Bovine RSV (Valarcher & Andreoletti, 2001); (Buchholz *et al.*, 1999; Larsen, 1998; Larsen *et al.*, 1998); ovine RSV (Alansari & Potgieter, 1994); pneumovirus of mice (Barr *et al.*, 1991).

6.3 RESULTS

6.3.1 RT-PCR

Specimens were selected from each of the subgroup A and B genotypes identified to date in South Africa, and the two prototype strains RSV A2 and B1. The RT-PCR resulted in a specific band of 944 bp for all subgroup A and 1,029 bp for all subgroup B specimens confirming the specificity and sensitivity of the selected primers (Figure 6.1).

6.3.2 Amino acid sequence comparison

The amino acid sequence of the N-protein of the selected subgroup A and B specimens were compared in alignments with the amino acid sequences of other members of the Pneumovirinae, including the pneumovirus genus (human RSV, bovine RSV, ovine RSV and pneumovirus of mice) and the metapneumovirus genus (avian- and human metapneumovirus). The average between-group amino acid P-distances (calculated with Mega version 2) are shown in the distance matrix in Table 6.1. The average P-distance between human RSV subgroup A and B was 5% while the distance between human RSV and bovine or ovine RSV was 6% or 7%, respectively. Much less conservation existed with other members of the Pneumovirinae. The distances of RSV to pneumovirus of mice and the metapneumoviruses were 38% and 59%, respectively. When taking only the human RSV genotypes into consideration, much more variation was found between the two subgroups at the nucleotide level (P=0.14) than at the amino acid level (0.05), with the average synonymous substitutions per synonymous site (Ks=0.23) greatly exceeding the nonsynonymous substitutions per nonsynonymous site (Ka=0.01) (Ka/Ks=0.04), and transitions exceeding transversions (Ts: Tv= 4) (calculated with Mega, version 2.1 (Kumar *et al.*, 2001).

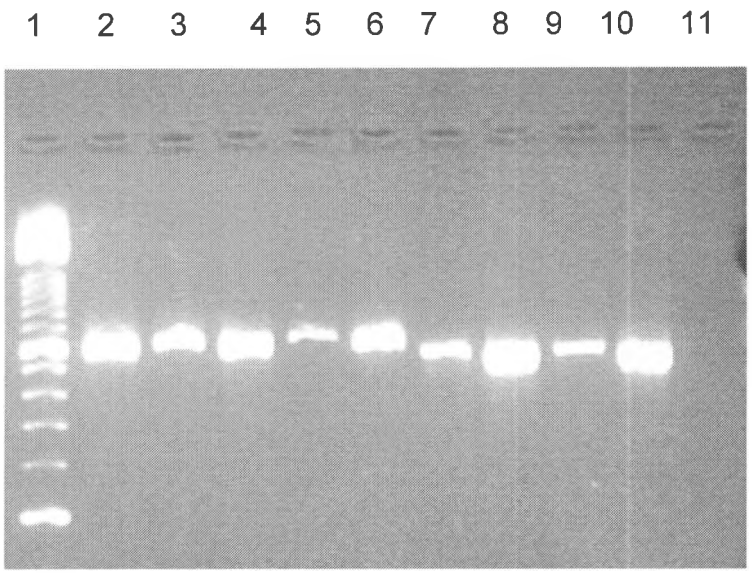


Figure 6.1: RT-PCR of RSV subgroup A and B nucleoprotein genes. 1. 100 bp ladder size marker, 2. SAB1 isolate Ab17B00, 3. SAB3 isolate Ab27CT00, 4. SAB3 isolate Ab5075B01, 5. SAB2 isolate SA99V800, 6. SAB2 isolate SA99Vr1325 7. GA5 isolate Ab20BI00, 8. GA2 isolate Ab4029BI01, 9. GA5 isolate Ab3-61CT01, 10. Cultured control A2 prototype, 11. Negative control.

Table 1: Distance matrix of the average between group amino acid P-distances of the members of the Pneumovirinae.

	Human RSV, subgroup A	Human RSV, subgroup B	Bovine RSV	Ovine RSV	Pneumovirus of mice	Human meta-pneumovirus	Avian meta-pneumovirus
hRSV-A							
HRSV-B	0.05						
Bovine RSV	0.06	0.06					
Ovine RSV	0.07	0.06	0.03				
Pneumovirus of mice	0.38	0.38	0.38	0.38			
Human meta-pneumovirus	0.59	0.59	0.59	0.60	0.55		
Avian meta-pneumovirus	0.59	0.59	0.59	0.59	0.56	0.12	

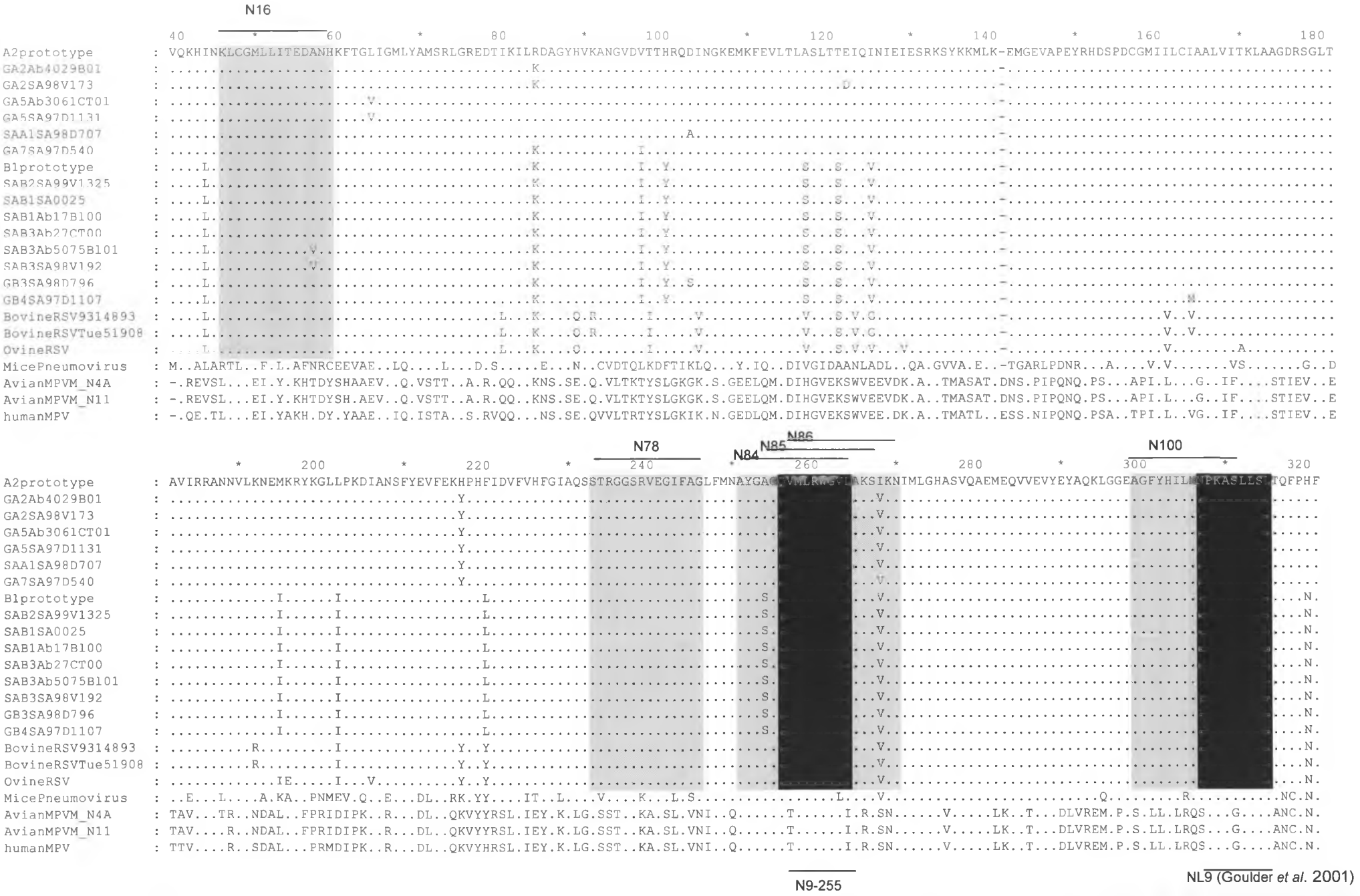


Figure 6.2: See legend on the following page.

The amino acid alignment of the N-proteins is shown in Figure 6.2. The positions of the epitopic domains identified in this study are highlighted. Epitopes mapped up to 14-mer level are shown in grey while the mapped 9-mer epitope, N9-255, and the epitope identified by Goulder *et al.* (2001) (NL9) are highlighted in black. The positions of the epitope containing peptides were mostly conserved, with the 14-mers N-78, N-100 and the mapped 9-mers N9-255 and NL9 (Goulder *et al.*, 2001) being identical between all human RSV genotypes in both subgroups as well as to the animal RSV (ovine and bovine RSV). One amino acid change was detected in peptide N-16 in genotype SAB3. Peptides N-84 and N-85 had a single amino acid difference between subgroup A and B that was conserved in the respective subgroups. One amino acid difference was detected in N-86 in all RSV specimens relative to the A2 prototype; however, this mutation resulted in amino acids of the same similarity groups (Ile and Val). The positions of the mapped epitopes N9-255 and NL9 were conserved between all human and animal RSV. The sequence of NL9 was also conserved in the pneumovirus of mice.

Figure 6.2: Amino acid alignment of the N protein sequences of human RSV subgroup A and B genotypes with bovine RSV, ovine RSV, pneumovirus of mice, and avian and human metapneumoviruses (MPV) relative to the A2 prototype. (See figure on previous page)

The genotype assignment of the specimens is indicated in the first letter of the name: subgroup A (GA2; GA5; SAA1; GA7); subgroup B (SAB1, SAB2, SAB3, GB3, and GB4). The RSV prototype strains A2 and B1 from subgroup A and B are included. The position of the six 14-mers that gave the dominant response in IFN-gamma ELIspot are highlighted in grey and the mapped 9-mer epitopes mapped in this and previous studies in black.

6.4 Discussion

Comparison of the N-protein amino acid sequences of RSV subgroup A and B specimens suggested that the epitopes N9-255 and NL9 identified in this and in previous studies (Goulder *et al.*, 2000b) were conserved in both subgroups in all RSV genotypes isolated between 1997-2001 in South Africa, and in both the historical prototype strains A2 (Wertz *et al.*, 1985) and B1 (Karron *et al.*, 1997b), suggesting a lack of immune selection in these epitopes. The specimens were selected to be as divergent as possible from all the different RSV genotypes identified over a period of five years. The fact that the amino acid sequences of the epitopes were also conserved between different hosts of RSV suggests that conservation of this region may be important to maintain an essential structural or functional role. The N-protein is essential for RNA replication, for fully processive transcription and for packaging (Collins *et al.*, 1999).

When looking at the phylogenetic relationship between the Pneumovirinae (Figure 6.3), hRSV subgroup A and B which have 5% amino acid differences are most closely related to bovine and ovine RSV, with only 7 and 8% differences in the N-proteins; this in comparison to 38% sequence dissimilarity with pneumovirus of mice and 59% with the metapneumoviruses. Several reports of positive selection in the G-proteins suggest that immunological pressure induced by the antibody response may cause immune evasion by selection for change in the surface proteins of RSV (Cane, 2001; Sullender, 2000). A point mutation is visible in the region of the N-protein sequence covered by N-16 in two SAB3 specimens which have a Val substituting the Ala found in other specimens seen in Figure 6.2 This prevents the complete exclusion of the possibility of immune selection in CTL epitopes, in specific regions of the N-protein of RSV. However, the identified 9-mer epitopes N9-255 and NL9 (Goulder *et al.*, 2000b) as well as epitopes on N-78, and N-100 are completely conserved between all genotypes in both subgroups. The substitutions visible in the region flanking N9-255 on N-84 in subgroup B as well as most other amino acid differences on the N-protein are conserved within the whole subgroup, which does not suggest that immune pressure to be involved. The conservation visible especially toward the C-terminal domain may be needed for virus replication, and so would prevent

immunological pressure from selecting for changes in the CTL epitopes. It will be of interest to determine if immune selection occurs in CTL epitopes of other proteins which are less conserved between the two subgroups, e.g., the F-protein (11% difference)(Johnson & Collins, 1988) and the SH-proteins (24%) (Collins *et al.*, 1990).

In HIV, reports of positive selection of HIV-1 proviral sequences encoding variants within a CTL epitope in Nef that diminished or escape CTL recognition, suggested that CTLs exert selection pressure on the viral population in acute HIV infection (Price *et al.*, 1997). Recently it was demonstrated that mutations in a dominant and highly conserved B27-restricted Gag CTL epitope that is associated with prolonged immune containment leads to loss of recognition and disease progression in adults. Mothers expressing HLA-B27 perinatally transmitted viruses encoding CTL escape variants leading to suboptimal epitope targeting and failure to contain replication in the infected infants. Furthermore, the CTL escape mutants remained stable without reversion in the absence of evolutionary pressure. This suggests that CTL escape mutations in epitopes associated with suppression of viraemia will accumulate as the epidemic progresses, and that epitopes currently controlling HIV-1 infection may not be relevant in viruses circulating the community in the future (Goulder *et al.*, 2001b). Concerns about the impact of viral escape from CTL recognition on the efficacy of AIDS vaccines have been confirmed in HIV-1 vaccine trails in Rhesus monkeys where CTL escape mutations resulted in eventual vaccine failure, followed by a burst of virus replication, clinical disease progression and death by AIDS-related complications (Barouch *et al.*, 2002).

Recent selection of CTL escape variants in influenza virus, which is an acute viral infection, has also been demonstrated. Studies in mice, transgenic for an influenza nucleoprotein $(NP)_{366-374}$ peptide- specific T-cell receptor, suggested that CD8 + CTL responses directed to an internal viral protein can exert selective pressure on the virus. In this study, variants with epitope-specific point mutations occurred that permited infected cells to escape CTL attack by interfering with T-cell receptor recognition, and allowed escape from polyclonal CTL recognition by affecting peptide binding to the MHC class I. This facilitated viral reemergence

and persistence weeks after the acute infection had apparently been cleared (Price *et al.*, 2000). Recent sequence variation was also described in a HLA B-35-restricted epitope in influenza A virus nucleoprotein that was associated with CTL escape. CTLs directed against more recent strains failed to recognise old strains suggesting escape from CTL immunity (Boon *et al.*, 2002). Comparison of influenza virus sequences obtained from the influenza virus database revealed amino acid variation in four positions in this epitope, and that the variants emerged in chronological order. These reports suggest antigenic drift resulting from immune pressure mediated by specific CTLs that can result in escape from CTL mediated immunity to influenza virus (Boon *et al.*, 2002).

