

**ROLE OF MOLECULAR EVOLUTION IN  
RESPIRATORY SYNCYTIAL VIRUS  
ANTIGENIC PROTEINS DURING ANNUAL  
EPIDEMICS IN SOUTH AFRICA**

Elizabeth Agenbach

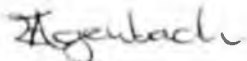
A Thesis submitted to the Faculty of Health Sciences  
University of the Witwatersrand, Johannesburg  
In fulfilment of the requirements for the  
Degree of Master of Science

Johannesburg 2004

---

## DECLARATION

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



.....  
Elizabeth Agenbach

On this 28 day of May, 2004.

**To Charl**

with sincere thanks  
for his love and encouragement  
during the writing of this thesis

**Science knows no country, because knowledge belongs to humanity, and is the torch which  
illuminates the world.**

-Louis Pasteur (1822-1895)

---

## **PUBLICATIONS**

1. E. Agenbach, C. T. Tiemessen, M. Venter. Antigenic variability within the Respiratory syncytial virus fusion protein during annual epidemics in South Africa. Submitted to Journal of General Virology.

## **PRESENTATIONS**

1. E. Agenbach, S. A. Madhi, C.T. Tiemessen, M. Venter. The role of molecular evolution in respiratory syncytial virus antigenic proteins during annual epidemics in South Africa. Oral/poster presentation, Medical Virology Congress, Kruger National Park, South Africa, May 2003.
2. E. Agenbach. The role of molecular evolution in respiratory syncytial virus antigenic proteins during annual epidemics in South Africa. Junior scientific lecture, National Institute for Communicable Diseases, Johannesburg, South Africa, July 2003.

---

## **Acknowledgements**

All the experimental work was conducted at the Molecular Virology Unit of the National Institute for Communicable Diseases, Johannesburg, South Africa.

My Sincere thanks to:

My supervisor, Marietjie Venter at the National Institute of Communicable Diseases, for her guidance, training, revision of this thesis and constant support and interest throughout my RSV studies.

My co-supervisor, Dr. Caroline Tiemessen at the National Institute for Communicable Diseases for revision of this thesis.

Dr. Shabir Madhi at the Chris Hani Baragwanath hospital for the provision of all of the specimens used in this study.

The Poliomyelitis Research Foundation, National Institute for Communicable Diseases and Dr. Shabir Madhi for financial support.

Lizelle Botha and Samantha Donninger at the National Institute for Communicable Diseases for their technical assistance in sequencing.

For Charl, my parents and friends for their encouragement and support.

---

## **Abstract**

Respiratory syncytial virus (RSV) is a major cause of pneumonia in children worldwide and there is currently no vaccine available. Molecular analysis of annual RSV epidemics in Soweto (1997 to 2000), revealed co-circulation and displacement of dominant genotypes. However in 1999 to 2000, one genotype GA2, predominated. To further elucidate the molecular epidemiology of RSV in South Africa the 2001 RSV epidemic in Soweto was characterised. Phylogenetic analysis of the G-protein revealed subtype B dominance (89%). Sequences clustered in genotypes GA2, GA5, SAA1, GB3, SAB1 and SAB3. SAB3 dominated (69%) followed by SAB1 (18%), thereby displacing GA2, which dominated the previous two seasons. Evidence of positive selection may account for the genetic variability observed and may contribute to the reestablishment of annual epidemics. To investigate if this influences evolution of other RSV antigenic proteins the F-protein of South African genotypes was characterised. Most amino acid differences identified within known neutralising and CTL epitopes were conserved within subtype A, and although this does not suggest immune selection these epitopes may not be recognised efficiently by antibodies or CTL specific to subtype B viruses.

## Table of Contents

|                                     |     |
|-------------------------------------|-----|
| Declaration.....                    | I   |
| Publications and Presentations..... | III |
| Acknowledgements.....               | IV  |
| Abstract.....                       | V   |
| List of Figures.....                | IX  |
| List of Tables.....                 | XI  |
| List of Abbreviations.....          | XII |

|                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------|----|
| <b>chapter 1</b> Literature Survey .....                                                                         | 1  |
| 1.1 Introduction .....                                                                                           | 1  |
| 1.2 Molecular biology of RSV .....                                                                               | 3  |
| 1.2.1 Classification .....                                                                                       | 3  |
| 1.2.2 RSV structure and replication .....                                                                        | 3  |
| 1.2.3 Protein function and importance .....                                                                      | 4  |
| 1.3 Epidemiology .....                                                                                           | 9  |
| 1.3.1 Subtypes of RSV .....                                                                                      | 10 |
| 1.3.2 Genotypes of RSV .....                                                                                     | 11 |
| 1.3.3 Molecular epidemiology .....                                                                               | 12 |
| 1.3.4 Evolution of RSV .....                                                                                     | 14 |
| 1.4 RSV disease .....                                                                                            | 15 |
| 1.5 Role of RSV proteins in disease and protection.....                                                          | 16 |
| 1.5.1 The surface glycoproteins.....                                                                             | 17 |
| 1.5.2 The nucleoprotein.....                                                                                     | 20 |
| 1.6 Vaccine development .....                                                                                    | 20 |
| 1.6.1 Immune responses important for protection .....                                                            | 21 |
| 1.6.2 Experimental vaccines against RSV: the importance of the G- and F-<br>protein in vaccine development ..... | 22 |
| 1.7 The role of bioinformatics in modeling antigens.....                                                         | 27 |
| 1.7.1 Importance of bioinformatics.....                                                                          | 27 |
| 1.7.2 Methods of predicting epitopes.....                                                                        | 28 |
| 1.7.3 Identification of immune selection.....                                                                    | 31 |
| 1.8 STUDY OBJECTIVES .....                                                                                       | 33 |

|                  |                                                                                                                    |    |
|------------------|--------------------------------------------------------------------------------------------------------------------|----|
| <b>chapter 2</b> | <b>Materials and Methods</b>                                                                                       | 35 |
| 2.1              | Patient specimens                                                                                                  | 35 |
| 2.2              | RNA extraction                                                                                                     | 35 |
| 2.3              | G-protein specific Reverse transcriptase-polymerase chain reaction (RT-PCR) and multiplex nested PCR for subtyping | 37 |
| 2.3.1            | RT-PCR                                                                                                             | 37 |
| 2.3.2            | Nested polymerase chain reaction (nested PCR)                                                                      | 37 |
| 2.4              | PCR product analysis                                                                                               | 39 |
| 2.5              | Sequencing                                                                                                         | 39 |
| 2.5.1            | PCR product purification                                                                                           | 39 |
| 2.5.2            | Cycle sequencing                                                                                                   | 40 |
| 2.5.3            | Sequence product purification                                                                                      | 40 |
| 2.5.4            | Automated sequencing                                                                                               | 41 |
| 2.6              | Phylogenetic analysis                                                                                              | 41 |
| 2.6.1            | Alignments                                                                                                         | 41 |
| 2.6.2            | Phylograms                                                                                                         | 41 |
| 2.6.3            | Genotype criteria                                                                                                  | 43 |
| 2.6.4            | Estimates of selection                                                                                             | 43 |
| 2.7              | Secondary structure predictions                                                                                    | 44 |
| 2.8              | Amplification of the F-protein by RT-PCR                                                                           | 44 |
| 2.8.1            | Primer design                                                                                                      | 44 |
| 2.8.2            | F-protein RT-PCR                                                                                                   | 45 |
| 2.9              | Sequences used in phylogenetic analysis                                                                            | 47 |
| 2.10             | Protein secondary structure predictions                                                                            | 47 |
| 2.11             | Prediction of antigenic sites on the F-protein                                                                     | 47 |
| 2.12             | Prediction of CTL epitopes                                                                                         | 48 |
| 2.12.1           | Published CTL epitopes                                                                                             | 48 |
| 2.13             | Amino acid sequence analysis of known neutralising epitopes                                                        | 48 |
| 2.14             | Escape mutants                                                                                                     | 48 |
| <b>chapter 3</b> | <b>Results A: Molecular epidemiological analysis of a RSV outbreak in Soweto during 2001</b>                       | 49 |
| 3.1              | Introduction                                                                                                       | 49 |

|                  |                                                                                                     |            |
|------------------|-----------------------------------------------------------------------------------------------------|------------|
| 3.2              | Results.....                                                                                        | 52         |
| 3.2.1            | Subtype-specific multiplex nested PCR.....                                                          | 52         |
| 3.2.2            | Subtyping of isolates from the 2001 RSV outbreak in Soweto .....                                    | 54         |
| 3.2.3            | Phylogenetic analysis .....                                                                         | 55         |
| 3.2.4            | Strain variation.....                                                                               | 58         |
| 3.2.5            | G-protein gene nucleotide mutation analysis during the 2001 outbreak.....                           | 62         |
| 3.2.6            | Estimates of selection .....                                                                        | 64         |
| 3.2.6            | Amino acid analysis of Soweto 2001 sequences.....                                                   | 66         |
| 3.2.7            | G-protein amino acid variation over consecutive epidemics.....                                      | 70         |
| 3.2.8            | Antigenic variation in the G-protein.....                                                           | 73         |
| 3.2.9            | Genotype displacement over 5 consecutive seasons in Soweto .....                                    | 75         |
| 3.3              | Discussion .....                                                                                    | 76         |
| <b>chapter 4</b> | <b>Results B: Antigenic variability within the RSV fusion protein during annual epidemics .....</b> | <b>82</b>  |
| 4.1              | Introduction.....                                                                                   | 82         |
| 4.2              | Results.....                                                                                        | 85         |
| 4.2.1            | Amplification of the F-protein .....                                                                | 85         |
| 4.2.2            | Sequence analysis .....                                                                             | 86         |
| 4.2.3            | Prediction of the secondary structure of the F-protein from the amino acid alignments.....          | 94         |
| 4.2.4            | Prediction of antigenic sites.....                                                                  | 100        |
| 4.2.5            | Epitope analysis .....                                                                              | 102        |
| 4.3              | Discussion .....                                                                                    | 107        |
| <b>chapter 5</b> | <b>Concluding Remarks.....</b>                                                                      | <b>116</b> |
| <b>chapter 6</b> | <b>References .....</b>                                                                             | <b>122</b> |
|                  | Appendix A.....                                                                                     | 141        |
|                  | Appendix B.....                                                                                     | 142        |
|                  | Appendix C.....                                                                                     | 143        |
|                  | Appendix D.....                                                                                     | 144        |

## **List of Figures**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1 An illustration of a <i>Paramyxovirus</i> particle. ....                                                                                                                                                                                                                                                                                                                                                                                                                           | 4  |
| Figure 2.1 Illustration of the sequence primer positions relative to the F-protein gene<br>for subtype A (A) and B (B), respectively. ....                                                                                                                                                                                                                                                                                                                                                    | 46 |
| Figure 3.1 A) RSV Subtype-specific multiplex RT-PCR, B) The subtype-specific<br>multiplex nested PCR using Expand high fidelity PCR enzyme (Roche,<br>Mannheim, Germany) and C) An illustration of hotstart PCR using HotStarTaq<br>DNA polymerase (QIAGEN, Hilden, Germany).....                                                                                                                                                                                                             | 53 |
| Figure 3.2 RSV subtype distribution at Chris Hani Baragwanath hospital during 1997<br>to 2001.....                                                                                                                                                                                                                                                                                                                                                                                            | 54 |
| Figure 3.3 Neighbor-joining trees for subtype A (A) and B (B),<br>respectively.....                                                                                                                                                                                                                                                                                                                                                                                                           | 57 |
| Figure 3.4 Frequency of specific types of transitions (AG+GA, CU+UC) and<br>transversions (AC+CA, UA+AU, CG+GC, UG+GU) within the G-protein of<br>Soweto 2001 sequences.....                                                                                                                                                                                                                                                                                                                  | 63 |
| Figure 3.5 The amino acid alignment of the G-protein's second variable region of<br>unique RSV sequence groups identified in the 2001 Soweto outbreak. ....                                                                                                                                                                                                                                                                                                                                   | 68 |
| Figure 3.6 Illustration of genotype-specific mutations and mutations that may affect<br>the secondary structure of the G-protein.....                                                                                                                                                                                                                                                                                                                                                         | 71 |
| Figure 3.7 The amino acid alignment of subtype A and B Soweto 2001 sequence<br>groups. A) The amino acid changes in strain-specific epitopes and natural<br>occurring strains which fail to react with monoclonal antibodies are indicated. B)<br>Four potential antigenic sites, predicted with Antheprot 2000 version 5.2<br>(Deleage <i>et al.</i> , 1988) ( <a href="http://antheprot-pbil.ibcp.fr">http://antheprot-pbil.ibcp.fr</a> ) are indicated on the subtype<br>B sequences. .... | 72 |
| Figure 3.8 Identification of antigenic sites within the subtype B isolates by means of<br>antigenicity and hydrophobicity plots .....                                                                                                                                                                                                                                                                                                                                                         | 74 |
| Figure 3.9 Genotype distribution during 1997 – 2001 at Chris Hani Baragwanath<br>hospital. ....                                                                                                                                                                                                                                                                                                                                                                                               | 75 |
| Figure 4.1 RT-PCR products of RSV fusion protein for both subtypes. ....                                                                                                                                                                                                                                                                                                                                                                                                                      | 85 |
| Figure 4.2 Amino acid alignment of the F-protein sequences of the South African<br>genotypes with bovine RSV and ovine RSV .....                                                                                                                                                                                                                                                                                                                                                              | 90 |
| Figure 4.3 Amino acid identity within the different RSV fusion protein domains.....                                                                                                                                                                                                                                                                                                                                                                                                           | 93 |

---

|                                                                                                                        |     |
|------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.4 Secondary structure prediction for the F-protein sequence of South African subtype A and B genotypes.....   | 95  |
| Figure 4.5 Prediction of coiled coils in the secondary structure of the fusion protein....                             | 99  |
| Figure 4.6 Prediction of possible antigenic sites within the fusion protein by means of physico-chemical profiles..... | 100 |
| Figure 4.7 Hydropathic profiles per amino acid residue of the F-protein.....                                           | 102 |
| Figure 4.8 The position of CTL and neutralising epitopes on the South African F-protein sequences.....                 | 106 |
| Figure 4.9 Phylogenetic analysis of the <i>Pneumovirinae</i> subfamily.....                                            | 115 |

## **List of Tables**

|                                                                                                                                                                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 2.1 RSV positive clinical specimens that were selected for fusion protein amplification.....                                                                                                                                                                                                                  | 36  |
| Table 2.2 Primers used in G-protein specific RT-PCR and nested PCR.....                                                                                                                                                                                                                                             | 39  |
| Table 2.3 Subtype-specific primers used for the sequencing of the F-protein.....                                                                                                                                                                                                                                    | 46  |
| Table 3.1 RSV clinical specimens analysed during the 2001 Soweto outbreak.....                                                                                                                                                                                                                                      | 59  |
| Table 3.2 Pairwise nucleotide distances (P-distance) for all unique subtype A (A) and B (B) Soweto 2001 sequences.....                                                                                                                                                                                              | 60  |
| Table 3.3 Average nucleotide P-distances within and between subtype A (A) and B (B) genotypes, respectively.....                                                                                                                                                                                                    | 61  |
| Table 3.4 Ratio of non-synonymous substitutions per non-synonymous site (Ka) to synonymous substitutions per synonymous site (Ks) for all the Soweto genotypes..                                                                                                                                                    | 65  |
| Table 3.5 Amino acid P-distance within and between Soweto genotypes. ....                                                                                                                                                                                                                                           | 69  |
| Table 3.6 Differences in termination codon usage by Soweto sequences (2001).....                                                                                                                                                                                                                                    | 69  |
| Table 3.7 Amino acid changes identified in RSV subtype A strain-specific epitopes residing in the second variable region of the G-protein.....                                                                                                                                                                      | 74  |
| Table 4.1 The nucleotide pairwise distance of all the South African F-protein isolates.....                                                                                                                                                                                                                         | 87  |
| Table 4.2 Ratio of non-synonymous substitutions per non-synonymous site (Ka) to synonymous substitutions per synonymous site (Ks) for the F-protein for subtype A (A), subtype B (B) and all the South African genotypes, relative to the prototype strains A2 and B1 for subtype A and B, respectively (A+B). .... | 88  |
| Table 4.3 Intergenotypic amino acid pairwise distance of the F-protein of the South African specimens relative to the prototype strains.....                                                                                                                                                                        | 89  |
| Table 4.4 Possible CTL epitopes predicted by SYFPHEITHI and BIMAS algorithms for the South African HLA types.....                                                                                                                                                                                                   | 102 |

---

**List of Abbreviations**

|        |                                                 |
|--------|-------------------------------------------------|
| ALRI   | Acute lower respiratory tract infections        |
| cDNA   | Copy Deoxyribonucleic acid                      |
| CTL    | Cytotoxic T lymphocyte                          |
| DNA    | Deoxyribonucleic acid                           |
| DTT    | Dithiothreitol                                  |
| dNTP   | Deoxynucleoside triphosphate                    |
| FI-RSV | Formalin-inactivated RSV vaccine                |
| H-2    | Histocompatibility antigens of mice             |
| HIV    | Human immunodeficiency virus                    |
| HLA    | Human leukocyte antigen                         |
| HPLC   | High Performance Liquid Chromatography          |
| IFA    | Immunofluorescence assay                        |
| Ig     | Immunoglobulin                                  |
| LRTI   | Lower respiratory tract infection               |
| MHC    | Major histocompatibility complex                |
| M      | Molar                                           |
| Mabs   | Monoclonal antibodies                           |
| mRNA   | Messenger ribonucleic acid                      |
| ng     | Nanogram                                        |
| NICD   | National Institute for Communicable Diseases    |
| ORF    | Open reading frame                              |
| PCR    | Polymerase chain reaction                       |
| RFLP   | Restriction fragment length polymorphism        |
| RNA    | Ribonucleic acid                                |
| Rnase  | Ribonuclease                                    |
| RSV    | Respiratory syncytial virus                     |
| RT-PCR | Reverse transcriptase-polymerase chain reaction |
| Th     | T-helper cells                                  |
| U      | Unit                                            |
| UK     | United Kingdom                                  |
| US     | United States                                   |
| μ      | Micro                                           |
| WHO    | World Health Organisation                       |

# CHAPTER 1

## LITERATURE SURVEY

### 1.1 INTRODUCTION

Respiratory syncytial virus (RSV) was originally identified in a laboratory chimpanzee in 1955 and called “chimpanzee coryza agent” (Morris *et al.*, 1956). The following year, an identical virus was detected in lung secretions of hospitalised infants with acute lower respiratory tract infection (ALRI) and the virus was renamed respiratory syncytial virus to reflect the giant syncytia, which developed during tissue culture growth (Chanock *et al.*, 1957). RSV is a pneumovirus and the most important viral cause of pneumonia in infants and young children in industrialised and developing countries (Selwyn, 1990; Simoes & Carbonell-Estrany, 2003). It is a significant cause of infant morbidity and mortality especially in developing countries (Stensballe *et al.*, 2003), and is a serious pathogen in immunocompromised adults (Englund *et al.*, 1988; Falsey *et al.*, 1992). It is also an important cause of community-acquired pneumonia among hospitalised adults of all ages (Meissner, 2003; Sullender, 2000), and a major nosocomial hazard in paediatric wards during annual outbreaks (Hall, 2000).

RSV is unusual in that it causes repeated infections throughout life in all age groups, reflecting incomplete immunity to this virus (Reviewed by Walsh & Graham, 1999). This is only one obstacle in vaccine development, others include, the lack of an effective animal model to study the disease, the challenge of successfully immunising very young infants who are immunologically immature, and that antigenic variation may play a role in the ability of RSV to evade the immune response (Crowe, 2002). RSV vaccine development has also been hampered by an incident that occurred in the 1960's where the administration of a formalin-inactivated RSV vaccine (FI-RSV) to infants resulted in enhanced lung disease and even in some deaths among the recipients (Kapikian *et al.*, 1969). There is currently no effective licensed RSV vaccine available, in part because the mechanism resulting in the FI-RSV vaccine-enhanced disease is not yet fully understood (Reviewed by Varga & Braciale, 2002).

Epidemiological analysis of RSV is important for providing necessary background information for the development of an effective vaccine (Sullender, 2000). The majority of molecular epidemiological studies conducted on RSV, have been carried out in industrialised countries (Reviewed by Cane, 2001), hence, there is limited data available about RSV epidemics in developing countries (Cane *et al.*, 1999; Roca *et al.*, 2001; Venter *et al.*, 2002; Venter *et al.*, 2001). RSV consists of two antigenic subtypes (Anderson *et al.*, 1985) each with distinct lineages or genotypes (Reviewed by Cane, 2001).

A previous study conducted in South Africa from 1997 to 2000 has shown that both RSV subtypes are important in local RSV outbreaks and that co-circulation, switching and displacement of dominant genotypes takes place from one season to the next (Venter *et al.*, 2001), consistent with findings from the Northern hemisphere (Peret *et al.*, 1998). However in 1999 to 2000, one genotype (GA2) became dominant. To determine if this was a once-off phenomenon that would eventually be overcome by herd immunity, phylogenetic analysis of the G-protein from the 2001 RSV outbreak was undertaken to study the transmission dynamics of RSV in a single community, the evolution rate in the G-protein as well as the role of this neutralisation antigen in reestablishment of annual epidemics. The correlation between the variations observed in the currently identified G-protein genotypes and changes in the antigenic structure of this protein, especially those documented in escape mutants (Cane & Pringle, 1995; Garcia *et al.*, 1994), were determined. Although an absence of positive selection has been previously found in the cytotoxic T lymphocytes (CTL) epitopes of the conserved RSV nucleoprotein (N-protein) (Venter *et al.*, 2003), immune selection does however occur in the CTL epitopes of the highly conserved influenza nucleoprotein (Boon *et al.*, 2002; Price *et al.*, 2000; Voeten *et al.*, 2000), which implies the need for regular updating of vaccines containing these CTL epitopes. To further look at evolution of RSV antigenic proteins, the F-protein of known South African genotypes, was studied to determine if immune selection occurs in both the neutralising antibody and CTL epitopes of the known South African RSV genotypes. This is the first study to investigate the F-protein of South African genotypes.

The following literature review will give an overview of RSV biology to provide an understanding of the molecular characteristics of the virus, the epidemiology, the role of antigenic proteins in disease and protection, and vaccine development. A short description about the potential role and importance of bioinformatics in the modelling of antigens as a predictive tool in vaccine development will also be given.

## 1.2 MOLECULAR BIOLOGY OF RSV

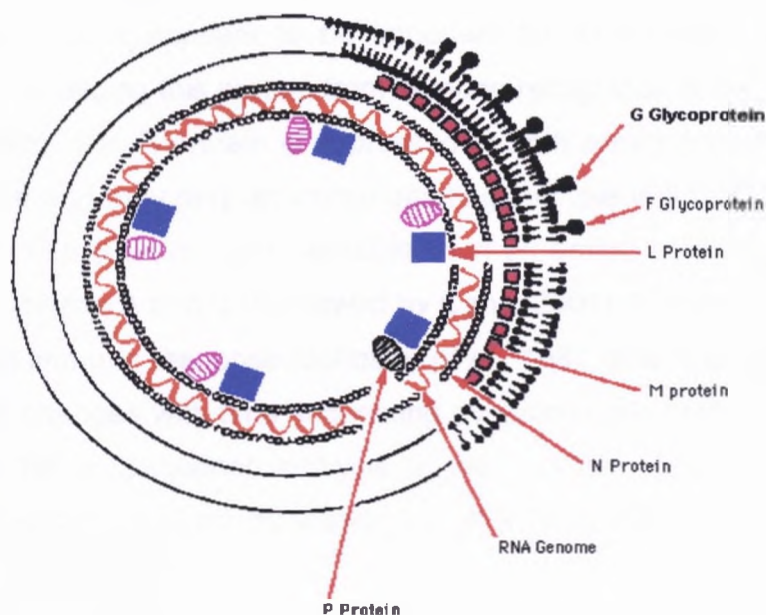
### 1.2.1 Classification

RSV belongs to the *Pneumovirus* genus in the subfamily *Pneumovirinae* in the family *Paramyxoviridae*. The *Paramyxoviridae* are a group of enveloped viruses characterised by a nonsegmented, single-stranded, negative-sense RNA genome (Collins *et al.*, 1996b). The *Pneumovirus* genus includes bovine, ovine and caprine RSV and pneumovirus of mice. The subfamily *Paramyxovirinae* includes the genus *Respirovirus* (Sendai virus, human parainfluenza virus types 1 and 3), *Morbillivirus* (measles virus) and *Rubulavirus* (mumps virus, simian virus 5, human parainfluenza virus type 2) genus (Reviewed by Collins *et al.*, 1999).

### 1.2.2 RSV structure and replication

RSV has a pleomorphic spherical structure (Bachi, 1988) and encodes 10 subgenomic mRNAs that are capped and polyadenylated. Every mRNA encodes a single viral protein, except for the M2 mRNA, which contains two overlapping open reading frames (ORFs), that encodes the M2-1 and M2-2 proteins (Collins *et al.*, 1996a). A lipid envelope surrounds the RNA genome, which is in close association with the nucleocapsid (N-protein). The nucleocapsid also includes the phosphoprotein (P-protein), the polymerase subunit (L-protein) and the transcription elongation factor (M2-1). The envelope contains protruding glycoprotein spikes, each consisting of the three transmembrane viral proteins, namely the attachment protein (G-protein), the fusion protein (F-protein) and the small hydrophobic protein (SH-protein). RSV has two matrix proteins (M1- and M2-proteins) and two non-structural proteins (NS1- and NS2-proteins) (Collins *et al.*, 1999). Figure 1.1 illustrates the viral structure of RSV.

Viral replication is restricted to the cytoplasm of the surface epithelium cells of the respiratory tract. RSV infection is initiated by virus attachment to an unknown cell receptor, followed by fusion at the cell surface, which allows the nucleocapsid to enter the cell. Transcription occurs step-wise from the 3' end by the gene start (GS) and gene-end (GE) transcription signals that flank each of the viral genes. A full-length, positive sense copy called the "antigenome" is synthesised, which serves as the template for the synthesis of progeny genome in the nucleocapsid. Viral assembly occurs in the cytoplasm and at the cell surface, after which virions bud from the cell membrane. The released virions infect adjacent epithelial cells (Bachi, 1988; Cane, 2001; Collins *et al.*, 1999).



**Figure 1.1** An illustration of a *Paramyxovirus* particle. See text for a detailed description of the function of the different viral proteins.

### 1.2.3 Protein function and importance

The importance and function of each RSV protein will be described in this section. The gene order of RSV proteins is as follows: 3' NS1, NS2, N, P, M1, SH, G, F, M2 and L (Reviewed by Cane, 2001).

#### 1.2.3.1 Surface glycoproteins

As mentioned above, the transmembrane proteins include the attachment (G-protein), the fusion (F-protein) and the small hydrophobic (SH-protein) proteins.

The **G-protein** mediates viral attachment and comprises of 289-299 amino acids, depending on the strain. The G-protein has an ectodomain, which consists of a highly conserved central region (13 amino acids) with four conserved cysteine residues (Sullender & Wertz, 1991a), flanked by two hypervariable regions (Johnson *et al.*, 1987). Figure 1.2 illustrates the structural features of the G-protein. The G-protein is unusual among paramyxoviruses because it is heavily glycosylated (Levine *et al.*, 1987; Walsh *et al.*, 1986a). The G-protein is modified by the addition of both N- and O-linked oligosaccharides, which is facilitated through the protein's high serine and threonine content (Gruber & Levine, 1985). The high serine, threonine and proline content, combined with the carbohydrate side chains are thought to form mucin-like structures similar to those of the mucinous proteins of the respiratory tract. It appears to be important for viral infectivity and antigenicity, possibly by shielding this protein from immune recognition (Lambert, 1988; Olmsted *et al.*, 1989b). The G-protein is expressed in both a membrane-bound form and a soluble form and may play an immunopathogenic role in infections (Johnson *et al.*, 1998). The G-protein is highly variable, with an amino acid similarity of only 53% between subtypes A and B (Reviewed by Cane, 2001). It is one of the targets for the neutralising immune response (Johnson *et al.*, 1987) and is believed to accumulate amino acid changes with time suggesting evolution under selective pressure (Cane & Pringle, 1995). Sequence analysis of the G-protein can therefore be useful in molecular epidemiological investigations of RSV epidemics (Reviewed by Cane, 2001; Sullender, 2000).

The **F-protein** promotes fusion of the viral and the cell membranes, viral entry and syncytia formation. The F-protein is synthesised as an N-glycosylated precursor, F0-protein, which consists of 574 amino acids (Walsh & Graham, 1999). The F-protein is activated by cleavage into two disulfide-linked fragments (F1 and F2) by a furin-type intracellular protease at cleavage-activation sites at amino acids 131-136 (Collins *et al.*, 1984) and amino acids 106-109 (González-Reyes *et al.*, 2001). This action exposes a hydrophobic domain on the F1-subunit, which may be important in membrane fusion, cell tropism and pathogenicity (Collins *et al.*, 1984). Multiple glycosylated forms of the F-protein are expressed in virus-infected cells. Although the reason for this is not yet clear, it has been suggested to affect some aspects of protein function (Rixon *et al.*, 2002). A sequence motif suggestive of coiled coil

structures is known as a heptad repeat (HR) (Buckland & Wild, 1989; Chambers *et al.*, 1990; Singh *et al.*, 1999). The F-protein contains three heptad repeat regions predicted to form coiled coils (HR1, HR2 and HR3). The HR1 core protein is located adjacent to the fusion peptide at the N-terminal of the F1-subunit, while HR2 is located adjacent to the transmembrane region of the F1-subunit. The HR3 region lies in the F2-subunit (Lambert, 1996). The HR1- and HR2-regions interact to form a stable  $\alpha$ -helix. Recent evidence suggests the F-protein core is a trimeric coiled coil (Matthews *et al.*, 2000; Zhao *et al.*, 2000). Figure 1.3 illustrates the structural features of the F-protein.

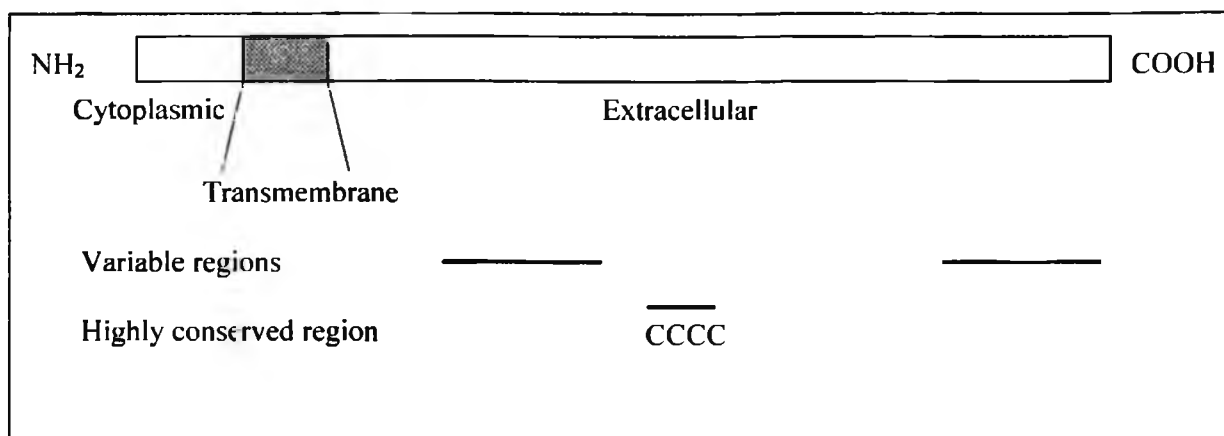
The F-protein, in contrast to the G-protein, is highly conserved between subtypes A and B, with an amino acid sequence identity of 91% (Johnson & Collins, 1988; Johnson *et al.*, 1987). Evidence suggests that the F-protein is the viral protein to which most neutralising monoclonal antibodies are directed (Walsh & Hruska, 1983; Walsh *et al.*, 1985), and it is an important target for RSV-specific CTL (Cherrie *et al.*, 1992). The F-protein is the major focus of prophylactic and subunit vaccine research.

The **SH-protein** exists as multimers in infected cells and purified virions (Anderson *et al.*, 1992; Olmsted & Collins, 1989a). The function of this viral protein is unknown, but a recombinant virus without the SH gene is fully viable in cell culture (Bukreyev *et al.*, 1997). Collectively, the SH-, G- and F-proteins, contribute to the fusion property of the virus (Heminway *et al.*, 1994). The SH-protein has an amino acid sequence identity of 43% between subtypes A and B (Collins *et al.*, 1999). It does not contain neutralising epitopes (Walsh & Graham, 1999) but induces a CTL response in humans (Cherrie *et al.*, 1992).

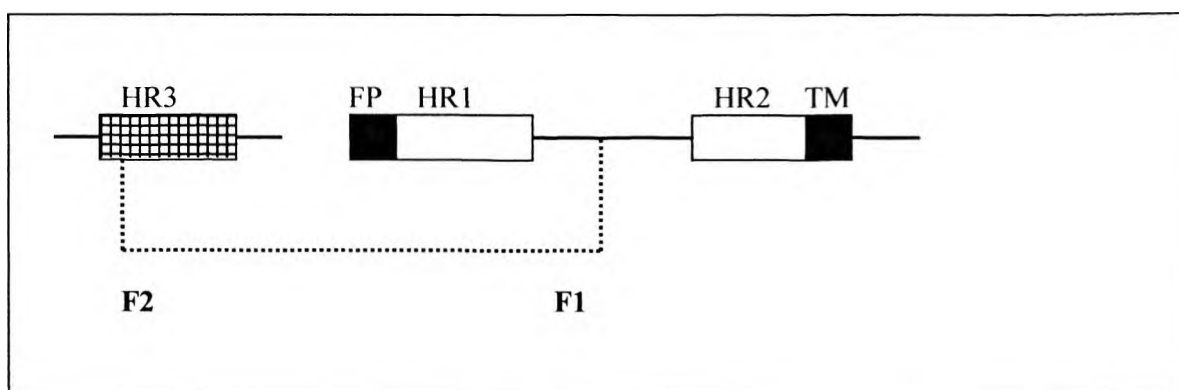
#### 1.2.3.2 *Nucleocapsid associated proteins*

The nucleoprotein (N-protein), the phosphorylated protein (P-protein) and the large polymerase protein (L-protein) are the replicative machinery of RSV (Collins, 1991; Garcia *et al.*, 1993; Garcia-Barreno *et al.*, 1996).

The **N-protein** binds to genomic and antigenomic RNA, forming the ribonucleoprotein template for RNA synthesis (Reviewed by Collins *et al.*, 1999).



**Figure 1.2 Structural features of the RSV G-protein.** The two variable regions of the G-protein ectodomain, flank the highly conserved central region that includes four cysteine residues. The cytoplasmic, transmembrane and extracellular regions are also indicated. Adapted from Cane (2001).



**Figure 1.3 Schematic diagram of the RSV F-protein.** The F1- and F2-subunits, the positions of the heptad repeat regions (HR1, HR2, HR3) as well as the fusion peptide (FP) and the transmembrane region (TM) are illustrated. A dashed line indicates the disulfide bond. Adapted from Matthews *et al.* (2000).

The amino acid terminus of the N-protein is important for nucleocapsid formation and may be essential for maintaining helical stability (Murphy *et al.*, 2003). The N-protein is highly conserved between both RSV subtypes with an amino acid sequence homology of 96% (Johnson *et al.*, 1998). It appears to be the major target for the memory CTL response in humans (Cannon *et al.*, 1987; Cherrie *et al.*, 1992).

The **P-protein** is believed to associate with free N- and L- proteins to sustain them in soluble form before their assembly into the nucleocapsid. The N- and P-protein are essential for the encapsidation of RNA and the P-protein is believed to be a polymerase co-factor (Reviewed by Collins *et al.*, 1999). The C-terminal of the P-protein is essential for the binding of NS1 (Hengst & Kiefer, 2000).

The **L-protein** is the major polymerase subunit and contains conserved polymerase motifs that may represent catalytic domains (Collins *et al.*, 1999). The N-, P- and L-proteins are required for RNA replication and are necessary and sufficient for a poorly processive form of transcription (Collins *et al.*, 1996a; Grosfeld *et al.*, 1995; Yu *et al.*, 1995).

The M2 mRNA encodes two proteins. The **M2-1 (22K)** promotes transcription elongation (Collins *et al.*, 1996a) and is thought to increase the frequency of read-through gene-end (GE) signals (Hardy & Wertz, 1998). The N, P, L and M2-1 proteins are necessary for fully processive transcription and direct the production of virus from cDNA (Collins *et al.*, 1996a; Collins *et al.*, 1995). M2-1 requires phosphorylation for protein function (Cartee & Wertz, 2001). The M2 protein is a CTL target, and provides transient protection to subtype A or B viruses in mice (Reviewed by Walsh & Graham, 1999). **M2-2** is a putative regulatory factor for replication and transcription (Collins *et al.*, 1996a).

#### 1.2.3.3 *Matrix protein*

The **M-protein** is an unglycosylated internal protein and may mediate association between the nucleocapsid and the envelope proteins (Teng & Collins, 1998). It has recently been shown that the transcriptase activity of ribonucleoproteins was enhanced by treatment with antibodies to the M-protein in a dose-dependent

manner, suggesting that the M-protein can inhibit virus transcription (Ghildyal *et al.*, 2002).

#### 1.2.3.4 *Nonstructural proteins*

RSV encodes two nonstructural proteins, **NS1** and **NS2**. NS1 is thought to be a negative regulatory factor for replication and transcription (Atreya *et al.*, 1998). It has been shown that in the absence of the NS1 gene, the recombinant virus is still viable in cell culture and that neither NS1 nor M2-2 is essential for RSV replication *in vivo*, although both are important for efficient replication (Teng *et al.*, 2000). Studies have shown that the NS2 protein is not necessary for RSV replication, although its presence improves virus growth in cell culture (Teng & Collins, 1999; Whitehead *et al.*, 1999). Evidence suggests that NS1, NS2, SH, and M2-2 provide supporting functions for efficient RSV replication (Jin *et al.*, 2000).

### 1.3 EPIDEMIOLOGY

Humans are the only recognised natural hosts of human RSV. RSV is usually transmitted via large respiratory droplets, fomites and hand contact with respiratory secretions. Precautions like the use of gloves, gowns, goggles and hand washing can limit virus spread (Reviewed by Walsh & Graham, 1999).

RSV is an important cause of morbidity in infants and young children, and is often associated with acute upper and lower respiratory tract infection, causing an estimated 125,000 hospitalisations and 4,500 deaths annually in the United States (Black, 2003). In Europe, RSV accounts for 42 to 45% of hospital admissions for lower respiratory tract infections in children younger than 2 years of age (Simoes & Carbonell-Estrany, 2003). It produces significant morbidity in young children, especially in those with underlying cardiopulmonary and cardiac manifestations (Hall, 2000). Deficiencies in cellular immunity appear to elicit severe RSV illness in children with malignancies, those undergoing chemotherapy and bone marrow transplant recipients (Meissner, 2003). RSV has the potential to cause disease in all age groups, including immunocompromised patients such as bone marrow transplant recipients, the elderly and those with cardiopulmonary diseases (Englund *et al.*, 1988; Falsey *et al.*, 1992). RSV mortality in developing countries is

significantly higher than in developed countries (Stensballe *et al.*, 2003). A study in South Africa has also shown RSV to be the single most important etiological agent of acute respiratory infection among children in South Africa (Madhi *et al.*, 2001). Madhi and colleagues have also found that HIV-positive children with viral associated ALRI, had a higher mortality rate (7.5%) than HIV-negative children (0%) (Madhi, 2000). Primary RSV infection occurs in 65% of infants in their first year of life, with 2-4 month old infants experiencing the most severe disease.

Primary infection can occur in young infants in the presence of maternally derived antibodies (Glezen *et al.*, 1981) and reinfection can occur throughout life, but successive infections are limited to the upper respiratory tract (Graham, 1996). Evidence suggests that ALRI is slightly more common in boys than in girls (1.5-2:1 (boys:girls)), in both developing and industrialised countries (Weber *et al.*, 1998).

RSV epidemics occur annually and outbreaks are often associated with the cold season in temperate climates and with the rainy season in the tropics (Weber *et al.*, 1998). RSV infections have been observed mainly during the rainy seasons in Asian, African and South American countries (Shek & Lee, 2003). Johannesburg in South Africa has a temperate climate and most outbreaks occur in autumn and winter when the rainfall is at its lowest (Madhi *et al.*, 2001; Joosting *et al.*, 1979).

### 1.3.1 Subtypes of RSV

Studies have shown that RSV strains can be divided into two antigenic subtypes, A and B, on the basis of their reactions with monoclonal antibody panels, directed against the P-, F- and G-proteins (Anderson *et al.*, 1985; Gimenez *et al.*, 1986; Mufson *et al.*, 1985). It has also been demonstrated that both subtypes are distinct at the genetic level. Both subtypes have been compared with the RNase A mismatch cleavage method, using probes corresponding to the P-, N-, M2-, F- and G-genes, and it was found that genetic heterogeneity exists within and between each subtype. Genetic variability is also observed among viruses of the same epidemic (Cristina *et al.*, 1990). Genetic variability in both subtypes has also been assessed by sequence analysis. Studies have shown that the F-protein and the N-protein are highly conserved between subtype A and B (amino acid identities of 91% and 96% for the F- and N-protein, respectively) (Johnson & Collins, 1988; Johnson

& Collins, 1989), while the G-protein has been found to be highly variable between the subtypes (53% amino acid similarity). A PCR with subtype-specific oligonucleotides has been developed for the G-protein, to differentiate between subtypes A and B (Sullender & Wertz, 1991c). The SH-protein is also found to be variable between subtypes (43% amino acid dissimilarity) (Collins *et al.*, 1996b).

Genetic and antigenic studies have shown that both subtypes co-circulate during most epidemics, with varying dominance each year (Reviewed by Sullender, 2000).

### 1.3.2 Genotypes of RSV

Antigenic and genetic variability was also detected within subtypes A and B (Cristina *et al.*, 1990). Antigenic variability of RSV strains within subtypes A and B has been demonstrated by differences in reactivity to monoclonal antibodies (Akerlind-Stopner *et al.*, 1990; Anderson *et al.*, 1991; Storch & Park, 1987). Garcia-Barreno and colleagues have found great antigenic variability in the G-protein of different viruses of the same subtype, that had limited amino acid sequence variability, and demonstrated the existence of monoclonal antibodies with strain-specific reactivity (Garcia-Barreno *et al.*, 1989).

Genetic variability between and within subtypes A and B, has been demonstrated using reverse transcriptase-polymerase chain reaction (RT-PCR) in combination with restriction fragment length polymorphism (RFLP) or nucleotide sequencing (Reviewed by Cane, 2001; Sullender, 2000) and ribonuclease (Rnase) A mismatch cleavage (Cristina *et al.*, 1991; Storch *et al.*, 1991). Sequence analysis of the G-protein showed differences of up to 20% among subtype A viruses (Cane & Pringle, 1991c), and differences of up to 12% among subtype B viruses (Sullender & Wertz, 1991a). A high frequency of nucleotide changes resulted in amino acid changes in the variable domains of the G-protein of both subtypes, suggesting there may be a selective pressure for change in this protein (Cane & Pringle, 1991c; Sullender & Wertz, 1991a). Sequence analysis of the G-protein can therefore be a useful molecular tool for the epidemiological investigation of RSV outbreaks.

A molecular study conducted in the USA over a five year period, showed that five subgroup A genotypes with 22 sub-lineages were present in a single community, as

well as four distinct subgroup B genotypes with six sub-lineages (Peret *et al.*, 1998). Peret *et al.*, (2000) have shown that during a single epidemic, seven subtype A genotypes (GA1-7) and four subtype B genotypes (GB1-4) were present in diverse locations in North America (Peret *et al.*, 2000). The method of genotyping used by these authors is therefore useful for the comparison of different strains from different parts of the world (Peret *et al.*, 2000). They have further shown that the second hypervariable region (270 nucleotides), which makes up the C-terminal of the G-protein, provides a reliable representation for the entire G-gene variability (Peret *et al.*, 1998; Peret *et al.*, 2000), and has subsequently been used in phylogenetic analysis for molecular epidemiological studies (Peret *et al.*, 1998; Peret *et al.*, 2000; Venter *et al.*, 2001; Venter *et al.*, 2002).

In the following section, molecular epidemiological studies using the RSV G-protein will be discussed.

### 1.3.3 Molecular epidemiology

Studies in the Northern hemisphere have shown that subtypes A and B co-circulate in each RSV outbreak with varying predominance each year (Reviewed by Sullender, 2000). A study conducted in Boston has shown that subtypes A and B circulated concurrently with temporal and geographic clustering of these subtypes (Hendry *et al.*, 1986). Over a two year period, alternating predominance of subtypes A and B has been observed in Finland (Waris, 1991), while a study conducted in Korea over nine consecutive epidemics has shown that subtype A was dominant during seven epidemics (Choi & Lee, 2000). Reports from developing countries showed similar results to those of industrialised countries. In a study conducted over four consecutive seasons in Gambia, it was found that subtype A dominated in 3 epidemics (Cane *et al.*, 1999). In South Africa a study conducted during 1997 to 2000, found that subtypes A and B circulated equally during 1997 and 1998, while subtype A predominated in 1999 and 2000 (Venter *et al.*, 2001). Subtype B dominated (85%) in an outbreak in Mozambique in 1999 (Roca *et al.*, 2001). In a recent study conducted over a two year period in India, it has also been found that subtype B dominated in both seasons (Rajala *et al.*, 2003).

In a study conducted in Birmingham (UK) over an 11 year period, it was shown that each epidemic was made up of a number of different genotypes. The authors suggested that each outbreak could be considered to be a collection of concurrent separate epidemics. In each epidemic different genotypes co-circulated, and the dominant genotype was replaced from one season to the next, with some genotypes reappearing and disappearing in subsequent years (Cane *et al.*, 1994). Similar results have been found in Korea (Choi & Lee, 2000) and in Rochester, USA (Peret *et al.*, 1998). Two non-contiguous epidemics (1990-1991 and 1997-1998) have been studied in Beijing. The authors found that both subtypes co-circulated and that very similar subtype A viruses were isolated 5 years previously, suggesting that the different genotypes in Beijing may be relatively stable, despite the strain variation observed in subtype A (Kong *et al.*, 2001). In South Africa it was found that subtype A was more variable and that displacement of genotypes occurred in most seasons while subtype B remained more consistent (Venter *et al.*, 2001). Peret *et al.* (2000) studied RSV variability during a single season in geographically diverse locations in North America and illustrated that each community showed a number of distinct genotypes, and that there was not a wide regional spread of one predominant genotype (Peret *et al.*, 2000). It has been suggested that the predominant genotype changes in each epidemic due to the accumulation of herd immunity to a particular predominant strain during the epidemic, which restricts circulation of that strain in subsequent epidemics (Cane, 2001).

Despite variability within and between local epidemics, it has been reported that virus strains of different parts of the world (UK, Uruguay, Malaysia, Germany, Finland, Austria, Australia, Spain and USA) circulating at similar times, were very similar with little evidence of geographical clustering (Cane *et al.*, 1992). Evidence also suggests that strains from single locations can be very diverse, while other strains could group together regardless of geographical location (Sullender, 2000). However, molecular epidemiological studies conducted in Africa (Gambia, Mozambique and South Africa), have showed multiple new virus strains that did not cluster with previously identified genotypes from the rest of the world (Cane *et al.*, 1999; Roca *et al.*, 2001; Venter *et al.*, 2001).

### 1.3.4 Evolution of RSV

Various molecular epidemiological studies have focussed on the variability of the G-protein, since this protein shows the highest level of antigenic and genetic variation between subtypes A and B, and it is a target for the neutralising and protective immune response (Melero *et al.*, 1997).

Evidence suggests that the majority of nucleotide changes in the G-protein resulted in amino acid changes, suggesting positive selective pressure (Cane & Pringle, 1995a; Garcia *et al.*, 1994). Woelk and Holmes (2001) have found six positively selected sites within the ectodomain of the G-protein of both subtypes and these sites have some association with antibody epitopes and O-glycosylation sites. The authors suggest that immune driven natural selection is an important determinant of RSV evolution and that positive selection differs amongst strains (Woelk & Holmes, 2001). A study in South Africa also provides evidence for positive selection in the second variable region of the G-protein (Venter *et al.*, 2001). Positive selection and changes in the antigenic structure of the G-protein may result in escape of the immune response (Melero *et al.*, 1997).

Cane and Pringle (1995) analysed the variability of the G-protein of subtype A over 38 years and found that the G-protein accumulates amino acid changes with time, which appears to be a result of evolution and survival of particular genotypes and the extinction of others. Variability may give rise to new strains that may be overcome by herd immunity in future epidemics (Cane & Pringle, 1995). Similar results for the G-protein were obtained by other studies conducted in the Northern hemisphere (Cane *et al.*, 1991; Choi & Lee, 2000; Peret *et al.*, 2000). However, in contrast to previous findings, a study conducted in Cuba (Valdes *et al.*, 1998) has shown a lack of genetic variation in the G-protein over two years (1994-1995). It has been found that the Cuban G-gene sequences had only five nucleotide differences with the prototype Long strain. Evidence also suggests a lack of genotype displacement as well as a lack of accumulated amino acid changes in the G-protein. This genetic stability may be due to travel restriction to and within the island and to the fact that RSV infections occur throughout the year (Valdes *et al.*, 1998).

Johansen *et al.* (1997) have studied the prevalence of the RSV subtypes from 1992 to 1995 in Denmark. RFLP and PCR analyses have shown that genetic variability was present in parts of the F- and G-proteins and that some genotypes appear to be related to strains isolated two decades ago in different parts of the world. This may indicate that temporal variation in predominant genotypes may be due to the favouring of strains from a pool of globally circulating, genetically stable genotypes (Johansen *et al.*, 1997). The genetic drift in closely related subtype A strains has been studied over a 5 year period in St. Louis by sequencing the membrane-associated genes, namely the M-, SH-, G-, F-, and M2-proteins. However, with the exception of the G- and SH-gene of some strains, the authors have been unable to demonstrate clear genetic drift in these genes over this short period (Zheng *et al.*, 1999).

Stine *et al.* (1997) have determined that the G-, F-, N- and P-protein of bovine RSV had antigenic differences among different strains. It has been found that the ectodomain of the G-protein of bovine RSV also contained areas of sequence variance (Stine *et al.*, 1997). Valarcher and colleagues have shown continuous evolution of the sequences of the N-, G-, and F-proteins of bovine RSV since 1967, and the exertion of a strong positive selective pressure on the mucin-like region of the G-protein and on certain sites of the N- and F-proteins (Valarcher *et al.*, 2000).

## 1.4 RSV DISEASE

RSV disease can range from rhinitis to severe bronchiolitis and pneumonia. A major part of RSV infection is due to virus-mediated destruction of epithelium and is thus determined by the extent of viral replication. It is likely that some components of natural RSV infection are due to the host immune response (Connors *et al.*, 1994; Graham, 1995; Graham *et al.*, 1993; Johnson *et al.*, 1998; Openshaw, 1995). However, the fact that RSV can be fatal in immunocompromised patients indicates that severe disease is not only dependent on a vigorous host immune response (Collins *et al.*, 1999).

An immune-mediated component of RSV disease, that does not involve natural infection alone, is illustrated by an unfortunate incident in the 1960s, when a formalin-inactivated RSV vaccine (FI-RSV) was administered to infants and young

children. Not only did the vaccine fail to induce protection against natural infection, but it also resulted in enhanced lung disease in vaccinees upon reexposure to RSV and in some deaths (Kapikian *et al.*, 1969). It has been hypothesised that formalin inactivation diminished the immunogenicity of the F-protein and that the virus spread by cell fusion was not inhibited (Merz *et al.*, 1980). Children who have experienced FI-RSV-enhanced disease had antibody responses to the F- and G-protein, but have lacked neutralising and antifusion responses and CTL responses (Murphy *et al.*, 1986).

Retrospective studies in experimental animals have shown that this enhanced disease was associated with a shift in the balance between Th1 and Th2 type cytokines and resulted in a shift towards a type 2 T-helper cell (Th2) response. It is believed that the FI-RSV primed for an increased CD4<sup>+</sup> T-cell mediated airway inflammation without inducing a normal CTL or protective antibody response (Collins *et al.*, 1996b; Graham, 1995; Murphy *et al.*, 1990). A recent study has indicated that the G-protein, specifically the G-protein CX3C chemokine motif, contributes to the inflammatory response to FI-RSV vaccination, possibly through the induction of pulmonary substance P. The G-protein CX3C peptides can compete with fractalkine, which is important for chemoattraction and activation of dendritic cells, for CX3CR1 binding and change fractalkine-mediated responses. The CX3C-CX3CR1 interaction may affect the molecular signals important for T-cell activation (Haynes *et al.*, 2003).

## 1.5 ROLE OF RSV PROTEINS IN DISEASE AND PROTECTION

The surface glycoproteins, G- and F-proteins, are the only RSV antigens that induce neutralising antibodies and are believed to be major protective antigens (Johnson *et al.*, 1987; Walsh & Hruska, 1983). However, CTL response also plays a role in protection. In RSV, the F- and N-proteins are the major targets for the CTL response (Cherrie *et al.*, 1992). RSV proteins of immunological importance will be discussed in terms of disease and protection.

---

## 1.5.1 The surface glycoproteins

### 1.5.1.1 The G-protein

As mentioned before, the G-protein is the most variable within and between RSV subtypes with 47% amino acid differences (Reviewed by Cane, 2001). However, the G-protein does contain some conserved epitopes (Garcia-Barreno *et al.*, 1990; Melero *et al.*, 1997; Rueda *et al.*, 1995).

The G-protein has been studied with respect to the antibody response; since there is little evidence of class I restricted CTL response in humans (Cherrie *et al.*, 1992). Using monoclonal antibodies, the existence of three types of epitopes have been illustrated on this protein. These include conserved epitopes present in all RSV strains; group-specific epitopes present in all viruses of the same subtype, and strain-specific or variable epitopes that are only present in some viral strains. The conserved and group-specific epitopes have been mapped to the central conserved region of the G-protein using escape mutants, while strain-specific epitopes reside in the C-terminal region (Garcia-Barreno *et al.*, 1990; Melero *et al.*, 1997; Rueda *et al.*, 1995). Evidence suggests that protective epitopes (protectopes) reside in the G-protein of subtype A (Long strain) and these protectopes appear to be conserved in humans (Huang *et al.*, 2003; Power *et al.*, 2001). All strain-specific anti-G murine monoclonal antibodies react with the mature form of the G-protein and a few react with synthetic peptides, suggesting that they recognise conformational rather than linear epitopes (Akerlind-Stopner *et al.*, 1990; Garcia-Barreno *et al.*, 1992). Studies with human convalescent sera demonstrated that several sera recognised the conserved region, but recognition in the C-terminal region was highly specific (Cane, 1997; Cane *et al.*, 1996; Norrby *et al.*, 1987). Linear epitopes recognised by primary infection sera are found to be concentrated in potential N-glycosylation sites, suggesting an extra method for the abolishment of, or the masking of antigenic sites from the immune response by the modulation of glycosylation sites (Cane, 1997). Since strain-specific epitopes reside in the C-terminal region, it has been suggested that RSV has the potential to change the antigenic profile of its G-protein in order to evade the immune response (Melero *et al.*, 1997). Amino acid substitutions, premature stop codons and multiple A-G hypermutations have been reported on the G-protein demonstrating the capacity of this protein to accommodate multiple

changes, especially within the antigenic C-terminal region (Garcia-Barreno *et al.*, 1990; Martinez *et al.*, 1997; Melero *et al.*, 1997; Rueda *et al.*, 1995).

Garcia *et al.* (1994) have found that there is a close correlation between genetic and antigenic relatedness of the G-protein from subtype A, and identified amino acid changes in the C-terminal of the G-protein that were associated with epitope loss in escape mutants and natural occurring strains (Garcia *et al.*, 1994). Similar results were reported by Cane and Pringle (Cane & Pringle, 1995). The evolution in subtype B viruses is very similar to that of subtype A viruses, with AG+GA transitions exceeding UC+CU transitions. Differences in protein lengths determined by the usage of alternative termination codons, are more prominent in subtype B viruses than in subtype A. Different termination codon usage due to certain mutations seems to play a role in the diversification of subtype B viruses (Martinez *et al.*, 1999). A region has been identified immediately preceding the C-terminal third of the G-protein, that is prone to polymerase errors in the mRNA (Cane *et al.*, 1993). These mRNAs have the potential to encode G-proteins with altered C-terminal regions.

Antigenic differences between RSV strains may influence the neutralising immune response in children (Cane *et al.*, 1996), which is predominately group-specific and reveals poor cross-reactivity (Langedijk *et al.*, 1997; Muelenaer *et al.*, 1991). Evidence suggests that monoclonal antibodies directed against the G-protein can inhibit RSV replication in mice and cotton rats (Taylor *et al.*, 1984; Walsh *et al.*, 1984). Immunisation with purified G-protein can partially reduce virus titers after challenge in cotton rats (Walsh *et al.*, 1987). Mice or cotton rats vaccinated with recombinant vaccinia virus carrying the G-gene have been protected against challenge only with the same subtype (Stott *et al.*, 1987). The G-protein has been associated with the induction of Th2 CD4+ lymphocytes (Alwan *et al.*, 1994). CD8+ CTL responses directed against the G-protein have never been detected in human or murine studies (Reviewed in Crowe, 2002).

#### 1.5.1.2      *The F-protein*

The highly conserved F-protein is an important protective antigen because it is the protein to which most neutralising monoclonal antibodies are directed (Walsh &

Hruska, 1983; Walsh *et al.*, 1985) and is an important target for CTL (Cherrie *et al.*, 1992).

Multiple antigenic sites are present on the F-protein (Anderson *et al.*, 1985) including certain conserved linear neutralisation epitopes (Walsh *et al.*, 1986b). Reports suggest that many neutralising epitopes on the F-protein are conformational in nature and require glycosylation and folding in order to be recognised (Beeler & Coelingh, 1989). Connor *et al.* (2001) have found limited sequence and antigenic variation between the F-protein genes from subtypes A and B although they did identify subtype-specific neutralising epitopes (Connor *et al.*, 2001). Arbiza *et al.* (1992) characterised two antigenic sites recognised by neutralising monoclonal antibodies directed against the F-protein in subtype A in position 255 to 272. A subtype B specific neutralisation epitope has been identified in position 205 to 225 of the F-protein (Corvaisier *et al.*, 1997). Escape mutants with sequence changes in the F-protein have been identified that resulted in epitope loss in conserved epitopes. Escape mutants with point mutations in a group-specific site in the F-protein of subtype A have also been identified (Lopez *et al.*, 1998).

The F-protein elicits CTL and an antibody response against RSV infection in BALB/c mice. Two RSV F-protein specific T-helper cell epitopes (Corvaisier *et al.*, 1993) as well as a MHC class I CD8+ H-2K(d)-restricted epitope, have been identified in mice (Chang *et al.*, 2001; Jiang *et al.*, 2002). Human HLA Class I-restricted CTL epitopes have also recently been identified on the F-protein of RSV. Brandenburg *et al.* (2002) identified HLA class I (B\*57 and Cw\*12)-restricted CTL epitopes in position 106 to 114 and 542 to 550, respectively, in the F-protein, while Rock and Crowe (2003) identified a HLA-A\*01-restricted CTL epitope in position 109 to 118 (Brandenburg *et al.*, 2000; Rock & Crowe, 2003).

Immunisation with the F-protein provides cross-protection against both RSV subtypes A and B in experimental animals (Taylor *et al.*, 1984; Walsh *et al.*, 1984). Evidence suggests that passive administration of neutralising monoclonal antibodies to the F-protein prevents the replication of RSV in mice (Taylor *et al.*, 1984). A study conducted on infants suggested that antibody to the F-protein is an important correlate of immunity, and that children with reduced RSV disease had significantly

higher anti-F immunoglobulin G (IgG) titers prior to infection (Kasel *et al.*, 1987). Although active immunisation with the F-protein or passive immunisation with anti-F monoclonal antibodies may be beneficial (Group, 1998), effectiveness in human infants has been less impressive than in rodents (Tristram *et al.*, 1994).

### 1.5.2 The nucleoprotein

Other RSV proteins that contain CTL epitopes, include the nucleoprotein (N), which is the most conserved between the subtypes, with 96% amino acid similarity, and the small hydrophobic (SH) and the M2-proteins (Reviewed by Cane, 2001). Two CTL epitopes have been defined to the N-protein to date, namely HLA B\*07 and HLA B\*08 epitopes (Goulder *et al.*, 2000; Venter *et al.*, 2003).

It has been shown that dominant CTL epitopes (HLA-B\*08- and HLA-B\*07-restricted) in the N-protein of human RSV remained conserved and that positive selection does not occur in the CTL epitopes of human RSV (Venter *et al.*, 2003). However, recent findings with influenza virus have suggested that immune selection occurs in the CTL epitopes of the highly conserved influenza virus nucleoprotein. This implies that regular updates will be needed in vaccines containing these epitopes (Price *et al.*, 2000). Boon *et al.* (2002) found that a CTL epitope in the influenza virus nucleoprotein had a high degree of variation at nonanchor residues, and that influenza virus variants emerged in chronological order over time. CTLs directed against old strains failed to recognise more recent strains of influenza virus, suggesting an escape from CTL immunity (Boon *et al.*, 2002).

Studies into the role of CTLs in protection, the molecular evolution of RSV and the mechanisms by which this virus evades the immune response, will make an important contribution to future vaccine research on RSV (Melero *et al.*, 1997).

## 1.6 VACCINE DEVELOPMENT

Multiple strategies have been used in RSV vaccine development and these include the generation of peptide, subunit and live virus vaccines. Only the inactivated subunit and live-attenuated RSV vaccines have been evaluated in clinical trials (Reviewed by Dudas & Karron, 1998). General information about the immune response following RSV infection, as well as vaccine strategies focussing on RSV

antigenic proteins (F- and G-proteins) and their importance, will be discussed in the following section.

### 1.6.1 Immune responses important for protection

The human immune system has two main branches, innate and adaptive immunity. Adaptive immunity in turn has 2 branches namely, humoral or antibody-mediated immunity and T-cell mediated immunity. Antibody-mediated immunity involves antibodies and the B-cells that produce them. Antigen presenting cells such as macrophages activated by the antibody response, ingest the antigen and antigen fragments are incorporated and displayed on the outer surface of the cell. On the antigen presenting cell surface, the major histocompatibility complex (MHC) class II presents the antigens to CD4+ T-cells. The CD4+ T-cell stimulates the B-cells to synthesise antibodies (Ig) against the antigen. Cell mediated immunity involves all the T-cells (CD4 and CD8 lymphocytes) and their respective effector functions. Cell-mediated immunity involves direct cell-cell interaction by cytotoxic T-cells. When a virus infects a cell, fragments of the virus are incorporated and presented on the MHC class I molecules on the cell surface. The cytotoxic T-cells identify and attach to infected cells, killing these cells by the secretion of perforin, which forms pores in the cell membrane (Reviewed in Clug & Cummings, 1997).

RSV can cause repeated infections throughout life (Walsh & Graham, 1999), however, partial immunity to reinfection does occur in humans (Hall *et al.*, 1991). Virus-specific immune responses are mainly responsible for protection and recovery from RSV infections. Immunity to RSV is mediated by means of humoral and cellular effectors. Humoral effectors involve secretory antibodies and serum antibodies (acquired as a result of infection or maternally derived in infants), while cellular effectors involve MHC class I-restricted CTL (Dudas & Karron, 1998; Meanwell *et al.*, 2000). Evidence suggests that activation of type 2 T-helper cells is a key factor in disease pathogenesis whereas protective immunity is associated with a type 1 T-helper cell response with the induction of CD8+ CTL (Graham, 1996).

Cell-mediated immunity plays a primary role in clearance of RSV infection in humans; however there is currently no evidence that increased cellular immune activity enhances protection against RSV infection (Crowe, 1998). Patients or experimental animals lacking T-cells reveal prolonged RSV shedding (Graham *et*

*al.*, 1991; King *et al.*, 1993). Memory CTL are induced in humans (Bangham *et al.*, 1986). T-helper cells are also found in the airways of individuals infected with RSV, producing Th1- and Th2-cytokines (Welliver, 1999). The balance between Th1 and Th2 responses are important in coordinating protective and immunopathogenic responses to RSV infection (Folkerts *et al.*, 1998). The Th1 response and CTL are associated with disease-free protection and clearance of infection (Reviewed by Walsh & Graham, 1999). FI-RSV-enhanced disease in rodents correlates with an abnormal induction of Th2 cytokines (Graham, 1996).

Evidence suggests that antibodies play a dominant role in the protection against RSV disease (Crowe, 1998). Antibodies are not required for the termination of viral replication after primary infection, however mice without antibodies had a more severe disease than mice with an intact antibody response (Walsh & Graham, 1999). Infants born with high levels of maternal neutralising immunoglobulin (IgG) developed less severe RSV pneumonia and have fewer infections than infants with lower titers (Glezen *et al.*, 1981; Kasel *et al.*, 1987). The antibody response to primary infection is short-lived and weak, since passively acquired maternal antibody probably suppresses the development of the infant's own immune response (Kaul *et al.*, 1981; Nadal *et al.*, 1990). Infants younger than 6-8 months have a poor antibody response to RSV infection compared to older infants and there is failure to develop a protective secretory IgA or a neutralising IgG response (Murphy *et al.*, 1986). Increased RSV-specific IgE in nasal secretions has been observed in wheezing children with severe RSV disease (Welliver *et al.*, 1980).

### **1.6.2 Experimental vaccines against RSV: the importance of the G- and F-protein in vaccine development**

Evidence suggests that during primary infection, infants produce secretory and serum neutralising antibody directed against both RSV G- and F-proteins. The F-protein induces preferentially a Th1-type response, while the G-protein induces mainly a Th2 response (Graham *et al.*, 1993).

The immune response to the F-protein is more influenced by the presence of maternally derived antibodies than age, while the opposite is true for the G-protein response (Reviewed by Walsh & Graham, 1999). Since the protective antibodies

are directed to the F- and G-protein, many investigators have studied the response to these proteins. It has been found that infection induces much higher titers of antibodies that bind to the F- and G-protein than that neutralise RSV *in vitro*. RSV-specific antibodies neutralise the virus and a smaller subset of these antibodies inhibit virus fusion *in vitro*. Evidence suggests that these neutralising antibodies are the key to protection against reinfection (Reviewed by Crowe, 2002).

In cotton rats, the F-protein provides almost complete protection against RSV subtypes A and B, while the G-protein provided group-specific protection, suggesting that a subunit vaccine based on the G-protein may need G-proteins of both RSV subtypes to provide broad protection, while a F-protein vaccine might equally protect against both subtypes (Sullender *et al.*, 1990). Brideau *et al.* (1989) showed that chimeric FG-protein used to immunise cotton rats reduced the viral titers after challenge with either subtype A or B (Brideau *et al.*, 1989). These studies emphasise the importance of the F- and G-proteins in future vaccine development.

There is currently no effective licensed RSV vaccine available, partly due to the fact that the FI-RSV vaccine-enhanced disease is not yet completely understood (Reviewed by Varga & Braciale, 2002). A few obstacles arise in attempts to immunise infants (< 3 months) against RSV infection. Nearly all infants have maternal antiviral antibodies that can interfere with infectivity of primary immunisation. The ability to respond to the surface glycoproteins (F- and G-protein) appears to be gradually acquired over the first six to nine months of life. The risk of immunising young infants is critical, therefore highly attenuated virus strains with minimal side effects are required. Lastly, since antigenic variability exists between virus strains, this may require that multivalent vaccines comprising multiple virus strains be developed and these be adjusted annually, similar to that for influenza virus vaccines (Reviewed by Crowe, 1998).

As mentioned before, clinical trials conducted in the 1960s using a formalin inactivated virus vaccine (FI-RSV) resulted in severe enhanced disease in immunised children during natural infection. The FI-RSV is probably due to an increased CD4+ T-cell mediated airway inflammation without inducing a normal CTL or protective antibody response (Collins *et al.*, 1996b; Murphy, 1990; Graham *et al.*,

1995). Due to the failure of the formalin inactivated vaccine, vaccine development was directed toward live-attenuated RSV mutants. Intranasal immunisation with a live-attenuated RSV vaccine should stimulate both systematic and local immunity and therefore protect against both upper and lower respiratory tract infection. Additionally, the live vaccine should closely resemble the response to natural infection and therefore be unlikely to produce enhanced disease (Crowe, 2002). Different live-attenuated RSV vaccine strategies have been investigated, including the creation of host range mutants, cold-passaged (*cp*) mutants, and temperature sensitive (*ts*) mutants. Most of these candidate vaccines were either over-attenuated (RSV *ts*-2) or under-attenuated (*cp*RSV and RSV *ts*-1) (Reviewed by Dudas & Karron, 1998). The most successful of these intranasally administered vaccines, was the RSV strain, cpts248/404, with eight nucleotide changes from the wild-type virus, and seven predicted amino acid changes (one in the N-protein, two in the F-protein and four in the L-protein) (Meanwell *et al.*, 2000). It will therefore be advantageous to use attenuated viruses with a high level of genetic stability of the principal attenuating mutations (Crowe, 2002).

Live vaccines derived via genetic engineering, especially reverse genetics, involve site-directed mutations into viruses. The combination of point mutations, gene insertions, gene deletions and chimeric viruses are currently being explored (Reviewed by (Crowe, 2002)). Whitehead *et al.* (1999) has created a chimeric virus by changing the RSV subtype through the replacement of the F- and G-proteins of subtype A with those of subtype B to create live-attenuated RSV subtype B candidates. The advantage is that a subtype A vaccine can be developed and if this vaccine works, the genes can be replaced with subtype B. This approach is currently under investigation (Whitehead *et al.*, 1999). Another approach involves the engineering of the human RSV G-protein on bovine RSV to produce a live-attenuated human RSV vaccine (Buchholz *et al.*, 2000). Live virus vectors that express RSV F- and G-proteins are also being tested in experimental animals. Vaccinia viruses that express the F-protein (vac-F) or G-protein (vac-G) are immunogenic and protective in rodents, however it is less immunogenic in chimpanzees and fails to protect the lower respiratory tract against wild-type RSV challenge (Crowe *et al.*, 1993).

Purified F- and G-proteins as well as novel chimeric FG glycoproteins (baculovirus-expressed fusion glycoprotein that contains ectodomains of RSV F- and G-protein), have been evaluated as potential subunit vaccines in rodent and primate models (Reviewed by Dudas & Karron, 1998). Purified F-protein and a chimeric FG-protein, induce an altered T-cell response similar to that of the FI-RSV, when used as the primary immunogen in RSV-naïve rodents (Connors *et al.*, 1992; Murphy *et al.*, 1990). Recently it has been found that type2 T-cell responses were induced by the nonglycosylated ectodomain of the G-protein in mice (Hancock *et al.*, 2003). When the chimeric FG-protein is delivered intramuscularly, it protects against lower but not against upper respiratory tract infection in rodents (Wathen *et al.*, 1989), while in African green monkeys, it induces low levels of neutralising antibody and provides minimal protection against the lower respiratory tract infections (Dudas & Karron, 1998). In mice, intranasal administration of chimeric FG-protein induces antibodies and protects the upper respiratory tract after RSV challenge (Oien *et al.*, 1994). Purified F-protein induces neutralising antibody and a type 1 T-helper cell responses in mice (Hancock *et al.*, 1995), however it only protects the lower respiratory tract after RSV challenge (Walsh *et al.*, 1987). Purified F-protein subunit vaccines are not particularly immunogenic in young infants, but they are moderately immunogenic in older children and immunocompromised adults. These vaccines may thus be suitable for the immunisation of previously infected patients who are at risk of enhanced disease during reinfection (Falsey & Walsh, 1997; Groothuis *et al.*, 1998; Piedra *et al.*, 1996; Piedra, 2003). Recently, women in their third trimester of pregnancy have been immunised with a purified F-protein subunit vaccine. A small increase in neutralisation antibodies have been observed in vaccine recipients and their infants after birth (Munoz *et al.*, 2003). Subunit vaccines may be useful as a booster in combination with live-attenuated vaccine candidates and are currently being evaluated in healthy young adults and the elderly (Gonzalez *et al.*, 2000).

F- and G-protein synthetic peptide vaccines based on bacterial expression in prokaryotes and immune stimulating complex-formulated vaccines, have been investigated in mice. However, it is unclear if this vaccine will offer an advantage over subunit vaccines (Reviewed by Dudas & Karron, 1998). Power *et al.* (1997) has found that the expression of a G-protein fragment in bacteria, is immunogenic and protective in rodents (Power *et al.*, 1997). Synthetic peptide epitope-based

vaccines have also been evaluated in mice. The immunisation of mice with peptides that mimic conformational and protective epitopes (mimotopes) of RSV F-protein, induce anti-RSV neutralising antibodies and reduced viral load *in vivo* (Chargelegue *et al.*, 1997). Steward *et al.* (2001) found that immunisation of mice with mimotopes of the F-protein, produced neutralising antibody titers equivalent to those in sera from RSV-infected animals that neutralised viral infection *in vitro*, and significantly reduced the viral titer *in vivo* (Steward, 2001).

Plasmid DNA vaccines have also been under investigation and may be useful for immunising infants when the maternal antibodies have decreased below levels needed for viral suppression. Li *et al.* have demonstrated that vaccination with plasmid vectors encoding the RSV F- or G-protein can induce a protective response in rodents (Li *et al.*, 1998; Li *et al.*, 2000).

The commercially available prophylaxis, Palivizumab (Synagis®), is a humanised immunoglobulin (IgG-I) monoclonal antibody that binds to the F-protein and is highly active *in vitro* against RSV subtype A and B. Monthly intramuscular administration of Palivizumab is safe and effective and reduces RSV hospitalisation by 55% in preterm infants (Groothuis & Nishida, 2002; Groothuis, 2003). Similar results have been found by other groups (Singleton *et al.*, 2003). However, this prophylaxis is expensive, costing approximately \$1216.58 per 100 mg (Meissner *et al.*, 1999). A cost-effectiveness analysis conducted in Germany showed that the widespread use of Palivizumab among premature infants is not feasible (Roeckl-Wiedmann *et al.*, 2003). This emphasises the importance of developing an effective, safe and cost-effective vaccine against RSV.

Despite all the research, the lack of a commercially available RSV vaccine is probably due to the fact that RSV vaccine development has been hampered by the lack of an ideal animal model to study the disease, as well as the challenge of effectively immunising very young infants who are immunologically immature, and the requirement for a bivalent vaccine (Crowe, 2002).

## **1.7 THE ROLE OF BIOINFORMATICS IN MODELING ANTIGENS**

The above literature review describes the importance of evolution of the RSV viral proteins in disease and protection for the development of a vaccine. One of the ways to study and predict the influence of these changes on the virus and protection is using bioinformatics to model the evolution of the gene and the resulting protein. Although this is theoretical science it may contribute to identifying and predicting components of successful vaccines.

It has been shown that gene and protein alignments can provide valuable insight to the evolutionary relationships among sequences (Baxevanis, 1998). The role of bioinformatics in molecular research and the prediction of epitopic domains by means of physico-chemical profiles and algorithms will be described in the following section.

### **1.7.1 Importance of bioinformatics**

The number of sequences available on Internet databases has increased dramatically in the last couple of years. The comparison of one sequence to an entire database has become an important and convenient technique available to molecular biologists to determine the degree of sequence evolution (Schuler, 1998).

Most alignment methods attempt to model molecular mechanisms by which the sequence evolves. However, while it is presumed that homologous sequences diverged from a common ancestral sequence through molecular changes, the actual ancestral sequence is unknown. Thus, in the discovery of new proteins, the new sequence is aligned with protein sequences of known function (Schuler, 1998). Whether the alignment is pairwise or involves multiple sequences, aligned sequences can help to study the phylogenetic relationship of proteins to realise the evolution of the protein (Baxevanis, 1998). Phylogenetic analysis is the means of estimating evolutionary relationships and is depicted as a tree-like diagram, which represents the lineage of the inherited relationships among organisms (Hershkovitz & Leipe, 1998). Bioinformatics is useful in the assessment of substitution models that are essential for phylogenetic analysis (Felsenstein, 1996). Bioinformatics can therefore be enlightening in the prediction of homology, identity and evolutionary relationships between sequences.

Bioinformatic software is important for interpretation and translation of genetic information and for protein sequence analysis. Using bioinformatics software for studying proteins, makes it possible to identify regions in the amino acid sequence or structure that has been better conserved than others during evolution, and that may thus usually be important for the function or the structure of the protein. All protein sequences, whether determined directly or through the translation of an open reading frame in a nucleotide sequence, contain valuable information in determining the structure and function of the protein. Software are available that predict the structure of the protein without the execution of labour-intensive techniques, like circular dichroism spectroscopy, optimal rotatory dispersion, X-ray crystallography and nuclear magnetic resonance (Reviewed by Baxevanis & Landsman, 1998). Algorithms can also aid in the prediction of antigenic sites and possible CTL epitopes from the amino acid sequence, and contribute to future vaccine research (Deleage *et al.*, 1988; Parker *et al.*, 1986; Rammensee *et al.*, 1999).

A major precaution with any predictive technique is that regardless of the method, the results are still just predictions. However, the proper use of these techniques along with primary biochemical data can provide valuable insights into the structure and function of the protein (Baxevanis & Landsman, 1998).

## 1.7.2 Methods of predicting epitopes

### 1.7.2.1 Antibody epitopes

Each of the 20 amino acid residues has a different chemical make-up; therefore each has the tendency to form protein structures of different types (Anfinsen *et al.*, 1961). The physical properties of each amino acid residue can be used to predict potential antigenic sites that contain potential B-cell epitopes (Baxevanis & Landsman, 1998). For example the computer software, Antheprot, can generate physico-chemical profiles from an amino acid sequence alignment. These profiles are calculated along the length of the protein and include various predictions of antigenicity, hydrophilicity, hydrophobicity and solvent accessibility (Deleage *et al.*, 1988). The hydropathic scale involves the assignment of a score reflecting its relative hydrophobicity based on physical characteristics of each amino acid residue (Kyte & Doolittle, 1982). The principle of the antigenicity calculation relies on atomic

flexibility, hydrophilicity and high performance liquid chromatography (HPLC) retention times of synthetic peptides (Parker *et al.*, 1986). The solvent accessibility method calculates the fraction of amino acids that are accessible to the solvent based on the probability that it would be located on the outside of the protein structure (Boger *et al.*, 1986). Evidence suggests that high levels of antigenicity, solvent accessibility and hydrophilicity are associated with potential antigenic sites, thus graphs generated with these profiles, can be used to determine the location of potential antigenic sites (Boger *et al.*, 1986; Kyte & Doolittle, 1982; Parker *et al.*, 1986). These predictions can aid in the identification and characterisation of antibody epitopes in future research.