It may be possible that the small number of individuals from diverse HLA types who were screened in the current study selected for the identification of conserved epitopes; however, the fact that epitope N9-255 is conserved both between human and animal RSV and that the mutations that occurred in pneumovirus of mice selected an amino acid from the same similarity group, suggests that conservation of this domain in the RSV nucleoprotein is important. The reports of antigenic drift occurring in influenza CTL epitopes makes one reluctant to completely exclude the possibility of this happening in RSV proteins, however, results obtained in this study suggest that the identified RSV N-protein specific CTL epitopes are conserved and can be used for studying the CTL response to RSV in any epidemic without the need for regular sequence updates. These findings have positive implications for vaccine development.

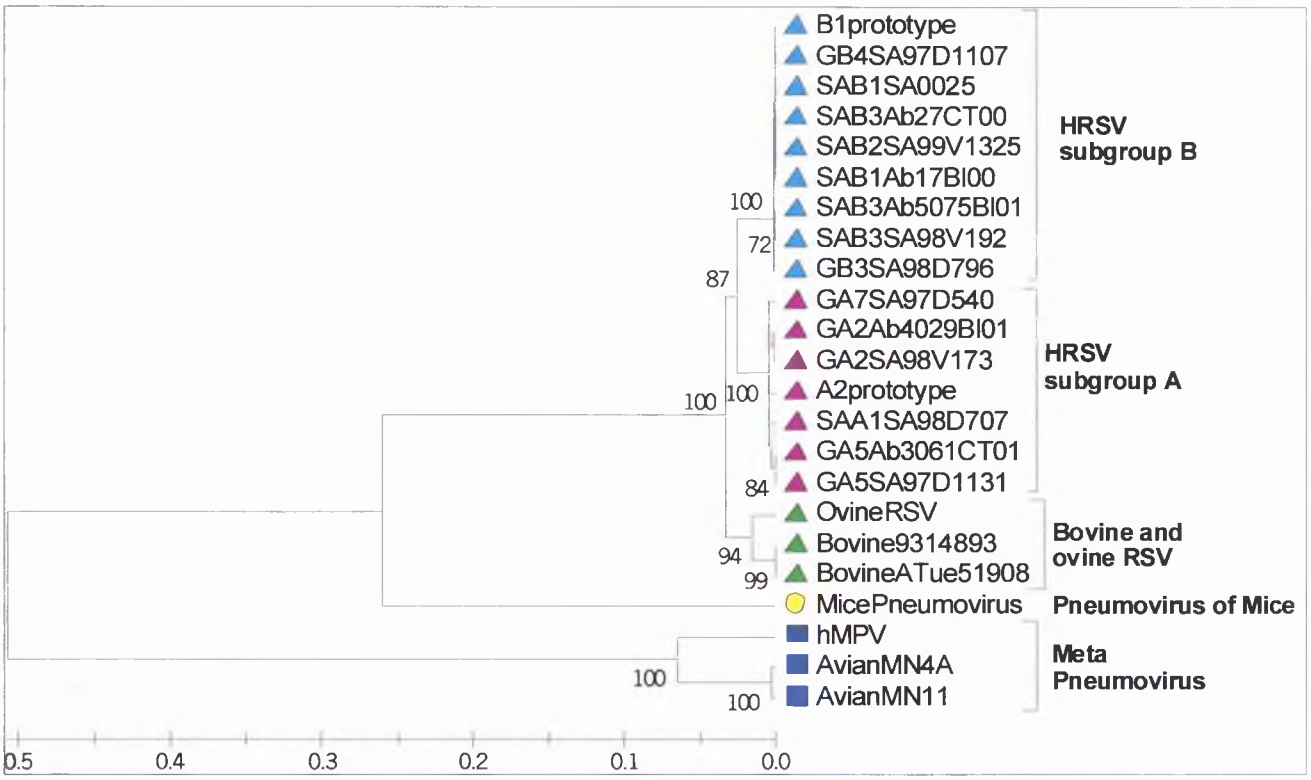


Figure 6.3: Phylogenetic relationship of members of the *Pneumovirinae*. The neighbor-joining tree was drawn with Mega version 2.1 (Kumar *et al.*, 2001) using the Amino acid Gamma model to compare the Pairwise distances between the nucleoprotein amino acid sequences. The phylogram shows the relationship of the nucleoproteins of human RSV subgroup A and B genotypes isolated in South African to bovine RSV, ovine RSV, pneumovirus of mice, and the metapneumoviruses (avian and human MPV). The linearised scale indicates the proportion of amino acid substitutions per site, since divergence time along the length of the tree. The different members of RSV are indicated by an ▲; pneumovirus of mice with a ● and the metapneumoviruses by a ■.

CHAPTER 7

Concluding remarks

Despite almost 50 years of research RSV still remains a major cause of morbidity world wide. Although it is infants, immunocompromised adults and the elderly that experience the most severe disease, RSV is also an important source of community-acquired pneumonia in patients of all age groups (Englund *et al.*, 1988; Falsey *et al.*, 1992; La Montagne, 1997). It is the major cause of hospitalisation for respiratory tract infections in young children in the USA and results in almost 100,000 hospitalisations at an estimated cost of $300,000,000 per annum (Heilman, 1990). Although the risk of mortality is low, the amount of morbidity and the cost to the health system makes RSV an important target for vaccine development. However the development of vaccines had been slowed by a lack of durable immunity even after natural infection, diversity of the virus and by fears of possible vaccine-induced pathogenesis (reviewed by Crowe, 2001). The availability of new technologies, and a better understanding of the pathogenesis, immunity and virus replication have, however, resulted in increased optimism that control is possible (Walsh & Graham, 1999). Although sterilising immunity may never be achieved, protection of the highest risk groups against lower respiratory tract infections may drastically decrease the disease burden caused by RSV (Dudas & Karron, 1998).

Limited data have so far been available about RSV in developing countries, especially from Africa (Weber *et al.*, 1998). RSV has been shown to be the most important cause of lower respiratory tract infections in children in South Africa (Madhi *et al.*, 2000). The high HIV prevalence in South Africa further increases the infection rate, the morbidity and mortality caused by RSV (Madhi *et al.*, 2000, Madhi *et al.*, 2001). Information about RSV molecular epidemiology and protective immunity in South Africa is needed to answer questions about pathogenesis, contribute to the development and identification of effective vaccines and plan for future vaccine programs. In this thesis the first studies of

the molecular epidemiology and immunology of RSV in South Africa were described. The following section will summarise the major conclusions from this study.

In the first study (Chapter 2) specimens that were collected over four consecutive seasons at an academic hospital serving Black urban South Africans in Soweto, were selected systematically to cover each of the four seasons and used for molecular epidemiological characterisations. Using these specimens, the techniques for subtyping were firstly established by adapting a published RT-PCR to be carried out in a one step reaction, and by developing a nested PCR for increasing the sensitivity and product yields for sequencing. These techniques made it possible to use clinical specimens directly for molecular analysis without the need to culture the virus first. This saves time, but also prevents the incorporation of possible mutations during the culturing process.

A total of 250 of these specimens were subtyped directly. It was shown that both subgroup A and B played an important role in RSV epidemics in South Africa and that vaccines to both subtypes will need to be incorporated into future vaccination programs. In the first two years (1997-1998) both subtypes co-circulated approximately equally while in the last two years (1999-2000) subgroup A was dominant. In 2000 almost all specimens were subgroup A. Subsequent subtyping of specimens collected in 2001 showed a switch back to subgroup B. There is currently no standardised method for genotyping RSV however a published method by members of the CDC (Peret *et al.,* 1998; 2000) was selected to facilitate comparison to previous studies. Sequence analysis of the C-terminal of the second hypervariable region of the G-protein followed by phylogenetic comparison to reference sequences published by the above-mentioned group, proved to be an effective and rapid procedure for genotyping RSV. Phylogenetic analysis revealed that some of the same genotypes identified in the the rest of the world also circulate South Africa, but also that certain genotypes are found here that have not been identified elsewhere. Some of these sequences were identified that were identical to South African specimens suggesting that specimens can spread worldwide. For subgroup A, one new genotype was identified that did not cluster with any published sequences on the Genbank, and

was named SAA1. Other subgroup A genotypes identified in South Africa included GA2, GA5 and GA7. The South African GA2 specimens cluster separately to the published GA2 specimens, but always clustered with an isolate from West Virginia (WV12342). One outlier was identified (SA98V137) that clustered with an isolate from Mozambique Moz-9-99 and separately from the other South African specimens. For Subgroup B two genotypes were identified in South Africa that have also been found in the the rest of the world, while 3 genotypes did not cluster with the reference sequences for genotypes identified by Peret *et al.* (2000). These were named SAB1, SAB2 and SAB3. No isolates could be identified on the Genbank that clustered with SAB1 and SAB3, however SAB2 specimens clustered with an isolate from Seoul (SEL93366) and one from Mozambique (Moz198-99). The isolate from Mozambique was identical to the South African isolate and both were isolated in 1999, suggesting possible cross-border transfer of strains. SAB3 specimens were most closely related to the GB3 specimens, however they grouped separately and some specimens had P-distances of more than 0.07 to members of the respective genotype. GB3 formed many subgenotypes and significant bootstrap values could not be obtained between South African specimens and members of GB3 identified by Peret *et al.* (2000).

Comparison of the P-distances between and within different genotypes suggested that members of the same genotype have P-distances of less than 0.07 to all other members of the same group. This was set as an additional criterion for genotyping, and assisted in making decisions about specimens that grouped with weak bootstrap statistics to specific genotypes. When further clinical specimens from the rest of the country were added to the analysis (Chapter 4) subgroup B genotype SAB1 had to be dissolved into two genotypes (SAB1 and SAB4) to facilitate genotyping of all clinical specimens. Analysis of all clinical specimens studied in chapters 2-4 suggested that the genotyping method is effective, but that especially subgroup B genotypes will probably evolve in future studies. Indications of positive selection were identified in the G-proteins of both subgroups. The G-protein is one of the targets of the neutralising antibody response and selection in these epitopes has also been reported in previous

studies (reviewed by Melero *et al.*, 1997). These results suggest evasion of the neutralising antibody response by RSV through epitope change.
Sufficient subgroup A and B clinical specimens were obtained in order to be able to assess the molecular epidemiology of both subgroups. Multiple genotypes co-circulated for both subgroups, however, subgroups A and B revealed different circulation patterns. Subgroup A genotypes were more variable between seasons than subgroup B. Different subgroup A genotypes became dominant, increasing in numbers before declining and eventually disappearing, showing a clear displacement of the dominant genotype, while two subgroup B genotypes remained co-dominant for all four years. It has been suggested that this higher variability may be advantageous to subgroup A and could contribute to its higher worldwide predominance (Coggins *et al.*, 1998; Hall *et al.*, 1990).

Various studies have shown that different genotypes are dominant each year (Cane *et al.*, 1994; Choi & Lee, 2000; Peret *et al.*, 1998). Although this was true for 3 of the studied seasons, one genotype (GA2) remained dominant for two years, and in the final year almost complete homogeneity existed with 78% of all clinical specimens belonging to GA2. When clinical specimens from the same year from the rest of the country were compared (Chapter 4) it was found that this genotype (GA2) had spread across the whole country and could be isolated in all the study sites. This is in contrast to findings of community-based epidemiology that have been identified in other countries (Anderson *et al.*, 1991; Peret *et al.*, 2000). It is unclear what the reason for these differences in epidemiology may be, however it may be attributed to frequent travelling of the workforce to the major cities, and transfer of virus between sites that are in close proximity to each other but have different climatic regions with differences in RSV seasons. The high HIV prevalence in South Africa that may result in prolonged shedding time or susceptibility to infections by the same RSV strains may also influence the epidemiology. It has been shown that isolation of RSV in HIV-positive children is not restricted to specific seasons as found in otherwise healthy children (Madhi *et al.*, 2000). Consideration should however also be given to the fact that 2000 was marked by some of the worst floods South Africa and Mozambique have experienced in years, which may have influenced normal disease epidemiology. Further work over consecutive seasons in different locations would have to be

carried out to determine how often genotypes circulate throughout the country or remain dominant in more than one season. Parts of the work described in Chapter 2 and 4 have been published in Venter *et al.* 2001, Madhi and Venter *et al.* 2001 and Madhi and Venter *et al.* 2002(in press).

In Chapter 3 RSV positive specimens from infants attending rural community clinics in the Limpopo Province of South Africa were compared, to clinical specimens from hospitalised infants in Soweto, near Johannesburg during the same period. Phylogenetic analysis indicated that identical viruses were found in both sites, suggesting wide distribution or transfer between communities. Analysis of the clinical manifestations caused by these viruses suggested that the same viruses could cause mild upper respiratory tract infections or severe lower respiratory tract infections, and that both subgroup A and B was an important cause of lower respiratory tract infections. Amino acid analysis identified one subgroup B community isolate that had a G-protein truncated by approximately 35 amino acids, however the amino acid sequence of other community RSV specimens were not significantly different from hospital specimens. Further investigation revealed that the isolate with the truncation in the G-protein had caused lower respiratory tract infection in a 32-month old infant, suggesting that deletion of the G-protein's C-terminus had not resulted in an attenuated phenotype. It has been shown that strain-specific epitopes are located at the C-terminal third of the G-protein and that mutations in this region that determined termination codon usage seem to have played an important role in the diversification of group B viruses (Martinez *et al.*, 1999). The fact that a G-protein C-terminal deletion of up to 35 amino acids did not alter the clinical disease manifestation of the RSV infection identified in the current study again confirms that RSV can accommodate drastic changes to the variable regions of the G-protein without being affected. The study described in Chapter 3 suggested that disease severity is most probably not associated with the infecting strain, but that other factors such as the individual's genetics and immune response may play a more important role in determining the clinical outcome of RSV infection. Chapter 3 has been published in Venter *et al.*, (2002).

One of the important determinants of RSV disease control is the cytotoxic T-cell response. Previous studies have suggested that the CTL response is essential for clearance of infection but also that it may induce immune pathology when passively transferred in large amounts to the lungs (Cannon *et al.*, 1988a). It is thought that a type 1 T-helper cell response with the induction of CTL will be crucial for the disease-free control of RSV by future vaccines (Graham, 1996). It is therefore essential to be able to monitor the CTL response induced by experimental vaccines in young children. In the past studies in humans were limited by the small sample volumes from infected children and the low frequency of memory cells in adults (Goulder *et al.*, 2000b). New highly sensitive technologies for studying the CTL response have now become available that will facilitate research into the role of CTL in disease and protection in children, however to use these techniques CTL epitopes and their HLA restriction elements in the target population have to be identified. In Chapter 5 the identification of memory CTL epitopes to RSV in healthy South African adults is described. Thirty-seven healthy adults representing various South African ethnic groups and a diverse HLA repertoire were screened by IFN-gamma ELIspot for memory CTL activity to peptides covering the complete N-protein of RSV. Responses of more than 40 sfu/million cells were detectable in 21 individuals. Six peptides that elicited reactions in the highest number of donors were selected for further characterization. Two 14-mers N-84 and N-85 were identified that induced strong CD8+ CTL responses in B*08-restricted target cells and weaker A*02- and B*15-restricted responses. Finer mapping identified a novel 9-mer RSV-specific CD8+ memory CTL epitope. This MHC class I epitope induced strong B*08-restricted lysis. The A*02- and B*15- restricted response could not be mapped. These HLA types are common in the South African population and will be useful for subsequent studies of the CTL response to RSV in humans. The epitope was detected in healthy adults and most likely reflect the response generated in the course of natural RSV infection. This is the first time that a dominant CTL response was characterised for a complete RSV protein. These are but the fourth human CTL epitope to be identified to RSV and the first to these HLA types. This study emphasises the sensitivity of the IFN-gamma ELIspot technique for detection of low frequency memory cells, but also that memory CTLs to RSV

exist in adults. This confirms that CTLs are elicited during RSV infection and that memory cells are activated upon subsequent contact with the virus.