Amino acid sequences can also be used to predict potential N- and O-glycosylation sites within a protein. In the G-protein of RSV, glycosylation contributes to its antigenic structure (Palomo *et al.*, 1991) and can also contribute to immune evasion by either abolishing G-protein recognition by antibodies or by masking antigenic sites (Melero *et al.*, 1997; Palomo *et al.*, 2000). Thus, in the RSV G-protein, it is important to determine if amino acid variation influences the proteins antigenic profile and thereby influences recognition by antibodies. Mucin type O-glycosylated serine and threonine residues can be predicted based on sequence context, secondary structure and surface accessibility with NetOglyc (Hansen *et al.*, 1998). The NetOglyc database (part of O-GLYCBASE database) consists of glycoproteins with experimentally verified O-linked glycosylation sites that have been compiled from protein sequence databases and literature. The O-glycosylated sites are predicted through amino acid comparison between the sequence of interest and the sequences residing in the database (Gupta *et al.*, 1999; Hansen *et al.*, 1998). ScanProsite (<http://ca.expasy.org/cgi-bin/scanprosite>) is a software tool available to scan the PROSITE database of different protein families. N-linked glycosylation sites including asparagine glycosylation, protein kinase C phosphorylation, casein kinase II phosphorylation, tyrosine kinase phosphorylation, N-myristoylation, can be identified using Scanprosite (Gattiker *et al.*, 2002).

Conformational changes can influence the antigenic structure of a protein (Melero *et al.*, 1997). Bioinformatics software exists for the prediction of  $\alpha$ -helices and  $\beta$ -strands using the amino acid sequence. This can be very helpful when no

information is available about the structure of a protein (Baxeavanis & Landsman, 1998). An  $\alpha$ -helix is a corkscrew-type structure with the hydrophilic main chain forming the backbone, while the side chains project outward from the helix (Pauling & Corey, 1951). Unlike  $\alpha$ -helices, a  $\beta$ -strand is an extended structure, forming bonds with one or more adjacent  $\beta$ -strands. A variant of the  $\beta$ -strand is the  $\beta$ -turn, where the polypeptide chain makes a hairpin bend (Pauling & Corey, 1951). An example of a user-friendly algorithm that predicts the secondary structure type for each amino acid residue is NNREDICT. This algorithm uses a two-layer, feed-forward neural network for the assignment of the predictions. It is useful for predicting  $\alpha$ -helices,  $\beta$ -strands or  $\beta$ -turn elements. It has been found to have a reliability of 79% for the class of all  $\alpha$ -proteins (Kneller *et al.*, 1990).

Specialised structures like coiled coils and transmembrane regions can also be accurately predicted from the amino acid sequence. Algorithms used for the prediction of coiled coils run the sequence in question against a database of proteins with known coiled coil structures. Based on the differences observed, the probability with which the query sequence can form a coiled coil region, is determined. Algorithms like LearnCoil-VMF, is designed to identify coiled-coil-like regions in viral membrane-fusion proteins (Singh *et al.*, 1999). LearnCoil-VMF predictions in retrovirus, paramyxovirus and filovirus membrane-fusion proteins have been found to be highly accurate (Singh *et al.*, 1999). The MultiCoil program predicts the probability that a fusion amino acid sequence forms a trimeric or a dimeric coiled coil and the sum of the probabilities represent the propensity for the region forming any sort of coil (Wolf *et al.*, 1997).

#### 1.7.2.2 CTL epitopes

CTL epitope prediction algorithm software programs like BIMAS and SYFPEITHI, can be used to assist in the selection of peptides for mapping CTL epitopes and to predict the HLA restriction of these epitopes (Parker *et al.*, 1994; Rammensee *et al.*, 1999). Both MHC class I- and class II-restricted epitopes can be predicted with these algorithms.

Parker *et al.* (1994) have developed the BIMAS algorithm based on experimental peptide binding data. This method predicts the relative binding strengths of all

possible nonapeptides to the MHC class I molecule. The authors found that for most peptides, each side-chain of the peptide contributes a specific amount to the stability of the HLA complex that is independent of the sequence of the peptide. A table containing multiple binding coefficients has been generated each representing the contribution of the individual amino acids in the peptide to the binding of the class I molecule. This algorithm can be used to score 8-mer, 9-mer or 10-mer peptides with a collection of HLA coefficients with a good score result ranging from 50 and higher (Parker *et al.*, 1994).

SYFPEITHI is a database of experimentally identified MHC ligands and peptide motifs and contains a collection of MHC class I and class II ligands and peptide motifs from humans, primates, cattle, chickens, and mice. With this algorithm it is possible to search for MHC alleles, MHC motifs, natural ligands, T-cell epitopes, source proteins/organisms and references. The scoring system of SYFPEITHI also evaluates every amino acid within a given peptide and the allocation of values is based on the frequency of each individual amino acid in natural ligands, T-cell epitopes or binding peptides. Epitope prediction (MHC Class I) of this algorithm, has a reliability of at least 80% according to the authors, and the naturally presented epitope should be amongst the top-scoring 2% of all the peptides predicted (Rammensee *et al.*, 1999).

### 1.7.3 Identification of immune selection

Studies have shown that certain regions of the RSV G-protein may be under selective pressure. This includes reports that glycosylation sites abolish or mask epitopes (Melero *et al.*, 1997; Palomo *et al.*, 2000), and that point mutations (Garcia *et al.*, 1994; Martinez *et al.*, 1997), hypermutations (Martinez *et al.*, 1997; Rueda *et al.*, 1994), frameshifts (Garcia-Barreno *et al.*, 1990) and alternative termination codons (Martinez *et al.*, 1999; Rueda *et al.*, 1995) might also change epitopes. Evidence for positive selection has also been found through simple pairwise comparisons of the numbers of nonsynonymous and synonymous substitutions per site (Reviewed by Melero *et al.*, 1997; Sullender, 2000).

Evidence suggests that the ratio of the number of nonsynonymous substitutions per nonsynonymous site ( $K_a$ ) to the number of synonymous substitutions per

synonymous site ( $K_s$ ) gives a measure of the amount of change of the sequence during evolution. When  $K_a$  is greater than  $K_s$  ( $K_a/K_s \gg 1$ ), it provides strong evidence that selection has acted to change the protein (positive selection). When selection is in favour of synonymous substitutions, it is referred to as purifying selection (Hurst, 2002; Nei *et al.*, 2000). There are many bioinformatic packages available to convert an alignment to  $K_a$  and  $K_s$  estimates, for example, the user-friendly MEGA2 program (Kumar *et al.*, 2000). This software package includes the Modified Nei-Gojobori method, which does not assume that the transitional and transversional substitutions occur with the same frequency. The transition/transversion (R) ratio has to be provided and can be calculated using TreePuzzle software (<http://www.nsc.liu.se/software/biology/puzzle5/>) (Nei *et al.*, 2000).

It is important to establish if positive selection occurs in viral proteins, for example, due to genetic drift in the influenza virus, regular updating of influenza vaccines are necessary (Price *et al.*, 2000). It is also important to establish if positive selection occurs, when antigens are selected to protect against all the different virus strains.

Bioinformatics is a useful tool for the critical evaluation and observation of evolution in proteins, and plays an increasingly important role in the prediction of protein function, identifying potential changes in proteins and selection of effective antigens for vaccine development.

---

## 1.8 STUDY OBJECTIVES

**Aims:** Firstly, to undertake an extensive phylogenetic analysis of the G-protein from the 2001 RSV outbreak that occurred at the Chris Hani Baragwanath hospital, in order to study the transmission dynamics of RSV in a single community, and to investigate the evolution rate in the G-protein. Secondly, to further look at evolution of RSV antigenic proteins, the F-protein of known South African genotypes, has been analysed to determine if immune selection occurs in the neutralising and CTL epitopes.

The study was divided into two sections with the following specific objectives:

**A: Molecular epidemiology of RSV in a single well-defined outbreak in South Africa:**

1. To characterise the molecular epidemiology of a RSV outbreak during 2001 at Chris Hani Baragwanath hospital. This data will be compared to outbreaks that occurred during the previous four seasons, to determine if the dominant outbreak genotype of 1999 and 2000 persisted in 2001, or if herd immunity may have selected for a new genotype.
2. To characterise the transmission dynamics and evolution of the G-protein during a single well-defined outbreak.
3. To determine if there is a correlation between the variations observed in the currently identified G-protein genotypes and changes in the antigenic structures.
4. To determine if immune selection occurred in known neutralising epitopes of the G-protein of South African genotypes.

**B: Molecular evolution of the RSV F-protein during annual epidemics in South Africa:**

1. To determine the extent of nucleotide and amino acid conservation in the F-protein between different South African genotypes.
2. To determine if immune selection occurs in the neutralising and CTL epitopes of the F-protein.
3. To analyse the secondary structure of the F-protein to determine the influence of amino acid variation on the antigenic structure of this protein.

In a study conducted at Chris Hani Baragwanath hospital, Soweto, over four consecutive seasons (1997-2000), co-circulation, switching and displacement of dominant RSV genotypes was found (Venter *et al.*, 2001). However in 1999 and 2000, one genotype (GA2) predominated. To determine if this was a once-off phenomenon that would eventually be overcome by herd immunity, extensive phylogenetic analysis of the G-protein from the 2001 outbreak was undertaken to elucidate the transmission dynamics of RSV during an epidemic in a single community, and to study the role of evolution in this neutralisation antigen in reestablishment of annual epidemics.

A recent study suggested that positive selection does not occur in the CTL epitopes of the highly conserved RSV nucleoprotein (N-protein) (Venter *et al.*, 2003). However, evidence of immune selection has been demonstrated in the CTL epitopes of the highly conserved influenza nucleoprotein (Voeten *et al.*, 2000; Boon *et al.*, 2002; Price *et al.*, 2000), and this implies that regular updating will be needed in vaccines containing these CTL epitopes. Since emphasis is placed on the F-protein in RSV vaccine development, it is important to determine if immune selection occurs in both the neutralising antibodies and CTL epitopes of different RSV genotypes. The secondary structure of the F-protein will be studied to determine the influence of amino acid variation on the antigenic structure of the F-protein. This is the first study to investigate the F-protein of South African genotypes.

## **CHAPTER 2**

### **MATERIALS AND METHODS**

#### **2.1 PATIENT SPECIMENS**

Nasal pharyngeal aspirates (NPA) were collected from children hospitalised for lower respiratory tract infection (LRTI) at Chris Hani Baragwanath hospital, Soweto, during March to June 2001. RSV-positive NPA were identified by immunofluorescence assay (IFA) and stored at  $-70^{\circ}\text{C}$  in viral transport media. Specimens were transported on ice to the National Institute for Communicable Diseases (NICD), Johannesburg, for subsequent genetic analysis. A total of 82 RSV-positive specimens were selected randomly for molecular epidemiological analysis of the G-protein of this outbreak. The HIV status of the patients was unknown and was not a criterion for selection in this study. RSV A2 and B1 prototype strains (Karron *et al.*, 1997; Wertz *et al.*, 1985) obtained from Dr. J. Beeler (Food and Drug Administration, USA) were cultured on Hep-2 cells at the NICD by Mrs. A. Buys, and used as positive controls.

A total of 18 clinical specimens that have been genotyped by means of their G-protein in previous studies in South Africa ((Venter *et al.*, 2001; Venter *et al.*, 2002; Madhi *et al.*, 2003), were selected for analysis of the F-protein. Representative specimens of all genotypes identified in South Africa to date were selected, for subtype A (GA2, GA5, GA7 and SAA1) and subtype B genotypes (GB3, GB4, SAB1, SAB2, SAB3 and SAB4) (Table 2.1).

#### **2.2 RNA EXTRACTION**

Frozen NPA specimens in viral transport media, were thawed on ice and immediately supplemented with 20 Units of RNase inhibitor (Roche molecular biochemicals, Mannheim, Germany) to prevent viral degradation. The specimens were vortexed vigorously to bring them into solution. RNA was extracted directly from NPA specimens as well as the cells of the cultured controls, using the High Pure™ Viral RNA kit (Roche molecular biochemicals, Mannheim, Germany), according to the manufacturer's instructions with minor modifications. In brief, 200 µl of NPA was added to 400 µl binding buffer (4.5 M guanidine hydrochloride, 50 mM

tris-HCl and 30% Triton X-100, pH 6.6 (25°C)) supplemented with 2 mg poly (A) carrier RNA, and vortexed briefly. This was followed by a 10 minute incubation step at room temperature to help increase the RNA yield. In specimens where a high mucous content resulted in high viscosity, the lysates were first homogenised using QIAshredder™ homogenizer columns (QIAGEN, Hilden, Germany), according to the manufacturer's recommendations. The lysates were then transferred to the High pure™ filter tubes containing glass fiber fleece, which promotes the binding of nucleic acid. Residual impurities were removed by rapid wash-and-spin steps first with 500 µl inhibitor removal buffer (5 M guanidine-HCl, 30 mM Tris-HCl, pH6.6 in 100% ethanol) followed by 450 µl wash buffer (20 mM NaCl and 2 mM Tris-HCl, pH 7.5 in 100% ethanol). For the elution of viral RNA, 50 µl of elution buffer (nuclease-free redistilled water) was added. After RNA elution, 20 U of RNase inhibitor (Roche molecular biochemicals, Mannheim, Germany) was added to prevent RNA degrading during storage at -70 °C.

**Table 2.1 RSV-positive clinical specimens that were selected for fusion protein amplification.** The abbreviated specimen name, date of collection, RSV subtype and genotype are specified. All these specimens were collected at the Chris Hani Baragwanath hospital, Soweto. Each genotype is indicated in a different colour.

|    | Specimen  | Date collected | Subtype | Genotype |
|----|-----------|----------------|---------|----------|
| 1  | 0157KS01  | 18-May-01      | A       | GA2      |
| 2  | VR173     | 13-July-98     | A       | GA2      |
| 3  | 0128KS01  | 30-April-01    | A       | GA5      |
| 4  | D1131     | 22-May-97      | A       | GA5      |
| 5  | D540      | 17-March-97    | A       | GA7      |
| 6  | VR360     | 23-March-99    | A       | GA7      |
| 7  | D707      | 20-March-98    | A       | SAA1     |
| 8  | VR341     | 3-August-98    | A       | SAA1     |
| 9  | 0258KS01  | 15-June-01     | B       | SAB1     |
| 10 | 00265KS01 | 08-June-01     | B       | SAB1     |
| 11 | V800      | 25-May-99      | B       | SAB2     |
| 12 | VR102     | 15-July-98     | B       | SAB3     |
| 13 | 0072K     | 17-April-01    | B       | SAB3     |
| 14 | D1888     | 22-May-98      | B       | SAB4     |
| 15 | D798      | 27-March-98    | B       | GB3      |
| 16 | D991      | 4-August-98    | B       | GB3      |
| 17 | D1107     | 20-July-97     | B       | GB4      |
| 18 | VR468     | 20-August-98   | B       | GB4      |

## **2.3 G-PROTEIN SPECIFIC RT-PCR (REVERSE TRANSCRIPTASE-POLYMERASE CHAIN REACTION) AND MULTIPLEX NESTED PCR FOR SUBTYPING**

To amplify the RNA extracted from the NPA, a multiplex nested RT-PCR was performed as described previously (Venter *et al.*, 2002). This combined a RT-PCR method (Sullender *et al.*, 1993) with a nesting step using primers specific to RSV subtypes A and B. The nested PCR was modified to a hotstart PCR during this study to increase product yield and improve specificity by reducing non-specific bands. A2 and B1 prototype strains were used as positive controls. The Titan™ One Tube RT-PCR system (Roche molecular biochemicals, Mannheim, Germany), Expand high fidelity PCR enzyme mix (Roche molecular biochemicals, Mannheim, Germany) and the hotstart polymerase, Fast Start Taq Polymerase (Roche molecular biochemicals, Mannheim, Germany) were selected for their proofreading functions, to avoid PCR errors. RT-PCR was performed directly on the clinical specimens and the PCR product sequenced to ensure that the DNA sequence reflected the dominant quasi-species present in the NPA.

### **2.3.1 RT-PCR**

A modified one step RT-PCR reaction using the Titan™ One Tube RT-PCR system (Roche molecular biochemicals, Mannheim, Germany) was used as described by Venter *et al.* (2002). The subtype-specific 5' primers and a universal subgroup A and B specific 3' primer amplify the C-terminal of the G-protein and were as published in Sullender *et al.* (1993) (Table 2.2). The RT-PCR reaction was carried out in a 50 µl reaction. 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, 10 U RNase inhibitor, and 1 µl of the Titan™ enzyme mix, and the final volume was adjusted with deionised distilled water. The PCR reaction was performed in a GeneAmp PCR system 9600 thermocycler (PE Biosystems, Foster City, USA) according to the following program: 50 °C for 30 minutes, 94°C for 2 minutes, (94 °C for 30 seconds, 52 °C for 30 seconds, 68 °C for 1minute); 35 cycles; 68 °C for 7 minutes.

### **2.3.2 Nested polymerase chain reaction (nested PCR)**

A nested PCR reaction was used to increase PCR product yield from clinical specimens. The nested primers used, are specific to the 5' ends of either RSV subtype A or B RT-PCR products and are summarised in Table 2.2. The anti-sense primer, F1, is specific to both RSV subtypes A and B (Peret *et al.*, 1998). The subtype B nesting primer (GB52) has been described before (Christensen *et al.*, 1999). If amplification was unsuccessful, the subtype B, G46B primer was used (M. Venter, unpublished data). The subtype A specific primer, G314A, was published in

Venter *et al.* (2002) and used in standard assays. Primer G283A was used where no PCR product was detected with G314A (Venter *et al.*, 2002). The use of different subtype-specific primers increased the number of isolates that could be subtyped and analysed.

During the course of this project, different polymerase enzymes were tested to optimise the subtype-specific nested PCR and the results were compared by assessing the specificity and PCR product yield. These enzymes included the Expand high fidelity PCR enzyme mix (Roche molecular biochemicals, Mannheim, Germany); the hotstart polymerases, Fast Start Taq Polymerase (Roche molecular biochemicals, Mannheim, Germany) and HotStarTaq DNA Polymerase (QIAGEN, Hilden, Germany).

The Expand high fidelity PCR system is composed of thermostable Taq DNA polymerase with proofreading functions. This mixture results in PCR products with high yield, high fidelity, and high specificity (Venter *et al.*, 2001). Reactions were conducted in 100 µl: In brief; 2 µl of the RT-PCR product was mixed with 0.2 mM of each deoxynucleotide (dTTP, dATP, dGTP, dCTP), 30 pmol of each primer (G52B, G283A, F1), 2.6 U Expand high fidelity PCR enzyme mix, and 10X Expand HF buffer containing 15 mM MgCl<sub>2</sub>. The following amplification cycle was used: 94 °C for 2 minutes; (94 °C for 30 seconds, 59 °C for 1 minutes, 72 °C for 1minutes) for 30 cycles; and finally 72 °C for 7 minutes.

To increase the specificity and sensitivity of the nested PCR, the PCR was modified to a hotstart PCR. Hotstart PCR makes use of a thermostable polymerase and prevents the formation of misprimed products and primer-dimers at low temperatures. The HotStarTaq DNA Polymerase system (QIAGEN, Hilden, Germany) gave the best results and was used routinely for the rest of the project. The enzyme is activated through incubation at 95 °C for 15 minutes. It provided high PCR specificity and often increased the yield of the specific PCR product. The hotstart PCR was conducted in a 50 µl reaction: 1 µl of the RT-PCR product was mixed with 0.2 mM of each deoxynucleotide (dTTP, dATP, dGTP, dCTP), 30 pmol of each primer (G52B, G314A, F1), 1.25 U HotStarTaq DNA Polymerase: 10X buffer with 15 mM MgCl<sub>2</sub>. The following reaction cycle was used for hotstart PCR: 95 °C for 15 minutes; (94 °C for 1 minute, 59 °C for 1 minute, 72 °C for 1 minute) for 35 cycles; and finally 72 °C for 10 minutes.

**Table 2.2 Primers used in G-protein specific RT-PCR and nested PCR.**

| Primer | PCR    | Orientation | Subgroup specificity | Sequence                                 | Author                             |
|--------|--------|-------------|----------------------|------------------------------------------|------------------------------------|
| F164   | RT     | 3'          | A & B                | GTTATGACACTGGTATACCAACC                  | (Sullender <i>et al.</i> , 1993)   |
| G267A  | RT     | 5'          | A                    | GATGCAACAAGCCAGATCAAG                    | (Sullender <i>et al.</i> , 1993)   |
| G32B   | RT     | 5'          | B                    | GCAACCATGTCCAAACACAAG                    | (Sullender <i>et al.</i> , 1993)   |
| G52B   | nested | 5'          | B                    | AATCAACGCACTGCCAGKACTC<br>(where K=G/T)) | (Christensen <i>et al.</i> , 1999) |
| G46B   | nested | 5'          | B                    | CAAGAATCAACGCACTGCCA                     | M. Venter, unpublished data        |
| G314A  | nested | 5'          | A                    | GAATCCCCAGCTTGAATCA                      | (Venter <i>et al.</i> , 2002)      |
| G283A  | nested | 5'          | A                    | CAAGAACACAACCCCAACAT                     | (Venter <i>et al.</i> , 2002)      |
| F1     | nested | 5'          | A & B                | CAACTCCATTGTTATTTGCC                     | (Peret <i>et al.</i> , 1998)       |

## 2.4 PCR PRODUCT ANALYSIS

PCR products were analysed by agarose gel electrophoresis. Subtyping was achieved through separation of different size PCR bands on a 2% agarose gel (1 X 0.4 M Tris Acetate EDTA). The DNA was visualised by addition of 6 µl of 10 mg/ml Ethidium Bromide. A 100 bp DNA ladder was used as the molecular weight marker (DNA molecular marker XIV, Roche diagnostics, Mannheim Germany). The gels were visualized and photographed on a UVP trans-illuminator using Grab IT annotating Grabber 2.51 software (UVP (Inc), USA).

## 2.5 SEQUENCING

### 2.5.1 PCR product purification

PCR products were purified with the Qiaquick PCR purification kit (QIAGEN, Hilden, Germany) according to the manufacturer's recommendations. The Qiaquick columns have a silica-gel membrane with selective binding properties. In short, five volumes of PB buffer (guanidine hydrochloride, isopropanol) were added to 1 volume of PCR reaction. Primers and impurities are removed by washing with 750 PE buffer containing 88% ethanol. PCR product was eluted with distilled water.

The concentration and quality of the recovered product was estimated by comparison to 1 µg of a 100 bp molecular weight marker (DNA molecular marker XIV, Roche diagnostics, Mannheim Germany) on a 2% agarose gel.

### 2.5.2 Cycle sequencing

The second hypervariable region (270 bp) at the C-terminal of the G-protein was sequenced. The primers used for sequencing are specific to the 5' ends of either RSV subtype A or B (nucleotide position 589 and 604 for subtype A and B, respectively) and were as follows: G598A (5' GGAAAGAAAACCACCACCAA) and G604B (5'AAACCAACCATCAAACCCACA3') for subtype A and B, respectively; 3' specific sequence primers to the F-G intergene region were FG-A: GATTCTGTTTGATTTGGTCAT; FG-B: CTTGCTTGATAGATCACGGTT, for subtype A and B, respectively (Venter *et al.*, 2002).

Nucleotide sequencing was carried out on the sense and anti-sense strands using the ABI Prism<sup>®</sup>BigDye<sup>™</sup> Terminator Cycle Ready Reaction kit v2.0 (PE Biosystems, Foster City, USA) according to the manufacturer's recommendations. The reaction was carried out in a 20 µl reaction: In brief, approximately 50 – 90 ng of PCR product was mixed with 4.0 µl Terminator Ready Reaction Mix (A, C, G and T-dye terminators; dNTPs, Tris-HCL (pH 9.0), MgCl<sub>2</sub>, thermal stable pyrophosphatase, and AmpliTaq DNA polymerase FS) (PE Biosystems, Foster City, USA), 2 µl of 5X Sequencing Buffer and 3.2 pmol primer. Cycle sequencing was performed according to the protocol specified for the GeneAmp 2400 thermocycler: 94 °C for 1 minute; (94 °C for 10 seconds, 50 °C for 5 seconds, 60 °C for 4 minutes) for 25 cycles; and finally 4 °C infinity.

### 2.5.3 Sequence product purification

Excess DyeDeoxy terminators were removed from the DNA sequencing products with Centri-Sep spin columns (PE Biosystems, Foster City, USA) according to the manufacturer's recommendations. The purified products were dried in a Speedivac at -60 °C, and stored at -20 °C.

## 2.5.4 Automated sequencing

Automated sequencing was carried out on an ABI 3100 Genetic Analyzer version 2 (PE Biosystems, Foster City, USA) by technologists at the NICD. Sequence analysis of the raw data was performed with Sequencher™ software version 4.0.5.

## 2.6 PHYLOGENETIC ANALYSIS

### 2.6.1 Alignments

The second hypervariable region of the **G-protein** was used for phylogenetic analysis and corresponds to nucleotides 649-918 of the A2 strain and nucleotides 652-921 of the 18537 strain of subtype A and B, respectively (Johnson *et al.*, 1987). Phylogenetic analysis was conducted separately for subtypes A and B. The nucleotide sequences were compared by alignment to G-protein nucleotide sequences selected to represent all genotypes identified in the Northern hemisphere and South America by Peret *et al.* (1998; 2000) and in Southern Africa by Venter *et al.* (2001). Sequences were obtained from the Genbank database and were aligned with the sequence alignment program, Clustal X 1.81 (Thompson *et al.*, 1997) using the multiple alignment mode.

The South African **F-protein** nucleotide sequences were compared to the A2 and B1 prototype F-protein sequences, for subtypes A and B, respectively (Collins *et al.*, 1984; Karron *et al.*, 1997). Sequences were aligned with Clustal X 1.81 (Thompson *et al.*, 1997).

### 2.6.2 Phylograms

Distance-based and character-based trees were constructed and compared in this study. Distance-based methods compare the number of differences (pairwise distance) between two aligned sequences to derive trees, while character-based methods assess the reliability of each base position on the basis of all the other nucleotide bases in an alignment (Li, 1997; Saitou, 1996; Swofford *et al.*, 1996). The most commonly applied distance-based methods include unweighted pair-group method with arithmetic mean (UPGMA), neighbor-joining (NJ) and minimum evolution (ME) methods. UPGMA method joins tree branches of the greatest similarity among pairs. Minimum evolution method finds the shortest tree that is

consistent with the path lengths between branches. The neighbor-joining algorithm generates an initial star tree, and each subsequent step generates one internal node at a time and estimates lengths of two branches connected at the node. The branches are inserted between the closest neighbor pairs and the remaining terminals in the tree (reviewed by Hershkovitz & Leipe, 1998). The most widely used character-based methods include maximum parsimony and maximum likelihood methods. The maximum likelihood method (ML) generates all possible trees and determines which tree best fits the available data. Maximum parsimony (MP) method finds the tree that needs the smallest amount of changes to explain the differences seen in taxa (Hershkovitz & Leipe, 1998).

Phylograms were constructed separately for subtypes A and B. Unrooted Maximum likelihood phylograms were constructed with DNAML of Phylip Version 3.5c (Felsenstein, 1993). Neighbor-joining trees were also generated with NEIGHBOR using the maximum likelihood distance matrix. Consensus trees that were generated from 100 bootstrap replicates of these trees correlated with the maximum likelihood trees. These bootstrap values are indicated on the maximum likelihood trees shown in Appendix A.

Bootstrap values represent datasets of the same size that are generated according to a specific model of sequence evolution and therefore estimates the reliability of the phylogram by computing the number that gives the proportion of times that a particular branch appears in a tree (Hershkovitz & Leipe, 1998).

Trees generated with Phylip Version 3.5c were visualised with Treeview version 1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) by using the prototype strains, A2 and B1 for outgroup rooting of subtypes A and B, respectively. Neighbor-joining trees were also constructed with MEGA version 2.1 (Kumar *et al.*, 2000) using the Kimura two-distance parameter (Kimura, 1980) under 100 bootstrapped replicates and midpoint rooted. The transition to transversion ratio was determined using Treepuzzle version 5.0 (<http://www.nsc.liu.se/software/biology/puzzle5/>).

The Neighbor-joining and Maximum likelihood methods constructed similar trees when bootstrapped by 100 replicates. This correlated with findings of Martinez *et al.*

(Martinez *et al.*, 1999). Distance-based methods were subsequently selected for further analysis since they are less computer intensive and faster to use.

Amino acid Gamma trees were constructed for the Pneumovirinae members and the F-protein sequences of the South African genotypes, and were bootstrapped by 100 replicates (MEGA version 2.1 (Kumar *et al.*, 2000).

### 2.6.3 Genotype criteria

The RSV subtype A and B sequences were assigned to genotypes by phylogenetic comparison to reference sequences from the genotypes identified by Peret *et al.* (1998; 2000) and Venter *et al.* (2001; 2002). Sequence groups that clustered together with a bootstrap statistic of 70 -100% and a P-distance of less than 0.7 to all other members of the same group, were considered to be a genotype (Venter *et al.*, 2001).

### 2.6.4 Estimates of selection

#### 2.6.4.1 *Proportion of differences*

The nucleotide alignments of the **G-protein** from subtypes A and B were analysed separately with MEGA version 2.1 (Kumar *et al.*, 2000) to identify specific mutations and positive selection. The proportion of differences (P-distance) was determined within and between genotypes for the **G- and the F-proteins**. The P-distance is the number of nucleotide/amino acid differences divided by the total number of nucleotides/amino acids compared.

#### 2.6.4.2 *Ka/Ks ratio*

The ratio of nonsynonymous substitutions per nonsynonymous site (Ka) to synonymous substitutions per synonymous site (Ks), were calculated for the **G- and F-proteins** according to the modified Nei-Gojobori method using MEGA version 2.1 (Kumar *et al.*, 2000). If the amount of nonsynonymous substitutions exceeds the synonymous substitutions, then strong evidence exists for positive selection ( $Ka/Ks > 1$ ) (Hurst, 2002; Nei *et al.*, 2000). The average Ka/Ks ratio was calculated for subtypes A and B separately, and relative to the prototype strains A2 and B1 (Karron *et al.*, 1997; Wertz *et al.*, 1985).

---

#### 2.6.4.3 *Amino acid analysis*

The nucleotide alignments of subtypes A and B were translated using MEGA version 2.1 (Kumar *et al.*, 2000) for both the **F- and G-proteins**. The amino acid P-distance was calculated separately for subtype A and B. Amino acid alignments were analysed with Genedoc version 2.6.002 software (<http://www.psc.edu/biomed/genedoc/>) to identify amino acid changes that may alter the secondary structure of the G-protein.

The subtype A and B sequences were compared to reference sequences from the genotypes previously identified in South Africa by Venter *et al.* (2001), to determine the extent of antigenic variation within the **G-protein**. N-glycosylation sites and threonine and serine rich regions were predicted with the software ScanProsite (<http://ca.expasy.org/cgi-bin/scanprosite>). The mucin type O-glycosylated serine and threonine residues were predicted with NetOglyc version 2 (Gupta *et al.*, 1999).

N-glycosylation sites and the mucin type O-glycosylated serine and threonine residues were also predicted for the **F-protein**.

## 2.7 SECONDARY STRUCTURE PREDICTIONS

Potential antigenic sites were predicted with Antheprot 2000 version 5.2 (Deléage *et al.*, 1988) (<http://antheprot-pbil.ibcp.fr>) using the antigenicity (Parker *et al.*, 1986), hydrophobicity (Kyte & Doolittle, 1982) and solvent accessibility (Boger *et al.*, 1986) site predictions.

## 2.8 AMPLIFICATION OF THE F-PROTEIN BY RT-PCR

Subtype-specific primers were designed for amplification of the F-proteins used in this study. The oligonucleotide primers were designed to be complementary to the 5' and 3' coding ends of the inactive precursor F-protein (F0).

### 2.8.1 Primer design

The PCR primers were designed with Dnasis for windows version 2.5 (Hitachi Software engineering Co, Ltd) using the prototype A2 and B1 sequences for subtypes A and B, respectively. The presence of hairpins, dimers and palindromes

in the primers were predicted using Netprimer (<http://www.premierbiosoft.com/netprimer/netprimer.html>). Primers used for PCR were as follows:

Subtype A sense primer: RSVFA(5654) 5'ggggcaaataacaatggagtt 3'  
 antisense primer: RSVFAR(7422) 5'cattgtaagaacatgattaggtgct 3'

Subtype B sense primer: RSVFB(5632) 5'aagcaagaacgaaattaaacctg 3'  
 antisense primer: RSVFBR(7383) 5'ctgaatgcaatattattgattcca 3'

Subtype-specific sequencing primers were designed with Webprimer (<http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer>) using the prototype A2 and B1 sequences for subtype A and B, respectively (Table 2.3). Hairpin and palindrome analysis, annealing and melting temperatures and the G/C ratio were determined with Cybergene (<http://www.cybergene.se/primer.html>). The primers were subtype-specific and were designed to amplify  $\pm$  500 bp fragments spanning the entire F-protein (Figure 2.1).

To ensure the effectiveness of the PCR primers, they were selected to fulfill the parameters described by (Rychlik, 1993). In brief, primers were selected to be free of complementarity at their 3' termini as this promotes primer-dimer formation that reduces the product yield. The GC/AT ratio of the primers was similar to or higher than that of the amplified template while the melting temperature ( $T_m$ ) was approximately similar for both the sense and antisense primers. Primers were selected to be stable at their 5' ends (C or G nucleotides) but unstable at their 3' ends (A or T nucleotides), in order to eliminate false priming.

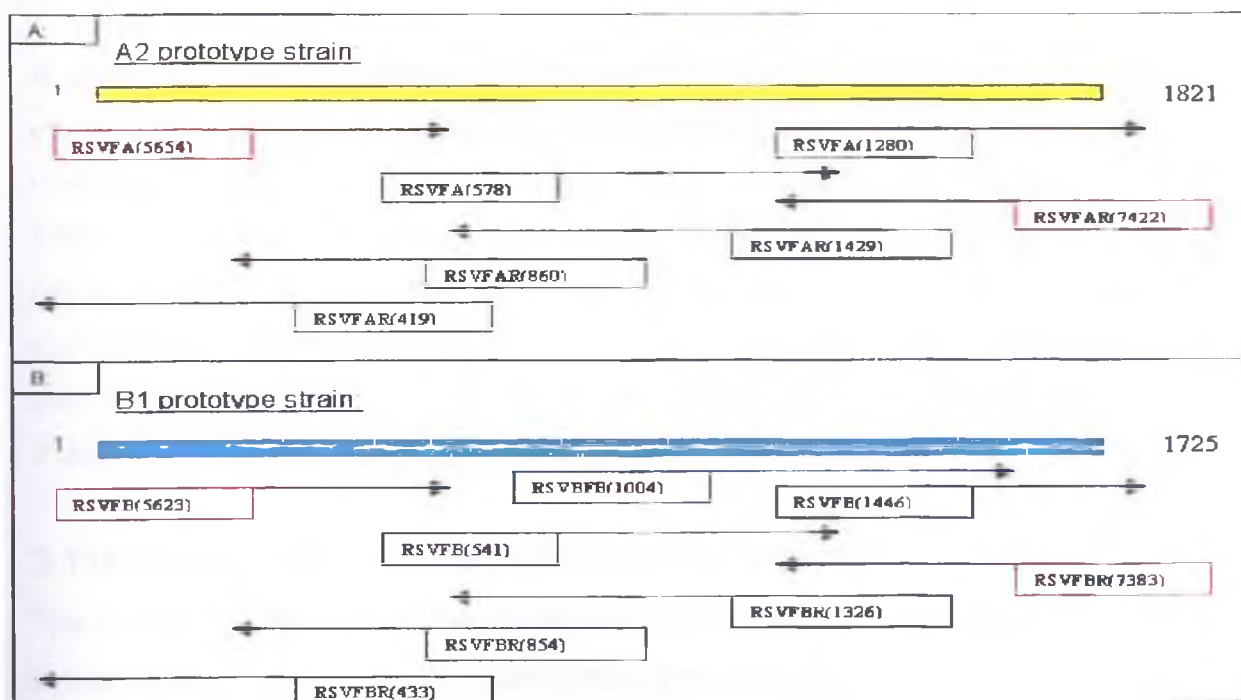
### 2.8.2 F-protein RT-PCR

The PCR reactions were performed separately for subtypes A and B using the Titan™ one tube RT-PCR system (Roche molecular biochemicals, Mannheim, Germany). The reaction was conducted in 50  $\mu$ l: 10  $\mu$ l of RNA was added to 10  $\mu$ l 5X reaction buffer, 10 mM of each dNTP, 20 pmol of each primer, 5 mM DTT-solution, 10 U RNase inhibitor, and 1  $\mu$ l the Titan™ enzyme mix. The following amplification cycle was used: 50 °C for 30 minutes, 94 °C for 2 minutes, (94 °C for 30 seconds, 52 °C for 30 seconds, 68 °C for 1minute) 39 cycles; 68 °C for 7 minutes.

**Table 2.3 Subtype-specific primers used for sequencing of the F-protein.** The primer name indicates subtype and position (column 1). All the sequence primers were designed with Webprimer (<http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer>) and analysed with Cybergene (<http://www.cybergene.se/primer.html>).

| Primer       | Subgroup (A or B) | Orientation | Sequence                  |
|--------------|-------------------|-------------|---------------------------|
| RSVFA(5654)  | A                 | Sense       | GGGGCAAATAACAATGGAGTT     |
| RSVFA(578)   | A                 | Sense       | CAGCAAAGTGTTAGACCTCAA     |
| RSVFA(1280)  | A                 | Sense       | CAGCATCCAATAAAAAATCGTG    |
| RSVFAR(419)  | A                 | Anti-sense  | CTTCTTTTCCTTTTCTTGCTTA    |
| RSVFAR(860)  | A                 | Anti-sense  | GTAACCTTGCTGTCTAACTATT    |
| RSVFAR(1429) | A                 | Anti-sense  | GTTCACCTTTTACATAGAGACT    |
| RSVFAR(7422) | A                 | Anti-sense  | CATTGTAAGAACATGATTAGGTGCT |
| RSVFB(5632)  | B                 | Sense       | AAGCAAGAACGAAATTAAACCTG   |
| RSVFB(541)   | B                 | Sense       | CTATCAAATGGGGTCAGTGTT     |
| RSVFB(1004)  | B                 | Sense       | CAAGGACTGATAGAGGATGGT     |
| RSVFB(1446)  | B                 | Sense       | GTTTCCTTCTGATGAGTTTGA     |
| RSVFBR(433)  | B                 | Anti-sense  | CTACACCTAACAGAAGCCCA      |
| RSVFBR(854)  | B                 | Anti-sense  | CTTTGTTGCCTTACTATCTGA     |
| RSVBR(1326)  | B                 | Anti-sense  | CACATAGTCACAACCATTAGA     |
| RSVFBR(7383) | B                 | Anti-sense  | CTGAATGCAATATTATTGATTCCA  |

**Figure 2.1 Illustration of the sequence primer positions relative to the F-protein gene for subtypes A (A) and B (B), respectively.** The numbers in brackets indicates the primer annealing location on the F-protein gene. The PCR primers are indicated in red and the number shown in brackets indicates the position on the RSV genome.



## 2.9 SEQUENCES USED IN PHYLOGENETIC ANALYSIS

The **G-protein** sequences used for phylogenetic analysis are summarised in Appendix B.

The **F-protein** sequences used for phylogenetic analysis included human RSV (A2 strain, AAB59858 (Collins *et al.*, 1984); B1 strain, AAB82436 (Karron *et al.*, 1997), bovine RSV (A51908 strain, AAL49410 (Yunus *et al.*, 2001)), ovine RSV (strain 28869, AAL91342 (Eleraky *et al.*, 2001)), mice pneumovirus (strain 15, BAA01902 (Chambers *et al.*, 1992)), avian metapneumovirus (MN-2a strain, AAF01772 (Seal *et al.*, 2000)) and human metapneumovirus (00-1 strain, NP\_690066 (van den Hoogen *et al.*, 2002)).

## 2.10 PROTEIN SECONDARY STRUCTURE PREDICTIONS

To investigate the influence of amino acid variability on the secondary structure of the F-protein, the secondary structure ( $\alpha$ -helices,  $\beta$ -strands and turns) for each amino acid residue were predicted with NNpredict (Kneller *et al.*, 1990) (<http://www.cmpharm.ucsf.edu/~nomi/nnpredict.html>). The coiled coil regions were previously predicted for the prototype A2 and Long strain for subtype A (Lambert *et al.*, 1996) and CH-18537 strain for subtype B (Matthews *et al.*, 2000). To determine if amino acid variation observed in the South African F-protein sequences might influence the coiled coil structure of the F-protein, heptad-repeat regions (which have been shown to be associated with coiled coil regions by (Buckland & Wild, 1989; Chambers *et al.*, 1990; Singh *et al.*, 1999)) were predicted for the South African sequences using LEARNCOIL-VM (Berger & Singh, 1997) (<http://learncoil-vmf.lcs.mit.edu/cgi-bin/vmf>). Dimeric and trimeric coiled coil regions were predicted with MULTICOIL (Wolf *et al.*, 1997) (<http://multicoil.lcs.mit.edu/cgi-bin/multicoil>). These results were compared to the prototype strains.

## 2.11 PREDICTION OF ANTIGENIC SITES ON THE F-PROTEIN

The F-protein contains both neutralising and CTL epitopes. Potential antigenic epitopic domains were identified using hydrophobic and hydrophilic profiles (Kyte & Doolittle, 1982). antigenic profiles (Parker *et al.*, 1986) and solvent accessibility profiles (Boger *et al.*, 1986) with Antheprot 2000 version 5.2 (Deleage *et al.*, 1988) (<http://antheprot-pbil.ibcp.fr>). The hydropathic profiles of individual amino acids were shaded with

Genedoc version 2.6.002. Hydropathic profiles (Kyte & Doolittle, 1982) of all the South African genotypes relative to the historic prototype strains (Collins *et al.*, 1984, Karron *et al.*, 1997) were constructed with <http://bioinformatics.weizmann.ac.il/hydroph/>. The hydropathic profiles per amino acid residue were created with Genedoc version 2.6.002 (<http://www.psc.edu/biomed/genedoc/>).

## **2.12 PREDICTION OF CTL EPITOPES**

BIMAS (<http://bimas.cit.nih.gov/>) (Parker *et al.*, 1994) and SYFPEITHI (<http://www.syfpeithi.de/>) (Rammensee *et al.*, 1999) CTL algorithms were used to predict possible CTL epitope position on the F-protein. Only HLA types shown to be dominant in at least 10% of the different ethnic groups of South Africa were used in the predictions (Venter *et al.*, 2003; Hammond *et al.*, 1997; personal communication Dr Adrian Puren, NICD). Only predicted epitopes that fell in the highest probability range were considered for mapping on the F-protein sequence map. A good score result for SYFPEITHI ranged between 15 and 30 (Rammensee *et al.*, 1999), while it ranged from 50 and higher in BIMAS (Parker *et al.*, 1994).

### **2.12.1 Published CTL epitopes**

The published CTL epitopes were manually shaded with Genedoc version 2.6.002 (<http://www.psc.edu/biomed/genedoc/>) on the South African F-protein amino acid alignments. The CTL epitopes in humans were identified by Brandenburg *et al.* (2000) (HLA B\*57 and Cw\*12-restricted) and by Rock and Crowe (2003) (HLA-A\*01-restricted). Two H-2K restricted T-helper cell epitopes identified in SCID mice, were also included (Corvaisier *et al.*, 1997; Corvaisier *et al.*, 1993).

## **2.13 AMINO ACID SEQUENCE ANALYSIS OF KNOWN NEUTRALISING EPITOPES**

The amino acid sequences of the South African F-proteins were compared to published neutralising epitopes. The following neutralising epitopes were studied, 7C2 (Trudel *et al.*, 1987), 47F (Lopez *et al.*, 1990), L4 (Martin-Gallardo *et al.*, 1991) and RS-348 (Corvaisier *et al.*, 1993).

## **2.14 ESCAPE MUTANTS**

Amino acid point mutations previously identified in escape mutants were also included in the analysis (Arbiza *et al.*, 1992; Lopez *et al.*, 1998).

## **CHAPTER 3**

### **RESULTS A:**

#### **MOLECULAR EPIDEMIOLOGICAL ANALYSIS OF A RSV OUTBREAK IN SOWETO DURING 2001**

##### **3.1 INTRODUCTION**

To date the majority of molecular epidemiological studies conducted on RSV, have been carried out in industrialised countries. Hence, there is limited data available about RSV epidemics in developing countries and only a few reports from Africa (Cane *et al.*, 1999; Roca *et al.*, 2001; Venter *et al.*, 2002; Venter *et al.*, 2001).

RSV has two antigenic subtypes, A and B (Anderson *et al.*, 1985), with distinct lineages within them. The major surface attachment protein (G) is the most variable within and between the subtypes (Reviewed by Melero *et al.*, 1997). The G-protein is one of the targets for the neutralising immune response (Johnson *et al.*, 1987), and is believed to accumulate amino acid changes with time suggesting evolution under selective pressure (Cane & Pringle, 1995). Evidence suggests that antigenic differences between the RSV strains may influence the neutralising immune response in children (Cane *et al.*, 1996), which is predominately group-specific and reveals poor cross-reactivity (Langedijk *et al.*, 1997; Muelenaer *et al.*, 1991). Therefore sequence analysis of the G-protein can be a useful molecular tool for the epidemiological investigation of RSV epidemics. It has been shown that the second hypervariable region (270 nucleotides), which makes up the C-terminal of the G-protein, provides a reliable representation for the entire G-gene variability (Peret *et al.*, 1998). This region has subsequently become a frequently used target for phylogenetic analysis for molecular epidemiological studies (Peret *et al.*, 2000; Peret *et al.*, 1998; Reviewed in Sullender, 2000).

Most RSV molecular epidemiological studies are based on subtype A viruses (Cane *et al.*, 1991; Cane & Pringle, 1995b; Cane, 1997; Cane *et al.*, 1999; Garcia *et al.*, 1994) although a limited number of studies have also looked at the extent of

variation in subtype B (Coggins *et al.*, 1998; Martinez *et al.*, 1999; Sullender *et al.*, 1990; Sullender *et al.*, 1991b). Various groups have demonstrated that genotype assignment to clusters of RSV strains, are helpful in the identification of the virus circulation patterns in communities (Peret *et al.*, 1998; reviewed by Sullender, 2000; reviewed by Cane, 2001). Although there is not yet a standardised method for RSV genotyping, a method published by Peret *et al.* of the CDC is now used frequently in molecular epidemiological studies (Peret *et al.*, 2000; Sullender, 2000; Venter *et al.*, 2002; Venter *et al.*, 2001).

Peret and colleagues found that RSV outbreaks are community-based in nature and that different genotypes co-circulated within the community. Local factors, possibly immunity to the previous year's strains, determine the predominant strains during a RSV epidemic (Peret *et al.*, 2000). The excess of coding over noncoding nucleotide changes in the G-protein gene sequences suggests that the variation observed between seasons may be a response to immune driven selective pressure (Cane *et al.*, 1992). Despite this variability within one community, very similar viruses appeared to be present globally (Cane *et al.*, 1992).

Melero *et al.* have identified residues within the C-terminal end of the G-protein essential for the integrity of strain-specific epitopes (Melero *et al.*, 1997). Martinez and colleagues showed a significant correlation between the use of alternative stop codons and changes in genotype-specific epitopes in subtype B viruses (Martinez *et al.*, 1999). These findings were supported by an evolutionary study conducted on subtype B in Japan, which also suggested linear genetic drift within genotypes that allow persistence over consecutive seasons. The authors suggest a more distinctive geographical and temporal clustering of subtype B strains than subtype A strains, and suggested that this may be due to reduced infectivity of subtype B viruses which limit the spread of subtype B viruses (Kamasaki *et al.*, 2001). Various authors suggest the disappearance and appearance of new genotypes within the same community in both subtypes, and found that subtype B strains were more stable than subtype A (Cane *et al.*, 1991; Choi & Lee, 2000; Peret *et al.*, 2000).

A study conducted at the Chris Hani Baragwanath hospital, Soweto, over four consecutive seasons (1997-2000), revealed that both subtypes are important in South African outbreaks and that positive selection occurs in the G-protein of both RSV subtypes. In each epidemic, co-circulation, switching and displacement of dominant genotypes was found (Venter *et al.*, 2001). However during 1999 and 2000, one genotype, GA2, became dominant and persisted as the dominant genotype in both seasons. When other studies are taken into consideration, it appears to be unusual for one genotype to dominate for more than one season. Furthermore, this genotype was also isolated across the whole country during 2000. To determine if this was a once-off phenomenon that would be eventually overcome by herd immunity, extensive phylogenetic analysis of the G-protein from the 2001 RSV outbreak was undertaken in this Chapter. This study demonstrates the transmission dynamics of RSV during an epidemic in a single community and the rate of evolution in the G-protein. This study will provide additional insight into the evolution of RSV during a single epidemic in South Africa.

---

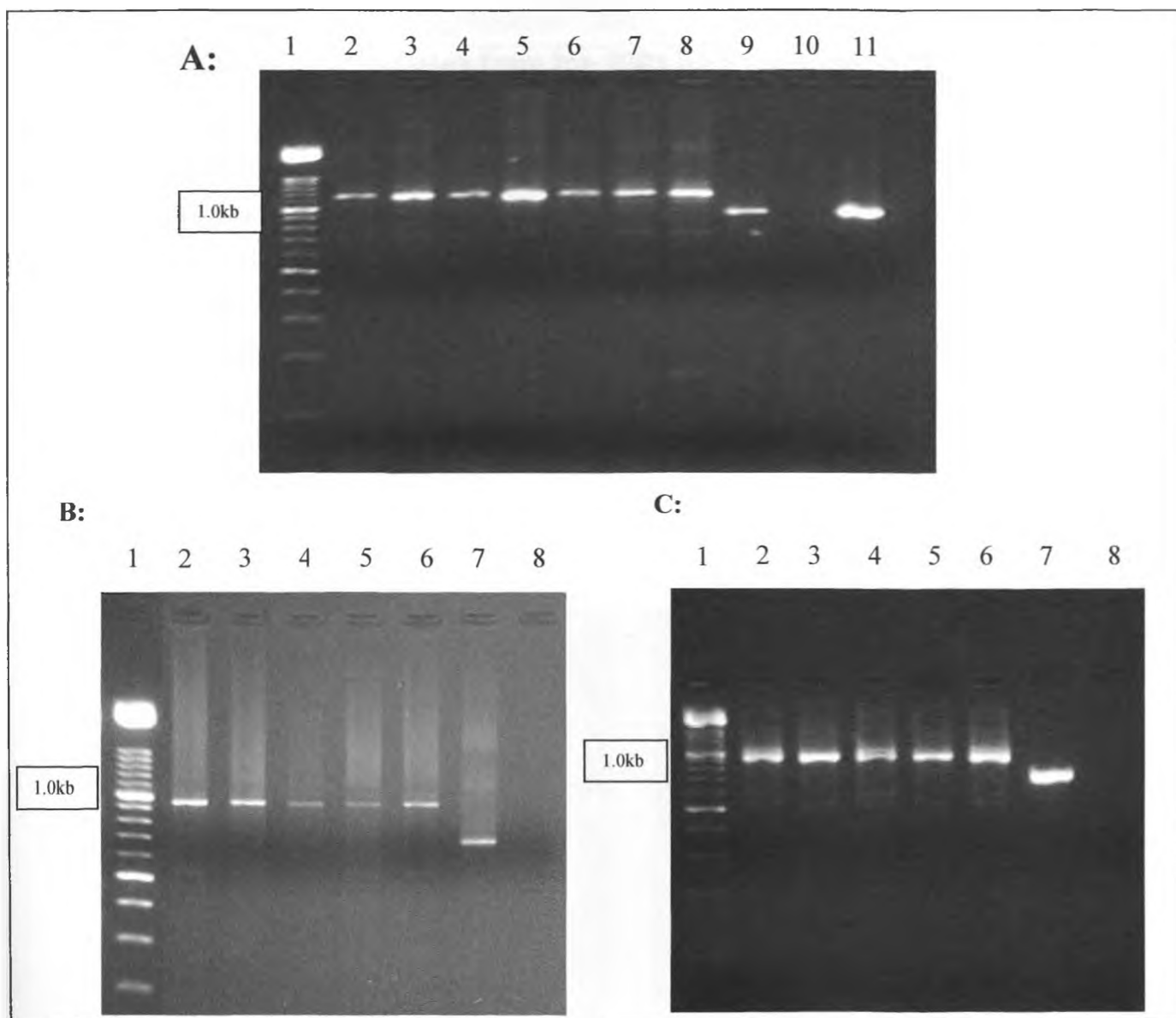
## 3.2 RESULTS

### 3.2.1 Subtype-specific multiplex nested PCR

In this Chapter a region of 290 bp at the C-terminal end of the G-protein gene of RSV was amplified by multiplex RT-PCR (Sullender *et al.*, 1993) combined with a nested-PCR that discriminates between RSV subtypes A and B (Venter *et al.*, 2002). The RT-PCR resulted in PCR products of 900 and 1100 bp for subtype A and B, respectively (Sullender *et al.*, 1993). The RT-PCR products were easily distinguished on a 2% agarose gel and could be used for subtype assignment (Figure 3.1 A).

Since nasal pharyngeal aspirates often contain low levels of RNA, a previously developed multiplex nested PCR (Venter *et al.*, 2002) was used to increase the product yield and sensitivity of the RT-PCR. The nested PCR primers were specific to the 5'ends of the RT-PCR product of either subtypes A or B, and resulted in PCR products of 728 and 946 bp for subtypes A and B, respectively, that were distinguishable on a 2% agarose gel (Figure 3.1 B).

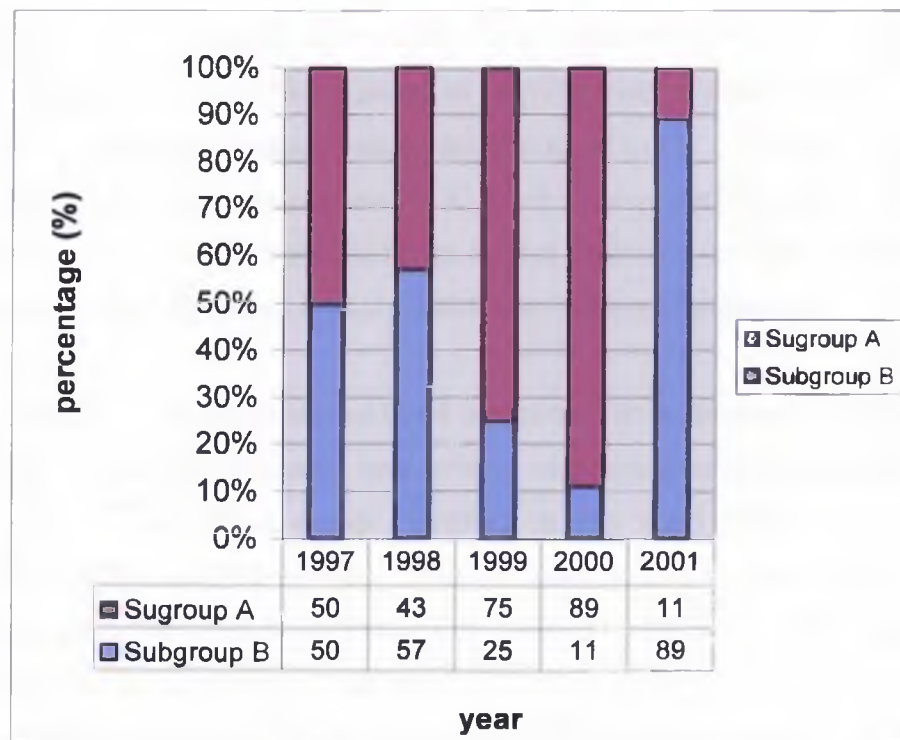
To increase the specificity and sensitivity of the nested PCR, this PCR was modified to a hotstart PCR using the same primers. The improvement in yield, sensitivity and specificity of the nested PCR is illustrated in Figure 3.1 C.



**Figure 3.1 A) RSV Subtype-specific multiplex RT-PCR.** Lane 1: Molecular weight size marker (100 bp)(Roche, Mannheim, Germany); Lanes 2-8: Subtype B RT-PCR products; Lane 9: Subtype A RT-PCR product; Lane 10. Negative control; Lane 11: Positive control (subtype A) **B) The subtype-specific multiplex nested PCR using Expand high fidelity PCR enzyme (Roche, Mannheim, Germany).** Lane 1: 100 bp molecular weight size marker; Lanes 2-6: Subtype B RT-PCR products; Lane 7: Positive control (subtype A); Lane 8: Negative control. **C) An illustration of hotstart PCR using HotStarTaq DNA polymerase (QIAGEN, Hilden, Germany).** Lane 1: 100 bp molecular weight size marker; Lanes 2-6: Subtype B RT-PCR products; Lane 7: Positive control (subtype A); Lane 8: Negative control.

### 3.2.2 Subtyping of isolates from the 2001 RSV outbreak in Soweto

A total of 82 clinical specimens from the 2001 RSV outbreak at Chris Hani Baragwanath hospital, Soweto, were randomly selected for subtyping. Specimens were available over a four month period from March to June 2001. The specimen collection dates corresponded to annual RSV outbreaks in South Africa (March to July). Of the 82 RSV specimens from 2001, only nine were subtype A (11%), while 73 (89%) belonged to subtype B. The subtype distribution of RSV subtypes at Chris Hani Baragwanath hospital during 1997 – 2000 (Venter *et al.*, 2001) and the findings of this study for 2001 are collectively illustrated in Figure 3.2.



**Figure 3.2 RSV subtype distribution at Chris Hani Baragwanath hospital during 1997 to 2001.** The findings of this study for the 2001 outbreak are combined with the subtypes circulation found during 1997 to 2000 as described by Venter *et al.* (2001).

### 3.2.3 Phylogenetic analysis

A total of 81 of the 82 specimens were sequenced and compared by phylogenetic analysis to specimens of previous seasons in South Africa. One subtype B isolate was of insufficient quality for sequencing and was not considered further.

Sequence analysis was conducted separately for the subtypes. The sequences were aligned with the prototype strains for subtype A (A2 strain) and B (B1 strain), respectively, using Clustal X 1.81 (Thompson *et al.*, 1997) (Appendix C).

The RSV subtype A and B sequences were assigned to genotypes by phylogenetic comparison to reference sequences from genotypes described by Peret *et al.* (1998; 2000) and Venter *et al.* (2001; 2002). Both distance- and character-based trees were constructed and the statistical significance analysed for 100 bootstrap replicates. Distance based neighbor-joining (Figure 3.3) and character-based maximum-likelihood trees (Appendix A) were similar, confirming results obtained by Martinez *et al.* (1999). Since distance-based methods are less computer intensive and faster to use, this method was used in subsequent analyses.

Neighbor-joining distance-based trees are shown for subtypes A and B, respectively (Figure 3.3). This tree includes only unique sequences chosen as representatives of each identical sequence group identified in the 2001 RSV outbreak in Soweto (indicated with a postfix of -S01). A selection of South African strains from each genotype described before (Venter *et al.*, 2001; Venter *et al.*, 2002), and sequences from the rest of the world that were shown to form unique clusters in previous studies (Garcia *et al.*, 1994; Karron *et al.*, 1997; Martinez *et al.*, 1999; Peret *et al.*, 2000; Roca *et al.*, 2001; Sullender *et al.*, 1991b; Wertz *et al.*, 1985), were included for genotyping the Soweto-2001 sequences.

#### Subtype A

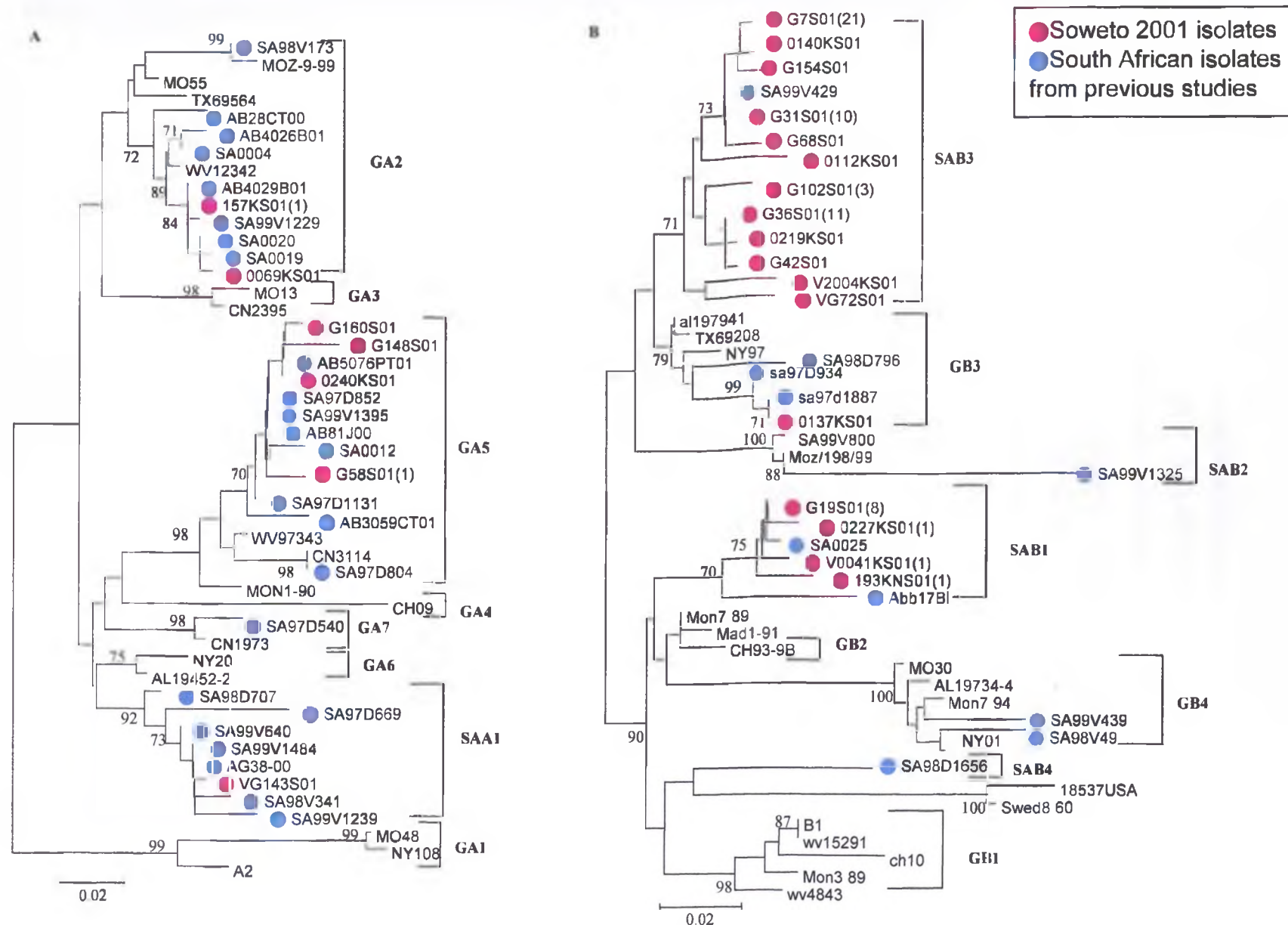
Nine subtype A specimens were identified in the 2001 outbreak. Subtype A isolates clustered with previously assigned genotypes GA5 (5 isolates), GA2 (3 isolates) and SAA1 (1 isolate) with bootstrap values of 70 – 99% (Figure 3.3 A). The GA2 Soweto isolate 157KS01, was identical to a previously identified South African isolate,

AB4026B01, from Bloemfontein, also isolated in 2001. The GA5 Soweto isolate 0240KS01 was also similar to a South African isolate AB5076PT01, circulating in the Pretoria area during the same year.

### Subtype B

The majority of clinical specimens from 2001 clustered with subtype B (n=73 of 81 isolates). The subtype B sequences clustered with the genotypes SAB3 (56 isolates), SAB1 (16 isolates) and GB3 (1 isolate), with bootstrap values of 70 – 100% (Figure 3.3 B). The SAB3 sequence G31S01, was identical to a Soweto strain circulating in 1999, SA99V429. The SAB3 genotype dominated this epidemic with 69% followed by SAB1 with 18%, replacing the genotype (GA2) that dominated the 1999 and 2000 RSV epidemics in the same community (Venter *et al.*, 2001).

**Figure 3.3 Neighbor-joining trees for subtype A (A) and B (B), respectively (see next page).** The trees were computed with the Kimura 2 distance matrix and midpoint rooted using Mega version 2 (Kumar *et al.*, 2000). Genotypes with bootstrap values above 70 are indicated in brackets next to the phylogram. Isolates from the Soweto 2001 outbreak are indicated by red bullets (•), where the "S" indicated Soweto and the postfix the year of isolation (2001). Previously identified South African isolates are indicated by blue bullets (•). These included isolates from Soweto, SA97-00 (1997-2000), (Venter *et al.*, 2001), isolates from rural community clinics in Agincourt in the Limpopo province, Ag98-00 (1998-2000), (Venter *et al.*, 2002) and private hospitals from Pretoria, Johannesburg, Bloemfontein, Cape Town and Durban (Ab-00) (Madhi *et al.*, 2003). GenBank sequences included isolates from the USA (AL, Alabama; CN, Canada; CH, Rochester; MO, Missouri; NY, New York; TX, Texas; WV, West Virginia) (Peret *et al.*, 2000), Uruguay (MON) (Garcia *et al.*, 1994; Martinez *et al.*, 1999), Mozambique (MOZ) (Roca *et al.*, 2001), Sweden (Swed) (Sullender *et al.*, 1991), Spain (MAD) (Martinez *et al.*, 1999; Garcia *et al.*, 1994) and the prototype strains (A2 (Wertz *et al.*, 1985) and B1 (Karron *et al.*, 1997) for subtypes A and B, respectively).



---

### 3.2.4 Strain variation

#### Subtype A

Table 3.1 summarises the unique sequence groups within the genotypes identified in the 2001 outbreak. The number of identical isolates and the date of first and last virus isolation are indicated. Seven unique sequence groups were identified among the subtype A genotypes. This included two different GA2 sequence groups (group I and II), four GA5 groups (group III to VI) and one SAA1 group (group VII). The different subtype A strains circulated throughout the outbreak, from the end of April until the end of June.

The mean nucleotide P-distance for the Soweto subtype A sequences was 0.07. The subtype A sequence groups identified in the 2001 outbreak, had pairwise distances (P-distances) ranging from 0.00 to 0.11 to each other (Table 3.2 A). The average intergenotypic P-distances between the Soweto 2001 sequences ranged from 0.073 (SAA1 and GA2) to 0.103 (GA2 and GA5) (Table 3.3 A). The intragenotypic P-distance for the Soweto isolates was 0.005 and 0.019 for GA2 and GA5, respectively. The intergenotypic P-distance ranged from 0.053 (SAA1 and GA6) to 0.160 (GA1 and GA5) when Soweto sequences and previously identified genotypes were considered.

**Table 3.1 RSV clinical specimens analysed by sequencing during the 2001 RSV outbreak in Soweto.** The unique sequence groups are indicated in roman numbers and the representative strain of each group which was isolated first, is indicated in bold. The first and last isolation date is also shown.

| RSV SUBTYPE A | CLINICAL SPECIMENS | GENOTYPE | UNIQUE SEQUENCE GROUP | NUMBER OF IDENTICAL SEQUENCES | DATE OF FIRST ISOLATION IN EACH SEQUENCE GROUP | DATE OF FINAL ISOLATION IN EACH SEQUENCE GROUP |
|---------------|--------------------|----------|-----------------------|-------------------------------|------------------------------------------------|------------------------------------------------|
|               | <b>0069KS01</b>    | GA2      | I                     | 1                             | 24-Apr-01                                      |                                                |
|               | <b>157KNSS01</b>   |          | II                    | 2                             | 11-May-01                                      | 15-May-01                                      |
|               | 0119KS01           |          |                       |                               | 15-May-01                                      |                                                |
|               | <b>0146KS01</b>    | GA5      | III                   | 1                             | 31-May-01                                      |                                                |
|               | <b>G148S01</b>     |          | IV                    | 1                             | 15-Jun-01                                      |                                                |
|               | <b>G160S01</b>     |          | V                     | 1                             | 22-Jun-01                                      |                                                |
|               | <b>G58S01</b>      |          | VI                    | 2                             | 10-Apr-02                                      | 30-Apr-01                                      |
|               | 0128KS01           |          |                       |                               | 30-Apr-01                                      |                                                |
|               | <b>VG143S01</b>    | SAA1     | VII                   | 1                             | 12-Jun-01                                      |                                                |
| RSV SUBTYPE B | <b>0137KS01</b>    | GB3      | VIII                  | 1                             | 03-May-01                                      |                                                |
|               | <b>G19S01</b>      | SAB1     | IX                    | 10                            | 27-Mar-01                                      | 11-May-01                                      |
|               | 1802737S01         |          |                       |                               | 01-Apr-01                                      |                                                |
|               | G40S01             |          |                       |                               | 09-Apr-01                                      |                                                |
|               | G51S01             |          |                       |                               | 09-Apr-01                                      |                                                |
|               | 0080KS01           |          |                       |                               | 12-Apr-01                                      |                                                |
|               | G107S01            |          |                       |                               | 30-Apr-01                                      |                                                |
|               | G104S01            |          |                       |                               | 30-Apr-01                                      |                                                |
|               | V0024KS01          |          |                       |                               | 04-May-01                                      |                                                |
|               | G197KS01           |          |                       |                               | 11-May-01                                      |                                                |
|               | G164S01            |          |                       |                               | 30-Apr-01                                      |                                                |
|               | <b>V0041KS01</b>   |          | X                     | 2                             | 05-Apr-01                                      | 04-May-01                                      |
|               | V0042KS01          |          |                       |                               | 04-May-01                                      |                                                |
|               | 193KS01            |          | XI                    | 2                             | 17-May-01                                      | 08-Jun-01                                      |
|               | 00265KS01          |          |                       |                               | 08-Jun-01                                      |                                                |
|               | <b>0227KS01</b>    |          | XII                   | 2                             | 01-Jun-01                                      | 15-Jun-01                                      |
|               | 0258KS01           |          |                       |                               | 15-Jun-01                                      |                                                |
|               | <b>G36S01</b>      | SAB3     | XIII                  | 12                            | 04-Mar-01                                      | 19-Jun-01                                      |
|               | G17S01             |          |                       |                               | 27-Mar-01                                      |                                                |
|               | G66S01             |          |                       |                               | 12-Apr-01                                      |                                                |
|               | 0068KS01           |          |                       |                               | 13-Apr-01                                      |                                                |
|               | G94S01             |          |                       |                               | 23-Apr-01                                      |                                                |
|               | G99S01             |          |                       |                               | 25-Apr-01                                      |                                                |
|               | G100S01            |          |                       |                               | 25-Apr-01                                      |                                                |
|               | VG46S01            |          |                       |                               | 04-May-01                                      |                                                |
|               | 0167KS01           |          |                       |                               | 11-May-01                                      |                                                |
|               | G131S01            |          |                       |                               | 05-Jun-01                                      |                                                |
|               | G141S01            |          |                       |                               | 12-Jun-01                                      |                                                |
|               | 0282KS01           |          |                       |                               | 19-Jun-01                                      |                                                |
|               | <b>G7S01</b>       |          | XIV                   | 22                            | 22-Mar-01                                      | 22-Jun-01                                      |
|               | G18S01             |          |                       |                               | 27-Mar-01                                      |                                                |
|               | G15S01             |          |                       |                               | 30-Mar-01                                      |                                                |
|               | G39S01             |          |                       |                               | 03-Apr-01                                      |                                                |
|               | G43S01             |          |                       |                               | 04-Apr-01                                      |                                                |
|               | 0064KS01           |          |                       |                               | 13-Apr-01                                      |                                                |
|               | G76S01             |          |                       |                               | 17-Apr-01                                      |                                                |
|               | G78S01             |          |                       |                               | 18-Apr-01                                      |                                                |
|               | G91S01             |          |                       |                               | 23-Apr-01                                      |                                                |
|               | G124S01            |          |                       |                               | 02-May-01                                      |                                                |
|               | 0093KS01           |          |                       |                               | 02-May-01                                      |                                                |
|               | G116S01            |          |                       |                               | 03-May-01                                      |                                                |
|               | G121S01            |          |                       |                               | 07-May-01                                      |                                                |
|               | 0136KS01           |          |                       |                               | 11-May-01                                      |                                                |
|               | 0190KS01           |          |                       |                               | 15-May-01                                      |                                                |
|               | 0007KS01           |          |                       |                               | 18-May-01                                      |                                                |
|               | 0153KS01           |          |                       |                               | 22-May-01                                      |                                                |
|               | 0239KS01           |          |                       |                               | 31-May-01                                      |                                                |
|               | 0263KS01           |          |                       |                               | 04-Jun-01                                      |                                                |
|               | G137S01            |          |                       |                               | 07-Jun-01                                      |                                                |
|               | G114S01            |          |                       |                               | 12-Jun-01                                      |                                                |
|               | G161S01            |          |                       |                               | 22-Jun-01                                      |                                                |
|               | <b>G31S01</b>      |          | XV                    | 11                            | 30-Mar-01                                      | 28-Jun-01                                      |
|               | 0072KS01           |          |                       |                               | 17-Apr-01                                      |                                                |
|               | 1601251S01         |          |                       |                               | 17-Apr-01                                      |                                                |
|               | 0079KS01           |          |                       |                               | 20-Apr-01                                      |                                                |
|               | G87S01             |          |                       |                               | 20-Apr-01                                      |                                                |
|               | 0092KS01           |          |                       |                               | 26-Apr-01                                      |                                                |
|               | 0209KS01           |          |                       |                               | 21-May-01                                      |                                                |
|               | 0181KS01           |          |                       |                               | 22-May-01                                      |                                                |
|               | 0207KS01           |          |                       |                               | 04-Jun-01                                      |                                                |
|               | G163S01            |          |                       |                               | 23-Jun-01                                      |                                                |
|               | G175S01            |          |                       |                               | 28-Jun-01                                      |                                                |
|               | <b>V2004KS01</b>   |          | XVI                   | 1                             | 02-Apr-01                                      |                                                |
|               | <b>G42S01</b>      |          | XVII                  | 1                             | 04-Apr-01                                      |                                                |
|               | G68S01             |          | XVIII                 | 1                             | 12-Apr-01                                      |                                                |
|               | VG72S01            |          | XIX                   | 1                             | 17-Apr-01                                      |                                                |
|               | 0112KS01           |          | XX                    | 1                             | 25-Apr-01                                      |                                                |
|               | <b>G103S01</b>     |          | XXI                   | 3                             | 26-Apr-01                                      | 07-Jun-01                                      |
|               | 0216KS01           |          |                       |                               | 23-May-01                                      |                                                |
|               | G135S01            |          |                       |                               | 07-Jun-01                                      |                                                |
|               | 0146KS01           |          | XXII                  | 1                             | 03-May-01                                      |                                                |
|               | <b>0219KS01</b>    |          | XXIII                 | 1                             | 23-May-01                                      |                                                |
|               | G154S01            |          | XXIV                  | 1                             | 19-Jun-01                                      |                                                |

**Table 3.2 Pairwise nucleotide distances (P-distance) for all unique subtype A (A) and B (B) Soweto 2001 sequences.** The sequence group number is indicated left of the representative specimen and genotype name.

**A:**

|                     | 157K<br>GA2 | 0069K<br>GA2 | G160<br>GA5 | 0204K<br>GA5 | G148<br>GA5 | G58<br>GA5 | VG143<br>SAA1 |
|---------------------|-------------|--------------|-------------|--------------|-------------|------------|---------------|
| 157KS01_{GA2}_II    |             |              |             |              |             |            |               |
| 0069KS01_{GA2}_I    | 0.01        |              |             |              |             |            |               |
| G160S01_{GA5}_V     | 0.10        | 0.10         |             |              |             |            |               |
| 0204KS01_{GA5}_III  | 0.10        | 0.10         | 0.00        |              |             |            |               |
| G148S01_{GA5}_IV    | 0.11        | 0.11         | 0.02        | 0.01         |             |            |               |
| G58S01_{GA5}_VI     | 0.10        | 0.10         | 0.02        | 0.02         | 0.03        |            |               |
| VG143S01_{SAA1}_VII | 0.07        | 0.08         | 0.08        | 0.08         | 0.09        | 0.09       |               |

**B:**

|                      | 0137K<br>GB3 | 0227K<br>SAB1 | 193K<br>SAB1 | G19<br>SAB1 | V0041K<br>SAB1 | 0112K<br>SAB3 | 0140K<br>SAB3 | 0219K<br>SAB3 | G102<br>SAB3 | G31<br>SAB3 | G36<br>SAB3 | G42<br>SAB3 | G68<br>SAB3 | G7<br>SAB3 | V2004K<br>SAB3 | VG72<br>SAB3 |
|----------------------|--------------|---------------|--------------|-------------|----------------|---------------|---------------|---------------|--------------|-------------|-------------|-------------|-------------|------------|----------------|--------------|
| 0137KS01_{GB3}_VIII  |              |               |              |             |                |               |               |               |              |             |             |             |             |            |                |              |
| 0227KS01_{SAB1}_XII  | 0.08         |               |              |             |                |               |               |               |              |             |             |             |             |            |                |              |
| 193KS01_{SAB1}_XI    | 0.08         | 0.02          |              |             |                |               |               |               |              |             |             |             |             |            |                |              |
| G19S01_{SAB1}_IX     | 0.07         | 0.01          | 0.01         |             |                |               |               |               |              |             |             |             |             |            |                |              |
| V0041KS01_{SAB1}_X   | 0.08         | 0.01          | 0.02         | 0.01        |                |               |               |               |              |             |             |             |             |            |                |              |
| 0112KS01_{SAB3}_XX   | 0.06         | 0.09          | 0.10         | 0.09        | 0.09           |               |               |               |              |             |             |             |             |            |                |              |
| 0140KS01_{SAB3}_XXII | 0.05         | 0.08          | 0.09         | 0.07        | 0.08           | 0.03          |               |               |              |             |             |             |             |            |                |              |
| 0219KS01_{SAB3}_XXII | 0.05         | 0.08          | 0.08         | 0.07        | 0.07           | 0.03          | 0.03          |               |              |             |             |             |             |            |                |              |
| G102S01_{SAB3}_XXI   | 0.05         | 0.08          | 0.08         | 0.07        | 0.08           | 0.03          | 0.03          | 0.02          |              |             |             |             |             |            |                |              |
| G31S01_{SAB3}_XV     | 0.04         | 0.07          | 0.08         | 0.07        | 0.07           | 0.03          | 0.01          | 0.02          | 0.02         |             |             |             |             |            |                |              |
| G36S01_{SAB3}_XIII   | 0.05         | 0.07          | 0.08         | 0.07        | 0.07           | 0.03          | 0.02          | 0.00          | 0.01         | 0.01        |             |             |             |            |                |              |
| G42S01_{SAB3}_XVII   | 0.05         | 0.08          | 0.08         | 0.07        | 0.07           | 0.03          | 0.03          | 0.01          | 0.02         | 0.02        | 0.00        |             |             |            |                |              |
| G68S01_{SAB3}_XVIII  | 0.05         | 0.08          | 0.09         | 0.07        | 0.08           | 0.03          | 0.01          | 0.03          | 0.03         | 0.01        | 0.02        | 0.02        |             |            |                |              |
| G7S01_{SAB3}_XIV     | 0.04         | 0.08          | 0.08         | 0.07        | 0.07           | 0.03          | 0.00          | 0.02          | 0.03         | 0.00        | 0.02        | 0.02        | 0.01        |            |                |              |
| V2004KS01_{SAB3}_XVI | 0.06         | 0.08          | 0.08         | 0.07        | 0.07           | 0.04          | 0.04          | 0.03          | 0.03         | 0.03        | 0.03        | 0.03        | 0.04        | 0.04       |                |              |
| VG72S01_{SAB3}_XIX   | 0.06         | 0.08          | 0.09         | 0.07        | 0.08           | 0.05          | 0.04          | 0.03          | 0.04         | 0.03        | 0.03        | 0.03        | 0.04        | 0.03       | 0.03           |              |

**Table 3.3 Average nucleotide P-distances within and between subtype A (A) and B (B) genotypes, respectively.** The matrix shows the average intergenotypic P-distances for the clinical specimens identified in the Soweto 2001 outbreak relative to genotypes identified in other studies. The genotypes identified in this study are marked in blue. The intragenotypic P-distance is shown on the diagonal. The same calculation is also shown when only the unique Soweto nucleotide sequence groups identified in this study are considered.