Recent reports have suggested that influenza virus may evade the immune response through positive selection in CTL epitopes. The first study that addresses sequence variation in human RSV CTL epitopes is described in Chapter 6. To determine if the CTL epitopes recognised by South African subjects were conserved between the virus genotypes and subgroups, viral specimens were selected from all viral genotypes identified in South Africa from 1997-2001, a major part of the N-protein gene sequenced and compared with historic RSV isolates. To determine the extent of conservation, the N-protein sequences were also compared with those of other members of the *Pneumovirinae*. Comparison of the N-protein amino acid sequences of RSV subgroup A and B specimens suggested that the epitopes N9-255 and NL9 identified in this and a previous study (Goulder *et al.*, 2000b) were conserved in all RSV genotypes isolated between 1997-2001 in South Africa, in both the historical prototype strains A2 and B1 and in bovine and ovine RSV. This suggests a lack of immune selection in these epitopes. The specimens were selected to be as divergent as possible from all the different RSV genotypes identified over a period of five years. The fact that the epitopes were also conserved between different host species of RSV suggests that conservation of this region may be important to maintain an essential structural or functional role. In contrast to findings for the G-protein, no signs of positive selection could be found in the nucleoprotein. This suggests that immune evasion through selection does not occur in the N-protein CTL epitopes, and that frequent updating would not be required when using this epitope to monitor the CTL response in vaccine trails. It would be of interest to determine if immune selection occurs in CTL epitopes of other proteins, which are less conserved between the two subgroups, e.g., the F-protein. Parts of Chapters 5 and 6 have been submitted for publication to J.Virol in Venter, Puren, Tiemessen, Rock and Crowe (outcome pending).

To summarise, the highly diverse nature of RSV and immune evasion of antibody-mediated immunity through positive selection in the G protein in epidemics in South Africa are described in Chapters 2-4. This diversity may

facilitate reinfection and reestablishment of annual epidemics however there appears to be no association between the infecting genotype and disease severity. Data from Chapter 6 however suggest that nucleoprotein-specific CTL-mediated immunity is not affected by immune selection. This suggests that although repeat infections may be facilitated through G-protein change, clearance of these infections by CTL-mediated immunity should not be affected. Recent studies in mice have, however, suggested that RSV may overcome the CTL response through interference with T-cell receptor-mediated signalling. The CTL epitopes identified in this (Chapter 5) and other studies will facilitate studies to determine if similar strategies exist in humans, and help clarify the role of CTL in protection and disease pathogenesis. Although South Africa has a highly diverse HLA repertoire and each individual has a different profile of RSV CTL epitope recognition, findings in Chapter 5 suggested that CTL activity to dominant epitopes exists that may be recognised by different HLA types. Thus, although it may be more difficult to monitor CTL activity in the outbred human population, surveillance of these dominant epitopes in vaccine trials will help to overcome this problem. The availability of the reverse genetics system for vaccine design will make it possible to include dominant CTL epitopes identified in this and other populations in future vaccines. Data collected in this study may contribute to the development of relevant vaccines and vaccine programs and to the clarification of the role of different genotypes and the CTL response in disease pathogenesis.

Future directions:

The molecular epidemiological data collected in this study will be useful for development and planning of relevant vaccine programs. Although it has been suggested that strain variation may contribute to immune evasion by RSV and reestablishment of infections, the extent of cross-protection between the identified genotypes has to be clarified. This will help to determine if more than one genotype of each subtype needs to be included in future vaccines. Further molecular epidemiological studies over consecutive seasons in different locations of South Africa will help to determine if genotypes frequently dominate for more than one season, and if the nation-wide distribution observed in this study was a once-off phenomena. This will also help to clarify the seasonality of RSV epidemics in the different climatic regions of South Africa. The influence of the

high HIV prevalence on RSV molecular epidemiology in South Africa should also be elucidated. Once a commercial vaccine becomes available studies of the molecular epidemiology may become indispensable for the success of vaccine programs. The technology established in this study will allow the rapid description of RSV epidemics in South Africa. CTL-epitopes identified in this study will be useful for monitoring the CTL responses elicited by experimental vaccines in infants in South Africa. These epitopes will make it possible to determine the role of CTLs in protection and disease pathogenesis in humans. It will also be of interest to determine the influence of HIV infections on the RSV-specific CTL response.

REFERENCES

Ada, G. L. & Jones, P. D. (1987). The cell-mediated and humoral responses to influenza virus infection in the mouse lung. *Adv Exp Med Biol* 216B, 1033-42.

Alansari, H. & Potgieter, L. N. (1994). Molecular cloning and sequence analysis of the phosphoprotein, nucleocapsid protein, matrix protein and 22K (M2) protein of the ovine respiratory syncytial virus. *J Gen Virol* 75, 3597-601.

Altfeld, M., Trocha, A., Eldridge, R., Rosenberg, E., Phillips, M., Addo, M., Sekaly, R., Kalams, S., Burchett, S., McIntosh, K., Walker, B. & Goulder, P. (2000). Identification of dominant optimal HLA-B60- and HLA-B61- restricted cytotoxic T-lymphocyte (CTL) epitopes: rapid characterisation of CTL responses by enzyme linked immunospot assay. *J Virol* 74, 8541-9.

Altman, J., Mossm, P., Goulder, P., Barouch, D., McHeyzer-Williams, M., Bell, J., McMichael, A. & Davis, M. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science* 4, 94-6.

Anderson, J. J., Norden, J., Saunders, D., Toms, G. L. & Scott, R. (1990). Analysis of the local and systemic immune responses induced in BALB/c mice by experimental respiratory syncytial virus infection. *J Gen Virol* 71, 1561-70.

Anderson, L., Hendry, R., Pierik LT, Tsou C & McIntosh, K. (1991). Multicenter study of strains of respiratory syncytial virus. *J Infect Dis,* 163, 687-92.

Anderson, L., Hierholzer, J., Tsou, C., Hendry, R., Fernie, B., Stone, Y. & McIntosh, K. (1985). Antigenic characterisation of respiratory syncytial virus strains with monoclonal antibodies. *J Infect Dis.*; 151, 626-33.

Apolloni, A., Moss, D., Stumm, R., Burrows, S., Suhrbier, A., Misko, I., Schmidt, C. & Sculley, T. (1992). Sequence variation of cytotoxic T cell epitopes in different isolates of Epstein-Barr virus. *Eur J Immunol* 22, 183-9.

Arai, K. I., Lee, F., Miyajima, A., Miyatake, S., Arai, N. & Yokota, T. (1990). Cytokines: coordinators of immune and inflammatory responses. *Ann Rev Biochem* 59, 783-836.

Arnold, R., Konig, B., Galatti, H., Werchau, H. & Konig, W. (1995). Cytokine (IL-8, IL-6, TNF-alpha) and soluble TNF receptor-I release from human peripheral blood mononuclear cells after respiratory syncytial virus infection. *Immunology* 85, 364-72.

Askonas, B. A. & Taylor, P. M. (1987). T cell mediated immunity in virus infection. *Immunol Lett* 16, 337-40.

Atreya, P. L., Peeples, M. E. & Collins, P. L. (1998). The NS1 protein of human respiratory syncytial virus is a potent inhibitor of minigenome transcription and RNA replication. *J Virol* 72, 1452-61.

Aung, S., Rutigliano, J. A. & Graham, B. S. (2001). Alternative mechanisms of respiratory syncytial virus clearance in perforin knockout mice lead to enhanced disease. *J Virol* 75, 9918-24.

Azuma, M., Cayabyab, M., Buck, D., Phillips, J. H. & Lanier, L. L. (1992a). CD28 interaction with B7 costimulates primary allogeneic proliferative responses and cytotoxicity mediated by small, resting T lymphocytes. *J Exp Med* 175, 353-60.

Azuma, M., Phillips, J. H. & Lanier, L. L. (1992b). CD28 co-stimulation of T-cell-mediated cytotoxicity. *Int J Cancer Suppl* 7, 33-5.

Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. (1985). Binding of immunogenic peptides to histocompatibility molecules. *Nature* 317, 359-61.

Bangham, C. R. & Askonas, B. A. (1986). Murine cytotoxic T cells specific to respiratory syncytial virus recognise different antigenic subtypes of the virus. *J Gen Virol* 67, 623-9.

Bangham, C. R., Cannon, M. J., Karzon, D. T. & Askonas, B. A. (1985). Cytotoxic T-cell response to respiratory syncytial virus in mice. *J Virol* 56, 55-9.

Bangham, C. R. & McMichael, A. J. (1986). Specific human cytotoxic T cells recognise B-cell lines persistently infected with respiratory syncytial virus. *Proc Natl Acad Sci U S A* 83, 9183-7.

Bangham, C. R., Openshaw, P. J., Ball, L. A., King, A. M., Wertz, G. W. & Askonas, B. A. (1986). Human and murine cytotoxic T cells specific to respiratory syncytial virus recognise the viral nucleoprotein (N), but not the major glycoprotein (G), expressed by vaccinia virus recombinants. *J Immunol* 137, 3973-7.

Barouch, D. H., Kunstman, J., Kuroda, M. J., Schmitz, J. E., Santra, S., Peyerl, F. W., Krivulka, G. R., Beaudry, K., Lifton, M. A., Gorgone, D. A., Montefiori, D. C., Lewis, M. G., Wolinsky, S. M. & Letvin, N. L. (2002). Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* 415, 335-9.

Barr, J., Chambers, P., Pringle, C. R. & Easton, A. J. (1991). Sequence of the major nucleocapsid protein gene of pneumonia virus of mice: sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. *J Gen Virol* 72, 677-85.

Beeler, J. A. & van Wyke Coelingh, K. (1989). Neutralization epitopes of the F glycoprotein of respiratory syncytial virus: effect of mutation upon fusion function. *J Virol* 63, 2941-50.

Belshe, R. B., Van Voris, L. P. & Mufson, M. A. (1982). Parenteral administration of live respiratory syncytial virus vaccine: results of a field trial. *J Infect Dis* 145, 311-9.

Bermingham, A. & Collins, P. L. (1999). The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proc Natl Acad Sci U S A* 96, 11259-64.

Bertoletti, A., Sette, A., Chisari, F. V., Penna, A., Levrero, M., De Carli, M., Fiaccadori, F. & Ferrari, C. (1994). Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* 369, 407-10.

Bertoni, R., Sette, A., Sidney, J., Guidotti, L. G., Shapiro, M., Purcell, R. & Chisari, F. V. (1998). Human class I supertypes and CTL repertoires extend to chimpanzees. *J Immunol* 161, 4447-55.

Boon, A. C., de Mutsert, G., Graus, Y. M., Fouchier, R. A., Sintnicolaas, K., Osterhaus, A. D. & Rimmelzwaan, G. F. (2002). Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *J Virol* 76, 2567-72.

Bossert, B. & Conzelmann, K. K. (2002). Respiratory syncytial virus (RSV) nonstructural (NS) proteins as host range determinants: a chimeric bovine RSV with NS genes from human RSV is attenuated in interferon-competent bovine cells. *J Virol* 76, 4287-93.

Bouvier, M. & Mwiley, D. C. (1994). Importance of peptide amino and carboxy termini stability of MHC class I molecules. *Science* 265, 398-402.

Bowyer, S. M. (2002). PhD Thesis: Molecular characterisation of the Hepatitis B virus in Southern African patients. In *Department of health science.* Johannesburg: The University of the Witwatersrand.

Brandenburg, A. H., de Waal, L., Timmerman, H. H., Hoogerhout, P., de Swart, R. L. & Osterhaus, A. D. (2000a). HLA class I-restricted cytotoxic T-cell epitopes of the respiratory syncytial virus fusion protein. *J Virol* 74, 10240-4.

Brandenburg, A. H., van Beek, R., Moll, H. A., Osterhaus, A. D. & Claas, E. C. (2000b). G protein variation in respiratory syncytial virus group A does not correlate with clinical severity. *J Clin Microbiol* 38, 3849-52.

Brandt, C. D., Kim, H. W., Arrobio, J. O., Jeffries, B. C., Wood, S. C., Chanock, R. M. & Parrott, R. H. (1973). Epidemiology of respiratory syncytial virus infection in Washington, D.C. 3. Composite analysis of eleven consecutive yearly epidemics. *Am J Epidemiol* 98, 355-64.

Buchholz, U. J., Finke, S. & Conzelmann, K. K. (1999). Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. *J Virol* 73, 251-9.

Buchholz, U. J., Granzow, H., Schuldt, K., Whitehead, S. S., Murphy, B. R. & Collins, P. L. (2000). Chimeric bovine respiratory syncytial virus with glycoprotein gene substitutions from human respiratory syncytial virus (HRSV): effects on host range and evaluation as a live-attenuated HRSV vaccine. *J Virol* 74, 1187-99.

Bukreyev, A., Whitehead, S. S., Murphy, B. R. & Collins, P. L. (1997). Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse. *J Virol* 71, 8973-82.

Bukreyev, A., Whitehead, S. S., Prussin, C., Murphy, B. R. & Collins, P. L. (2000). Effect of coexpression of interleukin-2 by recombinant respiratory syncytial virus on virus replication, immunogenicity, and production of other cytokines. *J Virol* 74, 7151-7.

Cane, P., Matthews, D. & Pringle, C. (1991). Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. *J Gen Virol,* 72, 2091-6.

Cane, P. & Pringle, C. (1995a). Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein. *J Virol,* 69, 2918-25.

Cane, P. & Pringle, C. (1995b). Molecular epidemiology of Respiratory syncytial virus: A review of the use of reverse transcription-polymerase chain reaction in the analysis of genetic variability. *Electrophoresis* 16, 329-333.

Cane, P. A. (1997). Analysis of linear epitopes recognised by the primary human antibody response to a variable region of the attachment (G) protein of respiratory syncytial virus. *J Med Virol* 51, 297-304.

Cane, P. A. (2001). Molecular epidemiology of respiratory syncytial virus. *Reviews in Medical Virology* 11, 103-116.

Cane, P. A., Matthews, D. A. & Pringle, C. R. (1992). Analysis of relatedness of subgroup A respiratory syncytial viruses isolated worldwide. *Virus Res* 25, 15-22.

Cane, P. A., Matthews, D. A. & Pringle, C. R. (1994). Analysis of respiratory syncytial virus strain variation in successive epidemics in one city. *J Clin Microbiol* 32, 1-4.

Cane, P. A. & Pringle, C. R. (1991). Respiratory syncytial virus heterogeneity during an epidemic: analysis by limited nucleotide sequencing (SH gene) and restriction mapping (N gene). *J Gen Virol* 72, 349-57.