**A:**

| Subtype A: | GA1   | GA2   | GA3   | GA4   | GA5   | GA6   | GA7   | SAA1  |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| GA1        | 0.051 |       |       |       |       |       |       |       |
| GA2        | 0.136 | 0.026 |       |       |       |       |       |       |
| GA3        | 0.149 | 0.067 | 0.015 |       |       |       |       |       |
| GA4        | 0.173 | 0.120 | 0.122 | n/c   |       |       |       |       |
| GA5        | 0.160 | 0.094 | 0.104 | 0.119 | 0.023 |       |       |       |
| GA6        | 0.133 | 0.060 | 0.065 | 0.106 | 0.074 | 0.019 |       |       |
| GA7        | 0.142 | 0.074 | 0.078 | 0.107 | 0.087 | 0.057 | 0.015 |       |
| SAA1       | 0.140 | 0.073 | 0.087 | 0.107 | 0.082 | 0.053 | 0.072 | 0.024 |

**Soweto 2001 sequences:**

|      | GA2   | GA5   | SAA1 |
|------|-------|-------|------|
| GA2  | 0.005 |       |      |
| GA5  | 0.103 | 0.019 |      |
| SAA1 | 0.073 | 0.084 | n/c  |

**B:**

| Subtype B: | GB1   | GB2   | GB3   | GB4   | SAB1  | SAB2  | SAB3  | SAB4 |
|------------|-------|-------|-------|-------|-------|-------|-------|------|
| GB1        |       |       |       |       |       |       |       |      |
| GB2        | 0.063 |       |       |       |       |       |       |      |
| GB3        | 0.082 | 0.045 |       |       |       |       |       |      |
| GB4        | 0.120 | 0.073 | 0.103 |       |       |       |       |      |
| SAB1       | 0.087 | 0.043 | 0.062 | 0.097 |       |       |       |      |
| SAB2       | 0.119 | 0.079 | 0.078 | 0.114 | 0.098 |       |       |      |
| SAB3       | 0.089 | 0.057 | 0.038 | 0.108 | 0.073 | 0.074 |       |      |
| SAB4       | 0.094 | 0.062 | 0.082 | 0.115 | 0.058 | 0.119 | 0.091 |      |

**Soweto 2001 sequences:**

|      | GB3   | SAB1  | SAB3  |
|------|-------|-------|-------|
| GB3  | n/c   |       |       |
| SAB1 | 0.077 | 0.015 |       |
| SAB3 | 0.051 | 0.078 | 0.025 |

Subtype B

Seventeen unique sequence groups were identified among the subtype B Soweto 2001 specimens (Table 3.1). This included one GA3 group (group VIII), four different SAB1 groups (group IX to XII) and 12 different SAB3 groups (group XIII to XXIV). Many sequence groups only infected one to three people (group VIII, X–XII, XVI–XXIV), however the dominant outbreak strains in sequence groups, IX (SAB1), XIII (SAB3), XIV (SAB3) and XV (SAB3) infected many individuals throughout the epidemic (March to May for strain IX (SAB1) and March to June for SAB3 sequence groups XIII, XIV and XV). These four sequence groups might be the initial SAB1 and SAB3 outbreak strains since they were isolated from the start to the end of the outbreak. The other SAB1 and SAB3 sequence groups may have emerged as a result of positive selection in the dominant SAB1 and SAB3 outbreak strains.

The subtype B sequence groups had P-distances ranging from 0.00 to 0.09 relative to each other, with a mean nucleotide P-distance for all Soweto subtype B genotypes of 0.05 (Table 3.2 B). The intergenotypic P-distances of the Soweto 2001 sequences ranged from 0.051 (GB3 and SAB3) to 0.078 (SAB1 and SAB3) (Table 3.3 B). The intragenotypic P-distance for SAB1 and SAB3 was 0.015 and 0.025, respectively. This suggests that the highest level of nucleotide variation occurred in SAB3, the dominant outbreak genotype. The intergenotypic P-distance ranged from 0.043 (SAB1 and GB2) to 0.119 (SAB2 and SAB4) when Soweto sequences and previously identified genotypes were considered.

### 3.2.5 G-protein gene nucleotide mutation analysis during the 2001 outbreak

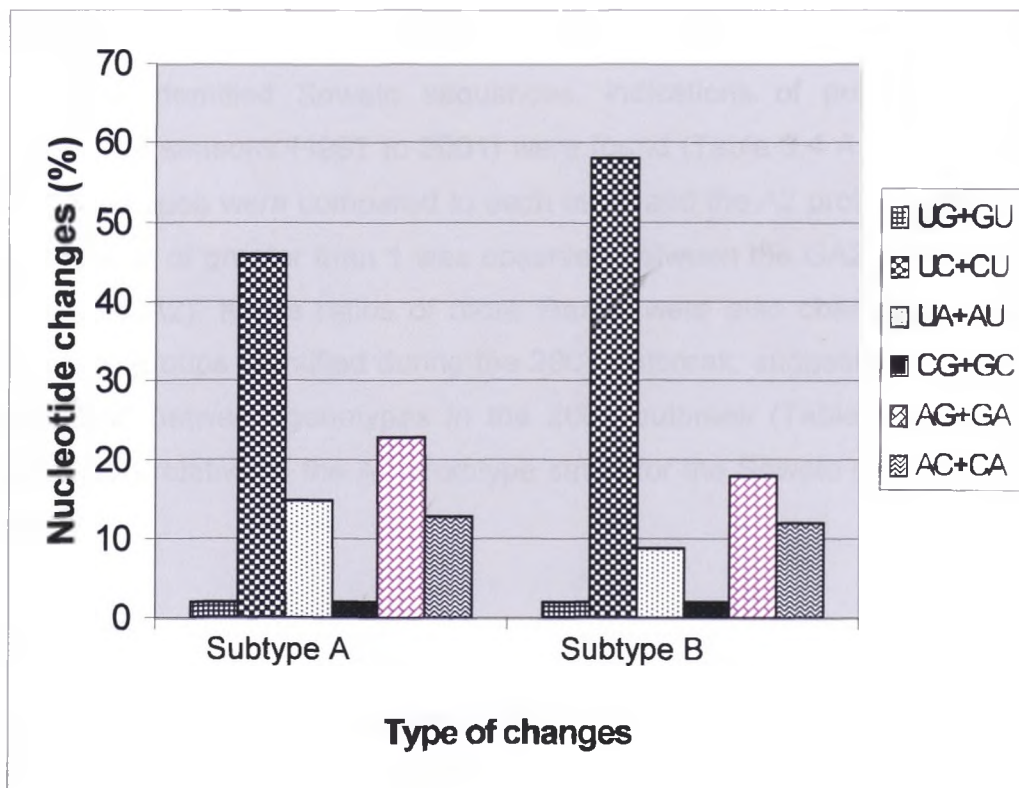
Subtype A

Appendix C shows the nucleotide variability observed within the unique nucleotide sequence groups found in subtype A during the 2001 outbreak. Most of the nucleotide changes observed in the subtype A sequences were genotype-specific. Point mutations were the only type of mutations observed. Transitions exceeded transversions (Ts:Tv = 4.59) with UC+CU transitions dominating AG+GA transitions in subtype A (Figure 3.4). Since the ancestral virus strain is unknown, the forward

and reverse nucleotide changes (UC and CU; AG and GA, etc.) were combined for these calculations.

### Subtype B

In Appendix C, the Soweto subtype B sequences are aligned to the B1 prototype strain. Mutations in subtype B were mostly genotype-specific point mutations. In subtype B, transitions also exceeded transversions (Ts:Tv = 6.15) with UC+CU transitions dominating AG+GA transitions (Figure 3.4).



**Figure 3.4** Frequency of specific types of transitions (AG+GA, UC+CU) and transversions (AC+CA, UA+AU, CG+GC, UG+GU) within the G-protein of Soweto 2001 sequences.

### 3.2.6 Estimates of selection

To estimate if positive selection played a role in the variation observed in subtypes A and B during the 2001 outbreak, the ratio of nonsynonymous substitutions per nonsynonymous site ( $K_a$ ) to the number of synonymous substitutions per synonymous site ( $K_s$ ), were calculated according to the modified Nei Gojobori method (Kumar *et al.*, 2000), where ratios of  $K_a/K_s \gg 1$  suggest possible positive selection (Hurst, 2002; Nei *et al.*, 2000). Table 3.4 indicates the  $K_a/K_s$  ratios calculated for the single 2001 outbreak (subtype A and B), as well as the  $K_a/K_s$  ratio of the G-proteins of Soweto sequences over five consecutive seasons, combining the 2001 outbreak with sequences previously identified in Soweto (1997 – 2000) (Venter *et al.*, 2001).

#### Subtype A

When comparing the nucleotide sequence of the Soweto 2001 sequences to previously identified Soweto sequences, indications of positive selection over consecutive seasons (1997 to 2001) were found (Table 3.4 A1). When the Soweto 2001 genotypes were compared to each other and the A2 prototype strain, average  $K_a/K_s$  ratios of greater than 1 was observed between the GA2 and GA5 genotypes (Table 3.4 A2).  $K_a/K_s$  ratios of more than 1 were also observed within the GA5 sequence groups identified during the 2001 outbreak, suggesting positive selection within and between genotypes in the 2001 outbreak (Table 3.4 A3). The mean  $K_a/K_s$  ratio relative to the A2-prototype strain for the Soweto subtype A sequences was 1.75.

#### Subtype B

The  $K_a/K_s$  ratios of all subtype B genotypes with the exception of SAB2, were greater than one ( $>1$ ), when compared to Soweto sequences from previous seasons (Table 3.4 B1). When the subtype B 2001 genotypes were compared to each other and the B1 prototype strain, the  $K_a/K_s$  ratios of all the Soweto genotypes (GB3, SAB1 and SAB3) was greater than 1 (Table 3.4 B2).  $K_a/K_s$  ratios of more than one ( $>1$ ) was also observed in the unique sequence groups identified in SAB3. This suggests possible positive selection within the 2001 outbreak (Table 3.4 B4) (Hurst, 2002). The mean  $K_a/K_s$  ratio relative to the prototype B1 strain for subtype B was 2.00.

**Table 3.4 Ratio of nonsynonymous substitutions per nonsynonymous site (Ka) to synonymous substitutions per synonymous site (Ks) for all the Soweto genotypes.** The average Ka/Ks ratios are calculated for all the identified Soweto subtype A genotypes (1997-2001) (Venter et al., 2001) relative to the A2 prototype strain (A1) as well as all the Soweto subtype B genotypes (1997-2001) relative to the B1 prototype strain (B1). Ka/Ks ratios are shown for the Soweto 2001 subtype A (genotype A2) and B (genotype B2) relative to the A2 and B1 prototype strains, respectively. The Ka/Ks ratios are calculated for the individual subtype A (Table A3: GA5; Table A4: GA2) and B (Table B3: SAB1; Table B4: SAB4) nucleotide sequence groups identified in the Soweto 2001 outbreak. Values greater or equal to one ( $\geq 1$ ) are highlighted in blue.

#### **SUBTYPE A**

##### **A1) Average Ka/Ks ratios for Soweto genotypes (1997-2001) relative to the A2 prototype strain:**

| Ka/Ks         | [1] GA1 | [2] GA2 | [3] GA5 | [4] GA7 | [5] SAA1 |
|---------------|---------|---------|---------|---------|----------|
| [1] A2 strain |         |         |         |         |          |
| [2] GA2       | 1.57    |         |         |         |          |
| [3] GA5       | 1.43    | 2.33    |         |         |          |
| [4] GA7       | 1.39    | 2.31    | 1.34    |         |          |
| [5] SAA1      | 0.78    | 0.82    | 0.68    | 0.73    |          |

##### **A2) Average Ka/Ks ratio for Soweto 2001 genotypes relative to the A2 prototype strain:**

| Ka/Ks         | [1] GA1 | [2] GA2 | [3] GA5 | [4] SAA1 |
|---------------|---------|---------|---------|----------|
| [1] A2 strain |         |         |         |          |
| [2] GA2       | 1.89    |         |         |          |
| [3] GA5       | 1.47    | 2.6     |         |          |
| [4] SAA1      | 0.73    | 0.79    | 0.66    |          |

##### **A3) Ka/Ks ratios for individual GA5 nucleotide sequence groups:**

| Ka/Ks | III       | IV   | V | VI |
|-------|-----------|------|---|----|
| III   |           |      |   |    |
| IV    | 0.5       |      |   |    |
| V     | 0.01/0.00 |      |   |    |
| VI    |           | 0.75 |   |    |

##### **A4) Ka/Ks ratio for individual GA2 nucleotide sequence groups:**

| Ka/Ks | I         | II |
|-------|-----------|----|
| I     |           |    |
| II    | 0.01/0.00 |    |

#### **SUBTYPE B**

##### **B1) Average Ka/Ks ratio for Soweto genotypes (1997-2001) relative to the B1 prototype strain:**

| Average Ka/Ks | [1] GB1 | [3] GB3 | [4] GB4 | [5] SAB1 | [6] SAB2 | [7] SAB3 | [8] SAB4 |
|---------------|---------|---------|---------|----------|----------|----------|----------|
| [1] B1 strain |         |         |         |          |          |          |          |
| [3] GB3       | 1.8     |         |         |          |          |          |          |
| [4] GB4       | 1.8     | 1.3     |         |          |          |          |          |
| [5] SAB1      | 1.8     | 1.8     | 1.8     |          |          |          |          |
| [6] SAB2      | 0.83    | 0.75    | 0.93    | 0.92     |          |          |          |
| [7] SAB3      | 1.4     | 1.67    | 1.89    | 1.8      | 0.64     |          |          |
| [8] SAB4      | 1.18    | 1.83    | 1.27    | 1.2      | 1        | 1.28     |          |

##### **B2) Average Ka/Ks ratios for Soweto 2001 genotypes relative to the B1 prototype strain:**

| Ka/Ks         | [1] GB1 | [2] GB3 | [3] SAB1 | [4] SAB3 |
|---------------|---------|---------|----------|----------|
| [1] B1 strain |         |         |          |          |
| [2] GB3       | 1.8     |         |          |          |
| [3] SAB1      | 1.8     | 1.8     |          |          |
| [4] SAB3      | 1.83    | 1.8     | 1.33     |          |

##### **B3) Ka/Ks ratio for individual SAB1 nucleotide sequence groups:**

| Ka/Ks | IX        | X         | XI        | XII |
|-------|-----------|-----------|-----------|-----|
| IX    |           |           |           |     |
| X     | 0.01/0.00 |           |           |     |
| XI    | 0.02/0.00 | 0.03/0.00 |           |     |
| XII   | 0.01/0.00 | 0.02/0.00 | 0.03/0.00 |     |

##### **B4) Ka/Ks ratio for individual SAB3 nucleotide sequence groups:**

| Ka/Ks | XIII   | XIV    | XV     | XVI | XVII | XVIII | XIX | XX   | XXI    | XXII | XXIII | XXIV |
|-------|--------|--------|--------|-----|------|-------|-----|------|--------|------|-------|------|
| XIII  |        |        |        |     |      |       |     |      |        |      |       |      |
| XIV   | 1      |        |        |     |      |       |     |      |        |      |       |      |
| XV    | 1      | 0.01/0 |        |     |      |       |     |      |        |      |       |      |
| XVI   | 1      | 0.75   | 0.75   |     |      |       |     |      |        |      |       |      |
| XVII  | 0      | 0.5    | 0.5    | 0.5 |      |       |     |      |        |      |       |      |
| XVIII | 0.5    | 1      | 1      | 0.6 | 1    |       |     |      |        |      |       |      |
| XIX   | 1      | 3      | 3      | 0.4 | 0.5  | 1.5   |     |      |        |      |       |      |
| XX    | 1      | 0.75   | 0.5    | 0.8 | 0.5  | 0.6   | 0.8 |      |        |      |       |      |
| XXI   | 0.02/0 | 3      | 2      | 2   | 2    | 1.5   | 2   | 1.6  |        |      |       |      |
| XXII  | 2      | 0.01/0 | 0.01/0 | 1   | 1    | 2     | 4   | 0.75 | 3      |      |       |      |
| XXIII | 0.01/0 | 2      | 1      | 1.5 | 1    | 1     | 1.5 | 1    | 0.02/0 | 2    |       |      |
| XXIV  | 0.5    | 0      | 1      | 0.6 | 0.25 | 0.5   | 1.5 | 0.6  | 1.5    | 1    | 1     |      |

### 3.2.6 Amino acid analysis of Soweto 2001 sequences

The mean amino acid P-distance for the G-protein of the Soweto 2001 sequences was 0.13 and 0.09 for subtype A and B, respectively, suggesting higher variability at amino acid level than at nucleotide level ( $P=0.07$  and  $0.05$  for subtype A and B, respectively), confirming findings of previous studies (Cane & Pringle, 1995; Martinez *et al.*, 1999; Sullender *et al.*, 1991b). Amino acid variation was higher in subtype A ( $P=0.13$ ) than in subtype B ( $P=0.09$ ). In total, 22 unique amino acid sequence groups could be identified, 7 for subtype A and 15 for subtype B, relative to 7 nucleotide sequence groups for subtype A and 17 nucleotide sequence groups for subtype B, suggesting some nucleotide changes were silent. In Figure 3.5 the amino acid alignment of the unique sequence groups identified in 2001 are shown relative to the prototype strains (A2 and B1 for subtypes A and B, respectively).

#### Subtype A

The intergenotypic amino acid P-distance for the Soweto 2001 subtype A sequences ranged from 0.117 (GA2 and SAA1) to 0.213 (GA2 and GA5). The intragenotypic P-distance of the Soweto 2001 sequences was 0.015 and 0.034 for GA2 and GA5, respectively (Table 3.5 A). The mean proportion of nucleotide substitutions resulting in amino acid changes among the Soweto 2001 sequences was 61% for subtype A.

The seven unique amino acid sequence groups identified within subtype A during the 2001 outbreak are indicated in Figure 3.5. Single amino acid substitutions were noted within the subtype A alignment, when sequence groups identified at the start of the 2001 epidemic (groups I and III) were compared to sequence groups identified at the end of the epidemic (groups II and IV-VII).

Martinez *et al.* (1999) have shown that changes in protein length correlates with the classification of RSV in different genotypes. The use of alternative stop codons has been associated with antigenic variation found in escape mutants that recognise strain-specific epitopes (Melero *et al.*, 1997). This phenomenon was more pronounced in the subtype B Soweto isolates than in subtype A. Table 3.6 illustrates the differences in stop codon usage that results in different amino acid sequence

lengths for different RSV sequence groups. The Soweto 2001 subtype A viruses had limited variation in the length of the G-protein, which corresponds with previous findings for subtype A (Garcia *et al.*, 1994). All the Soweto 2001 subtype A sequences terminated at the UAG triplet at nucleotide position 913 of the G-gene, resulting in a protein length of 299 amino acid residues (GA2 and GA5), with the exception of one SAA1 sequence which resulted in a protein length of 298 amino acid residues and used a UAG stop codon at nucleotide position 910.

#### Subtype B

The intergenotypic P-distance between genotypes identified in the 2001 epidemic ranged from 0.1 (GB3 and SAB3) to 0.161 (GB3 and SAB1). The intragenotypic P-distance within the subtype B genotypes was 0.04 and 0.044 for SAB1 and SAB3, respectively (Table 3.5 B). The proportion of nucleotide substitutions resulting in amino acid changes among the Soweto 2001 sequences was 57 % for subtype B.

Of the 22 unique amino acid sequence groups identified in this epidemic, fifteen belonged to subtype B (Figure 3.5). Single point mutations could also be identified for subtype B between the early strains identified in March (SAB1 (sequence group IX) and SAB3 (XIII, XIV and XV)) when compared to later sequence groups in the same genotype (SAB1 (sequence groups X-XII) and SAB3 (XVI-XXII)). Many of these amino acid changes resulted in amino acid residues that belong to different classification groups.

The subtype B sequences had more variation in stop codon usage than subtype A (Table 3.6). GB3, SAB1 and SAB3 G-protein translations resulted in amino acid sequences of 299, 295 and 295 or 299, respectively, with the exception of three SAB3 isolates (G36S01, G42S01, and 0219KS01) which each had 292 residues. When the codon triplet at position 892 of the G-gene was CAA instead of UAA, the UAG termination codon at position 901 was used, which was the case for all the SAB1 sequences and two of the SAB3 sequences (V2004KS01 and G102S01). Changes in the first and second termination codons in subtype B resulted in amino acids sequence lengths of 299, and the UAG stopcodon at position 913 was used. Therefore the total number of subtype B isolates using the first, second and third in-frame termination codons was 14, 21 and 37, respectively.

**A:**

```

                220          *          240          *          260          *          280          *          300
A2_GA1          : KKDPKPQTTSKEVPTTKPTEEPTINTTKTNIITLLTSNTTGNPELTSQMETFHSTSSSEGNPSPSQVSTTSEYPSQPSSPPNTPRQ [298]
0069KS01_GA2_I  : .....P...L.....K.....S.R.....H..LK..L...T...I.....Y.....L..SL..S..TKW [298]
157KS01_GA2_II  : .....P...L.....K.....S.R.....H..K..L...T...L.....Y.....L..SL..S..TKW [298]
0240KS01_GA5_III : ...L.....P..A.....K....I..P..R.....NS...L.H...E..L.....T...IY.....L...P..S.ITD. [298]
G148S01_GA5_IV  : ...L.....P..A.....K....I..P..R.....NS...L.H...E..L.....T...IH...K.L...P..S.ITD. [298]
G160S01_GA5_V   : ...L.....P..A.....K....I..P..R.....NS...L.H...E..L..L.....T...IY.....L...P..S.ITD. [298]
G58S01_GA5_VI   : ...L.....P..TE.....DK....I..P..R.....NS...L.H...E..L.....T...IY.....L...P..S.ITD. [298]
VG143S01_SAA1_VII : ...L.....P.....KS.....R.....N.....H..K..L.....Y.....L...P..S..TN- [297]

```

**B:**

```

                220          *          240          *          260          *          280          *          300
B1_GB1          : KRDPKTPAKTTKKETTNPTKKPTLTTERDTSTSQSTVLDTTTTLEHTIQQQSLHSTTPENTPNSTQTPTASEPSTSNSTQNTQSHA [299]
0137KS01_GB3_VIII : .....IL...L...N.....PK.....I.....SK.....K....S.....KL..Y. [299]
G19S01_SAB1_IX   : .....L.....L..K.....SK..TL.....F.....EA---- [295]
V0041KS01_SAB1_X : .....L.....L..K.....SI..AL.....F.....EA---- [295]
193KS01_SAB1_XI  : .....L.....L..K.....PK..TP.....L.....L....EA---- [295]
0227KS01_SAB1_XII : .....L.....L..K.....SK..TL.....K..F.....LEA---- [295]
G36S01_SAB3_XIII : .....L...L.....PK.....A.SG.....----- [292]
G7S01_SAB3_XIV   : .....L...L.....PK.....A.SG.....N.....I.....KL..Y. [299]
G31S01_SAB3_XV   : .....L...L.....PK.....A.SG.....N.....KL..Y. [299]
V2004KS01_SAB3_XVI : .....L.....SK.....P.....SG.....KL---- [295]
G68S01_SAB3_XVII : .....L...L.....IPK.....A.SG.....N.....KL..Y. [299]
VG72S01_SAB3_XVIII : .....L.....PK..G...P.....A.SG.....P.....KL---- [295]
0112KS01_SAB3_XIX : .....L...L.....PKI.....A.SG.....Y.....KL..Y. [299]
G102S01_SAB3_XX  : .....L...L.....PK.....A.PG.....K.....P.....KL---- [295]
0140KS01_SAB3_XXI : .....L...L.....PK.....A.SG.....N.....I.....KL..LY. [299]
0219KS01_SAB3_XXII : .....L...L.....PK.....A.SGY.....----- [292]

```

**Figure 3.5 The amino acid alignment of the G-protein's second variable region of the unique RSV sequence groups identified in the 2001 Soweto outbreak.** The alignments are shown relative to the prototype strains for subtype A (A2) and B (B1), respectively. The genotype and the unique sequence group (indicated in roman figures) are indicated next to the specimen name and the amino acid length is indicated at the end of each sequence.

**Table 3.5 Amino acid P-distances within and between Soweto genotypes.** The average amino acid P-distance for the Soweto 2001 sequences are relative to genotypes identified in other studies. The intragenotypic P-distance is shown on the diagonal.

A:

Soweto 2001 sequences:

|      | GA2   | GA5   | SAA1 |
|------|-------|-------|------|
| GA2  | 0.015 |       |      |
| GA5  | 0.213 | 0.034 |      |
| SAA1 | 0.117 | 0.141 | n/c  |

B:

Soweto 2001 sequences:

|      | GB3   | SAB1  | SAB3  |
|------|-------|-------|-------|
| GB3  | n/c   |       |       |
| SAB1 | 0.161 | 0.040 |       |
| SAB3 | 0.100 | 0.140 | 0.044 |

**Table 3.6 Differences in termination codon usage by Soweto sequences (2001).** The 7 C-terminal codons of the G-protein mRNA are shown for the prototype strains (A2 for subtype A, B1 for subtype B) and the unique amino acid sequence groups identified in Soweto during 2001. The sequences correspond to the amino acid positions 292 to 299, and the nucleotide position of 892 to 915 for the G-gene. The corresponding translated amino acid sequences of the prototype strains are shown above the nucleotide codons.

| Subtype A           |         |         |         |         |         |         |         |         |
|---------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| Amino acid position | 292     | 293     | 294     | 295     | 296     | 297     | 298     | 299     |
| Nucleotide position | 892-894 | 895-897 | 898-900 | 901-903 | 904-906 | 907-909 | 910-912 | 913-915 |
| A2 aa               | P       | P       | N       | T       | P       | R       | Q       | *       |
| A2 nt               | CCA     | CCC     | AAC     | ACA     | CCA     | CGC     | CAG     | UAG     |
| 0069K_GA2_I         | CCA     | UCC     | AAC     | ACA     | ACA     | AAG     | UGG     | UAG     |
| 157KS01_GA2_II      | CCA     | UCC     | AAC     | ACA     | ACA     | AAG     | UGG     | UAG     |
| 0240KS01_GA5_III    | CCA     | UCC     | AAC     | AUA     | ACA     | GAC     | CAG     | UAG     |
| G148S01_GA5_IV      | CCA     | UCC     | AAC     | AUA     | ACA     | GAC     | CAG     | UAG     |
| G160S01_GA5_V       | CCA     | UCC     | AAC     | AUA     | ACA     | GAC     | CAG     | UAG     |
| G58S01_GA5_VI       | CCA     | UCC     | AAC     | AUA     | ACA     | GAC     | CAG     | UAG     |
| VG143S01_SAA1_VII   | CCA     | UCC     | AAC     | ACA     | ACA     | AAC     | UAG     | UAG     |

| Subtype B           |         |         |         |         |         |         |         |         |
|---------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| Amino acid position | 292     | 293     | 294     | 295     | 296     | 297     | 298     | 299     |
| Nucleotide position | 892-894 | 895-897 | 898-900 | 901-903 | 904-906 | 907-909 | 910-912 | 913-915 |
| B1 aa               | Q       | N       | T       | Q       | S       | H       | A       | *       |
| B1 nt               | CAA     | AAU     | ACC     | CAA     | UCA     | CAU     | GCU     | UAG     |
| 0137KS01_GB3_VIII   | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |
| G19S01_SAB1_IX      | CAA     | GAA     | GCC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| V0041KS01_SAB1_X    | CAA     | GAA     | GCC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| 193KS01_SAB1_XI     | CAA     | GAA     | GCC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| 0227KS01_SAB1_XII   | CUA     | GAA     | GCC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| G36S01_SAB3_XIII    | UAA     | AAA     | CUC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| G7S01_SAB3_XIV      | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |
| G31S01_SAB3_XV      | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |
| V2004KS01_SAB3_XVI  | CAA     | AAA     | CUC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| G42S01_SAB3_XVII    | UAA     | AAA     | CUC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| G68S01_SAB3_XVIII   | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |
| VG72S01_SAB3_XIX    | CAA     | AAA     | CUC     | UAA     | UCA     | UAU     | GCU     | UAG     |
| 0112KS01_SAB3_XX    | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |
| G102S01_SAB3_XXI    | CAA     | AAA     | CUC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| 0140KS01_SAB3_XXII  | CAA     | AAA     | CUC     | CAG     | UUA     | UAU     | GCU     | UAG     |
| 0219KS01_SAB3_XXIII | UAA     | AAA     | CUC     | UAG     | UCA     | UAU     | GCU     | UAG     |
| G154S01_SAB3_XXIV   | CAA     | AAA     | CUC     | CAG     | UCA     | UAU     | GCU     | UAG     |

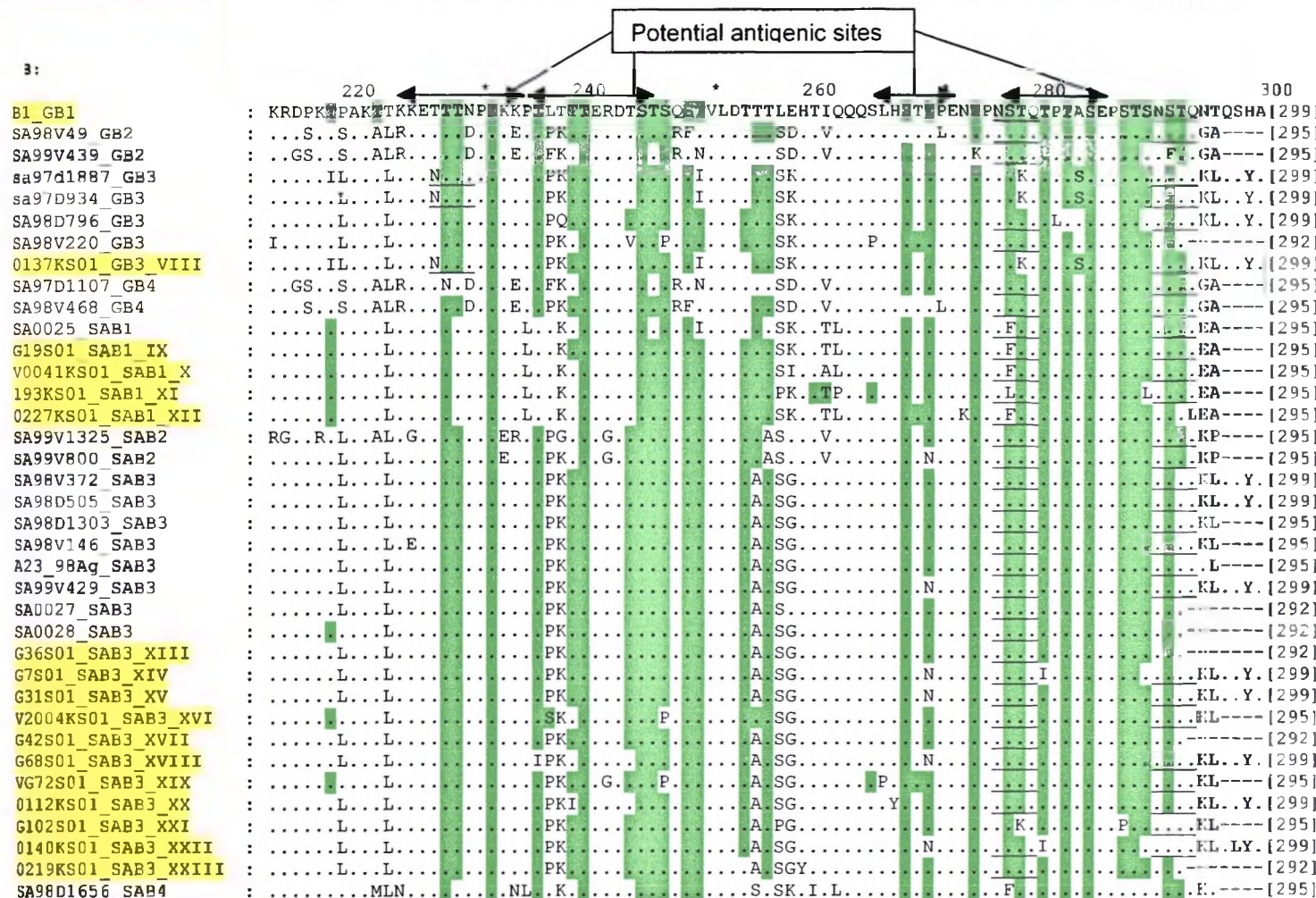
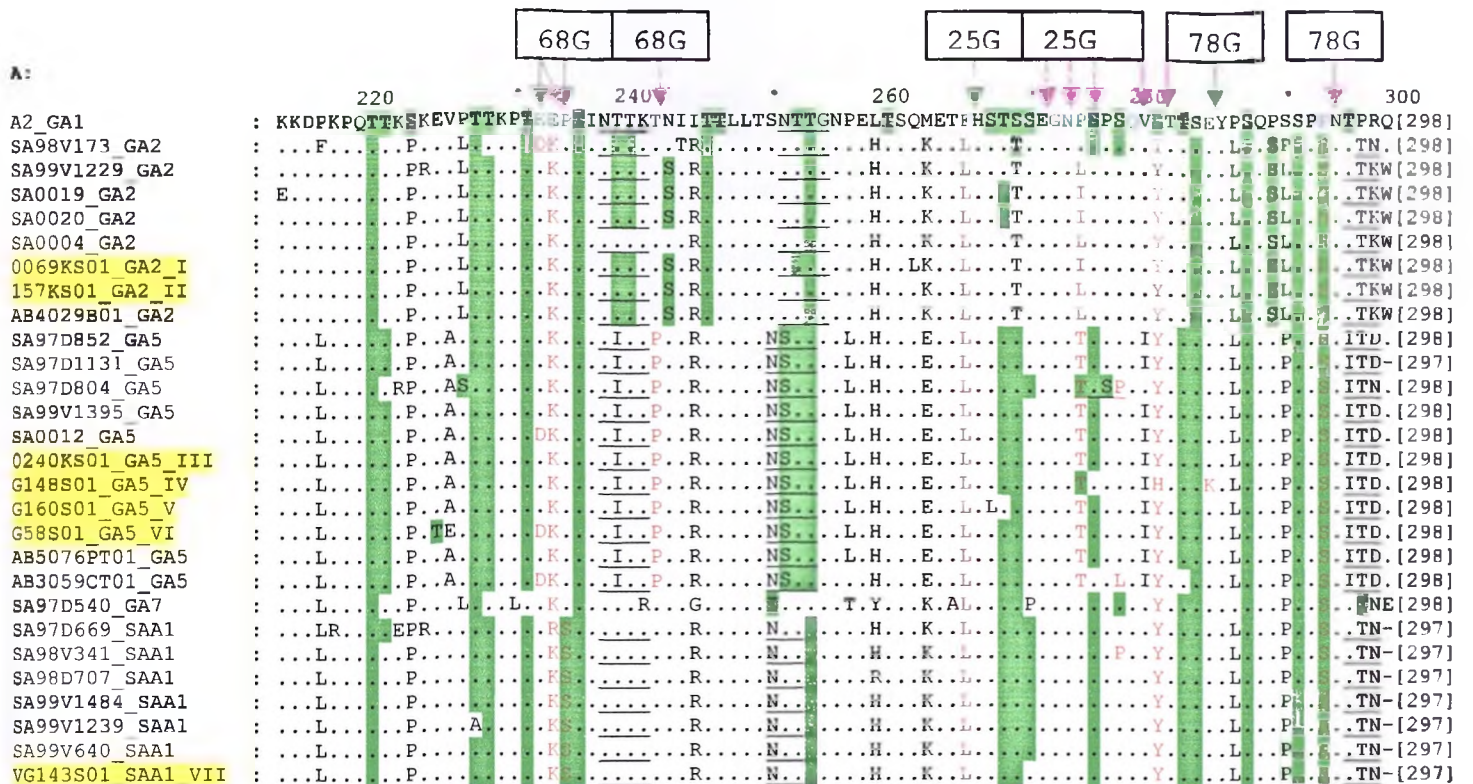
### 3.2.7 G-protein amino acid variation over consecutive epidemics

In Figure 3.6, the Soweto 2001 genotypes and previously identified Soweto genotypes (Venter *et al.*, 2001), are aligned relative to the overall consensus sequence. Genotype-specific amino acid substitutions could be identified within the Soweto 2001 sequences, confirming previous studies (Peret *et al.*, 1998; Venter *et al.*, 2001). To determine if genotype-specific amino acid substitutions might result in changes in the secondary protein structure, the hydrophilic, hydrophobic and special amino acid residues associated with secondary folding (cysteine, glycine and proline) (Lodish *et al.*, 1995), were colour-coded. Various amino acid changes associated with hydrophobic changes within the G-protein of the Soweto 2001 isolates are visible.

Lambert *et al.* (1988) found that RSV infectivity is associated with N- and O-linked sugars. In Figure 3.7, the N- and O-glycosylation sites are indicated for the Soweto G-protein sequences identified in 2001 and previous seasons. Three N-glycosylation sites were identified in the Soweto 2001 subtype A and B sequences. The N-glycosylation positions were conserved within subtype A with the exception of the second site that moved one position forward in SAA1 and GA5. In subtype B, one N-glycosylation site was strain-specific and two were conserved in most subtype B sequences.