Cane, P. A. & Pringle, C. R. (1992). Molecular epidemiology of respiratory syncytial virus: rapid identification of subgroup A lineages. *J Virol Methods* 40, 297-306.

Cane, P. A. & Pringle, C. R. (1995c). Evolution of subgroup A respiratory syncytial virus: evidence for progressive accumulation of amino acid changes in the attachment protein. *J Virol* 69, 2918-25.

Cane, P. A., Thomas, H. M., Simpson, A. F., Evans, J. E., Hart, C. A. & Pringle, C. R. (1996). Analysis of the human serological immune response to a variable region of the attachment (G) protein of respiratory syncytial virus during primary infection. *J Med Virol* 48, 253-61.

Cane, P. A., Weber, M., Sanneh, M., Dackour, R., Pringle, C. R. & Whittle, H. (1999). Molecular epidemiology of respiratory syncytial virus in The Gambia. *Epidemiol Infect* 122, 155-60.

Cannon, M. J. & Bangham, C. R. (1989). Recognition of respiratory syncytial virus fusion protein by mouse cytotoxic T cell clones and a human cytotoxic T cell line. *J Gen Virol* 70, 79-87.

Cannon, M. J., Openshaw, P. J. & Askonas, B. A. (1988). Cytotoxic T cells clear virus but augment lung pathology in mice infected with respiratory syncytial virus. *J Exp Med* 168, 1163-8.

Cannon, M. J., Stott, E. J., Taylor, G. & Askonas, B. A. (1987). Clearance of persistent respiratory syncytial virus infections in immunodeficient mice following transfer of primed T cells. *Immunology* 62, 133-8.

Cartee, T. L. & Wertz, G. W. (2001). Respiratory syncytial virus M2-1 protein requires phosphorylation for efficient function and binds viral RNA during infection. *J Virol* 75, 12188-97.

Chang, J. & Braciale, T. J. (2002). Respiratory syncytial virus infection suppresses lung CD8+ T-cell effector activity and peripheral CD8+ T-cell memory in the respiratory tract. *Nat Med* 8, 54-60.

Chanock, R. M. R., B. Myers, R. (1957). Recovery from infants with respiratory illness of a virus related to chimpanzee coryza agent: Isolation, properties and characterisation. *Am J Hyg* 66, 281-290.

Cherrie, A., Anderson, K., Wertz, G. & Openshaw, P. (1992a). Human cytotoxic T-cells stimulated by antigen on dendritic cells recognise the N, SH, F, M, 22k and 1b proteins of Respiratory syncytial virus. *J Virol.* 66, 2012-2110.

Cherrie, A. H., Anderson, K., Wertz, G. W. & Openshaw, P. J. (1992b). Human cytotoxic T cells stimulated by antigen on dendritic cells recognise the N, SH, F, M, 22K, and 1b proteins of respiratory syncytial virus. *J Virol* 66, 2102-10.

Chiba, Y., Higashidate, Y., Suga, K., Honjo, K., Tsutsumi, H. & Ogra, P. L. (1989). Development of cell-mediated cytotoxic immunity to respiratory syncytial virus in human infants following naturally acquired infection. *J Med Virol* 28, 133-9.

Chin, J., Magoffin, R. L., Shearer, L. A., Schieble, J. H. & Lennette, E. H. (1969). Field evaluation of a respiratory syncytial virus vaccine and a trivalent parainfluenza virus vaccine in a pediatric population. *Am J Epidemiol* 89, 449-63.

Choi, E. & Lee, H. (2000). Genetic diversity and molecular epidemiology of the G protein of subgroups A and B of respiratory syncytial viruses isolated over 9 consecutive epidemics in Korea. *J Infect Dis.* May; 181, 1547-56.

Christensen, L. S., Larsen, L. B., Johansen, J., Andersen, E. A., Wejse, C., Klug, B. & Hornsleth, A. (1999). The fluctuating pattern of various genome types of respiratory syncytial virus in Copenhagen and some other locations in Denmark. *Apmis* 107, 843-50.

Clayton, J. & Lonjou, C. (1997). Allele and haplotype frequencies for HLA loci in various ethnic groups. In *The Twelfth International Histocompatibility Workshop and Conference.*, pp. 665-820. Edited by D. Charron. Paris, France.

Coggins, W. B., Lefkowitz, E. J. & Sullender, W. M. (1998). Genetic variability among group A and group B respiratory syncytial viruses in a children's hospital. *J Clin Microbiol* 36, 3552-7.

Collins, P. L., Anderson, K., Langer, S. J. & Wertz, G. W. (1985). Correct sequence for the major nucleocapsid protein mRNA of respiratory syncytial virus. *Virology* 146, 69-77.

Collins, P. L., Hill, M. G., Camargo, E., Grosfeld, H., Chanock, R. M. & Murphy, B. R. (1995). Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5' proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development. *Proc Natl Acad Sci U S A* 92, 11563-7.

Collins, P. L., Hill, M. G., Cristina, J. & Grosfeld, H. (1996a). Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc Natl Acad Sci U S A* 93, 81-5.

Collins, P. L., McIntosh, K. M. & Chanock, R. M. (1996b). Respiratory syncytial virus. In *Virology*, 3rd edn, pp. 1313-1352. Edited by F. B.N., K. D.M. & H. P.M. Philadelphia: Lippincott-Raven.

Collins, P. L., Olmsted, R. A. & Johnson, P. R. (1990). The small hydrophobic protein of human respiratory syncytial virus: comparison between antigenic subgroups A and B. *J Gen Virol* 71, 1571-6.

Collins, P. L., Whitehead, S. S., Bukreyev, A., Fearns, R., Teng, M. N., Juhasz, K., Chanock, R. M. & Murphy, B. R. (1999). Rational design of live-attenuated recombinant vaccine virus for human respiratory syncytial virus by reverse genetics. *Adv Virus Res* 54, 423-51.

Connors, M., Collins, P. L., Firestone, C. Y. & Murphy, B. R. (1991). Respiratory syncytial virus (RSV) F, G, M2 (22K), and N proteins each induce resistance to RSV challenge, but resistance induced by M2 and N proteins is relatively short-lived. *J Virol* 65, 1634-37.

Connors, M., Collins, P. L., Firestone, C. Y., Sotnikov, A. V., Waitze, A., Davis, A. R., Hung, P. P., Chanock, R. M. & Murphy, B. R. (1992). Cotton rats

previously immunized with a chimeric RSV FG glycoprotein develop enhanced pulmonary pathology when infected with RSV, a phenomenon not encountered following immunisation with vaccinia-RSV recombinants or RSV. *Vaccine* 10, 475-84.

Cristina, J., Lopez, J. A., Albo, C., Garcia-Barreno, B., Garcia, J., Melero, J. A. & Portela, A. (1990). Analysis of genetic variability in human respiratory syncytial virus by the RNase A mismatch cleavage method: subtype divergence and heterogeneity. *Virology* 174, 126-34.

Crowe, J. E., Jr. (1995). Current approaches to the development of vaccines against disease caused by respiratory syncytial virus (RSV) and parainfluenza virus (PIV). A meeting report of the WHO Programme for Vaccine Development. *Vaccine* 13, 415-21.

Crowe, J. E., Jr. (2001). Respiratory syncytial virus vaccine development. *Vaccine* 20 Suppl 1, S32-7.

Daponte, B. (1995). Results from the 1995 Soweto population survey: Estimates of the population of towns in Soweto 1995, Human rights Institute of South Africa.

Domachowske, J. B., Bonville, C. A., Mortelliti, A. J., Colella, C. B., Kim, U. & Rosenberg, H. F. (2000). Respiratory syncytial virus infection induces expression of the anti-apoptosis gene IEX-1L in human respiratory epithelial cells. *J Infect Dis* 181, 824-30.

du Toit, E. D., MacGregor, K. J., Taljaard, D. G. & Oudshoorn, M. (1988). HLA-A, B, C, DR and DQ polymorphisms in three South African population groups: South African Negroes, Cape Coloureds and South African Caucasoids. *Tissue Antigens* 31, 109-25.

Dudas, R. A. & Karron, R. A. (1998). Respiratory syncytial virus vaccines. Clin Microbiol Rev 11, 430-9.

Elias, J. A., Zheng, T., Einarsson, O., Landry, M., Trow, T., Rebert, N. & Panuska, J. (1994). Epithelial interleukin-11. Regulation by cytokines, respiratory syncytial virus, and retinoic acid. *J Biol Chem* 269, 22261-8.

Englund, J. A., Sullivan, C. J., Jordan, M. C., Dehner, L. P., Vercellotti, G. M. & Balfour, H. H. (1988). Respiratory syncytial virus infection in immunocompromised adults. *Ann Intern Med* 109, 203-8.

Falsey, A., Treanor, J., Betts, R. & Walsh, E. (1992). Viral respiratory infections in the institutionalized elderly: clinical and epidemiologic findings. *J Am Geriatr Soc* 40, 115-9.

Falsey, A. R. & Walsh, E. E. (1997). Safety and immunogenicity of a respiratory syncytial virus subunit vaccine (PFP-2) in the institutionalized elderly. *Vaccine* 15, 1130-2.

Fishaut, M., Tubergen, D. & McIntosh, K. (1980). Cellular response to respiratory viruses with particular reference to children with disorders of cell-mediated immunity. *J Pediatr* 96, 179-86.

Fletcher, J. N., Smyth, R. L., Thomas, H. M., Ashby, D. & Hart, C. A. (1997). Respiratory syncytial virus genotypes and disease severity among children in hospital. *Arch Dis Child* 77, 508-11.

Fremont, D. H., Hendrickson, W. A., Marrack, P. & Kappler, J. (1996). Structures of an MHC class II molecule with covalently bound single peptides. *Science* 272, 1001-4.

Garcia, O., Martin, M., Dopazo, J., Arbiza, J., Frabasile, S., Russi, J., Hortal, M., Perez-Brena, P., Martinez, I., Garcia-Barreno, B. & *et al.* (1994). Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. *J Virol* 68, 5448-59.

Garcia-Beato, R., Martinez, I., Franci, C., Real, F. X., Garcia-Barreno, B. & Melero, J. A. (1996). Host cell effect upon glycosylation and antigenicity of human respiratory syncytial virus G glycoprotein. *Virology* 221, 301-9.

Garcia-Beato, R. & Melero, J. A. (2000). The C-terminal third of human respiratory syncytial virus attachment (G) protein is partially resistant to protease digestion and is glycosylated in a cell-type-specific manner. *J Gen Virol* 81, 919-27.

Garofalo, R., Mei, F., Espejo, R., Ye, G., Haeberle, H., Baron, S., Ogra, P. L. & Reyes, V. E. (1996). Respiratory syncytial virus infection of human respiratory epithelial cells up-regulates class I MHC expression through the induction of IFN-beta and IL-1 alpha. *J Immunol* 157, 2506-13.

Geer, L., Gjertson, D. W. & Terasaki, P. I. (1998). Chapter 7: Gene Frequencies and World Maps of their distribution. In *HLA:1998*, pp. 327-363. Edited by D. W. Gjertson & P. I. Terasaki. Lenexa, Kansas: American Society for Histocompatibility and Immunogenetics.

Ghildyal, R., Mills, J., Murray, M., Vardaxis, N. & Meanger, J. (2002). Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells. *J Gen Virol* 83, 753-7.

Glezen, W. P., Paredes, A., Allison, J. E., Taber, L. H. & Frank, A. L. (1981). Risk of respiratory syncytial virus infection for infants from low- income families in relationship to age, sex, ethnic group, and maternal antibody level. *J Pediatr* 98, 708-15.

Glezen, W. P., Taber, L. H., Frank, A. L. & Kasel, J. A. (1986). Risk of primary infection and reinfection with respiratory syncytial virus. *Am J Dis Child* 140, 543-6.

Goulder, P. J., Addo, M. M., Altfeld, M. A., Rosenberg, E. S., Tang, Y., Govender, U., Mngqundaniso, N., Annamalai, K., Vogel, T. U., Hammond, M., Bunce, M., Coovadia, H. M. & Walker, B. D. (2001a). Rapid definition of five novel

HLA-A*3002-restricted human immunodeficiency virus-specific cytotoxic T-lymphocyte epitopes by elispot and intracellular cytokine staining assays. *J Virol* 75, 1339-47.

Goulder, P. J., Brander, C., Annamalai, K., Mngqundaniso, N., Govender, U., Tang, Y., He, S., Hartman, K. E., O'Callaghan, C. A., Ogg, G. S., Altfeld, M. A., Rosenberg, E. S., Cao, H., Kalams, S. A., Hammond, M., Bunce, M., Pelton, S. I., Burchett, S. A., McIntosh, K., Coovadia, H. M. & Walker, B. D. (2000a). Differential narrow focusing of immunodominant human immunodeficiency virus gag-specific cytotoxic T-lymphocyte responses in infected African and caucasoid adults and children. *J Virol* 74, 5679-90.

Goulder, P. J., Brander, C., Tang, Y., Tremblay, C., Colbert, R. A., Addo, M. M., Rosenberg, E. S., Nguyen, T., Allen, R., Trocha, A., Altfeld, M., He, S., Bunce, M., Funkhouser, R., Pelton, S. I., Burchett, S. K., McIntosh, K., Korber, B. T. & Walker, B. D. (2001b). Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature* 412, 334-8.

Goulder, P. J., Lechner, F., Klenerman, P., McIntosh, K. & Walker, B. D. (2000b). Characterization of a novel respiratory syncytial virus-specific human cytotoxic T-lymphocyte epitope. *J Virol* 74, 7694-7.

Graham, B. (1996). Immunological determinants of disease caused by respiratory syncytial virus. *Trends Microbiol* 4, 290-3.

Graham, B. S., Bunton, L. A., Rowland, J., Wright, P. F. & Karzon, D. T. (1991a). Respiratory syncytial virus infection in anti-mu-treated mice. *J Virol* 65, 4936-42.

Graham, B. S., Bunton, L. A., Wright, P. F. & Karzon, D. T. (1991b). Reinfection of mice with respiratory syncytial virus. *J Med Virol* 34, 7-13.

Graham, B. S., Bunton, L. A., Wright, P. F. & Karzon, D. T. (1991c). Role of T lymphocyte subsets in the pathogenesis of primary infection and rechallenge with respiratory syncytial virus in mice. *J Clin Invest* 88, 1026-33.

Graham, B. S., Henderson, G. S., Tang, Y. W., Lu, X., Neuzil, K. M. & Colley, D. G. (1993). Priming immunisation determines T helper cytokine mRNA expression patterns in lungs of mice challenged with respiratory syncytial virus. *J Immunol* 151, 2032-40.

Graham, B. S., Perkins, M. D., Wright, P. F. & Karzon, D. T. (1988). Primary respiratory syncytial virus infection in mice. *J Med Virol* 26, 153-62.

Groothuis, J. R., King, S. J., Hogerman, D. A., Paradiso, P. R. & Simoes, E. A. (1998). Safety and immunogenicity of a purified F protein respiratory syncytial virus (PFP-2) vaccine in seropositive children with bronchopulmonary dysplasia. *J Infect Dis* 177, 467-9.

Groothuis, J. R. & Nishida, H. (2002). Prevention of respiratory syncytial virus infections in high-risk infants by monoclonal antibody (palivizumab). *Pediatr Inf* 44, 235-41.

Grosfeld, H., Hill, M. G. & Collins, P. L. (1995). RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. *J Virol* 69, 5677-86.

Gupta, R., Birch, H., Rapacki, K., Brunak, S. & Hansen, J. E. (1999). O-GLYCBASE version 4.0: a revised database of glycosylated proteins. *Nucleic Acids Res.* 27, 370-372.

Hall, C. B. (2000). Nosocomial respiratory syncytial virus infections: the "Cold War" has not ended. *Clin Infect Dis* 31, 590-6.

Hall, C. B. (2001). Respiratory syncytial virus and parainfluenza virus. *N Engl J Med* 344, 1917-28.

Hall, C. B., Walsh, E. E., Long, C. E. & Schnabel, K. C. (1991). Immunity to and frequency of reinfection with respiratory syncytial virus. *J Infect Dis* 163, 693-8.

Hall, C. B., Walsh, E. E., Schnabel, K. C., Long, C. E., McConnochie, K. M., Hildreth, S. W. & Anderson, L. J. (1990). Occurrence of groups A and B of respiratory syncytial virus over 15 years: associated epidemiologic and clinical characteristics in hospitalised and ambulatory children. *J Infect Dis* 162, 1283-90.

Hammond, M. G., du Toit, E. D., Sanchez-Mazas, A., Andrien, M., Coluzzi, M., de Pablo, M., de Stefano, G., Kaplan, C., Kennedy, L. J., Louie, L. & Migot, F. (1997). HLA in sub-Saharan Africa (12th International Histocompatibility Workshop SSAF report). In *The 12th International Histocompatibility Workshop and conference.*, pp. 345-353. Edited by D. Charron. Paris: EDK.

Hardy, R. W. & Wertz, G. W. (2000). The Cys(3)-His(1) motif of the respiratory syncytial virus M2-1 protein is essential for protein function. *J Virol* 74, 5880-5.

Department of Health, Report (2000). National HIV sero-prevalence survey of women attending public antenatal clinics in South Africa, 1999: Department of Health of South Africa.

Heilman, C. A. (1990). From the National Institute of Allergy and Infectious Diseases and the World Health Organization. Respiratory syncytial and parainfluenza viruses. *J Infect Dis* 161, 402-6.

Henderson, F., Collier, A., Clyde, W. J. & Denny, F. (1979). Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children. *N Engl J Med,* 8, 530-4.

Hendricks, D. A., McIntosh, K. & Patterson, J. L. (1988). Further characterisation of the soluble form of the G glycoprotein of respiratory syncytial virus. *J Virol* 62, 2228-33.

Hendry, R., Talis, A., Godfrey, E., Anderson, L., Fernie, B. & McIntosh, K. (1986). Concurrent circulation of antigenically distinct strains of respiratory syncytial virus during community outbreaks. *J Infect Dis,* 153, 291-7.

Hendry, R. M., Burns, J. C., Walsh, E. E., Graham, B. S., Wright, P. F., Hemming, V. G., Rodriguez, W. J., Kim, H. W., Prince, G. A., McIntosh, K. & *et al.* (1988). Strain-specific serum antibody responses in infants undergoing primary infection with respiratory syncytial virus. *J Infect Dis* 157, 640-7.

Heusel, J. W., Wesselschmidt, R. L., Shresta, S., Russell, J. H. & Ley, T. J. (1994). Cytotoxic lymphocytes require granzyme B for the rapid induction of DNA fragmentation and apoptosis in allogeneic target cells. *Cell* 76, 977-87.

Hill, A. V., Elvin, J., Willis, A. C., Aidoo, M., Allsopp, C. E., Gotch, F. M., Gao, X. M., Takiguchi, M., Greenwood, B. M., Townsend, A. R. & *et al.* (1992). Molecular analysis of the association of HLA-B53 and resistance to severe malaria. *Nature* 360, 434-9.

Holberg, C. J., Wright, A. L., Martinez, F. D., Ray, C. G., Taussig, L. M. & Lebowitz, M. D. (1991). Risk factors for respiratory syncytial virus-associated lower respiratory illnesses in the first year of life. *Am J Epidemiol* 133, 1135-51.

Hornsleth, A., Klug, B., Nir, M., Johansen, J., Hansen, K. S., Christensen, L. S. & Larsen, L. B. (1998). Severity of respiratory syncytial virus disease related to type and genotype of virus and to cytokine values in nasopharyngeal secretions. *Pediatr Infect Dis J* 17, 1114-21.

Huang, Y. T., Collins, P. L. & Wertz, G. W. (1985). Characterization of the 10 proteins of human respiratory syncytial virus: identification of a fourth envelope-associated protein. *Virus Res* 2, 157-73.

Hurst, L. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 18, 486.

Hussell, T. & Openshaw, P. J. (1998). Intracellular IFN-gamma expression in natural killer cells precedes lung CD8+ T cell recruitment during respiratory syncytial virus infection. *J Gen Virol* 79 (Pt 11), 2593-601.

Hussey, G. D., Apolles, P., Arendse, Z., Yeates, J., Robertson, A., Swingler, G. & Zar, H. J. (2000). Respiratory syncytial virus infection in children hospitalised with acute lower respiratory tract infection. *S Afr Med J* 90, 509-12.

Ingram, J. M., Rakes, G. P., Hoover, G. E., Platts-Mills, T. A. & Heymann, P. W. (1995). Eosinophil cationic protein in serum and nasal washes from wheezing infants and children. *J Pediatr* 127, 558-64.

Isaacs, D., Bangham, C. R. & McMichael, A. J. (1987). Cell-mediated cytotoxic response to respiratory syncytial virus in infants with bronchiolitis. *Lancet* 2, 769-71.

Janeway, C. A., Jr., Travers, P., Walport, M. & Capra, J. (1999). Immuno Biology. Edited by P. L. Austin, E. New York: Elsevier Science Ltd/Garland Publishing.

Jiang, S., Borthwick, N. J., Morrison, P., Gao, G. F. & Steward, M. W. (2002). Virus-specific CTL responses induced by an H-2K(d)-restricted, motif-negative 15-mer peptide from the fusion protein of respiratory syncytial virus. *J Gen Virol* 83, 429-38.

Johnson, P. J., Olmsted, R., Prince, G., Murphy, B., Alling, D., Walsh, E. & Collins, P. (1987a). Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity. *J Virol,* 61, 3163-6.

Johnson, P. R. & Collins, P. L. (1988). The fusion glycoproteins of human respiratory syncytial virus of subgroups A and B: sequence conservation provides a structural basis for antigenic relatedness. *J Gen Virol,* 69, 2623-8.

Johnson, P. R. & Collins, P. L. (1989). The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial virus (RSV) of antigenic subgroups A and B: sequence conservation and divergence within RSV genomic RNA. *J Gen Virol,* 70, 1539-47.

Johnson, P. R., Spriggs, M. K., Olmsted, R. A. & Collins, P. L. (1987b). The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc Natl Acad Sci U S A* 84, 5625-9.

Johnson, T. R. & Graham, B. S. (1999). Secreted respiratory syncytial virus G glycoprotein induces interleukin-5 (IL5), IL13, and eosinophilia by an IL4-independent mechanism. *J Virol* 73, 8485-95.

Johnson, T. R., Johnson, J. E., Roberts, S. R., Wertz, G. W., Parker, R. A. & Graham, B. S. (1998). Priming with secreted glycoprotein G of respiratory syncytial virus (RSV) augments interleukin-5 production and tissue eosinophilia after RSV challenge. *J Virol* 72, 2871-80.

Jones, L. P., Zheng, H. Q., Karron, R. A., Peret, T. C., Tsou, C. & Anderson, L. J. (2002). Multiplex assay for detection of strain-specific antibodies against the two variable regions of the G protein of respiratory syncytial virus. *Clin Diagn Lab Immunol* 9, 633-8.

Joosting, A. C., Harwin, R. M., Orchard, M., Martin, E. & Gear, J. H. (1979). Respiratory viruses in hospital patients on the Witwatersrand. A 7-year study. *S Afr Med J* 55, 403-8.

Kapikian, A. Z., Mitchell, R. H., Chanock, R. M., Shvedoff, R. A. & Stewart, C. E. (1969). An epidemiologic study of altered clinical reactivity to respiratory syncytial (RS) virus infection in children previously vaccinated with an inactivated RS virus vaccine. *Am J Epidemiol* 89, 405-21.

Karron, R. A., Buonagurio, D. A., Georgiu, A. F., Whitehead, S. S., Adamus, J. E., Clements-Mann, M. L., Harris, D. O., Randolph, V. B., Udem, S. A., Murphy, B. R. & Sidhu, M. S. (1997a). Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication *in vitro*: clinical evaluation and molecular characterisation of a cold-passaged, attenuated RSV subgroup B mutant. *Proc Natl Acad Sci U S A* 94, 13961-6.

Karron, R. A., Wright, P. F., Crowe, J. E., Jr., Clements-Mann, M. L., Thompson, J., Makhene, M., Casey, R. & Murphy, B. R. (1997b). Evaluation of two live, cold-passaged, temperature-sensitive respiratory syncytial virus vaccines in chimpanzees and in human adults, infants, and children. *J Infect Dis* 176, 1428-36.

Kim, H. W., Canchola, J. G., Brandt, C. D., Pyles, G., Chanock, R. M., Jensen, K. & Parrott, R. H. (1969). Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *Am J Epidemiol* 89, 422-34.

King, J. C., Jr., Burke, A. R., Clemens, J. D., Nair, P., Farley, J. J., Vink, P. E., Batlas, S. R., Rao, M. & Johnson, J. P. (1993). Respiratory syncytial virus illnesses in human immunodeficiency virus- and noninfected children. *Pediatr Infect Dis J* 12, 733-9.

Klenerman, P., Rowland-Jones, S., McAdam, S., Edwards, J., Daenke, S., Lalloo, D., Koppe, B., Rosenberg, W., Boyd, D., Edwards, A. & *et al.* (1994). Cytotoxic T-cell activity antagonized by naturally occurring HIV-1 Gag variants. *Nature* 369, 403-7.

Kulkarni, A. B., Collins, P. L., Bacik, I., Yewdell, J. W., Bennink, J. R., Crowe, J. E., Jr. & Murphy, B. R. (1995). Cytotoxic T cells specific for a single peptide on the M2 protein of respiratory syncytial virus are the sole mediators of resistance induced by immunisation with M2 encoded by a recombinant vaccinia virus. *J Virol* 69, 1261-4.

Kulkarni, A. B., Morse, H. C., 3rd, Bennink, J. R., Yewdell, J. W. & Murphy, B. R. (1993). Immunisation of mice with vaccinia virus-M2 recombinant induces epitope- specific and cross-reactive Kd-restricted CD8+ cytotoxic T cells. *J Virol* 67, 4086-92.

Kumar, S., Tamura, K., Jakobsen, I. & Nei, M. (2000). MEGA: Molecular evolutionary genetics analyses. Version 2.0. University Park: Pennsylvania State University.

Kumar, S., Tamura, K., Jakobsen, I. & Nei, M. (2001). MEGA2: Molecular Evolutionary Genetics Analysis software., 2.1 edn. Tempe, Arizona. USA: Arizona State University.

La Montagne, J. R. (1997). RSV pneumonia, a community-acquired infection in adults. *Lancet* 349, 149-50.

Lambert, D. M. (1988). Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins. *Virology* 164, 458-66.

Larsen, L. E., Gottshalch,E. and Blixenkrone-Moller,M. (1998). Bovine respiratory syncytial virus 9314893 nonstructural protein 1 (NS1), nonstructural protein 2 (NS2), and nucleocapsid protein (N) genes, complete cds.

Larsen, L. E., Uttenthal, A., Arctander, P., Tjornehoj, K., Viuff, B., Rontved, C., Ronsholt, L., Alexandersen, S. & Blixenkrone-Moller, M. (1998). Serological and genetic characterisation of bovine respiratory syncytial virus (BRSV) indicates that Danish isolates belong to the intermediate subgroup: no evidence of a selective effect on the variability of G protein nucleotide sequence by prior cell culture adoption and passages in cell culture or calves. *Vet Microbiol* 62, 265-79.

Lepage, P., Spira, R., Kalibala, S., Pillay, K., Giaquinto, C., Castetbon, K., Osborne, C., Courpotin, C. & Dabis, F. (1998). Care of human immunodeficiency virus-infected children in developing countries. International Working Group on Mother-to-Child Transmission of HIV. *Pediatr Infect Dis J* 17, 581-6.

Levely, M. E., Bannow, C. A., Smith, C. W. & Nicholas, J. A. (1991). Immunodominant T-cell epitope on the F protein of respiratory syncytial virus recognised by human lymphocytes. *J Virol* 65, 3789-96.

Liu, C. C., Walsh, C. M. & Young, J. D. (1995). Perforin: structure and function. *Immunol Today* 16, 194-201.

Looney, R. J., Falsey, A. R., Walsh, E. & Campbell, D. (2002). Effect of aging on cytokine production in response to respiratory syncytial virus infection. *J Infect Dis* 185, 682-5.

Madhi, S., Venter, M., Alexandra, R., Lewis, H., Kara, Y., Karshagen, W., Greef, M. & Lassen, C. (2002). Respiratory syncytial virus associated illness in high risk children and national characterisation of the circulating virus genotype in South Africa. *J.Clin. Virol.* (In press).

Madhi, S. A., Schoub, B., Simmank, K., Blackburn, N. & Klugman, K. P. (2000). Increased burden of respiratory viral associated severe lower respiratory tract infections in children infected with human immunodeficiency virus type-1. *J Pediatr* 137, 78-84.

Madhi, S. A., Venter, M., Madhi, A., Petersen, M. K. & Klugman, K. P. (2001). Differing manifestations of respiratory syncytial virus-associated severe lower respiratory tract infections in human immunodeficiency virus type 1-infected and uninfected children. *Pediatr Infect Dis J* 20, 164-70.

Marsh, S. (1998). Nomenclature for factors of the HLA system. WHO nomenclature committee for factors of the HLA system. *Hum. Immunol.* 59, 256-257.

Martin, A. J., Gardner, P. S. & McQuillin, J. (1978). Epidemiology of respiratory viral infection among paediatric inpatients over a six-year period in north-east England. *Lancet* 2, 1035-8.

Martinez, I., Dopazo, J. & Melero, J. (1997). Antigenic structure of the human respiratory syncytial virus G glycoprotein and relevance of hypermutation events for the generation of antigenic variants. *J Gen Virol* 78, 2419-29.

Martinez, I., Valdes, O., Delfraro, A., Arbiza, J., Russi, J. & Melero, J. A. (1999). Evolutionary pattern of the G glycoprotein of human respiratory syncytial viruses from antigenic group B: the use of alternative termination codons and lineage diversification. *J Gen Virol* 80, 125-30.