The amino acid sequences of both subtypes A and B are threonine rich (residues 220 to 269, for subtype A and residues 217 to 294 for subtype B). It has been suggested that the high serine and threonine content provide potential sites for O-glycosylation and that this may influence antibody recognition of certain variable epitopes (Garcia-Beato *et al.*, 1996). Mucin type O-glycosylated serine and threonine residues were therefore predicted for both subtypes using NetOglyc version 2 (Gupta *et al.*, 1999). In the subtype A and B sequences (including prototype strains), 37 and 35 O-linked sites were identified, respectively (Figure 3.7). In the subtype A sequences, 5 sites were conserved in all sequences and 15 were genotype-specific. In subtype B, 4 sites were conserved in all sequences. Most O-glycosylated sites were strain-specific, but 9 were genotype-specific.





**Figure 3.7 The amino acid alignment of subtype A and B Soweto 2001 sequence groups. Both alignments are relative to the prototype strains, A2 for subtype A and B1 for subtype B. The strain (indicated with roman figures) and the genotype is indicated next to the isolate number. The N-glycosylation sites are underlined in the sequences and were predicted by ProSite. The O-glycosylation sites predicted with NetOGlyc 2.0 are highlighted in green. Differences in amino acid sequence lengths are indicated by semi-colons at the end of each sequence. A) The position and names of strain-specific epitopes are indicated by blue residues and in the blocks above subtype A, respectively. The position of amino acid changes in the South African isolates (indicated by red residues) in strain-specific epitopes are indicated by green arrows and the position of natural occurring strains which fail to react with monoclonal antibodies are indicated with pink arrows. B) Because of a lack of published G-protein epitopes for subtype B, four potential antigenic sites, predicted with Antheprot 2000 version 5.2 (Deleage *et al.*, 1988) (<http://antheprot-pbil.ibcp.fr>) are indicated above the subtype B sequences to show the correspondence to positions of glycosylation.**

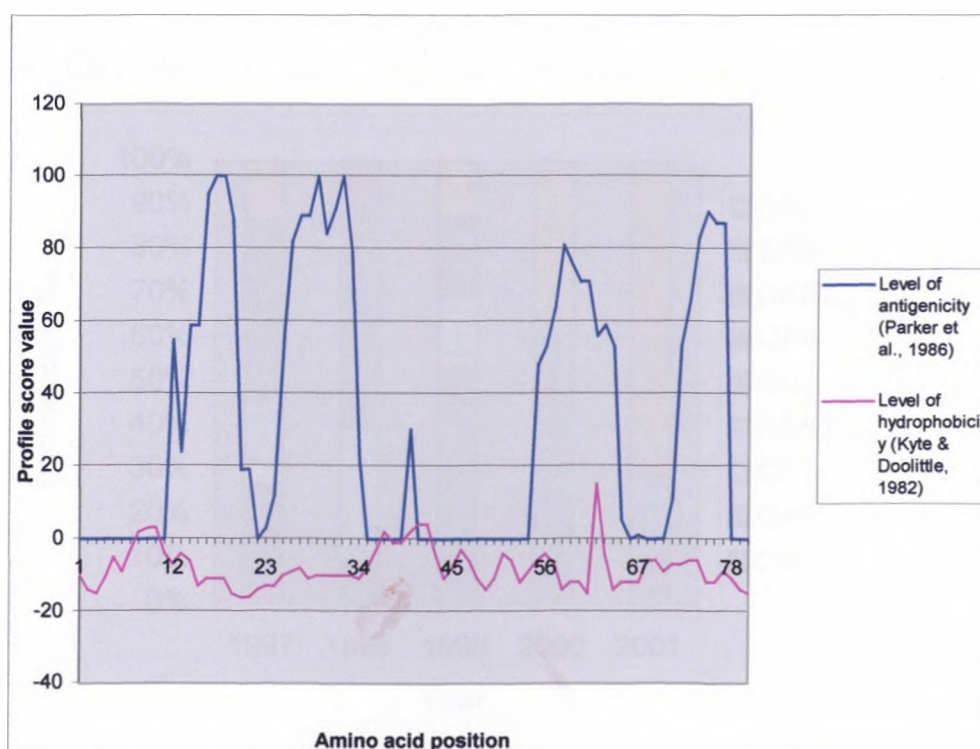
### 3.2.8 Antigenic variation in the G-protein

Strain-specific epitopes have been mapped within the C-terminal hypervariable region of subtype A sequences (Reviewed by Melero *et al.*, 1997). The positions of epitopes 68G, 78G and 25G that were identified from either escape mutants or natural occurring strains that failed to react with G-protein specific monoclonal antibodies raised against the subtype A prototype strain, are indicated in Figure 3.8 (Cane & Pringle, 1995; Garcia *et al.*, 1994). Some of the amino acid changes identified within the Soweto sequences are associated with epitope loss as illustrated with escape mutants and natural occurring strains (summarised in Table 3.7). Two amino acid changes associated with reactivity loss with monoclonal antibody 25G (amino acid position 265) and 78G (position 293), respectively, were conserved in all Soweto sequences.

Since no information was available about amino acid changes associated with epitope loss in subtype B, potential epitopic sites were predicted with Antheprot 2000 version 5.2 (Deleage *et al.*, 1988) (<http://antheprot-pbil.ibcp.fr>). Antigenicity (Parker *et al.*, 1986) and hydrophobicity (Kyte & Doolittle, 1982) plots are shown in Figure 3.8 for the Soweto subtype B sequences. Using these plots, 4 potential antigenic sites could be identified in subtype B (amino acid residues 223 to 232, 234 to 245, 266 to 270 and 276 to 282). Amino acid variations are evident within all four predicted antigenic regions in Figure 3.7. This suggests that potential epitopes residing in these predicted antigenic regions, may also be affected by amino acid variation.

**Table 3.7 Amino acid changes identified in RSV subtype A strain-specific epitopes residing in the second variable region of the G-protein.** The known amino acid mutations associated with epitope loss previously identified in escape mutants (dark gray) and natural occurring strains that fail to react with monoclonal antibodies (light gray) (Garcia *et al.*, 1994; Cane & Pringle, 1995), are indicated.

| Epitope | Amino acid change | Isolate/genotype                                                                      |
|---------|-------------------|---------------------------------------------------------------------------------------|
| 68G     | 232 E - D         | SA98V173 GA2, SA0012 GA5, G58S01 GA5, AB3059CT01 GA5                                  |
|         | 233 E - K/R       | All genotypes / SA97D669 SAA1                                                         |
|         | 234 P - ■         | SAA1                                                                                  |
|         | 244 T - P         | GA5                                                                                   |
| 78G     | 280 S - Y / H     | All genotypes/ G148S01 GA5 IV                                                         |
|         | 284 E - K         | G148S01 GA5 IV                                                                        |
|         | 293 P - S         | All genotypes                                                                         |
| 25G     | 265 F - L         | All genotypes                                                                         |
|         | 274 P - L / I / T | SA99V1229 GA2, SA0004 GA2, 157KS01 GA2 II /A0019 GA2, SA0020 GA2, 0069KS01 GA2 I /GA5 |



**Figure 3.8 Identification of antigenic sites within the subtype B isolates by means of antigenicity and hydrophobicity plots.** Physicochemical profiles were created with Antheprot 2000 version 5.2 (Deléage *et al.*, 1988). The amino acid position is relative to the sequenced fragment of the second variable region of the G-protein (amino acids 835 to 915 of the G-protein).

### 3.2.9 Genotype displacement over 5 consecutive seasons in Soweto

In a study of the genotype circulation over four consecutive RSV seasons at Chris Hani Baragwanath hospital, Venter *et al.* (2001) demonstrated displacement of the dominant genotypes from season to season, which corresponded to findings in other studies (Cane, 2001; Choi & Lee, 2000; Peret *et al.*, 2000). However, during 1999, one genotype became dominant and persisted in the next season as the dominant genotype (GA2). To further clarify the genotype distribution and switching over consecutive seasons in Soweto, the results of the 2001 genotyping data was combined with the previous seasons described by Venter *et al.* (2001). Figure 3.9, illustrates the genotype distribution for subtypes A and B at the Chris Hani Baragwanath hospital from 1997-2001. In 1997 GB3 and GA5 co-dominated, but both genotypes gradually declined until 2001 when only a few specimens of each were found. In 1998 SAA1 and GB3 were the dominant genotypes. GA2 was the dominant genotype during 1999 and 2000, but was clearly replaced by SAB3 in 2001. SAB1 became a prominent genotype in 2001, but was only present at a low frequency in 2000. SAA1 also gradually declined from 1998 when it was dominant until 2001 when it made up only 1% of the specimens.

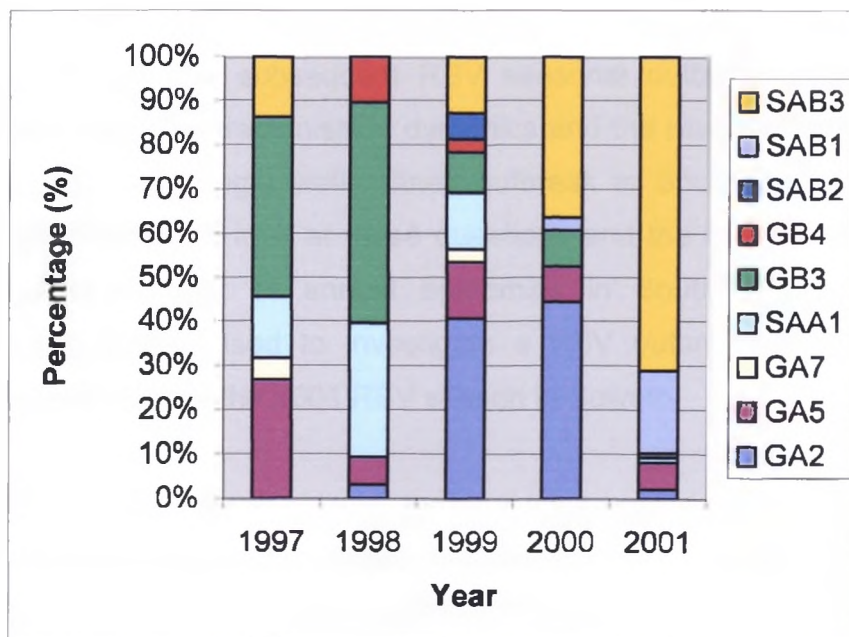


Figure 3.9 Genotype distribution of RSV subtypes A and B during 1997 – 2001 at Chris Hani Baragwanath hospital.

### 3.3 DISCUSSION

Limited data have so far been available about the molecular epidemiology of RSV outbreaks in developing countries, especially Africa (Cane *et al.*, 1999; Roca *et al.*, 2001; Venter *et al.*, 2002; Venter *et al.*, 2001). A previous molecular epidemiological study conducted over four consecutive seasons (1997-2000) in Soweto, South Africa, suggested both subtypes can circulate concurrently, and that switching of the predominant genotypes occur (Venter *et al.*, 2001). The authors showed that in 1997 and 1998, subtypes A and B co-circulated equally while subtype A predominated during 1999 and 2000. In 1999, the GA2 genotype became dominant with 42% of specimens falling in this group and increased even further to 78% in 2000 (Figure 3.9). The authors suggested that it is unusual for one genotype to predominate for more than one season, when considering other studies conducted in the rest of the world. During 2000 this genotype (GA2) was also isolated in different locations across the country, suggesting that highly effective and fit strains may have the ability to spread nationwide (Venter *et al.*, 2002; Venter *et al.*, 2001).

To determine if this was a once-off phenomenon and if herd immunity would eventually overcome the effectiveness of this genotype to reestablish annual epidemics, the subsequent RSV seasonal outbreak (2001) also needed to be analysed. The transmission dynamics and the rate of molecular evolution in the G-protein in a single well-defined outbreak in South Africa has also not yet been determined. To look at these questions and the role of G-protein variability in the reestablishment of annual epidemics in South Africa, intensive phylogenetic analysis was used to investigate a RSV outbreak at Chris Hani Baragwanath hospital during the 2001 RSV season in Soweto.

Sequence analysis of the C-terminal hypervariable region of the G-protein has been reported to provide a reliable representation of the entire G-gene variability (Peret, *et al.*, 1998) and so was used for this analysis. A nested PCR that was developed by Venter *et al.* (2002) was further optimised by using a hotstart step to increase the sensitivity and specificity for subtyping of RSV subtypes. This resulted in higher PCR product yield and was subsequently used as a standard methodology throughout this project.

A total of 82 RSV-positive nasopharyngeal aspirates were obtained from children hospitalised at Chris Hani Baragwanath hospital in Soweto during a RSV outbreak that occurred from March to June 2001. It was shown that both RSV subtypes co-circulated, but that subtype B dominated the outbreak with 89%, suggesting a definite displacement of the dominant subtypes from 2000 to 2001. Subtype A and B circulated equally in 1997, subtype B dominated in 1998, while subtype A predominated in 1999 and 2000 (Venter *et al.*, 2001). These data demonstrate an equally significant role of both subtypes in South African RSV outbreaks. Figure 3.2 shows the distribution pattern of RSV subtypes over five consecutive seasons at the Chris Hani Baragwanath hospital.

It has been suggested that variability in the G-protein sequence may play a role in the reestablishment of annual epidemics (Reviewed by Cane, 2001). To look at this, the genetic variability of the RSV G-protein during the 2001 and previous seasons in the same community was therefore characterised using distance-based (neighbor-joining) and character-based (maximum likelihood) trees. Genotyping methods described by (Peret *et al.*, 2000; Peret *et al.*, 1998; Venter *et al.*, 2002; Venter *et al.*, 2001) were used to assign the Soweto sequences to genotypes. Three subtype A genotypes (GA2, GA5 and SAA1) and three subtype B genotypes (GB3, SAB1 and SAB3) co-circulated during the 2001 Soweto outbreak. However, the subtype A genotype that dominated the previous two seasons (GA2) was clearly replaced by the SAB3 genotype, that dominated the 2001 outbreak with 69% followed by SAB1 with 18%.

The SAB1 genotype has not been prominent during previous seasons; in fact only one SAB1 isolate has been identified in Soweto before. This may indicate that rare strains from one season may become prominent outbreak genotypes in the following season in an immunological naïve population. The appearance and disappearance of genotypes has also been observed in previous seasons in Johannesburg (Venter *et al.*, 2001), for example SAA1 was a prominent genotype in 1998 but gradually declined until only one isolate was identified in 2001. The same could be said about GB3, which was common in 1997 although only one isolate was identified in 2001 (Figure 3.9). All this corresponds with the notion that rare

genotypes detected in some years can reappear and become prominent outbreak genotypes in future epidemics possibly when immunity declines in the population, while other previously dominant outbreak genotypes can disappear altogether (Cane *et al.*, 1994). It has been suggested that genotype switching in each epidemic is driven by herd immunity to the particular dominant strain. This may subsequently restrict the circulation of the particular virus strain (Peret *et al.*, 2000).

Martinez *et al.* (1999;1997) found that AG+GA transitions dominated in subtype A and B sequences. Roca and colleagues found that AG+GA transitions were frequent among subtype A and UC+CU was common in subtype B (Roca *et al.*, 2001). It was found that transitions exceeded transversions with UC+CU transitions dominating AG+GA transitions in both subtype A and B in the Soweto 2001 outbreak. A total of 24 unique nucleotide sequence groups could be identified within the 2001 outbreak, seven for subtype A and 17 for subtype B, demonstrating the high level of genetic variability that may occur during a single confined outbreak. Sequence analysis confirmed the previously described higher variability found for subtype A (Choi & Lee, 2000; Coggins *et al.*, 1998).

Many of the strains present in the outbreak circulated concurrently throughout the epidemic. Most of the identified strains infected only one or two individuals, however the outbreak was clearly dominated by four strains (one SAB1 and three SAB3), which circulated up until the end of the epidemic (March to May for SAB1 strain IX and March to June for SAB3 strains XIII, XIV and XV). This may suggest that these four strains might have been the initial SAB1 and SAB3 outbreak strains, since they were isolated in the beginning of the outbreak and infected many individuals throughout the entire epidemic. The other SAB1 and SAB3 strains might have been a result of possible positive selection in the outbreak strains. It has been suggested that the observed co-circulation of genotypes during an outbreak could be considered to be concurrent separate epidemics taking place at the same time (Cane *et al.*, 1994).

To determine what role positive selection plays in the G-protein variability observed during the 2001 outbreak, the nucleotide and amino acid P-distances were

compared as well as the Ka/Ks ratio (the number of nonsynonymous substitutions per nonsynonymous site (Ka) to the number of synonymous substitutions per synonymous site (Ks)). When comparing the individual genotypes identified in the 2001 outbreak to each other and to the prototype strains, nonsynonymous substitutions exceeded the synonymous substitutions in the 2001 genotypes GA2, GA5, GB3, SAB1 and SAB3 (Ka/Ks>1). When comparing the unique sequence groups identified within the individual genotypes, Ka/Ks ratios of greater than one (>1) was also identified for the unique sequence groups of GA5 and SAB3, providing strong evidence that positive selection occurred in this single outbreak (Hurst, 2002). In both subtypes transitions were also more prevalent than transversions. Sequence variation was also much higher at amino-acid level (mean amino acid P-distance 0.13 and 0.09 for subtype A and B, respectively) than at nucleotide level (mean nucleotide P-distance 0.07 and 0.05 for subtype A and B, respectively). Collectively these findings suggest that positive immune selection within genotypes may account for the emergence of altered strains observed later in each genotype and may contribute to the reestablishment of epidemics.

To further look at antigenic changes that may result from evolution in the G-protein, the amino acid translations of the sequences identified during the 2001 outbreak were analysed. Twenty-two different amino acid sequence groups could be identified, suggesting some nucleotide changes were silent, since 24 nucleotide sequence groups were found. Genotype-specific point mutations could be identified between the early SAB1 (IX) and SAB3 strain (XIII, XIV, XV) and the later strains (X to XII for SAB1, XVI to XXII for SAB3) of the same genotypes. Substitutions were also found among the subtype A strains, especially in the GA5 genotype. Many of these resulted in amino acid changes that may affect the antigenic structure of the G-protein (Figure 3.6). The subtype A isolates showed limited variation in the G-protein length, which corresponds with previous evolutionary studies in subtype A (Garcia *et al.*, 1994; Martinez *et al.*, 1999), while subtype B proteins were more variable in length (Martinez *et al.*, 1999). Protein lengths of 298 (8 sequences) and 297 (1 sequence) amino acids were identified and corresponded to the difference in termination codon usage between the strains of subtype A. Protein lengths of 299 (37 sequences), 295 (21 sequences) and 292 (14 sequences) amino acids were

observed in subtype B. The sequence changes in the termination codons may affect strain-specific epitopes residing in the C-terminal (Rueda *et al.*, 1991; Rueda *et al.*, 1995; Martinez *et al.*, 1997).

Palomo *et al.* (2000) found that carbohydrates in the C-terminal end of the G-protein influence the expression of certain epitopes by either masking it or contributing to antibody recognition. It has been reported that virus infectivity is sensitive to the removal of N- and O-linked sugars, indicating that carbohydrates are needed for the function of the G-protein (Lambert, 1988). High serine and threonine content provide potential sites for glycosylation, and antibody recognition is influenced by these carbohydrates in certain variable epitopes (Garcia-Beato *et al.*, 1996). The N-glycosylation sites in the Soweto 2001 sequences were mostly conserved within subtypes while the O-glycosylation sites were genotype- and sequence-specific. The amino acid sequences of both subtype A and B were found to be threonine rich.

A close correlation between genetic and antigenic variation has been demonstrated in the G-protein, which suggested antigenic drift over time (Cane *et al.*, 1995; Garcia *et al.*, 1994). Amino acid changes that are thought to be associated with loss in monoclonal antibody recognition were presents in some of the Soweto 2001 subtype A sequences. Amino acid variability was also noted in predicted antigenic sites within the subtype B sequences. Alterations in the antigenic structure of the G-protein might provide a way for the virus to effectively evade the immune response. This study illustrated the importance of genetic and antigenic variability within RSV subtypes and its potential role in immune evasion. This knowledge may contribute to the development of effective vaccines and vaccination strategies.

In summary, extensive genetic and antigenic analysis was conducted on the G-protein of RSV clinical specimens from an outbreak at Chris Hani Baragwanath hospital in Soweto during 2001, to further clarify the molecular epidemiology of RSV in South African outbreaks. Subtype co-circulation and switching and displacement of the dominant genotypes of the previous season were found. The dominant outbreak genotype from the previous two seasons (GA2) was replaced by a new genotype in 2001. This may be a result of herd immunity that accumulated in the

community over the previous two seasons. Extensive variability was seen between and within both subtypes, confirming previous molecular studies (Reviewed by Sullender, 2000). Results here also strengthen the argument that rare strains from previous seasons may become prominent outbreak genotypes in future epidemics as demonstrated for SAB1. The high ratio of transitions to transversions, an excess of nonsynonymous substitutions to synonymous substitutions and the fact that amino acid variability exceeded nucleotide variability, during the single outbreak were all indications of positive selection in the G-protein. This illustrates that positive selection occurs at a high rate throughout the epidemic and from one season to the next. Different amino acid lengths were also identified as a result of alternative termination codon usage. Martinez and colleagues have shown that this may affect strain-specific epitopes (Martinez *et al.*, 1997). Changes in antigenic structure of the G-protein might provide a way for the virus to effectively evade the immune response. Demonstration of mutations and glycosylation in specific antigenic and epitopic domains of RSV subtypes A and B further strengthen this argument. This study gives insight about molecular evolution occurring in RSV in a single well-defined outbreak and may contribute to the development of effective vaccines.

## **CHAPTER 4**

### **RESULTS B:**

### **ANTIGENIC VARIABILITY WITHIN THE RSV FUSION PROTEIN DURING ANNUAL EPIDEMICS**

#### **4.1 INTRODUCTION**

The surface glycoproteins of RSV, the attachment (G-protein) and fusion (F-protein), are both targets for the neutralising antibody response (Johnson *et al.*, 1987; Lopez *et al.*, 1988). In Chapter 3 it was shown that positive selection in the variable G-protein may account for the extent of amino acid variability observed within each genotype. It is hypothesised that the resulting changes in important epitopes of this protein may give rise to new outbreak strains in future epidemics. The F-protein which is highly conserved between the subtypes provides cross-protection against both RSV subtypes A and B in experimental animals (Taylor *et al.*, 1984; Walsh *et al.*, 1984). Evidence suggests that the F-protein is the major target for the protective immune response following RSV infection (Lopez *et al.*, 1988). The majority of RSV-specific neutralising monoclonal antibodies are directed to the F-protein (Walsh & Hruska, 1983; Walsh *et al.*, 1985). The F-protein is also a target for RSV-specific CTL (Cherrie *et al.*, 1992; Pemberton *et al.*, 1987). These properties make the F-protein a good candidate for subunit vaccine development. Therefore, an important objective is to determine if positive selection occurs in the F-protein of RSV genotypes identified over consecutive outbreaks in South Africa.

Evidence suggests that the F-protein is highly conserved between subtypes A and B, with an amino acid sequence homology of 91% (Johnson & Collins, 1988; Johnson *et al.*, 1987). The F-protein is synthesised as an N-glycosylated precursor, the F0-protein, and cleaved into two disulfide-linked fragments (F1 and F2) of approximately 50 kDa and 20 kDa, respectively (Walsh & Graham, 1999). The F0-protein is cleaved by a furin-type intracellular protease at cleavage-activation sites (amino acids 131-136 (Collins *et al.*, 1984) and amino acids 106-109 (González-Reyes *et al.*, 2001)), exposing a hydrophobic domain on the F1-subunit, which may be important in membrane fusion, cell tropism and pathogenicity (Collins *et al.*,

1984). A sequence motif suggestive of coiled coil structures is known as a heptad repeat (HR) (Buckland & Wild, 1989; Chambers *et al.*, 1990; Singh *et al.*, 1999). The F-protein contains three heptad repeats predicted to form coiled coils (HR1, HR2 and HR3). Recently, it has been found that the F-protein core is a trimeric coiled coil (Matthews *et al.*, 2000; Zhao *et al.*, 2000).

Escape mutants have been identified with group-specific point mutations in an antigenic site (subtype A) in the F1 chain of the F-molecule (Lopez *et al.*, 1998). Reports suggest that many neutralising epitopes on the F-protein are conformational in nature and require glycosylation and folding in order to be recognised (Beeler & Coelingh, 1989). Multiple antigenic sites have been identified on the F-protein (Anderson *et al.*, 1985), which includes certain conserved linear neutralisation epitopes related to the fusion function of the F-protein (Beeler & Coelingh, 1989; Martin-Gallardo *et al.*, 1991; Walsh *et al.*, 1986b) as well as a conformational neutralising epitope (Lopez *et al.*, 1990). HLA Class I-restricted CTL epitopes have also been identified in humans (Brandenburg *et al.*, 2000; Rock & Crowe, 2003).

In a study conducted on RSV in Birmingham, the nucleotide and amino acid sequences of the F genes of isolates were found to be similar to published subtype A viruses from diverse geographical locations. No sequence changes could be identified in any known antibody epitopes. The authors found that the F2 subunit was more variable than the F1 subunit, which may be because of the greater number of potential glycosylation sites in this subunit (Plows & Pringle, 1995).

The RSV F-protein is highly conserved and not thought to undergo positive selection (Zheng *et al.*, 1999). Evidence suggests that positive selection does not occur in the CTL epitopes of the highly conserved RSV nucleoprotein (N-protein) (Venter *et al.*, 2003). However, recent findings in influenza virus have suggested that immune selection does occur in the CTL epitopes of the highly conserved influenza nucleoprotein (Boon *et al.*, 2002; Price *et al.*, 2000; Voeten *et al.*, 2000). This implies that regular updating will be needed in vaccines containing these CTL epitopes. The emphasis that is placed on the F-protein in RSV vaccine development

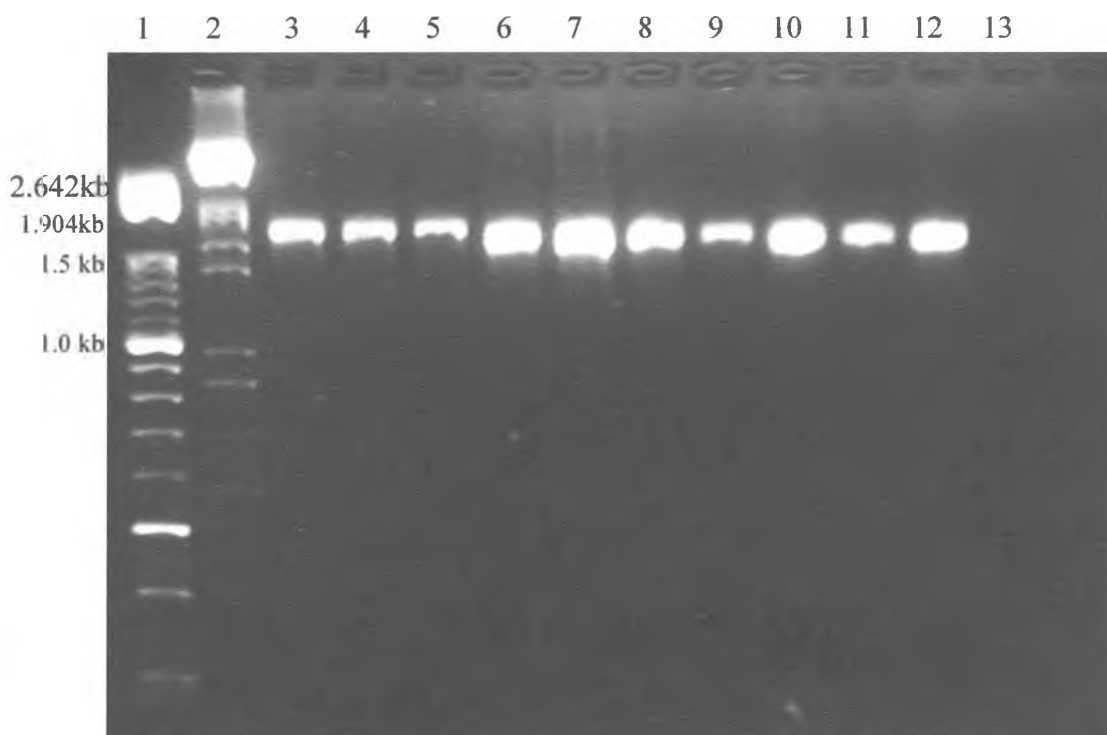
---

makes it important to determine if immune selection occurs in both the neutralising antibodies and CTL epitopes of different RSV genotypes.

Therefore in this Chapter the F-protein of representatives of each genotype previously identified in South Africa, was analysed to determine the extent of nucleotide and amino acid variation between genotypes and to determine if positive selection occurs within the known antibody and CTL epitopes. To further look at the extent of conservation in the F-protein, the F-protein sequences were compared to other members of the *Pneumovirinae* subfamily. Since conformational epitopes exist in the F-protein, the secondary structure was analysed to determine the influence of amino acid variation on important epitopes and the antigenic structure of the F-protein. This is the first study that investigates the F-protein of South Africa isolates.

### 4.2.1 Amplification of the F-protein

Clinical specimens (n=18) were selected from each of the subtype A and B genotypes identified in South Africa to date (Chapter 2, Table 2.1). The mRNA of the F-protein is approximately 1899 nucleotides long and encodes a protein of 574 amino acids (Collins *et al.*, 1984). A RT-PCR using subtype-specific primers was developed (Chapter 2, Table 2.3) for the amplification of the entire precursor F-protein (1 – 1729 bp) and resulted in PCR products of 1821 bp and 1760 bp for subtype A and B, respectively (Figure 4.1). The PCR was optimised for maximum product yield and specificity and could amplify all RSV genotypes.



**Figure 4.1 RT-PCR products of RSV fusion protein for both subtypes.** The RT-PCR resulted in products of 1821 and 1760 bp for subtypes A and B, respectively. Lane 1: Molecular weight size marker (100bp)(Roche, Mannheim, Germany); Lane 2: Molecular weight size marker (0.12-21.2 kbp) (Roche, Mannheim, Germany); Lane 3: Specimen VR173 (GA2); Lane 4: Specimen 0128K (GA5); Lane 5: Specimen D540 (GA7); Lane 6: Specimen D707 (SAA1); Lane 7: Specimen 0258K (SAB1); Lane 8: Specimen VR192 (SAB3); Lane 9: Specimen D1656 (SAB4); Lane 10: Specimen D941 (GB3); Lane 11: specimen D1107 (GB4); Lane 12: Positive control (subtype A); Lane 13: Negative control.

---

## 4.2.2 Sequence analysis

Sequencing primers were designed for the F-protein of subtypes A and B, respectively. These primers were useful in sequencing the F-protein genes of all the genotypes identified in South African to date. Primers were designed from the F-protein alignments of the prototype A2 (Collins *et al.*, 1984), subtype A Long (Lopez *et al.*, 1988), B1 (Karron *et al.*, 1997) and subtype B 18537 (Johnson & Collins, 1988) sequences obtained from the GenBank.

### 4.2.2.1 Nucleotide analysis

The nucleotide pairwise distance (P-distance) of the sequenced F-protein of RSV ranged from 0.02 to 0.03 between the South African subtype A genotypes and from 0.00 to 0.02 between the subtype B genotypes. The South African subtype A and B genotypes had 82% identity (P-distance 0.18). In Table 4.1, the P-distances between the F-proteins of individual specimens are indicated and ranged from 0.00 to 0.18. The overall P-distance within subtype A and B was 0.02 and 0.01, respectively, suggesting that the F-protein sequences were more variable for subtype A than for subtype B.

Transitions exceeded transversions (transition:transversion = 7.59 and 4.96 for subtypes A and B, respectively) and the synonymous substitutions per synonymous site (Ks) exceeded the nonsynonymous substitutions per nonsynonymous site for the F-protein. The overall Ka/Ks ratio was 0.07 for all subtype A genotypes and 0.09 for all subtype B genotypes, while the combined Ka/Ks ratio for subtype A and B was 0.12 (Table 4.2). This suggested an absence of positive selection within the F-protein gene sequences identified in South Africa over 5 consecutive seasons and relative to the historic prototype strains isolated 42 years ago (1961).

**Table 4.1 The nucleotide pairwise distance of the F-proteins of all the South African genotypes.**

| SPECIMEN<br>GENOTYPE | A2<br>GA1 | VR173<br>GA2 | 157K<br>GA2 | D1131<br>GA5 | 128K<br>GA5 | VR360<br>GA7 | D540<br>GA7 | B1<br>GB1 | D941<br>GB3 | D796<br>GB3 | VR468<br>GB4 | D1107<br>GB4 | VR341<br>SAA1 | D707<br>SAA1 | 00256K<br>SAB1 | 0258K<br>SAB1 | V800<br>SAB2 | VR192<br>SAB3 | 0072K<br>SAB3 | D1656<br>SAB4 |
|----------------------|-----------|--------------|-------------|--------------|-------------|--------------|-------------|-----------|-------------|-------------|--------------|--------------|---------------|--------------|----------------|---------------|--------------|---------------|---------------|---------------|
| [A2GA1]              |           |              |             |              |             |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [VR173GA2]           | 0.04      |              |             |              |             |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [157KGA2]            | 0.04      | 0.01         |             |              |             |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [D1131GA5]           | 0.03      | 0.02         | 0.03        |              |             |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [128KGA5]            | 0.04      | 0.02         | 0.03        | 0.00         |             |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [VR360GA7]           | 0.04      | 0.02         | 0.03        | 0.03         | 0.03        |              |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [D540GA7]            | 0.04      | 0.02         | 0.03        | 0.02         | 0.02        | 0.00         |             |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [B1GB1]              | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        |           |             |             |              |              |               |              |                |               |              |               |               |               |
| [D941GB3]            | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      |             |             |              |              |               |              |                |               |              |               |               |               |
| [D796GB3]            | 0.18      | 0.17         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.00        |             |              |              |               |              |                |               |              |               |               |               |
| [VR468GB4]           | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.00        | 0.01        |              |              |               |              |                |               |              |               |               |               |
| [D1107GB4]           | 0.18      | 0.18         | 0.18        | 0.18         | 0.18        | 0.17         | 0.17        | 0.01      | 0.01        | 0.01        | 0.01         |              |               |              |                |               |              |               |               |               |
| [VR341SAA1]          | 0.04      | 0.03         | 0.03        | 0.02         | 0.03        | 0.03         | 0.03        | 0.17      | 0.17        | 0.17        | 0.17         | 0.18         |               |              |                |               |              |               |               |               |
| [D707SAA1]           | 0.04      | 0.02         | 0.03        | 0.02         | 0.02        | 0.03         | 0.02        | 0.17      | 0.17        | 0.17        | 0.17         | 0.18         | 0.01          |              |                |               |              |               |               |               |
| [0025KSAB1]          | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.01        | 0.02        | 0.02         | 0.01         | 0.18          | 0.18         |                |               |              |               |               |               |
| [0258KSAB1]          | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.01        | 0.01        | 0.02         | 0.01         | 0.18          | 0.17         | 0.00           |               |              |               |               |               |
| [V800SAB2]           | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.01        | 0.01        | 0.01         | 0.02         | 0.17          | 0.17         | 0.02           | 0.02          |              |               |               |               |
| [VR192SAB3]          | 0.18      | 0.17         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.01      | 0.00        | 0.01        | 0.01         | 0.01         | 0.17          | 0.17         | 0.01           | 0.01          | 0.01         |               |               |               |
| [0072KSAB3]          | 0.18      | 0.18         | 0.18        | 0.17         | 0.17        | 0.17         | 0.17        | 0.02      | 0.00        | 0.01        | 0.01         | 0.01         | 0.17          | 0.17         | 0.01           | 0.01          | 0.01         | 0.01          | 0.00          |               |
| [D1656SAB4]          | 0.18      | 0.18         | 0.18        | 0.18         | 0.18        | 0.17         | 0.17        | 0.02      | 0.02        | 0.02        | 0.02         | 0.01         | 0.18          | 0.18         | 0.01           | 0.01          | 0.02         | 0.01          | 0.02          |               |

**Table 4.2 Ratio of nonsynonymous substitutions per nonsynonymous site (Ka) to synonymous substitutions per synonymous site (Ks) for the F-proteins of subtype A (A), subtype B (B) and all the South African genotypes combined, relative to the prototype strains A2 and B1 for subtype A and B, respectively (A+B).**

**A:**

| KA/KS | A2    | GA2   | GA5   | GA7   | SAA1 |
|-------|-------|-------|-------|-------|------|
| A2    |       |       |       |       |      |
| GA2   | 0.079 |       |       |       |      |
| GA5   | 0.140 | 0.037 |       |       |      |
| GA7   | 0.100 | 0.013 | 0.071 |       |      |
| SAA1  | 0.128 | 0.035 | 0.079 | 0.048 |      |

**B:**

| KA/KS | B1    | GB3   | GB4   | SAB1  | SAB2  | SAB3  | SAB4 |
|-------|-------|-------|-------|-------|-------|-------|------|
| B1    |       |       |       |       |       |       |      |
| GB3   | 0.095 |       |       |       |       |       |      |
| GB4   | 0.122 | 0.120 |       |       |       |       |      |
| SAB1  | 0.122 | 0.049 | 0.103 |       |       |       |      |
| SAB2  | 0.094 | 0.095 | 0.114 | 0.060 |       |       |      |
| SAB3  | 0.098 | 0.083 | 0.077 | 0.050 | 0.091 |       |      |
| SAB4  | 0.146 | 0.093 | 0.125 | 0.150 | 0.096 | 0.072 |      |

**A+B:**

| KA/KS | A2    | GA2   | GA5   | GA7   | SAA1  | B1    | GB3   | GB4   | SAB1  | SAB2  | SAB3  | SAB4 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|
| A2    |       |       |       |       |       |       |       |       |       |       |       |      |
| GA2   | 0.079 |       |       |       |       |       |       |       |       |       |       |      |
| GA5   | 0.140 | 0.037 |       |       |       |       |       |       |       |       |       |      |
| GA7   | 0.100 | 0.013 | 0.071 |       |       |       |       |       |       |       |       |      |
| SAA1  | 0.128 | 0.035 | 0.079 | 0.048 |       |       |       |       |       |       |       |      |
| B1    | 0.126 | 0.112 | 0.123 | 0.116 | 0.121 |       |       |       |       |       |       |      |
| GB3   | 0.131 | 0.116 | 0.129 | 0.120 | 0.126 | 0.095 |       |       |       |       |       |      |
| GB4   | 0.123 | 0.111 | 0.120 | 0.115 | 0.119 | 0.122 | 0.120 |       |       |       |       |      |
| SAB1  | 0.128 | 0.114 | 0.126 | 0.118 | 0.122 | 0.122 | 0.049 | 0.103 |       |       |       |      |
| SAB2  | 0.131 | 0.118 | 0.130 | 0.122 | 0.128 | 0.094 | 0.095 | 0.114 | 0.060 |       |       |      |
| SAB3  | 0.128 | 0.113 | 0.126 | 0.119 | 0.125 | 0.098 | 0.083 | 0.077 | 0.050 | 0.091 |       |      |
| SAB4  | 0.126 | 0.119 | 0.129 | 0.123 | 0.126 | 0.146 | 0.093 | 0.125 | 0.150 | 0.096 | 0.072 |      |

#### 4.2.2.2 RSV F-protein amino acid analysis

The amino acid P-distances within the South African subtype A and B genotypes, ranged from 0.00 to 0.01. When comparing the South African F-protein sequences from both subtypes to the prototype strains (strain A2 and B1 for subtype A and B, respectively), the intergenotypic amino acid P-distances ranged from 0.00 to 0.11 (Table 4.3). Subtype A and B shared 90% F-protein amino acid identity (P-distance=0.1). Sequence variation in the South African F-protein sequences was therefore higher at nucleotide level (P=0.18) than at an amino acid level (P=0.1), suggesting most mutations were silent.