Matsuda, K., Tsutsumi, H., Okamoto, Y. & Chiba, C. (1995). Development of interleukin 6 and tumor necrosis factor alpha activity in nasopharyngeal secretions of infants and children during infection with respiratory syncytial virus. *Clin Diagn Lab Immunol* 2, 322-4.

Mbawuike, I. N., Wells, J., Byrd, R., Cron, S. G., Glezen, W. P. & Piedra, P. A. (2001). HLA-restricted CD8+ cytotoxic T lymphocyte, interferon-gamma, and interleukin-4 responses to respiratory syncytial virus infection in infants and children. *J Infect Dis* 183, 687-96.

McMichael, A. (1994). Cytotoxic T lymphocytes specific for influenza virus. *Curr Top Microbiol Immunol* 189, 75-91.

Meissner, H. C., Welliver, R. C., Chartrand, S. A., Law, B. J., Weisman, L. E., Dorkin, H. L. & Rodriguez, W. J. (1999). Immunoprophylaxis with palivizumab, a humanized respiratory syncytial virus monoclonal antibody, for prevention of respiratory syncytial virus infection in high risk infants: a consensus opinion. *Pediatr Infect Dis J* 18, 223-31.

Melero, J. A., Garcia-Barreno, B., Martinez, I., Pringle, C. R. & Cane, P. A. (1997). Antigenic structure, evolution and immunobiology of human respiratory syncytial virus attachment (G) protein. *J Gen Virol* 78, 2411-8.

Meric, C., Spehner, D. & Mazarin, V. (1994). Respiratory syncytial virus nucleocapsid protein (N) expressed in insect cells forms nucleocapsid-like structures. *Virus Res* 31, 187-201.

Mills, J. t., Van Kirk, J. E., Wright, P. F. & Chanock, R. M. (1971). Experimental respiratory syncytial virus infection of adults. Possible mechanisms of resistance to infection and illness. *J Immunol* 107, 123-30.

Morris Jr, J., Blunt, R. & Savage, R. (1956). Recovery of cytopathogenic agent from chimpanzees with coryza. *Proc Soc Exp Biol Med* 92, 544-550.

MRC, Report. (1978). Respiratory syncytial virus infection: admissions to hospital in industrial, urban, and rural areas. Report to the Medical Research Council Subcommittee on Respiratory Syncytial Virus Vaccines. In *Br Med J* 2, 796-8.

Mufson, M., Belshe, R., Orvell, C. & Norrby, E. (1988). Respiratory syncytial virus epidemics: variable dominance of subgroups A and B strains among children, 1981-1986. *J Infect Dis* 157, 143-8.

Munoz, J. L., McCarthy, C. A., Clark, M. E. & Hall, C. B. (1991). Respiratory syncytial virus infection in C57BL/6 mice: clearance of virus from the lungs with virus-specific cytotoxic T cells. *J Virol* 65, 4494-7.

Murali-Krishna, K., Altman, J. D., Suresh, M., Sourdive, D., Zajac, A. & Ahmed, R. (1998). In vivo dynamics of anti-viral CD8 T cell responses to different epitopes. An evaluation of bystander activation in primary and secondary responses to viral infection. *Adv Exp Med Biol* 452, 123-42.

Murphy, B. R., Alling, D. W., Snyder, M. H., Walsh, E. E., Prince, G. A., Chanock, R. M., Hemming, V. G., Rodriguez, W. J., Kim, H. W., Graham, B. S. & *et al.* (1986a). Effect of age and preexisting antibody on serum antibody response of infants and children to the F and G glycoproteins during respiratory syncytial virus infection. *J Clin Microbiol* 24, 894-8.

Murphy, B. R., Graham, B. S., Prince, G. A., Walsh, E. E., Chanock, R. M., Karzon, D. T. & Wright, P. F. (1986b). Serum and nasal-wash immunoglobulin G and A antibody response of infants and children to respiratory syncytial virus F and G glycoproteins following primary infection. *J Clin Microbiol* 23, 1009-14.

Murphy, B. R., Sotnikov, A. V., Lawrence, L. A., Banks, S. M. & Prince, G. A. (1990). Enhanced pulmonary histopathology is observed in cotton rats immunized with formalin-inactivated respiratory syncytial virus (RSV) or purified F glycoprotein and challenged with RSV 3-6 months after immunisation. *Vaccine* 8, 497-502.

Nakayama, T., Sonoda, S., Urano, T., Sasaki, K., Maehara, N. & Makino, S. (1993). Detection of alpha-interferon in nasopharyngeal secretions and sera in children infected with respiratory syncytial virus. *Pediatr Infect Dis J* 12, 925-9.

Nei, M. & Kumar, S. (2000). Molecular Evolution and Phylogenetics: Oxford University press.

Nicholas, J. A., Rubino, K. L., Levely, M. E., Adams, E. G. & Collins, P. L. (1990). Cytolytic T-lymphocyte responses to respiratory syncytial virus: effector cell phenotype and target proteins. *J Virol* 64, 4232-41.

Nicholas, J. A., Rubino, K. L., Levely, M. E., Meyer, A. L. & Collins, P. L. (1991). Cytotoxic T cell activity against the 22-kDa protein of human respiratory syncytial virus (RSV) is associated with a significant reduction in pulmonary RSV replication. *Virology* 182, 664-72.

O'Donnell D, R., Milligan, L. & Stark, J. M. (1999). Induction of CD95 (Fas) and apoptosis in respiratory epithelial cell cultures following respiratory syncytial virus infection. *Virology* 257, 198-207.

Oldstone, M. B. (1997). How viruses escape from cytotoxic T lymphocytes: molecular parameters and players. *Virology* 234, 179-85.

Openshaw, P., Anderson, K., Wertz, G. & Askonas, B. (1990). The 22,000-kilodalton protein of respiratory syncytial virus is a major target for Kd-restricted cytotoxic T lymphocytes from mice primed by infection. *J Virol* 64, 1683-9.

Openshaw, P. J. (1995). Immunity and immunopathology to respiratory syncytial virus. The mouse model. *Am J Respir Crit Care Med* 152, S59-62.

Openshaw, P. J. M., Culley, F. J. & Olszewska, W. (2002). Immunopathogenensis of vaccine-enhanced RSV disease. *Vaccine* 20, 27-31.

Palomo, C., Cane, P. & Melero, J. (2000). Evaluation of the antibody specificities of human convalescent-phase sera against the attachment (G) protein of human respiratory syncytial virus:influence of strain variation and carbohydrate side chains. *J Med Virol* 60, 468-74.

Palomo, C., Garcia-Barreno, B., Penas, C. & Melero, J. A. (1991). The G protein of human respiratory syncytial virus: significance of carbohydrate side-chains and the C-terminal end to its antigenicity. *J Gen Virol* 72, 669-75.

Panuska, J. R., Hertz, M. I., Taraf, H., Villani, A. & Cirino, N. M. (1992). Respiratory syncytial virus infection of alveolar macrophages in adult transplant patients. *Am Rev Respir Dis* 145, 934-9.

Panuska, J. R., Merolla, R., Rebert, N. A., Hoffmann, S. P., Tsivitse, P., Cirino, N. M., Silverman, R. H. & Rankin, J. A. (1995). Respiratory syncytial virus induces interleukin-10 by human alveolar macrophages. Suppression of early cytokine production and implications for incomplete immunity. *J Clin Invest* 96, 2445-53.

Panuska, J. R., Midulla, F., Cirino, N. M., Villani, A., Gilbert, I. A., McFadden, E. R., Jr. & Huang, Y. T. (1990). Virus-induced alterations in macrophage production of tumor necrosis factor and prostaglandin E2. *Am J Physiol* 259, L396-402.

Parker, K. C., M. A., Bednarek, M.A. & Coligan.J.E. (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J Immunol* 152.

Patel, J. A., Kunimoto, M., Sim, T. C., Garofalo, R., Eliott, T., Baron, S., Ruuskanen, O., Chonmaitree, T., Ogra, P. L. & Schmalstieg, F. (1995). Interleukin-1 alpha mediates the enhanced expression of intercellular adhesion molecule-1 in pulmonary epithelial cells infected with respiratory syncytial virus. *Am J Respir Cell Mol Biol* 13, 602-9.

Peret, T. C., Hall, C. B., Hammond, G. W., Piedra, P. A., Storch, G. A., Sullender, W. M., Tsou, C. & Anderson, L. J. (2000). Circulation patterns of group A and B human respiratory syncytial virus genotypes in 5 communities in North America. *J Infect Dis* 181, 1891-6.

Peret, T. C., Hall, C. B., Schnabel, K. C., Golub, J. A. & Anderson, L. J. (1998). Circulation patterns of genetically distinct group A and B strains of human respiratory syncytial virus in a community. *J Gen Virol* 79, 2221-9.

Potts, W. K. & Slev, P. R. (1995). Pathogen-based models favoring MHC genetic diversity. *Immunol Rev* 143, 181-97.

Power, U. F., Plotnicky-Gilquin, H., Goetsch, L., Champion, T., Beck, A., Haeuw, J. F., Nguyen, T. N., Bonnefoy, J. Y. & Corvaia, N. (2001). Identification and characterisation of multiple linear B cell protectopes in the respiratory syncytial virus G protein. *Vaccine* 19, 2345-51.

Price, D. A., Goulder, P. J., Klenerman, P., Sewell, A. K., Easterbrook, P. J., Troop, M., Bangham, C. R. & Phillips, R. E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A* 94, 1890-5.

Price, G. E., Ou, R., Jiang, H., Huang, L. & Moskophidis, D. (2000). Viral escape by selection of cytotoxic T cell-resistant variants in influenza A virus pneumonia. *J Exp Med* 191, 1853-67.

Prince, G. A., Horswood, R. L., Berndt, J., Suffin, S. C. & Chanock, R. M. (1979). Respiratory syncytial virus infection in inbred mice. *Infect Immun* 26, 764-6.

Prince, G. A., Horswood, R. L. & Chanock, R. M. (1985). Quantitative aspects of passive immunity to respiratory syncytial virus infection in infant cotton rats. *J Virol* 55, 517-20.

Prince, G. A., Jenson, A. B., Horswood, R. L., Camargo, E. & Chanock, R. M. (1978). The pathogenesis of respiratory syncytial virus infection in cotton rats. *Am J Pathol* 93, 771-91.

Pringle, C. R. & Easton A.J., (1997). Monopartite negative strand RNA genome. *Semin Virol* 8, 49-57.

Rammensee, H. G. (1995). Chemistry of peptides associated with MHC class II molecules and I. *Curr Opin Immunol* 7, 85-86.

Ramshaw, I., Ruby, J., Ramsay, A., Ada, G. & Karupiah, G. (1992). Expression of cytokines by recombinant vaccinia viruses: a model for studying cytokines in virus infections *in vivo*. *Immunol Rev* 127, 157-82.

Roberts, S. R., Lichtenstein, D., Ball, L. A. & Wertz, G. W. (1994). The membrane-associated and secreted forms of the respiratory syncytial virus attachment glycoprotein G are synthesized from alternative initiation codons. *J Virol* 68, 4538-46.

Roca, A., Loscertales, M. P., Quinto, L., Perez-Brena, P., Vaz, N., Alonso, P. L. & Saiz, J. C. (2001). Genetic variability among group A and B respiratory syncytial viruses in Mozambique: identification of a new cluster of group B isolates. *J Gen Virol* 82, 103-111.

Rudd, C. E. (1996). Upstream-downstream: CD28 cosignaling pathways and T cell function. *Immunity* 4, 527-34.

Sanz, M., Kew, O. & Anderson, L. (1994). Genetic heterogeneity of the attachment glycoprotein G among group A respiratory syncytial viruses. *Virus Res* 33, 203-17.

Schlender, J., Bossert, B., Buchholz, U. & Conzelmann, K. K. (2000). Bovine respiratory syncytial virus nonstructural proteins NS1 and NS2 cooperatively antagonize alpha/beta interferon-induced antiviral response. *J Virol* 74, 8234-42.

Schwarze, J., Cieslewicz, G., Hamelmann, E., Joetham, A., Shultz, L. D., Lamers, M. C. & Gelfand, E. W. (1999a). IL5 and eosinophils are essential for the development of airway hyperresponsiveness following acute respiratory syncytial virus infection. *J Immunol* 162, 2997-3004.

Schwarze, J., Cieslewicz, G., Joetham, A., Ikemura, T., Hamelmann, E. & Gelfand, E. W. (1999b). CD8 T cells are essential in the development of respiratory syncytial virus-induced lung eosinophilia and airway hyperresponsiveness. *J Immunol* 162, 4207-11.

Schwarze, J., Makela, M., Cieslewicz, G., Dakhama, A., Lahn, M., Ikemura, T., Joetham, A. & Gelfand, E. W. (1999c). Transfer of the enhancing effect of respiratory syncytial virus infection on subsequent allergic airway sensitization by T lymphocytes. *J Immunol* 163, 5729-34.

Selwyn, B. (1990). The epidemiology of acute respiratory tract infection in young children: comparison of findings from several developing countries. *Rev. Infect. Dis.* 12, 870-888.

Sette, A. & Sidney, J. (1998). HLA supertypes and supermotifs: a functional perspective on HLA polymorphism. *Curr Opin Immunol* 10, 478-82.

Sette, A. & Sidney, J. (1999a). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50, 201-212.

Sette, A. & Sidney, J. (1999b). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics* 50, 201-12.

Shepherd, J. C., Schumacher, T. N., Ashton-Rickardt, P. G., Imaeda, S., Ploegh, H. L., Janeway, C. A., Jr. & Tonegawa, S. (1993). TAP1-dependent peptide translocation *in vitro* is ATP dependent and peptide selective. *Cell* 74, 577-84.

Shin, H. J., Cameron, K. T., Jacobs, J. A., Turpin, E. A., Halvorson, D. A., Goyal, S. M., Nagaraja, K. V., Kumar, M. C., Lauer, D. C., Seal, B. S. & Njenga, M. K. (2002). Molecular epidemiology of subgroup C avian pneumoviruses isolated in the United States and comparison with subgroup a and B viruses. *J Clin Microbiol* 40, 1687-93.

Simoes, E. (1999). Review: Respiratory syncytial virus infection. *Lancet.* 354, 847-52.

Smith, K. J., Reid, S. W., Harlos, K., McMichael, A. J., Stuart, D. I., Bell, J. I. & Jones, E. Y. (1996). Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* 4, 215-28.

Smyth, R. L., Mobbs, K. J., O'Hea, U., Ashby, D. & Hart, C. A. (2002). Respiratory syncytial virus bronchiolitis: Disease severity, interleukin-8, and virus genotype. *Pediatr Pulmonol* 33, 339-46.

Song, R. & Harding, C. V. (1996). Roles of proteasomes, transporter for antigen presentation (TAP), and beta 2-microglobulin in the processing of bacterial or particulate antigens via an alternate class I MHC processing pathway. *J Immunol* 156, 4182-90.

Southwood, S., Sidney, J., Kondo, A., del Guercio, M., Appella, E., Hoffman, S., Kubo, R., Chesnut, R. W., Grey, H. M. & Alessandro, S. (1998). Several common HLA-DR types share largely overlapping peptide binding repertoires. *The Journal of Immunology* 160, 3363-3386.

Squier, M. K. & Cohen, J. J. (1994). Cell-mediated cytotoxic mechanisms. *Curr Opin Immunol* 6, 447-52.

Stark, J. M., Godding, V., Sedgwick, J. B. & Busse, W. W. (1996). Respiratory syncytial virus infection enhances neutrophil and eosinophil adhesion to cultured respiratory epithelial cells. Roles of CD18 and intercellular adhesion molecule-1. *J Immunol* 156, 4774-82.

Stark, J. M., McDowell, S. A., Koenigsknecht, V., Prows, D. R., Leikauf, J. E., Le Vine, A. M. & Leikauf, G. D. (2002). Genetic susceptibility to respiratory syncytial virus infection in inbred mice. *J Med Virol* 67, 92-100.

Storch, G., Anderson, L., Park, C., Tsou, C. & Dohner, D. (1991). Antigenic and genomic diversity within group A respiratory syncytial virus. *J Infect Dis* 163, 858-61.

Suda, T., Takahashi, T., Golstein, P. & Nagata, S. (1993). Molecular cloning and expression of the Fas ligand, a novel member of the tumor necrosis factor family. *Cell* 75, 1169-78.

Sugden, B. & Mark, W. (1977). Clonal transformation of adult human leukocytes by Epstein-Barr virus. *J. Virol.* 23, 503-508.

Sullender, W. M. (2000). Respiratory syncytial virus genetic and antigenic diversity. *Clin Microbiol Rev* 13, 1-15.

Sullender, W. M., Anderson, K. & Wertz, G. W. (1990a). The respiratory syncytial virus subgroup B attachment glycoprotein: analysis of sequence, expression from a recombinant vector, and evaluation as an immunogen against homologous and heterologous subgroup virus challenge. *Virology* 178, 195-203.

Sullender, W. M., Anderson, L. J., Anderson, K. & Wertz, G. W. (1990b). Differentiation of respiratory syncytial virus subgroups with cDNA probes in a nucleic acid hybridization assay. *J Clin Microbiol* 28, 1683-7.

Sullender, W. M., Mufson, M. A., Anderson, L. J. & Wertz, G. W. (1991). Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. *J Virol* 65, 5425-34.

Sullender, W. M., Sun, L. & Anderson, L. J. (1993). Analysis of respiratory syncytial virus genetic variability with amplified cDNAs. *J Clin Microbiol* 31, 1224-31.

Takeuchi, R., Tsutsumi, H., Osaki, M., Haseyama, K., Mizue, N. & Chiba, S. (1998). Respiratory syncytial virus infection of human alveolar epithelial cells enhances interferon regulatory factor 1 and interleukin-1beta-converting enzyme gene expression but does not cause apoptosis. *J Virol* 72, 4498-502.

Tang, J., Costello, C., Keet, I. P., Rivers, C., Leblanc, S., Karita, E., Allen, S. & Kaslow, R. A. (1999). HLA class I homozygosity accelerates disease progression in human immunodeficiency virus type 1 infection. *AIDS Res Hum Retroviruses* 15, 317-24.

Teng, M. N. & Collins, P. L. (1998). Identification of the respiratory syncytial virus proteins required for formation and passage of helper-dependent infectious particles. *J Virol* 72, 5707-16.

Teng, M. N. & Collins, P. L. (2002). The central conserved cystine noose of the attachment G protein of human respiratory syncytial virus is not required for efficient viral infection *in vitro* or *in vivo*. *J Virol* 76, 6164-71.

Teng, M. N., Whitehead, S. S., Bermingham, A., St Claire, M., Elkins, W. R., Murphy, B. R. & Collins, P. L. (2000). Recombinant respiratory syncytial virus that does not express the NS1 or M2-2 protein is highly attenuated and immunogenic in chimpanzees. *J Virol* 74, 9317-21.

Teng, M. N., Whitehead, S. S. & Collins, P. L. (2001). Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication *in vitro* and *in vivo*. *Virology* 289, 283-96.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876-82.

Tollman, S., Herbst, K., Garenne, M., Gear, J. & Kahn, K. (1999). The Agincourt demographic and health study--site description, baseline findings and implications. *S Afr Med J* 89, 858-64.

Tosato, G., Pike, S., Koski, I. & Blaese, R. (1982). Selective inhibition of immunoregulatory cell functions by cyclosporin A. *J Immunol* 128, 1986-91.

Tsutsumi, H., Takeuchi, R. & Chiba, S. (2001). Activation of cellular genes in the mucosal epithelium by respiratory syncytial virus: implications in disease and immunity. *Pediatr Infect Dis J* 20, 997-1001.

Tsutsumi, H., Takeuchi, R., Ohsaki, M., Seki, K. & Chiba, S. (1999). Respiratory syncytial virus infection of human respiratory epithelial cells enhances inducible nitric oxide synthase gene expression. *J Leukoc Biol* 66, 99-104.

Valarcher, J. F., Bourhy, H., Lavenu, A., Bourges-Abella, N., Roth, M., Andreoletti, O., Ave, P. & Schelcher, F. (2001). Persistent infection of B lymphocytes by bovine respiratory syncytial virus. *Virology* 291, 55-67.

Valdes, O., Martinez, I., Valdivia, A., Cancio, R., Savon, C., Goyenechea, A. & Melero, J. (1998). Unusual antigenic and genetic characteristics of human respiratory syncytial viruses isolated in Cuba. *J Virol* 72, 7589-92.

van den Hoogen, B. G., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. (2002). Analysis of the genomic sequence of a human metapneumovirus. *Virology* 295, 119-32.

Venter, M., Madhi, S. A., Tiemessen, C. T. & Schoub, B. D. (2001). Genetic diversity and molecular epidemiology of respiratory syncytial virus over four consecutive seasons in South Africa: identification of new subgroup A and B genotypes. *J Gen Virol* 82, 2117-24.

Venter, M., Collinson, M., and Schoub, B. (2002). Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: comparison of viruses and genotypes causing different disease manifestations. *J Med Virol* 68,452-61.

Voeten, J. T., Bestebroer, T. M., Nieuwkoop, N. J., Fouchier, R. A., Osterhaus, A. D. & Rimmelzwaan, G. F. (2000). Antigenic drift in the influenza A virus (H3N2) nucleoprotein and escape from recognition by cytotoxic T lymphocytes. *J Virol* 74, 6800-7.

Walsh, E. & Graham, B. (1999). Respiratory Syncytial Viruses. In *Viral Infections of the Respiratory Tract*, pp. pages 161-203. Edited by E. Dolin R. Wright PF. New York, NY.: Marcel Dekker, Inc.

Walsh, E. E., Brandriss, M. W. & Schlesinger, J. J. (1987). Immunological differences between the envelope glycoproteins of two strains of human respiratory syncytial virus. *J Gen Virol* 68, 2169-76.

Walsh, E. E., Cote, P. J., Fernie, B. F., Schlesinger, J. J. & Brandriss, M. W. (1986). Analysis of the respiratory syncytial virus fusion protein using monoclonal and polyclonal antibodies. *J Gen Virol* 67, 505-13.

Walsh, E. E., McConnochie, K., Long, C. E. & Hall, C. B. (1997). Severity of Respiratory syncytial virus infection is related to virus strain. *Journal of Inf Dis* 175, 814-820.

Wang, E. E., Law, B. J. & Stephens, D. (1995). Pediatric Investigators Collaborative Network on Infections in Canada (PICNIC) prospective study of risk factors and outcomes in patients hospitalised with respiratory syncytial viral lower respiratory tract infection. *J Pediatr* 126, 212-9.

Waris, M. E., Tsou, C., Erdman, D. D., Zaki, S. R. & Anderson, L. J. (1996). Respiratory synctial virus infection in BALB/c mice previously immunized with formalin-inactivated virus induces enhanced pulmonary inflammatory response with a predominant Th2-like cytokine pattern. *J Virol* 70, 2852-60.

Weber, M., Mulholland, E. & Greenwood, B. (1998). Review: Respiratory syncytial virus infection in tropical and developing countries.. *Trop Med Int Health* 3, 268-80.

Welliver, R. C., Kaul, T. N., Sun, M. & Ogra, P. L. (1984). Defective regulation of immune responses in respiratory syncytial virus infection. *J Immunol* 133, 1925-30.

Wertz, G. W., Collins, P. L., Huang, Y., Gruber, C., Levine, S. & Ball, L. A. (1985). Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc Natl Acad Sci U S A* 82, 4075-9.

Whitehead, S. S., Hill, M. G., Firestone, C. Y., St Claire, M., Elkins, W. R., Murphy, B. R. & Collins, P. L. (1999). Replacement of the F and G proteins of respiratory syncytial virus (RSV) subgroup A with those of subgroup B generates chimeric live attenuated RSV subgroup B vaccine candidates. *J Virol* 73, 9773-80.

WHO Department of Vaccines and Biologicals (V&B). (2001). Report of a workshop on the epidemiology of respiratory syncytial virus (RSV) in developing countries. Johannesburg, South Africa: World Health Organisation.

Woelk, C. H. & Holmes, E. C. (2001). Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *J Mol Evol* 52, 182-92.

Wright, P. & Cutts, F. (2000). Generic protocol to examine the incidence of lower respiratory infection due to respiratory syncytial virus in children less than five years of age, field test version. In *Document WHO/V&B/00.08*. Geneva: World Health Organization.

Wright, P. F., Karron, R. A., Belshe, R. B., Thompson, J., Crowe, J. E., Jr., Boyce, T. G., Halburnt, L. L., Reed, G. W., Whitehead, S. S., Anderson, E. L., Wittek, A. E., Casey, R., Eichelberger, M., Thumar, B., Randolph, V. B., Udem, S. A., Chanock, R. M. & Murphy, B. R. (2000). Evaluation of a live, cold-passaged, temperature-sensitive, respiratory syncytial virus vaccine candidate in infancy. *J Infect Dis* 182, 1331-42.

Wright, P. F., Shinozaki, T., Fleet, W., Sell, S. H., Thompson, J. & Karzon, D. T. (1976). Evaluation of a live, attenuated respiratory syncytial virus vaccine in infants. *J Pediatr* 88, 931-6.

Yellen-Shaw, A. J., Wherry, E. J., Dubois, G. C. & Eisenlohr, L. C. (1997). Point mutation flanking a CTL epitope ablates *in vitro* and *in vivo* recognition of a full-length viral protein. *J Immunol* 158, 3227-34.

Yu, Q., Hardy, R. W. & Wertz, G. W. (1995). Functional cDNA clones of the human respiratory syncytial (RS) virus N, P, and L proteins support replication of RS virus genomic RNA analogs and define minimal trans-acting requirements for RNA replication. *J Virol* 69, 2412-9.

Zinkernagel, R. M. & Doherty, P. C. (1974). Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* 248, 701-2.

Zinkernagel, R. M. & Hengartner, H. (1994). T-cell mediated immunopathology versus direct cytolysis by virus: Implication for HIV and AIDS. *Immunology Today* 15, 262-268.

Appendix A:

Fisher's exact test for positive selection in Subgroup A and B RSV isolates described in Chapter 4.

Title: Fischer's Exact test for selection: Subgroup A

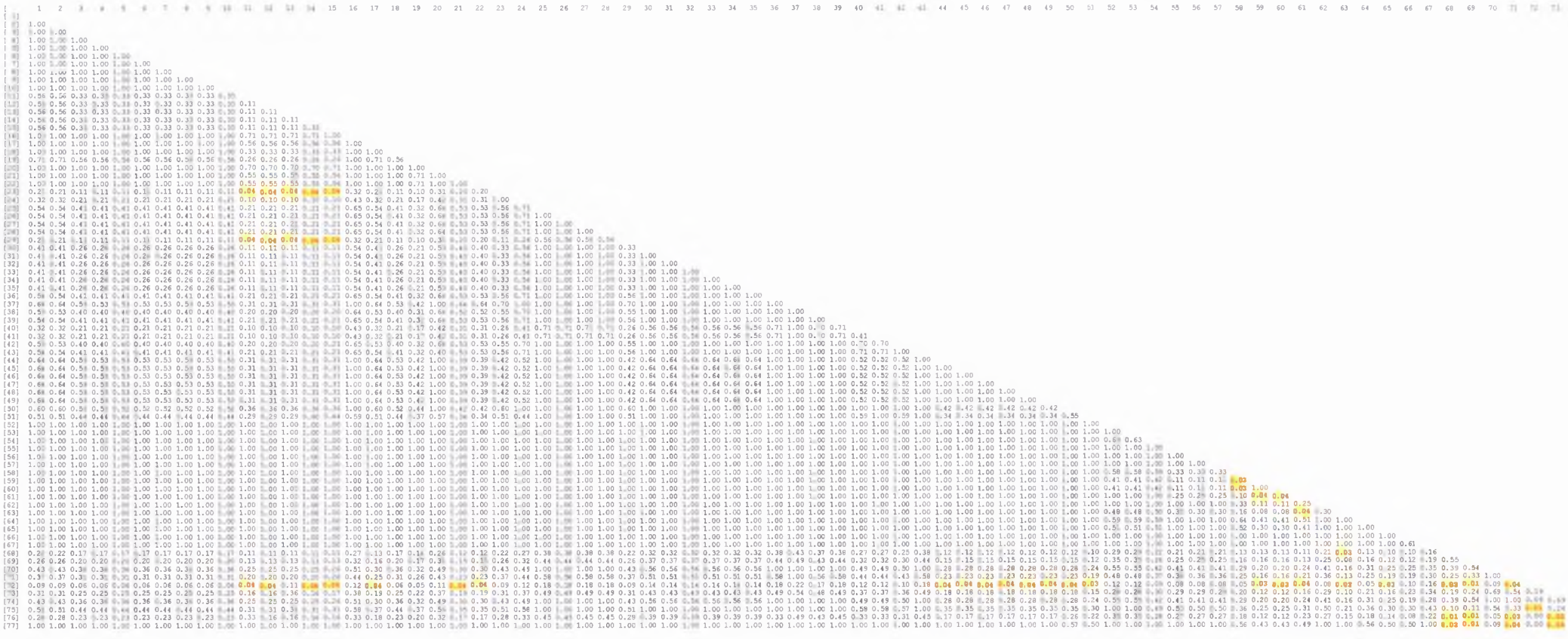
No. of Taxa : 77

Distance method : Codon: Modified Nei-Gojobori (Number of differences) [Pairwise distances]

Prob (black) : Probability computed (must be <0.05 for hypothesis rejection at 5% level [yellow])

If the number of synonymous substitutions are > than nonsynonymous substitutions the probability is set to 1