**Table 4.3 Intergenotypic amino acid pairwise distance of the F-protein of the South African specimens relative to the prototype strains.**

|      | A2   | GA2  | GA5  | GA7  | SAA1 | B1   | GB3  | GB4  | SAB1 | SAB2 | SAB3 | SAB4 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| A2   |      |      |      |      |      |      |      |      |      |      |      |      |
| GA2  | 0.02 |      |      |      |      |      |      |      |      |      |      |      |
| GA5  | 0.03 | 0.01 |      |      |      |      |      |      |      |      |      |      |
| GA7  | 0.02 | 0.00 | 0.01 |      |      |      |      |      |      |      |      |      |
| SAA1 | 0.03 | 0.01 | 0.01 | 0.01 |      |      |      |      |      |      |      |      |
| B1   | 0.10 | 0.09 | 0.09 | 0.09 | 0.09 |      |      |      |      |      |      |      |
| GB3  | 0.10 | 0.09 | 0.10 | 0.09 | 0.09 | 0.01 |      |      |      |      |      |      |
| GB4  | 0.10 | 0.09 | 0.10 | 0.09 | 0.09 | 0.01 | 0.01 |      |      |      |      |      |
| SAB1 | 0.11 | 0.09 | 0.10 | 0.09 | 0.09 | 0.01 | 0.01 | 0.01 |      |      |      |      |
| SAB2 | 0.11 | 0.09 | 0.10 | 0.09 | 0.09 | 0.01 | 0.01 | 0.01 | 0.10 |      |      |      |
| SAB3 | 0.10 | 0.09 | 0.10 | 0.09 | 0.09 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 |      |      |
| SAB4 | 0.10 | 0.10 | 0.10 | 0.10 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |      |

#### 4.2.2.3 Influence of Amino acid variability on the structural features of the F-protein

Figure 4.2 shows the F-protein amino acid alignment of the South African genotypes and different members of the *Pneumovirinae* subfamily, and indicates their common structural features. The alignment shows that the signal domain differs between the human RSV subtypes and the existence of subtype-specific point mutations across the protein. A few genotype-specific mutations also occurred among the South African isolates relative to the prototype strains. Amino acid changes resulting in amino acids of different biochemical classification groups, which might affect the secondary structure of the F-protein, were identified. In all the subtype A genotypes, amino acid changes resulting in different chemical classes relative to the prototype

strain (A2), were noted at residues 25 (G→S), 102 (P→A), 103 (T→A) and 122 (A→T). In subtype B an amino acid change resulting in a different chemical class relative to the prototype strain (B1) was also identified at residue 529 (T→A).

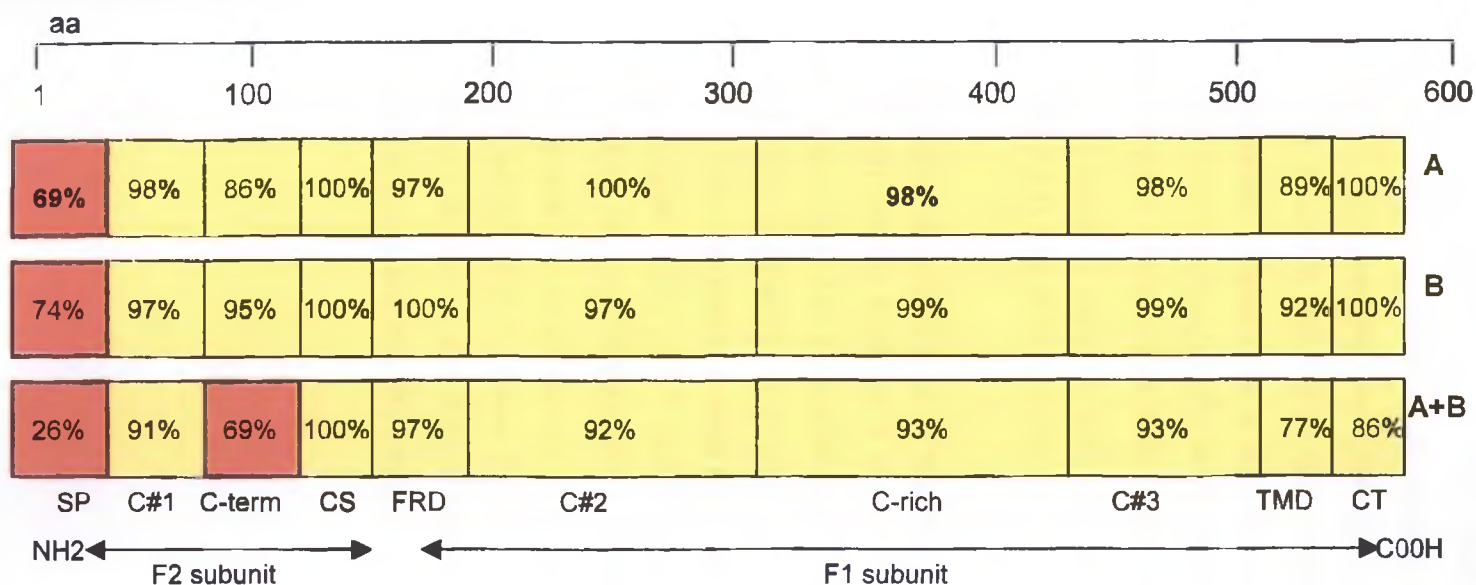
Most glycine (G) residues, which may destabilise  $\alpha$ -helical and  $\beta$ -sheet conformations, were conserved within the South African genotypes and the prototype strains with the exception of one at position 25 (G→S). These residues were found at intervals on the F-protein. All the cysteine residue positions were conserved between human, ovine and bovine RSV. It has been shown that glycosylation of the F0-protein of RSV, is inhibited completely in the presence of tunicamycin, suggesting that carbohydrate side chains are exclusively attached through N-linkage to asparagines residues (Huang *et al.*, 1984; Huang Wertz, 1983). All of the carbohydrate side chains predicted for the F-protein of the South African genotypes using ScanProsite (<http://ca.expasy.org/cgi-bin/scanprosite>) were attached through N-linkage to asparagine (N) amino acid residues. Four to five, N-linked glycosylation sites were identified on the F2-subunit of each isolate and one on the F1-subunit (ScanProsite (<http://ca.expasy.org/cgi-bin/scanprosite>)). One conserved mucin-type O-glycosylated serine residue was predicted with NetOglyc version 2 (Gupta *et al.*, 1999) for subtype A and B (position 405). A putative O-glycosylated threonine was identified in the sequence of specimen D707SAA1 (position 13). The F1-subunit terminus (F-L-G) at amino acid position 137-140, was conserved between human, bovine and ovine RSV.

---

**Figure 4.2 Amino acid alignment of the F-protein sequences of the South African genotypes with bovine and ovine RSV (see Figure on following page).** The structural features common to Pneumovirus fusion proteins are indicated above the sequence (Chambers *et al.*, 1992). The N-linked glycosylation sites predicted with ScanProsite (<http://ca.expasy.org/cgi-bin/scanprosite>), are indicated with an arrow and the cysteine residues are shown with a green circle symbol (●). The O-linked glycosylation site is indicated with a blue square symbol (■). The cleavage sites are boxed in black and the F1-subunit terminus (F-L-G) is boxed in red.



In Figure 4.3, the amino acid sequence identity within the two F-protein subunits (F1 and F2) are indicated for RSV subtypes A and B. The F2 subunit of the F-protein consists of a signal domain, a conserved region, the C-terminal and two cleavage sites (Chambers *et al.*, 1992; Collins *et al.*, 1984; González-Reyes *et al.*, 2001). When both RSV subtypes were taken into consideration, the signal domain (amino acid residues 1-23) was the most variable region within the F-protein, with an overall sequence identity of 26%. The conserved region (residues 24 – 87) was 91% identical, while the C-terminal (residues 88 – 130) was 69% identical. The cleavage site at amino acid position 131 to 136 was 100% conserved between subtypes, while the cleavage site within the C-terminal (residues 106 – 109) was 84% conserved. The F1-subunit consists of the fusion related domain (residues 137 – 166), a conserved region (residues 167-306), a cysteine rich domain (residues 307 - 144), a second conserved region (residues 442-524), a transmembrane domain (residues 525 –551) and a cytoplasmic tail (residues 552 – 574) (Chambers *et al.*, 1992). The amount of sequence conservation for the F1-subunit was 97% (fusion related domain), 92% (first conserved region), 93% (cysteine rich domain), 93% (second conserved region), 77% (transmembrane domain) and 86% (cytoplasmic tail), respectively. This suggests that the F1-subunit is more conserved than the F2 subunit.



**Figure 4.3 Amino acid identity within the different RSV fusion protein domains. The percentages of identical amino acid residues are indicated within each region. A) Amino acid identities within RSV subgroup A relative to the prototype strain, A2. B) Amino acid identity within RSV subgroup B relative to the prototype strain, B1. A+B) Percentage of identical sites within both RSV subgroups. Regions with identity higher than 75%, are indicated in yellow; while regions with an identity less than or equal to 75% are indicated in red. The scale indicates the amino acid positions. The signal peptide (SP), conserved region (C#1), C-terminal (C-term), cleavage site (CS), fusion related domain (FRD), conserved region (C#2), cysteine-rich region (C-rich), conserved region (C#3), transmembrane domain (TMD) and the cytoplasmic tail (CT) (Chambers *et al.*, 1992, Plows & Pringle, 1995), as well as the F-protein subunits, are indicated below the graph.**

### **4.2.3 Prediction of the secondary structure of the F-protein from the amino acid alignments**

#### **4.2.3.1 Conformation prediction: $\alpha$ -helices, $\beta$ -strands and turns**

The secondary structure ( $\alpha$ -helices,  $\beta$ -strands and turns) was predicted for each South African F-protein amino acid sequence to identify changes in their conformational status that may result from the limited amount of variation. In general, the conformation prediction within the subtype A genotypes was very similar, but differences were observed relative to the A2 prototype strain (Figure 4.4 A). The structure prediction was identical for GA2 and GA7. In GA5, an  $\alpha$ -helix conformation changed to a  $\beta$ -strand (residue 547) as a result of the amino acid change from leucine (L) (A2 prototype) to phenylalanine (F) (GA5). In D707SAA1,  $\alpha$ -helices were predicted at amino acid 344 that were absent in other genotypes. This was the result of aspartic acid (D) (A2 prototype) that changed to glutamic acid (E) (D707SAA1) that did not occur in the other genotypes.

Only a few conformational changes were noted between the B1 prototype strain and the South African subtype B sequences (Figure 4.4 B). Conformation variation was especially evident in the signal peptide domain (residues 1-23). This is probably due to the high frequency of amino acid variation observed within this region (Figure 4.2). In SAB1,  $\beta$ -strands were predicted at residue 312 relative to turn predictions in B1 prototype strain, which corresponded to the amino acid change from proline (P) to histidine (H). In VR192SAB3,  $\beta$ -strands were predicted at the amino acid change from alanine (A) to valine (V) (residue 39) while  $\alpha$ -helices were found in the other genotypes that did not have this mutation. In D1107GB4, an  $\alpha$ -helix was predicted relative to a turn in the B1-prototype, in the vicinity of the amino acid change of threonine (T) to alanine (A) (residue 518).



VR341SAA1:

**B:**

0258KSAB1:

V800SAB2:

0072KSAB3:

VR192SAB3:

```
Secondary structure prediction ( $H$  =  $\alpha$ -helix,  $E$  =  $\beta$ -strand, - = turns)
--EEHHHNNHHNHHHHHHHH-----HHHHHH[60]
-----HHHHHHHHHHHHHHHHHHHHHH-----H-HNH-----H-[120]
-----HHHH-EE-----EHHHHHHHHHH-NHNH-NHNH-----EEN[180]
-----E-H-----EH-----EHHHHHHHH-NHNH[240]
-----HHHHH-NHNH-----HH-NHNHHHHHHHH-EEEEHHHHHHHHHH[300]
-E--E-----EE[360]
--H-----E-----EEE-----EEEE[420]
-----EE-----E-----HHN-----E-----EEE[480]
-----HH-----HHHHHHHHHHHH-N-H-----EEEEEEEEEEEEEEEE[540]
EEEEENEHH[573]
```

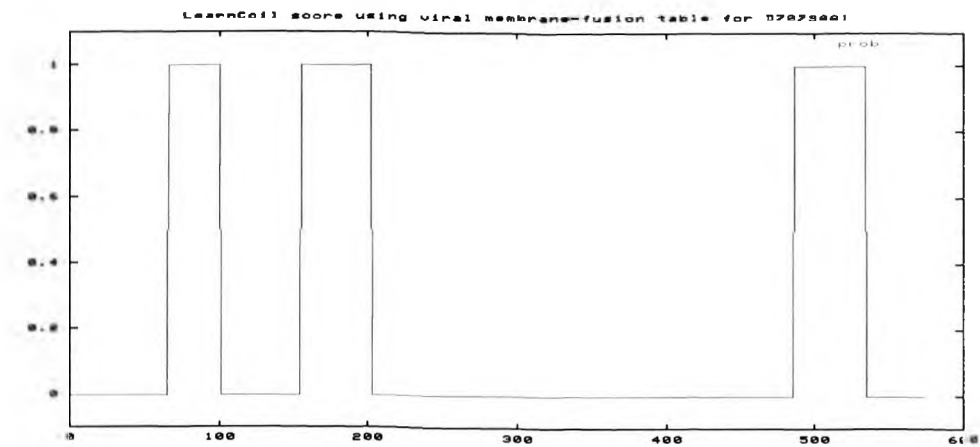
turns)

#### 4.2.3.2 Coiled coil prediction

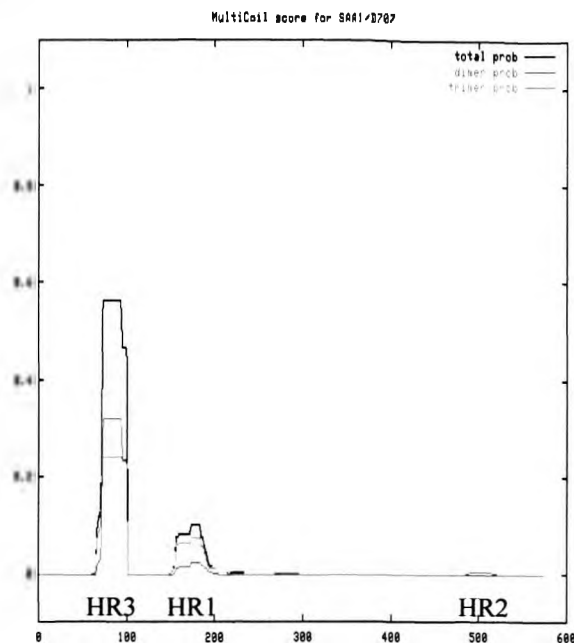
Previous studies suggested that the F-protein is a trimeric coiled coil based on analysis of the prototype strains A2, Long (subtype A) and CH-18537 (subtype B) (Lambert *et al.*, 1996; Matthews *et al.*, 2000). Heptad repeat (HR) regions are sequence motifs suggestive of coiled coil structures (Buckland & Wild, 1989; Chambers *et al.*, 1990; Singh *et al.*, 1999). To determine if the amino acid variation observed in the F-protein (of the South African genotypes relative to the prototype strains) influence the secondary structure of the F-protein, the coiled coil regions were predicted.

Three HR regions with probabilities of 100% were identified in the F-protein using the LearnCoil-VMF prediction algorithm (Berger & Singh, 1997; Singh *et al.*, 1999) (Figure 4.5). One HR was situated in the F2-subunit C-terminal (residues 66 to 100) (HR3), while two were located in the F1-subunit, one next to the fusion peptide (residues 155 to 202) (HR1) and one next to the transmembrane region (residues 486-534) (HR2). HR1 was predicted to be trimeric (probability = 10%), HR2 was trimeric (probability < 10%), while there were strong indications that HR3 was dimeric in subtype A (probability = 30%) and trimeric in subtype B (probability = 20%), using MultiCoil (Wolf *et al.*, 1997). Matthews and colleagues found that HR1, HR2 and HR3 were trimeric, dimeric/trimeric and dimeric, respectively, for the CH-18537 strain (Matthews *et al.*, 2000). Three amino acid changes that were conserved within each subtype were noted in HR3 among the South African strains. These changes were located at amino acid position 67 (asparagine (N) for subtype A and threonine (T) for subtype B), position 74 (alanine (A) for subtype A and threonine (T) for subtype B) and position 99 (serine (S) for subtype A and asparagine (N) for subtype B). These results suggested that there might be a slight difference within the F-protein coiled coil structure between subtypes A and B, probably due to subtype-specific amino acid variation.

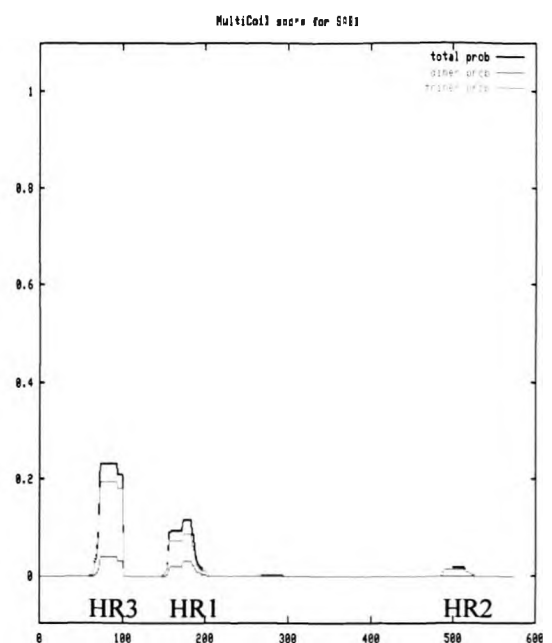
**A:**



**B:**



**C:**

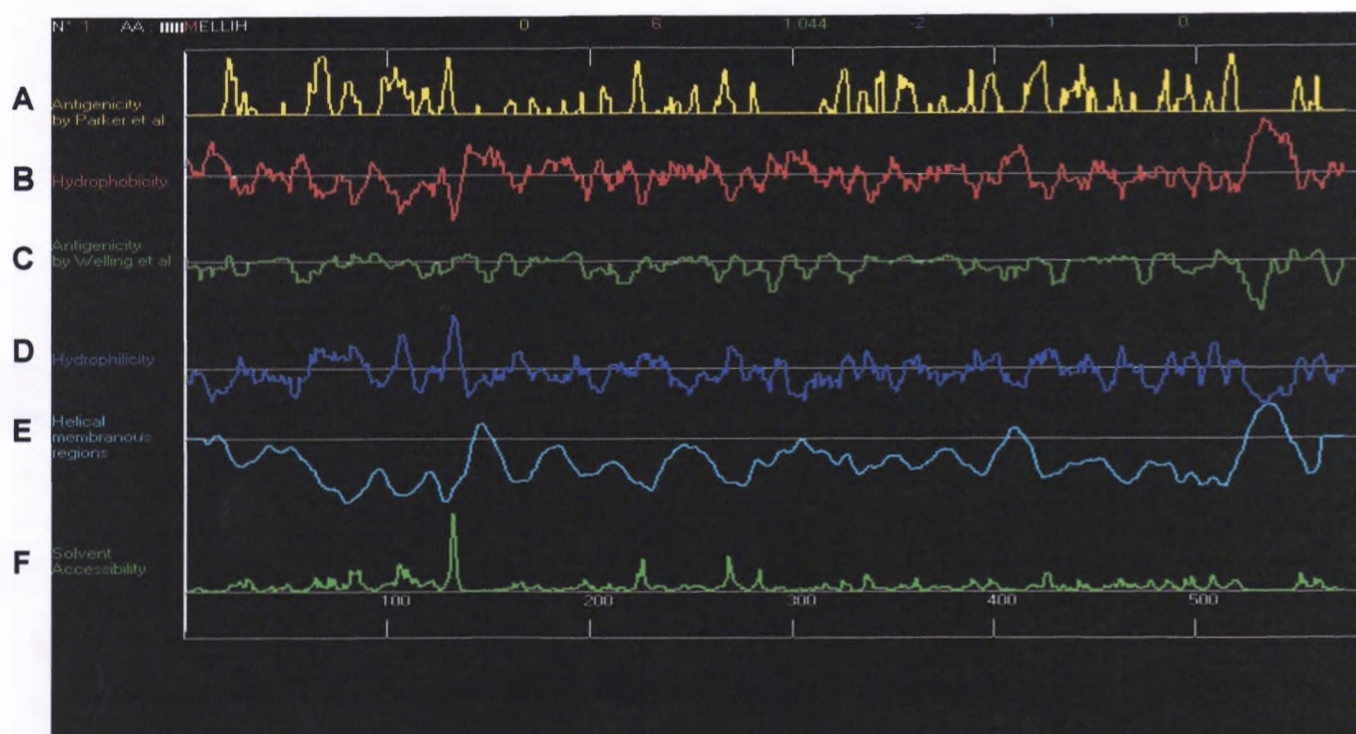


**Figure 4.5 Prediction of coiled coils in the secondary structure of the fusion protein.** A) The LearnCoil program (Singh *et al.*, 1999) was used to predict possible heptad repeat regions. The x-axis shows the amino acid residues and the y-axis indicates the probability of a coiled region. B) To predict the probability that the fusion amino acid sequence forms a trimeric or a dimeric coiled coil, the MultiCoil program (Wolf *et al.*, 1997) was used. The x-axis shows the residue location and the y-axis represents the probability that the residues form a coiled coil (a probability of 1 equals 100%). The probability of a dimeric coiled coil is shown in blue, while the trimeric probability is shown in red. The total coiled coil probabilities, (the sum of the dimeric and trimeric probability), is plotted in black. The three heptad repeat regions (HR) are indicated below the graph. All the predictions were identical within subtypes A and B, thus Figure A represents the analysis for both subtypes, while B and C represents the analysis of genotypes SAA1, and SAB1, respectively.

#### 4.2.4 Prediction of antigenic sites

The positions of possible antigenic sites within the F-protein were predicted with Antheprot 2000 version 5.2 (Deleage *et al.*, 1988) (Figure 4.6) and by the hydrophilic/hydrophobic status of the individual amino acid residues (Figure 4.7). In Figure 4.6, multiple antigenic (Parker *et al.*, 1986), hydrophilic (Kyte & Doolittle, 1982) and solvent accessibility peaks (Boger *et al.*, 1986) are distributed evenly throughout the F-protein alignment. Two hydrophilic peaks were slightly higher and more distinct, namely the two peaks between amino acids 97 and 137. However, all these regions suggested the presence of multiple antigenic sites in these regions. These regions were also shown to be hydrophilic in the alignment in Figure 4.7.

In Figure 4.7, only a few amino acid changes within the South African genotypes relative to the prototype strains resulted in hydropathic changes.



**Figure 4.6 Prediction of possible antigenic sites within the fusion protein by means of physicochemical profiles.** The different physicochemical profiles were created with Antheprot 2000 version 5.2 (Deleage *et al.*, 1988) with a window size of 7 amino acid residues. These included predictions of the antigenicity (Parker *et al.*, 1986, Welling *et al.*, 1985), hydrophobicity (Kyte & Doolittle, 1982), hydrophilicity (Hopp and Woods, 1981), transmembrane regions (Von Heijne, 1992) and solvent accessibility (Boger *et al.*, 1986). Each method is indicated next to the graph. A: antigenicity by Parker *et al.* (1986); B: hydrophobicity; C antigenicity by Welling *et al.* (1985); D: hydrophilicity; E: helical membranous regions; F: Solvent accessibility.



**Figure 4.7 Hydropathic profiles per amino acid residue (see previous page).** The profiles were created with Genedoc version 2.6001 ([www.psc.edu/biomed/genedoc](http://www.psc.edu/biomed/genedoc)). The amino acid alignment is relative to the A2 prototype strain (Collins *et al.*, 1984). The hydrophobic residues (L, I, V, M, A) are indicated in red, hydrophilic residues (D, E, H, K, R, N, Q, S, T) in blue and special amino acids involved in secondary folding (C, G, P) are indicated in yellow. The specimen name and specific genotype are shown on the left of the sequences. Structural features common to *Pneumoviruses* are indicated above the sequence.

## 4.2.5 Epitope analysis

### 4.2.5.1 Prediction of CTL epitopes

CTL algorithms (BIMAS CTL algorithms (BIMAS (Parker *et al.*, 1994) and SYFPEITHI (Rammensee *et al.*, 1999)) were used to predict possible CTL epitope positions within the South African F-protein sequences. Only the HLA types shown to be dominant in at least 10% of the different ethnic groups of South Africa were used in the predictions (Venter *et al.*, 2003; Hammond *et al.*, 1997; personal communication, Dr A Puren, NICD). In Table 4.4, the predicted epitopes with the highest score-results for each algorithm are shown for the different HLA types. Many of the predicted epitopes were located in overlapping regions. In Figure 4.8, the epitopes predicted with the BIMAS algorithm are plotted on the F-protein amino acid alignment of the South African and prototype strains. Six of these predicted epitopic sites were conserved between RSV subtypes, and the changes noted in the other predicted areas were mostly subtype-specific. A few genotype-specific amino acid changes were also identified within the amino acid sequence of the predicted epitopes: for GA7 at position 129, SAB1 at position 312, SAB2 at position 290 and SAB4 at position 312.

**Table 4.4 Possible CTL epitopes predicted by SYFPEITHI and BIMAS algorithms for the South African HLA types (See Table on the following page).** Only the top scoring nine-mer peptides are included in this Table for the different South African HLA types (A good score result in CTL prediction for SYFPEITHI ranges between 15 and 30 (Rammensee *et al.*, 1999), while a good score in BIMAS ranges from 50 (Parker *et al.*, 1994). Only HLA types were included in the predictions that were present in at least 10% of the major ethnic groups of South Africa (Venter *et al.*, 2003), personal communication, Dr A Puren, NICD).

Table 4.4. See previous page for legend

| HLA TYPE | ETHNIC GROUP       | ALGORITHM | SCORE RESULT | START POSITION | SEQUENCE   |
|----------|--------------------|-----------|--------------|----------------|------------|
| B*3501   | Caucasian          | BIMAS     | 60           | 111            | LPRFMNYTL  |
| Cw*0401  | African/ Caucasian | BIMAS     | 240          | 249            | TYMLTINSEL |
| A*68:1   | African            | BIMAS     | 120          | 126            | NVTLSKKRR  |
| A*68:1   | African            | BIMAS     | 120          | 413            | IVSCYGTKK  |
| A*68:1   | African            | BIMAS     | 100          | 127            | VTLSSKKRR  |
| A*68:1   | African            | BIMAS     | 60           | 57             | ITIELSNIK  |
| B*07     | Caucasian          | BIMAS     | 800          | 111            | LPRFMNYTL  |
| B*07     | Caucasian          | BIMAS     | 120          | 551            | KARSTFVTL  |
| A*68:1   | African            | BIMAS     | 180          | 218            | ETVIEFQOK  |
| B*3501   | Caucasian          | BIMAS     | 180          | 388            | NPKYDCKIM  |
| A*68:1   | African            | BIMAS     | 240          | 543            | AVGLLLYCK  |
| B*4403   | Caucasian          | BIMAS     | 480          | 109            | RELPRFMNY  |
| B*5801   | African            | BIMAS     | 88           | 129            | LSKKRRRRF  |
| B*5801   | African            | BIMAS     | 79           | 106            | RARRELPRF  |
| Cw*0301  | Caucasian          | BIMAS     | 125          | 301            | VQLFLYGV   |
| Cw*0301  | Caucasian          | BIMAS     | 50           | 373            | LTLPSEINL  |
| Cw*0401  | African/ Caucasian | BIMAS     | 440          | 85             | KYKNAVTEL  |
| A*01     | Caucasian          | BIMAS     | 113          | 470            | KGEPIINFY  |
| B*08     | Caucasian          | BIMAS     | 80           | 164            | VNKKISALL  |
| A*0201   | African/ Caucasian | BIMAS     | 81           | 273            | LMSNNVQIV  |
| A*01     | Caucasian          | BIMAS     | 50           | 383            | NVDIFNPKY  |
| A*0201   | African/ Caucasian | BIMAS     | 267          | 191            | KVLDLKNYI  |
| A*0201   | African/ Caucasian | BIMAS     | 257          | 170            | ALLSTNKAV  |
| A*0201   | African/ Caucasian | BIMAS     | 178          | 140            | FLLGVGSAI  |
| A*0201   | African/ Caucasian | BIMAS     | 138          | 536            | VILLSLIAV  |
| A*0201   | African/ Caucasian | BIMAS     | 135          | 272            | KLMSNNVQI  |
| A*0201   | African/ Caucasian | BIMAS     | 129          | 250            | YMLTINSELI |
| A*68:1   | African            | BIMAS     | 150          | 356            | ETCKVQSNR  |
| A*0201   | African/ Caucasian | BIMAS     | 84           | 187            | VLTSKVLDL  |
| Cw*0401  | African/ Caucasian | BIMAS     | 200          | 222            | EFQOKNNRL  |
| A*0201   | African/ Caucasian | BIMAS     | 81           | 525            | IMITTIIV   |
| A*0201   | African/ Caucasian | BIMAS     | 77           | 451            | SVGNLTYYV  |
| A*0201   | African/ Caucasian | BIMAS     | 64           | 498            | KINCSLAFI  |
| A*0201   | African/ Caucasian | BIMAS     | 57           | 114            | FMNYTLNNT  |
| A*0205   | African/ Caucasian | BIMAS     | 54           | 191            | KVLDLKNYI  |
| A*03     | Caucasian          | BIMAS     | 180          | 77             | KLIKQELDK  |
| A*68:1   | African            | BIMAS     | 240          | 307            | GVIDTPCWK  |
| A*0201   | African/ Caucasian | BIMAS     | 124          | 394            | KIMTSKTDV  |
| Cw*0401  | African/ Caucasian | BIMAS     | 200          | 504            | AFIRKSDEL  |
| Cw*0401  | African/ Caucasian | BIMAS     | 88           | 111            | LPRFMNYTL  |
| Cw*0401  | African/ Caucasian | BIMAS     | 75           | 487            | EFDASISQV  |
| Cw*0401  | African/ Caucasian | BIMAS     | 50           | 298            | AYVVQLPLY  |
| Cw*0401  | African/ Caucasian | BIMAS     | 50           | 136            | RFLGFLLGV  |
| Cw*0401  | African/ Caucasian | BIMAS     | 380          | 365            | VFCDTMNSI  |
| A*01     | Caucasian          | SYFPEITHI | 25           | 541            | LIAVGLLLY  |
| A*0201   | African/ Caucasian | SYFPEITHI | 26           | 187            | VLTSKVLDL  |
| A*0201   | African/ Caucasian | SYFPEITHI | 26           | 525            | IMITTIIV   |
| A*0201   | African/ Caucasian | SYFPEITHI | 26           | 536            | VILLSLIAV  |
| A*0201   | African/ Caucasian | SYFPEITHI | 25           | 173            | STNKAVVSL  |
| A*0201   | African/ Caucasian | SYFPEITHI | 25           | 538            | LLSLIAVGL  |
| A*0201   | African/ Caucasian | SYFPEITHI | 25           | 540            | SLIAVGLLL  |
| A*01     | Caucasian          | SYFPEITHI | 29           | 470            | KGEPIINFY  |
| A*01     | Caucasian          | SYFPEITHI | 28           | 383            | NIDIFNPKY  |
| A*0201   | African/ Caucasian | SYFPEITHI | 26           | 159            | HLEGEVNKI  |
| A*01     | Caucasian          | SYFPEITHI | 26           | 90             | VTELOLLMQ  |
| B*08     | Caucasian          | SYFPEITHI | 25           | 466            | SLYVKGEPI  |
| B*0702   | Caucasian          | SYFPEITHI | 22           | 111            | LPRFMNYTL  |
| A*01     | Caucasian          | SYFPEITHI | 27           | 450            | VSVGNLTYY  |
| A*0201   | African/ Caucasian | SYFPEITHI | 26           | 10             | AITAILAAV  |
| A*0201   | African/ Caucasian | SYFPEITHI | 28           | 533            | VIVILLSL   |
| A*0201   | African/ Caucasian | SYFPEITHI | 28           | 180            | SLSNGVSVL  |
| B*4402   | Caucasian          | SYFPEITHI | 26           | 109            | RELPRFMNY  |
| B*08     | Caucasian          | SYFPEITHI | 25           | 73             | DAKVKLIKQ  |
| B*08     | Caucasian          | SYFPEITHI | 26           | 152            | VAVSKVLHL  |
| B*08     | Caucasian          | SYFPEITHI | 26           | 121            | NAKKTNVTI  |
| B*08     | Caucasian          | SYFPEITHI | 28           | 187            | VLTSKVLDL  |
| B*08     | Caucasian          | SYFPEITHI | 31           | 164            | VNKKISALL  |
| B*08     | Caucasian          | SYFPEITHI | 32           | 130            | SKKKRRRRFI |
| A*03     | Caucasian          | SYFPEITHI | 25           | 512            | LLHNVNAGK  |
| A*03     | Caucasian          | SYFPEITHI | 26           | 307            | GVIDTPCWK  |
| A*03     | Caucasian          | SYFPEITHI | 29           | 413            | IVSCYGTKK  |
| A*03     | Caucasian          | SYFPEITHI | 29           | 77             | KLIKQELDK  |
| A*0201   | African/ Caucasian | SYFPEITHI | 28           | 530            | IIIVIIIVL  |

#### 4.2.5.2 *Published CTL epitopes*

The F-protein CTL epitopes identified in humans to date include HLA\*B57 and Cw\*12-restricted epitopes (amino acids 106-114, 542-550, respectively) (Brandenburg *et al.*, 2000) and HLA-A\*01-restricted (amino acids 109-119) (Rock & Crowe, 2003). Possible HLA-restricted CTL epitopes tested by Rock and Crowe (2003) for which the restriction has not been determined, include amino acids 8-18, 93-103, 260-270, 273-283, 285-295, 374-384, 388-398, 519-520, 521-531 (Rock & Crowe, 2003). These are indicated in pink in Figure 4.8. A H-2K-restricted CTL epitope (amino acid 92 –106) and two H-2-restricted T-helper cell epitopes identified in mice (amino acids 205-225, 255-278) (Corvaisier *et al.*, 1997; Corvaisier *et al.*, 1993) were also included and are indicated by yellow.

The RSV F-protein specific HLA B\*57-restricted (Brandenburg *et al.*, 2000), and HLA-A\*01-restricted (Rock & Crowe, 2003) CTL epitopes, were conserved for subtype A, but differed in at least 3 positions between subtype A and B. A single amino acid change was also noted at position 113 (glutamine (Q) changed to histidine (H)) in one GB3 isolate (GB3SA98D941) relative to the other subtype B sequences. This mutation resulted in amino acids of different biochemical classification groups. A genotype-specific amino acid mutation was identified in subtype A sequences of the GA5 genotype in the HLA Cw\*12-restricted CTL epitope (Brandenburg *et al.*, 2000) that changed a leucine (L) to phenylalanine (F) (GA5) at position 547 relative to the prototype strain. A single subtype-specific mutation is also visible in all subtype B isolates in position 544 that changed a valine (V) in subtype A to an isoleucine (I) in subtype B. Most of the putative CTL epitopes identified by Rock and Crowe (Rock & Crowe, 2003) were conserved within subtype A with single amino acid changes in all subtype B isolates with the exception of two that were located at amino acids 8 to 18 and 93 to 103, that differed between the South African sequences and the prototype strains.

The H-2K-restricted CTL epitope identified in BALB/c mice (Jiang *et al.*, 2002) also contained conserved subtype-specific changes (amino acid 92 –106). Both an antibody epitope and a T-helper cell epitope have been identified between amino acids, 205 to 225 in BALB/c mice (Corvaisier *et al.*, 1997). Subtype-specific amino acid changes are visible at position 209 (K→Q) and 213 (S→R) between subtype A and B, which may affect recognition of these epitopes. The T-helper cell epitope at position 255-278 (Corvaisier *et al.*, 1993) was conserved between both subtypes.



**Figure 4.8 The position of CTL and neutralising epitopes on the South African F-protein sequences (see previous page).** Published neutralising epitopes (7C2, RS-348, 47F, L4) are shown in green (Trudel *et al.*, 1987; Lopez *et al.*, 1990; Martin-Gallardo *et al.*, 1991; Corvaisier *et al.*, 1993) and HLA B\*57, Cw\*12 and HLA-A\*01-restricted CTL epitopes are shown in blue (Brandenburg *et al.*, 2000; Rock & Crowe, 2003). T-cell helper epitopes are indicated in yellow (Corvaisier *et al.*, 1993; Corvaisier *et al.*, 1997) and possible CTL epitopes are indicated in pink (Rock & Crowe, 2003). The epitope identification name is shown in bold between hurdle bars. Amino acid positions identified in escape mutants are shown in red. Potential CTL epitope sites predicted with BIMAS are boxed and the HLA-restrictions are indicated above the sequence (see Table 4.4 for more detail). The isolate name and genotype are shown at the left of the sequences.

#### 4.2.5.3 *Amino acid sequence analysis within known neutralising epitopes*

To further look at evolution in epitopic domains, the position of published neutralising epitopes were plotted on the South African F-protein alignment (Figure 4.8). The following F-protein specific neutralising epitopes were included, 7C2 (amino acids 221 to 236) (Trudel *et al.*, 1987), 47F (amino acids 262 to 268) (Lopez *et al.*, 1990), L4 (amino acids 289 to 298) (Martin-Gallardo *et al.*, 1991) and RS-348 (amino acids 205 to 225) (Corvaisier *et al.*, 1993).

The analysis shows that RS-348 and 7C2 were conserved within the RSV subtypes and 47F and L4 were highly conserved between and within both subtypes. In L4, a single amino acid change from isoleucine (I) to valine (V) was identified in SAB2 at position 291. However, these amino acids have the same biochemical classification group, and may not necessarily alter epitope recognition.

#### 4.2.5.4 *Escape mutants*

In Figure 4.8, the amino acid point mutations previously identified in escape mutants (Arbiza *et al.*, 1992, Lopez *et al.*, 1998) are plotted. However none of the South African F-protein sequences were changed in these positions relative to the prototype strains.

### 4.3 DISCUSSION

In this Chapter, subtype-specific primers were designed for the amplification of the precursor F-protein by RT-PCR. The F-protein of each G-protein genotype identified in South Africa to date, was amplified, sequenced and compared to the prototype strains A2 (Collins *et al.*, 1984) and B1 (Karron *et al.*, 1997) for subtypes A and B, respectively.

Evidence from studies conducted on isolates from the rest of the world, suggested that the F-protein is highly conserved between subtypes A and B, with an amino acid sequence homology of 91% (Johnson & Collins, 1988; Johnson *et al.*, 1987). Because the F-protein is the focus of prophylactic antibody and vaccine research, it is important to establish the extent of conservation between the various genotypes identified in South Africa to date and to determine if immune selection occurs in the antibody and CTL epitopes. In general, the F-protein was found to be highly conserved within the South African genotypes. Subtypes A and B shared 82% nucleotide identity within the F-protein of the South African genotypes. More nucleotide variation was present within the subtype A genotypes ( $P=0.02$ ) than the subtype B genotypes ( $P=0.01$ ). For the F-protein, transitions were more prevalent than transversions and in contrast to the G-protein, synonymous substitutions (Ks) exceeded nonsynonymous substitutions (Ka), with an overall Ka/Ks ratio of 0.12 for both subtypes. This suggested an absence of positive selection within this antigenic protein.