[1] SA0018_[GA2]
[2] AgA48_99_[GA2]
[3] Ab105B00_[GA2]
[4] Ab83J00_[GA2]
[5] Ab104B00_[GA2]
[6] Ab58Pt00_[GA2]
[7] Ab65Pt00_[GA2]
[8] Ab39Ct00_[GA2]
[9] AgJ9-00_[GA2]
[10] SA0004_[GA2]
[11] Ab52Pt00_[GA2]
[12] SA99V601_[GA2]
[13] SA0001_[GA2]
[14] SA0022_[GA2]
[15] Ab47Pt00_[GA2]
[16] SA0024_[GA2]
[17] AgB3-00_[GA2]
[18] SA0023_[GA2]
[19] SA0013_[GA2]
[20] Ab4026B01_[GA2]
[21] CJ-00AG_[GA2]
[22] Ab56Pt00_[GA2]
[23] SA0002_[GA2]
[24] SA0011_[GA2]
[25] Ab54Pt00_[GA2]
[26] Ab60Pt00_[GA2]
[27] SA99V1627_[GA2]
[28] Ab21B100_[GA2]
[29] SA0014_[GA2]
[30] SA0009_[GA2]
[31] Ab71D00_[GA2]
[32] Ab11B00_[GA2]
[33] Ab55Pt00_[GA2]
[34] Ab84J00_[GA2]
[35] Ab4029B01_[GA2]
[36] SA99V1229_[GA2]
[37] SA0019_[GA2]
[38] SA0020_[GA2]
[39] Ab86J00_[GA2]
[40] Ag32-00_[GA2]
[41] SA0005_[GA2]
[42] Ab31Ct00_[GA2]
[43] Ab43Ct00_[GA2]
[44] Ab78D00_[GA2]
[45] Ab24Ct00_[GA2]
[46] Ab42Ct00_[GA2]
[47] Ab29Ct00_[GA2]
[48] Ab28Ct00_[GA2]
[49] Ab35Ct00_[GA2]
[50] SA98V173_[GA2]
[51] MOZ-9-99_[GA2]
[52] SA99V1165_[GA5]
[53] SA97D989_[GA5]
[54] SA0007_[GA5]
[55] SA99V1395_[GA5]
[56] SA97D852_[GA5]
[57] Ab81J00_[GA5]
[58] Ab5076Pt01_[GA5]
[59] Ab20B100_[GA5]
[60] Ab164D00_[GA5]
[61] SA0012_[GA5]
[62] SA97D1131_[GA5]
[63] Ab3059CT01_[GA5]
[64] Moz-170-99_[GA5]
[65] Ab3061CT01_[GA5]
[66] SA99V360_[GA7]
[67] SA97D540_[GA7]
[68] SA98V341_[SAA1]
[69] SA98D612_[SAA1]
[70] AgK61-00_[SAA1]
[71] Ag38-00_[SAA1]
[72] SA99V1239_[SAA1]
[73] SA97D749_[SAA1]
[74] Ab34Pt00_[SAA1]
[75] SA99V640_[SAA1]
[76] SA97D669_[SAA1]
[77] SA98D707_[SAA1]



Title: Fischer's Exact test for selection Subgroup B

Distance method : Codon: Modified Nei-Gojobori (Number of differences) [Pairwise distances]

Prob (black) : Probability computed (must be <0.05 for hypothesis rejection at 5% level [red])

If the number of synonymous substitutions are > than nonsynonymous substitutions the probability is set to 1

```
[ 1] SA0025_{SAB1}
[ 2] AB17BL00_{SAB1}
[ 3] AB5078P01_{SAB3}
[ 4] SA98V372_{SAB3}
[ 5] AG_J15_99_{SAB3}
[ 6] SA0027_{SAB3}
[ 7] SA0028_{SAB3}
[ 8] AB3064C01_{SAB3}
[ 9] SA98D1303_{SAB3}
[10] AgA23_98_{SAB3}
[11] SA98V146_{SAB3}
[12] AB41CT_{SAB3}
[13] AB27CT00_{SAB3}
[14] SA98D1239_{SAB3}
[15] SA98D505_{SAB3}
[16] SA99V429_{SAB3}
[17] AB5075P01_{SAB3}
[18] SA0026_{GB3}
[19] K6_99AG_{GB3}
[20] J16_99AG_{GB3}
[21] MOZ50_99_{GB3}
[22] MOZ26_99_{GB3}
[23] AG21_00_{GB3}
[24] AgK53_00_{GB3}
[25] AgA88_00_{GB3}
[26] AgC13_00_{GB3}
[27] AB3062C01_{GB3}
[28] SA98V220_{GB3}
[29] SA99V800_{SAB2}
[30] MOZ198_99_{SAB2}
[31] SA99V1325_{SAB2}
[32] SA98D1656
[33] SA99V439_{GB4}
[34] SA97D1107_{GB4}
[35] SA98V602_{GB4}
[36] SA98V468_{GB4}
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
[1]																																				
[2]	1.00																																			
[3]	0.51	1.00																																		
[4]	0.58	1.00	0.17																																	
[5]	0.32	0.39	0.04	0.11																																
[6]	1.00	1.00	0.42	0.55	0.11																															
[7]	1.00	1.00	0.42	0.55	0.11	1.00																														
[8]	0.57	0.37	0.52	0.42	0.13	1.00	0.52																													
[9]	1.00	1.00	0.32	0.33	0.03	1.00	1.00	0.64																												
[10]	1.00	1.00	0.43	0.55	0.11	1.00	1.00	1.00	1.00																											
[11]	1.00	1.00	0.26	0.25	0.04	0.70	0.70	0.52	0.55	0.70																										
[12]	1.00	1.00	0.42	0.55	0.11	1.00	1.00	1.00	1.00	1.00	0.70																									
[13]	1.00	1.00	0.42	0.55	0.11	1.00	1.00	1.00	1.00	1.00	0.70	1.00																								
[14]	1.00	1.00	0.25	0.25	0.04	0.69	0.69	0.51	0.55	0.70	0.40	0.33	0.33																							
[15]	1.00	1.00	0.08	0.03	0.00	0.25	0.25	0.25	0.11	0.25	0.11	0.25	0.25	0.11																						
[16]	1.00	1.00	0.25	0.25	0.04	0.70	0.70	0.52	0.55	0.70	0.40	0.69	0.69	0.39	0.11																					
[17]	1.00	1.00	0.34	0.40	0.09	1.00	1.00	0.61	0.70	1.00	0.53	1.00	1.00	0.52	0.20	1.00																				
[18]	1.00	1.00	1.00	0.57	0.30	1.00	1.00	0.56	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00																			
[19]	1.00	1.00	1.00	0.51	0.25	0.60	0.59	0.50	1.00	1.00	0.59	1.00	1.00	0.57	1.00	1.00	1.00	1.00																		
[20]	0.57	1.00	1.00	0.44	0.19	0.51	0.51	0.43	0.60	1.00	0.51	1.00	1.00	0.50	0.60	1.00	1.00	1.00	1.00																	
[21]	1.00	1.00	0.51	0.28	0.09	0.34	0.52	0.44	0.43	0.52	0.36	0.51	0.51	0.34	0.43	1.00	1.00	1.00	1.00	1.00																
[22]	1.00	1.00	0.44	0.52	0.20	0.64	1.00	1.00	1.00	1.00	0.62	1.00	1.00	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00															
[23]	0.32	0.39	0.05	0.05	0.01	0.06	0.14	0.13	0.09	0.06	0.08	0.13	0.13	0.07	0.09	0.23	0.28	0.18	0.25	0.18	0.09	0.20														
[24]	0.37	0.44	0.07	0.06	0.01	0.07	0.16	0.16	0.11	0.16	0.09	0.16	0.16	0.09	0.11	0.28	0.34	0.35	0.29	0.22	0.11	0.25	0.55													
[25]	0.31	0.49	0.09	0.09	0.02	0.11	0.23	0.20	0.16	0.23	0.14	0.22	0.22	0.13	0.16	0.36	0.42	0.43	0.22	0.16	0.16	0.34	0.09	0.11												
[26]	0.58	1.00	0.26	0.29	0.11	0.36	0.51	0.44	0.44	0.51	0.37	0.50	0.50	0.35	0.44	1.00	1.00	1.00	0.50	0.42	0.44	1.00	0.29	0.36	0.70											
[27]	0.57	1.00	0.17	0.19	0.06	0.23	0.35	0.31	0.29	0.37	0.25	0.35	0.35	0.23	0.29	0.51	0.57	0.43	0.35	0.43	0.16	0.52	0.06	0.08	0.11	0.31										
[28]	1.00	1.00	1.00	1.00	0.44	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.44	0.52	0.60	1.00	1.00									
[29]	0.57	0.51	0.17	0.19	0.06	0.23	0.37	0.20	0.29	0.37	0.25	0.35	0.35	0.23	0.11	0.36	0.42	0.43	0.38	0.31	0.20	0.37	0.04	0.05	0.08	0.22	0.14	0.58								
[30]	0.51	0.44	0.14	0.14	0.04	0.16	0.29	0.16	0.23	0.29	0.19	0.29	0.29	0.19	0.08	0.44	0.51	0.37	0.32	0.26	0.16	0.29	0.03	0.03	0.05	0.17	0.11	0.51	1.00							
[31]	0.26	0.23	0.07	0.07	0.03	0.08	0.14	0.08	0.11	0.14	0.05	0.14	0.14	0.09	0.04	0.21	0.25	0.19	0.16	0.13	0.08	0.14	0.02	0.03	0.04	0.05	0.26	0.21	0.31	0.25						
[32]	1.00	1.00	0.38	0.43	0.22	0.49	0.48	0.43	0.56	1.00	0.49	1.00	1.00	1.00	0.56	1.00	1.00	1.00	0.54	0.48	1.00	1.00	0.49	0.56	0.49	1.00	1.00	1.00	0.49	0.43	0.38					
[33]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00				
[34]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
[35]	1.00	1.00	0.33	0.37	0.37	0.43	0.41	0.46	0.48	0.48	0.43	0.52	0.52	0.41	0.48	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.38	0.33	0.38	0.48	0.48	1.00	1.00	0.54	0.22	1.00	1.00	1.00		
[36]	1.00	1.00	0.22	0.23	0.23	0.27	0.27	0.33	0.32	0.32	0.28	0.37	0.37	0.28	0.32	0.48	0.54	1.00	0.54	1.00	0.43	1.00	0.24	0.20	0.24	0.33	0.33	1.00	1.00	1.00	0.25	0.56	0.59	0.51	0.54	

APPENDIX B South African RSV sequences submitted to the genbank.

Partial G-protein genes (Chapter 2)

Subgroup A			Subgroup B		
Isolate	**Accession no.**	**Genotype**	**Isolate**	**Accession no**	**Genotype**
SA97D804	AF348802	GA5	SA98V192	AF348811	SAB3
SA97D1289	AF348803	GA5	SA0028	AF348812	SAB3
SA99V360	AF348804	GA7	SA99V429	AF348813	SAB3
SA0003	AF348805	GA2	SA99V470	AF348814	GB3
SA98V173	AF348806	GA2[1]	SA98V153	AF348815	GB3
SA98V603	AF348807	SAA1	SA98D661	AF348816	GB3
SA99V1239	AF348808	SAA1	SA97D934	AF348817	GB3
SA97D669	AF348809	SAA1	SA98V268	AF348818	GB3
SA98D707	AF348810	SAA1	SA98D941	AF348819	GB3
			SA98V220	AF348820	GB3
			SA99V800	AF348821	SAB2
			A99V1325	AF348822	SAB2
			SA99V439	AF348823	GB4
			SA98V602	AF348824	GB4
			SA0025	AF348825	SAB1
			SA98D1656	AF348826	SAB4[2]

Partial G-protein Genes (Chapter 3)					
Ag38-00	AF548797	GA5	Ag21-00	AF548803	GB3
AgK28-00	AF548798	GA2	AgA88-00	AF548804	GB3
AgK23-00	AF548799	GA2	AgC13-00	AF548805	GB3
AgJ6_98	AF548800	GA2	AgK6-99	AF548806	GB3
Ag_A48_99	AF548801	GA2	J16_99Ag	AF548807	GB3
Ag32-00	AF548802	GA2	AgA23_98	AF548808	SAB3
			AgJ15-99	AF548809	SAB3

1. SA98V173 group separately from all other GA2 isolates.
2. SA98D1656 was genotyped as SAB1 in Chapter 2 but was re-typed as SAB4 when more isolates were added to the analysis in Chapter 4.

Partial G-protein genes (Chapter 4)

Isolate	Accession no.	Genotype	Isolate	Accession no	Genotype
Ab78D00	AY146411	GA2	Ab17BL00	AY146438	SAB1
Ab71D00	AY146412	GA2	Ab41CT00	AY146439	SAB3
Ab39Ct00	AY146413	GA2	Ab27CT00	AY146440	SAB3
Ab56Pt00	AY146414	GA2	Ab5078P01	AY146441	SAB3
Ab20Bl00	AY146415	GA2	Ab5075P01	AY146442	SAB3
Ab65Pt00	AY146416	GA2	Ab3064C01	AY146443	SAB3
Ab24Ct00	AY146417	GA2	Ab3062C01	AY146444	GB3
Ab43Ct00	AY146418	GA2			
Ab54Pt00	AY146419	GA2			
Ab21Bl00	AY146420	GA2			
Ab31Ct00	AY146421	GA2			
Ab11B00	AY146422	GA2			
Ab34Pt00	AY146423	SAA1			
Ab55Pt00	AY146424	GA2			
Ab104B00	AY146425	GA2			
Ab81J00	AY146426	GA5			
Ab84J00	AY146427	GA2			
Ab86J00	AY146428	GA2			
Ab47Pt00	AY146429	GA2			
Ab83J00	AY146430	GA2			
Ab105B00	AY146431	GA2			
Ab164D00	AY146432	GA5			
Ab3059CT01	AY146433	GA5			
Ab3061CT01	AY146434	GA5			
Ab4026B01	AY146435	GA2			
Ab4029B01	AY146436	GA2			
Ab5076Pt01	AY146437	GA5			

Partial Nucleoprotein genes (Chapter 6)

Isolate	Accession	Genotype	Isolate	Accession	Genotype
GA2sa98V173	AY151194	GA2	SAB2sa99V1325	AY151200	SAB2
SAA1sa98D707	AY151195	SAA1	SAB3Ab5075Bl01	AY151201	SAB3
GA7sa97D540	AY151196	GA7	SAB3Ab27CT00	AY151202	SAB3
GA5sa97D1131	AY151197	GA5	SAB1Ab17Bl00	AY151203	SAB1
GA5Ab3061CT01	AY151198	GA5	SAB3SA98V192	AY151204	SAB3
GA2Ab4029Bl01	AY151199	GA2	GB3SA98D796	AY151205	GB3
			GB4SA97D1107	AY151206	GB4
			SAB1SA0025	AY151207	SAB1

APPENDIX C Ethics clearance

Ethical clearance has been obtained from the COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL), University of the Witwatersrand, Johannesburg for:

1.) The Project: Molecular Epidemiology or Respiratory Syncytial virus in South Africa, under reference number R14/49Venter and protocol number M01-09-35, on the 28th of September 2001.

2.) The Project: Defining Epitope Specificity and HLA Restriction Patterns in Respiratory Syncytial Virus (RSV) induced CD8+ T-cells in South African Adults, under reference number R14/49 Venter and Protocol number M15, on the 28th of January 2001.