Most nucleotide changes resulted in silent mutations and the amino acid sequence identity was 90% between RSV subtypes A and B, which corresponded with previous reports (Johnson *et al.*, 1987). The higher level of nucleotide variation compared to the amino acid variation in the South African F-proteins, suggested that silent mutations were prevalent within this protein. The proportion of nucleotide substitutions resulting in amino acid changes among the F-protein sequences was 10% and 3% for subtypes A and B, respectively.

Phylogenetic analysis between members of the subfamily, *Pneumovirinae*, showed a close relationship between human and ovine or bovine RSV, with 18-19% amino acid differences between the the F-proteins, of the different host restricted RSVs

(Figure 4.9). The phylogenetic analysis combined with the distance data of the F-protein, showed a close relationship between human, ovine and bovine RSV (bootstrapped values of 81 – 100%). The avian and human metapneumoviruses clustered together with a bootstrap value of 100%. Mice pneumovirus grouped with the metapneumoviruses. When looking at structural features common to *Pneumovirinae* fusion proteins (Chambers *et al.*, 1992, Plows & Pringle, 1995), it was apparent that the signal domain (residues 1 – 23) was the most variable (26% identity) between subtypes A and B and the fusion related domain was the most conserved (97% identity), confirming previous reports (Chambers *et al.*, 1992, Plows & Pringle, 1995). The F1-subunit also proved to be more conserved than the F2-subunit in this analysis. Plows and Pringle (1995) suggested that the variability in the F2-subunit may be because of the greater number of potential glycosylation sites within this subunit.

Gruber and Levine (1983) had shown that the inactive precursor F-protein is glycosylated and proteolytically processed to generate the two subunits (F1 and F2) which are linked by sulphide bridges (Gruber & Levine, 1983). Rixon *et al.* (2002) showed that multiple glycosylated forms of the RSV F-protein are expressed in virus infected cells. The biological significance of this is unclear, although the authors suggested that these subtle differences could affect some aspects of the F-protein function (Rixon *et al.*, 2002). When the South African amino acid sequences were compared to the A2 (Collins *et al.*, 1984) and B1 (Karron *et al.*, 1997) prototype strains, subtype-specific mutations and a few intergenotypic mutations were observed across the F-protein. In the South African subtype A and B genotypes, amino acid changes occurred relative to the prototype strains (A2 and B1, respectively). These changes at residues 25 (G→S), 102 (P→A), 103 (T→A) and 122 (A→T) in subtype A and at residue 529 (T→A) in subtype B, might potentially affect the secondary structure since the mutations resulted in amino acids with different biochemical classification groups. Similar amino acid changes were also reported in the F-proteins of isolates from South Birmingham (Plows & Pringle, 1995). In the South African genotypes, all the cysteine residues positions, as well as the six N-glycosylation sites (one in the F1-subunit and five in F2-subunit) were conserved within the subtypes. The F1 terminus was conserved between both subtypes.

A few amino acid changes resulted in conformational changes ( $\alpha$ -helices,  $\beta$ -strands and turns) within the F-proteins of South African genotypes. The conformation predictions in the signal domain varied between the South African genotypes, probably due to the high level of amino acid variation present in this region. The amino acid mutations in position 344 of specimen D707SAA1 (D $\rightarrow$ E) and in position 518 of specimen D1107GB4 (A $\rightarrow$ T), resulted in putative predicted changes from turns to  $\alpha$ -helices relative to the prototype strain. The mutations in position 547 of specimens 0128KGA5 and D1131GA5 (L $\rightarrow$ F), and in position 39 of specimen VR192SAB3 (A $\rightarrow$ V) resulted in changes from an  $\alpha$ -helix to a  $\beta$ -strand relative to the prototype strains. In SAB1, the substitution of (P $\rightarrow$ H) in position 312 relative to the prototype resulted in a change from a turn to a  $\beta$ -strand. These changes in conformation of the F-protein correlated with changes observed in the hydropathic profiles. These results suggested that the minor amino acid changes observed between the F-proteins of the different genotypes may result in subtle changes in the secondary structure of the protein.

Fusion proteins of various viruses have trimeric structures that are centered around a trimeric coiled coil of three HR regions. These include gp41 of the simian immunodeficiency virus (SIV) (Caffrey *et al.*, 1998) and human immunodeficiency virus (HIV) (Chan *et al.*, 1997), hemagglutinin (HA) from influenza A (Bullough *et al.*, 1994), F from simian virus 5 (SV5) (Baker *et al.*, 1999), etc. HR regions are sequence motifs suggestive of coiled coil structures (Buckland & Wild, 1989; Chambers *et al.*, 1990; Singh *et al.*, 1999) that can associate with lipids and might assist in the fusion mechanism (Peisajovich *et al.*, 2000). In these proteins the HR2 region interacts in an antiparallel manner with the HR1 region. Previous studies also suggested that the F-protein of RSV is a trimeric coiled coil for the prototype strains A2, Long and CH-18537 (Lambert *et al.*, 1996; Matthews *et al.*, 2000; Zhao *et al.*, 2000). To determine if the amino acid variation observed in the South African F-protein sequences relative to the prototype strains, influence the structure of the F-protein, the position of coiled coil regions were predicted with MULTICOIL (Wolf *et al.*, 1997). The position of the HR regions were also characterised within the South African F-protein sequences. Consistent with findings from previous structural studies for RSV, one HR region was identified in the F2-subunit (HR3) and two in the F1-subunit (HR1 and HR2) (Matthews *et al.*, 2000; Zhao *et al.*, 2000). The HR1

was predicted to be trimeric and contained a significant amount of  $\alpha$ -helices, which corresponds with reports that the core of the F-protein is a trimeric coiled coil (Matthews *et al.*, 2000; Zhao *et al.*, 2000). The HR2 was dimeric in both subtypes, but HR3 was strongly dimeric in subtype A and trimeric in subtype B, confirming previous reports (Matthews *et al.*, 2000). This might be due to subtype-specific amino acid variation within this region. According to Matthews *et al.* (2000), even though the HR3 region does not contribute to the core of the protein it may still have a structural function, possibly through interaction with non-HR regions (Matthews *et al.*, 2000). These results may suggest the existence of slight structural differences in the F-proteins of subtypes A and B.

The F-protein is the major target for the RSV protective immune response (Lopez *et al.*, 1988) and provides cross-protection against both RSV subtypes A and B in experimental animals (Taylor *et al.*, 1984; Walsh *et al.*, 1984). It is also an important target for cytotoxic T lymphocytes (Cherrie *et al.*, 1992). These properties make this protein an important target for vaccine development. Various potential antigenic sites that may contain possible B-cell epitopes, could be predicted within the South African F-protein sequences (Figure 4.6). The predicted areas had high scores in antigenicity and solvent accessibility prediction plots, and low levels of hydrophobicity (Boger *et al.*, 1986; Kyte & Doolittle, 1982; Parker *et al.*, 1986). Antigenic sites predicted between amino acids 200 and 300, correspond with the locations of known B-cell epitopes (47F epitope (Lopez *et al.*, 1990), L4 epitope (Martin-Gallardo *et al.*, 1991), RS-348 epitope (Corvaisier *et al.*, 1993) and 7C2 epitope (Trudel *et al.*, 1987)). Evidence suggests that the F1-subunit is more antigenic than the F2-subunit (Plows & Pringle, 1995).

Previous reports suggested that a single amino acid substitution can influence the neutralisation response of foot-and-mouth disease virus (Hernandez *et al.*, 1992) hepatitis B virus (Howard, 1995) and human T-cell leukaemia virus type 1 (Blanchard *et al.*, 1999), or the CTL response in HIV (Takahashi *et al.*, 1989), lymphocytic choriomeningitis virus (Salvato *et al.*, 1991) and influenza virus (Terajima *et al.*, 1999). It is therefore important to investigate amino acid changes within the F-protein epitopes of RSV. To investigate immunological pressure on specific epitopes, the position of F-protein cytotoxic T-cell epitopes were first

predicted, using CTL algorithms (BIMAS and SYFPEITHI (Parker *et al.*, 1994; Rammensee *et al.*, 1999)) to identify the position of potential nine-mer CTL epitopes within known South African HLA types (Table 4.4) (Venter *et al.*, 2003; Hammond *et al.*, 1997, personal communication, Dr A Puren, NICD). The predicted CTL epitopes identified with the BIMAS algorithm are shown in Figure 4.8. Many of the predicted nine-mers overlapped and most of these predicted epitopes were located in the highly conserved F1 subunit.

Previously identified CTL epitopes were also plotted on the South African F-protein sequence alignment in Figure 4.8. Three CTL epitopes previously identified in subtype A (restricted to HLA B\*57, HLA Cw\*12 (Brandenburg *et al.*, 2000) and HLA-A\*01 (Rock & Crowe, 2003)) were mostly conserved for subtypes A although subtype-specific changes were visible in subtype B. A genotype-specific amino acid substitution was identified at position 547 of the GA5 genotype of the HLA Cw\*12-restricted CTL epitope which is located at an anchor residue of this epitope. The substitution of a phenylalanine (L→F), which contains an aromatic ring, in this anchor position may influence the binding of this CTL epitope by the MHC class I molecule. At least 3 amino acid changes are visible in subtype B relative to subtype A in the HLA B\*57, and HLA-A\*01 restricted epitopes. A point mutation that resulted in amino acids with different hydrophobic character, was also identified in the HLA B\*57- and HLA-A\*01-restricted epitopes at position 113 (Q→H) in subtype B. The HLA Cw\*12-restricted CTL epitope contains a subtype-specific change in all subtype B strains that may also influence binding of this epitope to the MHC class I molecule. These results suggest that the CTL directed against subtype A-specific epitopes may not necessarily recognize subtype B strains. The GA5 genotype-specific amino acid change in the Cw\*12 restricted epitope may also reduce the efficiency of recognition of this epitope by CTL specific to other subtype A genotypes and prevents the complete exclusion of the possibility of immune selection within this CTL epitope. Most of the putative CTL epitopes identified by Rock and Crowe (Rock & Crowe, 2003) were conserved within subtype A, but had subtype-specific changes in subtype B. The H-2K-restricted CTL epitope identified in BALB/c mice (amino acid 92 –106) (Jiang *et al.*, 2002) was also only conserved for subtype A with subtype-specific changes in subtype B. These findings were also reflected in the putative CTL epitopes predicted for common South African HLA

types, suggesting that reinfection with different subtypes may evade memory CTL to certain F-protein specific CTL epitopes.

These results suggests that CTL epitopes that have sofar been mapped using subtype A-specific peptides may not be recognised as efficiently in subtype B viruses. This should be taken into consideration when these epitopes are used to measure the CTL response during vaccine trials. Future F-protein CTL mapping studies should also be directed towards the subtype B viruses or attempt to identify peptides that are conserved between the subtypes.

When looking at previously identified B-cell epitopes, the conformational 47F epitope (Lopez *et al.*, 1990) and the linear L4 epitope (Martin-Gallardo *et al.*, 1991) were highly conserved between both subtypes. A genotype-specific amino acid substitution (I→V) is visible in the L4 epitope for SAB2 (Figure 4.8), however this mutation resulted in an amino acid of the same biochemical classification group and may not necessarily interfere with epitope recognition. Since the L4 and especially the 47F epitope was highly conserved between subtypes, these epitopes may be useful for the development of synthetic mimotope vaccines or the development of prophylactic monoclonal antibodies. The RS-348 epitope (Corvaisier *et al.*, 1993) and linear 7C2 epitope (Trudel *et al.*, 1987)) were conserved within the subtype A genotypes. Two amino acid substitutions are visible in the RS-348 epitope of all subtype B genotypes at position 209 (K→Q) and at position 213 (S→R). These mutations resulted in amino acids of different biochemical classification groups. Both substitutions at position 209 and 213, may therefore alter the antibody binding site of epitope RS-348 which induces a subtype B-specific neutralising antibody response (Corvaisier *et al.*, 1997). This illustrates one of the potential risks of a subunit vaccine against RSV that does not include both subtypes. Connor *et al.* (2001) found that subtype-specific antibody responses may be due to differences between subtype A and B F-protein sequences at residues 200, 201, 209 and 213 that are clustered within a 16 amino acid loop (200 –216 aa), and that at least two mutations are necessary for antibody escape (Connor *et al.*, 2001). In support of this finding, when examining the South African F-protein sequence conformational predictions, the amino acid change in the F-protein sequence between subtypes A

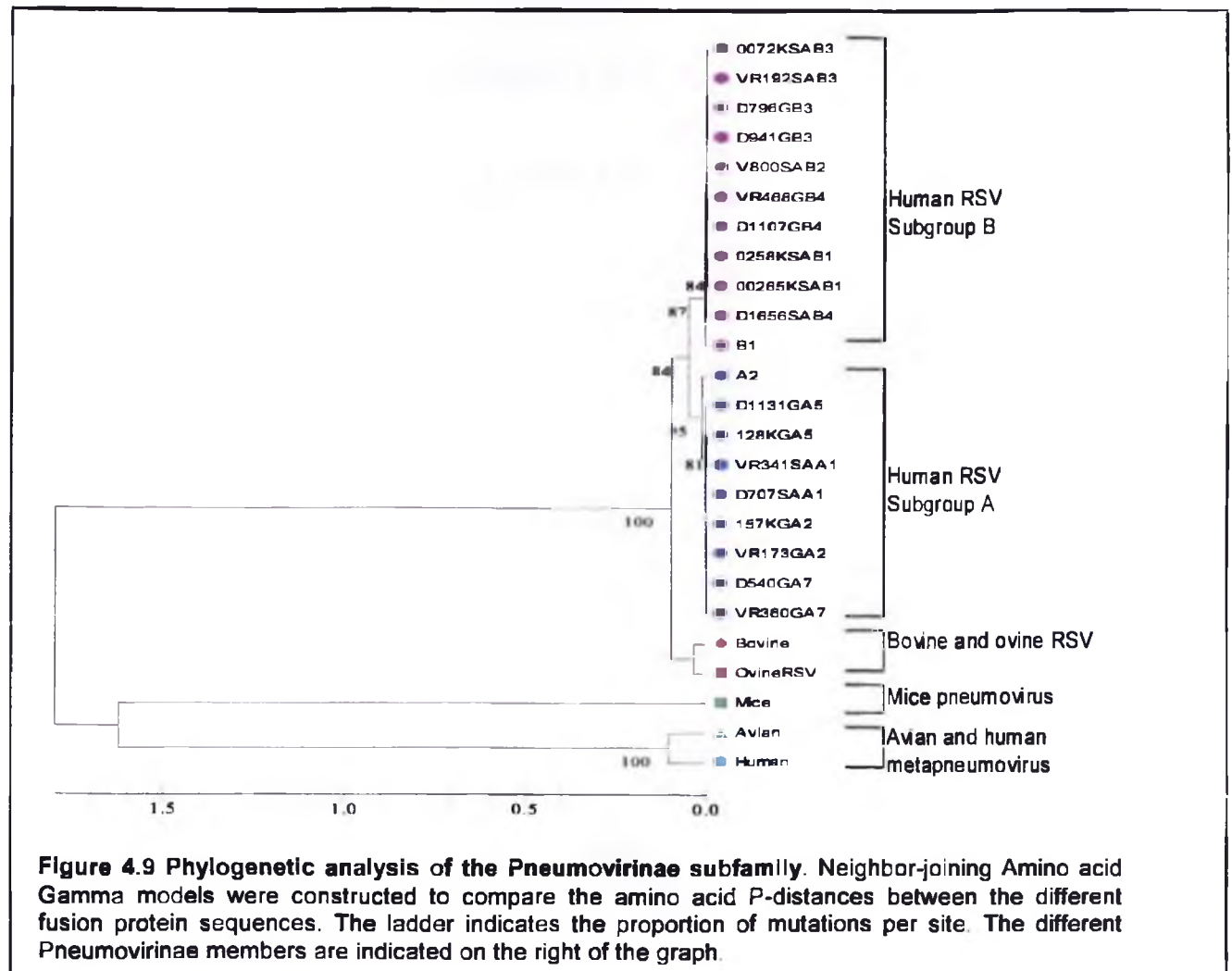
and B at position 200 (D→N) resulted in an  $\alpha$ -helix in subtype A and a turn in subtype B.

The B-cell and a T-helper cell epitopes identified between amino acids 205 and 225 (Corvaisier *et al.*, 1997), were conserved within the South African subtype A and B genotypes, respectively. Corvaisier *et al.* (1993) found that the amino acid substitutions identified at position 209 (K→Q) and 213 (S→R) in subtype B, may not be critical for the maintenance of the T-cell epitope, since a strong cross-reactivity was obtained with *in vitro* lymphocyte proliferation assays. The authors also found that at least one of the amino acids in 205 to 209 or 210 to 214 is necessary for the epitope function in subtype A (Corvaisier *et al.*, 1993). None of the amino acid changes identified in escape mutants (Arbiza *et al.*, 1992; Lopez *et al.*, 1998) was found within the South African F-protein sequences. The T-helper cell epitope residing between amino acids 255 to 278 (Corvaisier *et al.*, 1993) was found to be conserved between all genotypes of both subtypes (Figure 4.8).

Previous reports suggest that positive selection occurs in the G-protein of RSV over consecutive seasons in South Africa (Venter *et al.*, 2002; Venter *et al.*, 2001). In contrast to this sequence analysis of the N-protein suggested that recently identified dominant CTL epitopes remain conserved and that positive selection does not occur in the CTL epitopes of RSV (Venter *et al.*, 2003). Evidence of immune driven selective pressure by CTL in the Nef gene of the viral population of acutely infected HIV-1 patients has been reported (Price *et al.*, 1997). It has been reported that immune selection does occur in the CTL epitopes of the highly conserved influenza virus nucleoprotein (Boon *et al.*, 2002; Price *et al.*, 2000; Voeten *et al.*, 2000) and this implies that regular updating will be needed in vaccines containing these CTL epitopes. With the exception of a few mutations in known epitopes, most of the amino acid differences identified in this Chapter, in both the B-cell and T-cell epitopes of the F-protein were conserved within the subtypes, which does not suggest immune selection. However, the conformational changes identified within the South African F-protein sequences relative to the prototype strains, may influence conformational epitopes residing in the F-protein. Although the F-protein is much more conserved than the G-protein, group-specific neutralisation epitopes

have also been identified (Corvaisier *et al.*, 1997) which may affect subunit vaccine development based on the F-protein.

In summary, the F-protein sequences were highly conserved within and between South African genotypes, with most variability occurring at the nucleotide level. The prevalence of synonymous substitutions indicated the absence of positive selection within the protein. The F2-subunit was more variable than the F1-subunit and this might be due to the greater number of potential glycosylation sites in the F2-subunit (Plows & Pringle, 1995). The secondary structure predictions suggested that the protein structure might slightly differ between subtype A and B F-proteins and amino acid substitutions were identified that may result in conformational changes that could affect neutralisation epitopes. Most of the amino acid differences identified within neutralising and CTL epitopes were conserved within the subtypes, and therefore does not indicate immune selection. However, one genotype-specific amino acid difference was identified within one CTL epitope that may require further investigation. This event prevents the complete exclusion of the possibility of immune selection in this particular epitope. Further analysis over more seasons may be needed to clarify this. These results also suggest that most of the identified CTL epitopes are subtype A-specific and may not be recognised by subtype B viruses. This is the first study conducted in South Africa to determine the antigenic variation within the F-protein and might contribute to vaccine development.



**Figure 4.9** Phylogenetic analysis of the *Pneumovirinae* subfamily. Neighbor-joining Amino acid Gamma models were constructed to compare the amino acid P-distances between the different fusion protein sequences. The ladder indicates the proportion of mutations per site. The different *Pneumovirinae* members are indicated on the right of the graph.

## CHAPTER 5

### CONCLUDING REMARKS

There is currently no effective vaccine available for RSV, and even though the prophylaxis, Palivizumab (Synagis), has been shown to be highly effective in reducing RSV hospitalisations, administration is too costly for developing countries (Meissner *et al.*, 1999). RSV is responsible for one-fifth of lower respiratory tract infections in infants worldwide (Stensballe *et al.*, 2003). In general, the mortality rate is low, however, in developing countries, RSV mortality is significantly higher than in developed countries (Stensballe *et al.*, 2003). Primary RSV infection can occur in young infants in the presence of maternally derived antibodies (Glezen *et al.*, 1981) and reinfection can occur throughout life, probably due to incomplete protection (Graham, 1996). The role of antigenic diversity in RSV epidemiology is not yet completely understood, however it may partly explain the susceptibility to reinfection throughout life and the variation in the severity of epidemics within communities each year (Reviewed by Dudas & Karron, 1998).

Limited data about the molecular epidemiology of RSV are available from developing countries, especially Africa (Cane *et al.*, 1999; Roca *et al.*, 2001; Venter *et al.*, 2002; Venter *et al.*, 2001). Venter *et al.* (2001) characterised the molecular epidemiology of RSV over four consecutive seasons (1997-2000) at the Chris Hani Baragwanath hospital in Soweto South Africa. One of the questions that arose from this study was that during 1999 and 2000, one genotype (GA2) remained dominant over both seasons which appeared to be unusual when considering other studies conducted worldwide that have shown that different genotypes dominate each year (Cane *et al.*, 1994; Choi & Lee, 2000; Peret *et al.*, 1998).

In this thesis, the phylogenetic analysis of the 2001 outbreak in Soweto was described to elucidate the transmission dynamics of RSV during an epidemic in a single community, and further elucidate the molecular epidemiology over consecutive seasons in South Africa. The evolution of RSV antigenic proteins (F- and G-protein) of known South African genotypes, were analysed to determine if

immune selection occurred in the epitopic domains of these proteins. The results obtained in this thesis will be summarised in the following section.

In **Chapter 3**, the molecular epidemiology of a RSV outbreak that occurred at the Chris Hani Baragwanath hospital in Soweto 2001 was described. A total of 82 RSV-positive specimens were subtyped and sequenced. It was found that both subtypes A and B co-circulated during 2001, however subtype B dominated with 89%. The genetic variability of the RSV G-protein was characterised using distance-based (neighbor-joining) and character-based (maximum likelihood) trees. Genotyping methods previously described (Peret *et al.*, 2000; Peret *et al.*, 1998; Venter *et al.*, 2002; Venter *et al.*, 2001) were used to assign the Soweto 2001 sequences to genotypes. It was found that three subtype A genotypes (GA2, GA5 and SAA1) and three subtype B genotypes (GB3, SAB1 and SAB3) were present in the 2001 Soweto outbreak. The dominant subtype A genotype (GA2) from 1999 and 2000 was replaced by the SAB3 genotype during 2001, which dominated with 69% followed by SAB1 with 18%. Only one SAB1 isolate had been identified in Soweto before in the year 2000. This may suggest that rare strains from one season may become prominent outbreak strains in later seasons in an immunological naïve population. Cane (2001) suggested that genotype switching in each epidemic is driven by herd immunity to the particular dominant strain.

A total of 24 unique nucleotide sequence groups (seven subtype A and 17 subtype B) could be identified within the Soweto 2001 outbreak, demonstrating the high level of genetic variability that may occur during a single confined outbreak. The outbreak was dominated by four strains (one SAB1 and three SAB3), which circulated from the start up until the end of the epidemic. These four strains might have been the initial SAB1 and SAB3 outbreak strains, since they were isolated in the beginning of the outbreak and infected numerous individuals throughout the entire epidemic. The other SAB1 and SAB3 strains may have resulted because of possible immune driven selection in the dominant outbreak strains.

The role of positive selection in the G-protein was further investigated by analysing the nucleotide and amino acid sequences of the different outbreak strains. It was found that in both RSV subtypes, the mean amino acid P-distance (0.13 and 0.09

for subtype A and B, respectively) exceeded the mean nucleotide P-distance (0.07 and 0.05 for subtype A and B, respectively). Nonsynonymous substitutions exceeded synonymous substitutions (Ka/Ks ratios of 1.75 and 2 for subtype A and B, respectively) and transitions were more prevalent than transversions. Collectively these results suggest that positive immune selection occurred within the outbreak. This may account for the emergence of different strains later observed in each genotype. These results agree with findings from studies conducted in other parts of the world (Reviewed by Cane, 2001, Sullender, 2000).

G-protein amino acid sequence alignments revealed the existence of 22 different amino acid sequence groups in the 2001 outbreak, relative to 24 nucleotide sequence groups, suggesting some of the nucleotide mutations resulted in similar amino acids. Differences in the length of the G-protein were also observed with subtype A sequences showing limited variation in protein length, while subtype B proteins had more variable protein lengths, corresponding to previous findings (Martinez *et al.*, 1999). Changes in the termination codons are thought to affect strain-specific epitopes residing in the C-terminal of the G-protein (Rueda *et al.*, 1991; Rueda *et al.*, 1995; Martinez *et al.*, 1997). Strain-specific epitopes have been mapped within the C-terminal of the G-protein of subtype A sequences (Reviewed by Melero *et al.*, 1997). Amino acid changes associated with epitope loss identified in either escape mutants or natural occurring strains (Cane & Pringle, 1995; Garcia *et al.*, 1994), were present in some of the Soweto 2001 subtype A sequences. Alterations in the antigenic structure of the G-protein may provide a way for the virus to effectively evade the immune recognition (Melero *et al.*, 1997).

To further investigate the evolution in RSV antigenic proteins, the antigenic variability within the F-protein of each South African G-protein genotype, was studied in **Chapter 4**. The F-protein is the major focus of prophylactic antibody and subunit or recombinant vaccine research (Reviewed by Meanwell *et al.*, 2000). It is therefore important to determine the extent of variation between the various genotypes identified in South Africa to date and to determine if immune selection occurs in the antibody and CTL epitopes.

A total of 18 clinical specimens were selected from each of the subtype A and B genotypes identified in South Africa to date. A RT-PCR using subtype-specific primers was developed for the amplification of the precursor F-protein (1-1729bp) and successfully amplified all the South African genotypes. Nucleotide sequence analysis showed that the F-protein of subtypes A and B shared 82% identity, which agreed with previous findings (Johnson & Collins, 1988; Johnson *et al.*, 1987). Transitions were more prevalent than transversions and in contrast to the G-protein, for the F-protein synonymous substitutions (Ks) exceeded nonsynonymous substitutions (Ka) (Ka/Ks=0.12 for both subtypes). The Ka/Ks ratios suggested an absence of positive selection within this antigenic protein (Hurst, 2002).

Most nucleotide changes resulted in silent mutations and the amino acid sequence identity was 90% between both subtypes. When considering structural features common to *Pneumovirinae* fusion proteins (Chambers *et al.*, 1992; Plows & Pringle, 1995), results obtained here suggested that the F2-subunit of the F-protein was more variable than the F1-subunit for South African specimens, confirming the studies of the prototype isolates (Chambers *et al.*, 1992; Plows & Pringle, 1995). It has been suggested that the higher variability in the F2-subunit may be due to the greater number of potential glycosylation sites within this subunit (Plows & Pringle, 1995).

When the secondary structure ( $\alpha$ -helices,  $\beta$ -strands and turns) was predicted for the F-protein of each South African genotype, it was found that some amino acid differences resulted in conformational changes, which might affect the antigenic structure of the F-protein. Previous findings showed that the F-protein of RSV exists as a trimeric coiled coil for the prototype strains A2, Long and CH-18537 (Lambert *et al.*, 1996; Matthews *et al.*, 2000; Zhao *et al.*, 2000). Similar results were obtained for the F-protein sequences of the South African genotypes, regardless of the amino acid variation observed relative to the prototype strains.

The F-protein is the major target for the protective immune response (Lopez *et al.*, 1988) and is an important target for cytotoxic T lymphocytes (Cherrie *et al.*, 1992) and neutralising antibody response (Walsh *et al.*, 1985). When four previously identified B-cell epitopes were plotted on the South African F-protein sequences,

two epitopes (47F (Lopez *et al.*, 1990), L4 (Martin-Gallardo *et al.*, 1991)) were highly conserved between both subtypes, while the other two epitopes (RS-348 (Corvaisier *et al.*, 1993), and 7C2 (Trudel *et al.*, 1987)) were conserved only within the subtype A genotypes.

When three CTL epitopes previously identified in subtype A (restricted to HLA B\*57, HLA Cw\*12 (Brandenburg *et al.*, 2000) and HLA-A\*01 (Rock & Crowe, 2003)) were plotted on the South African F-protein sequences, all three were conserved within subtype A and B, although subtype specific differences were identified. One genotype-specific amino acid mutation (L→F) was observed at position 547 of GA5 in the HLA Cw\*12-restricted CTL epitope. This observation mutation prevents the complete exclusion of the possibility of immune selection within this CTL epitope. These results suggest that most of the identified CTL epitopes are subtype A-specific and may not be recognised by subtype B viruses, while the HLA Cw\*12 restricted epitope may also not be recognized efficiently in GA5 strains. This is the first study that addresses variation in CTL epitopes of the F-protein.

Of the two T-helper cell epitopes identified in mice, one (amino acids 255 to 278) (Corvaisier *et al.*, 1993) was found to be conserved between both subtypes A and B, while the other epitope (amino acids 205 and 225) (Corvaisier *et al.*, 1997), was conserved within the subtype A and B, respectively. None of the amino acid changes previously identified in escape mutants of RSV (Arbiza *et al.*, 1992; Lopez *et al.*, 1998) were found within the F-proteins of South African clinical specimens.

Previous studies conducted over consecutive seasons in South Africa (Venter *et al.*, 2002; Venter *et al.*, 2001); as well as extensive analysis of a single epidemic in this study (Chapter 3) suggested that positive selection occurs in the G-protein, which is one of the neutralising antigens of RSV. However, most of the amino acid differences identified in both the B-cell and T-cell epitopes of the F-protein from the South African genotypes, were conserved within the subtypes, which suggest that immune selection does not occur in this important antigen (Chapter 4). Some subtype-specific differences in the antibody and CTL epitopes do however suggest that both subtypes A and B should be represented in subunit vaccines for effective protection. One genotype-specific mutation was observed in a CTL epitope that may

require further investigation over more seasons to completely rule out the possibility of selection in these epitopes over the long term. An absence of immune selection in the CTL epitopes of the N-protein has also been demonstrated (Venter *et al.*, 2003).

Since the protective immune response is directed to the F- and G-proteins, these antigens are the major targets for subunit vaccine development (Crowe, 2002). The development of a synthetic peptide epitope-based (mimotopes) vaccine to induce virus-neutralising antibodies against the more conserved RSV F-protein is also under investigation. However, the potential role of antigenic variation in RSV's ability to escape the immune response and to establish infections as illustrated in this and previous studies emphasises the importance of evolutionary investigations of RSV antigenic proteins (F- and G-protein) for future vaccine development.

## CHAPTER 6 REFERENCES

- Akerlind-Stopner, B., Utter, G., Mufson, M., Orvell, C., Lerner, R. & Norrby, E. (1990). A subgroup-specific antigenic site in the G protein of respiratory syncytial virus forms a disulphide-bonded loop. *Journal of Virology* 64, 5143-5148.
- Alwan, W., Kozłowska, W. & Openshaw, P. (1994). Distinct types of lung disease caused by functional subsets of antiviral T cells. *J Exp Med* 179, 81-9.
- Anderson, K., King, A., Lerch, R. & Wertz, G. (1992). Polylactosaminoglycan modification of the respiratory syncytial virus small hydrophobic (SH) protein: a conserved feature among human and bovine respiratory syncytial viruses. *Virology* 191, 417-430.
- Anderson, L., Hendry, R., Pierik LT, Tsou C & K., M. (1991). Multicenter study of strains of respiratory syncytial virus. *J Infect Dis.* Apr;163, 687-92.
- Anderson, L., Hierholzer, J., Tsou, C., Hendry, R., Fernie, B., Stone, Y. & McIntosh, K. (1985). Antigenic characterization of respiratory syncytial virus strains with monoclonal antibodies. *J Infect Dis.* Apr;151, 626-33.
- Anfinsen, C., Haber, E., Sela, M. & White, F. (1961). The kinetics of the formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* 47, 1309-1314.
- Arbiza, J., Taylor, G., Lopez, J. A., Furze, J., Wyld, S., Whyte, P., Stott, E. J., Wertz, G., Sullender, W., Trudel, M. & et al. (1992). Characterization of two antigenic sites recognized by neutralizing monoclonal antibodies directed against the fusion glycoprotein of human respiratory syncytial virus. *J Gen Virol* 73 ( Pt 9), 2225-34.
- Atreya, P., Peebles, M. & Collins, P. (1998). The NS1 protein of human respiratory syncytial virus is a potent inhibitor of minigenome transcription and RNA replication. *J Virol* 71, 8973-8982.
- Bachi, T. (1988). Direct observation of the budding and fusion of an enveloped virus by video microscopy of viable cells. *J Cell Biol* 107, 1689-95.
- Baker, K., Dutch, R., Lamb, R. & Jardetzky, T. (1999). Structural basis for paramyxovirus-mediated membrane fusion. *Mol Cell* 285, 309-319.
- 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 recognize the viral nucleoprotein (N), but not the major glycoprotein (G), expressed by vaccinia virus recombinants. *J Immunol* 137, 3973-7.
- Baxeavanis, A. (1998). Practical aspects of multiple sequence alignment. In *Bioinformatics: A practical guide to the analysis of genes and proteins*. Edited by A. Baxeavanis & B. Ouellette. New York: Wiley-Liss. Inc.

- Baxevanis, A. & Landsman, D. (1998). Predictive methods using protein sequences. In *Bioinformatics: A practical guide to the analysis of genes and proteins*. Edited by A. Baxevanis & B. Ouellette. New York: Wiley-Liss. Inc.
- Beeler, J. & Coelingh, K. V. W. (1989). Neutralization epitopes of the F glycoprotein of respiratory syncytial virus: effect of mutation upon fusion function. *J Virol* 63, 2941-2950.
- Beem, M. (1967). Repeated infections with respiratory syncytial virus. *J Immunol* 98, 1115-22.
- Berger, B. & Singh, M. (1997). An iterative method for improved protein structural motif recognition. *Journal of Computational Biology* 4, 261-273.
- Black, C. P. (2003). Systematic review of the biology and medical management of respiratory syncytial virus infection. *Respir Care* 48, 209-31; discussion 231-3.
- Blanchard, S., Astier-Gin, T., Tallet, B., Moynet, D., Londos-Gagliardi, D. & Guillemain, B. (1999). Amino acid changes at positions 173 and 187 in the human T-cell leukemia virus type 1 surface glycoprotein induce specific neutralizing antibodies. *J Virol* 73, 9369-76.
- Boger, J., Emini, E. & Schmidt, A. (1986). Surface probability profile. An heuristic approach to the selection of synthetic peptide antigens. *Reports on the Sixth International Congress in Immunology (Toronto)*, 250.
- Boon, A. C., de Mutsert, G., Graus, Y. M., Fouchier, R. A., Sintricolaas, 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.
- Brandenburg, A. H., de Waal, L., Timmerman, H. H., Hoogerhout, P., de Swart, R. L. & Osterhaus, A. D. (2000). HLA class I-restricted cytotoxic T-cell epitopes of the respiratory syncytial virus fusion protein. *J Virol* 74, 10240-4.
- Brideau, R., Walters, R., Stier, M. & Wathen, M. (1989). Protection of cotton rats against human respiratory syncytial virus by vaccination with a novel chimeric FG glycoprotein. *J. Gen. Virol.* 70, 2637-2644.
- Buchholz, U., Granzow, H., Schuldt, K., Whitehead, S., Murphy, B. & Collins, P. (2000). Chimeric bovine respiratory syncytial virus with glycoprotein gene substitutions from human respiratory syncytial virus: effects on host-range and evaluation as a live attenuated HRSV vaccine. *J Virol* 74, 1187-1199.
- Buckland, R. & Wild, F. (1989). Leucine zipper motif extends. *Nature* 338, 547.
- 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.
- Bullough, P., Hughson, F., Skehel, J. & Wiley, D. (1994). Structure of the influenza haemagglutinin at the pH of membrane fusion. *Nature* 371, 37-43.

- Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M. & Clore, G. M. (1998). Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *Embo J* 17, 4572-84.
- 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*. Sep;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* May;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. *Rev Med Virol* 11, 103-16.
- 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. (1993). Frequent polymerase errors observed in a restricted area of clones derived from the attachment (G) protein gene of respiratory syncytial virus. *J Virol* 67, 1090-3.
- 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. (1991c). 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., 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., 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.

- Chambers, P., Pringle, C. & Easton, A. (1992). Sequence analysis of the gene encoding the fusion glycoprotein of pneumonia virus of mice suggests possible conserved secondary structure elements in paramyxovirus fusion glycoproteins. *J. Gen. Virol.* 73, 1717-1724.
- Chambers, P., Pringle, C. R. & Easton, A. J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J Gen Virol* 71 ( Pt 12), 3075-80.
- Chan, D., Fass, D., Berger, J. & Kim, P. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89, 263-273.
- Chang, J., Srikiatkachorn, A. & Braciale, T. J. (2001). Visualization and characterization of respiratory syncytial virus F-specific CD8(+) T cells during experimental virus infection. *J Immunol* 167, 4254-60.
- Chanock, R., Roizman, B. & Myers, R. (1957). Recovery from infants with respiratory illness of a virus related to Chimpanzee Coryza Agent (CCA). *Am J Hyg* 66, 281-290.
- Chargelegue, D., Obeid, O. E., Shaw, D. M., Denbury, A. N., Hobby, P., Hsu, S. C. & Steward, M. W. (1997). Peptide mimics of a conformationally constrained protective epitopes of respiratory syncytial virus fusion protein. *Immunol Lett* 57, 15-7.
- Cherrie, A. H., Anderson, K., Wertz, G. W. & Openshaw, P. J. (1992). Human cytotoxic T cells stimulated by antigen on dendritic cells recognize the N, SH, F, M, 22K, and 1b proteins of respiratory syncytial virus. *J Virol* 66, 2102-10.
- 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(5):1547-56., 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.
- Clug, W. & Cummings, M. (1997). Concepts of Genetics, fifth edition edn, pp. 614-635. Edited by S. Snaveley. New Jersey: Prentice-Hall, Inc.
- Coates, H., Kendrick, L. & Chanock, R. (1963). Antigenic differences between two strains of respiratory syncytial virus. *Proc. Soc. Exp. Biol. Med.* 12, 958-964.
- 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. (1991). The Molecular Biology of Human Respiratory Syncytial Viruses (RSV) of the Genus Pneumovirus. In *The Paramyxoviruses*, pp. 103-162. Edited by K. D. ed. New York: Plenum Press.
- Collins, P., Hill, M., Chrisina, J. & Grosfeld, H. (1996a). Transcription elongation factor for respiratory syncytial virus, a nonsegmented negative strand RNA virus. *Proc Natl Acad Sci U S A* 93, 81-85.

- Collins, P., McIntosh, K. & Chanock, R. (1996b). Respiratory syncytial virus. In *Virology*, 3rd ed edn, pp. 1313-1352: B.N. Fields, D.M Knipe, P.M. Howley, eds.
- Collins, P., Whitehead, S., Bukreyev, A., Fearn, R., Teng, M., Juhasz, K., Chanock, R. & Murphy, B. (1999). Rational design of live attenuated recombinant vaccine virus for human respiratory syncytial virus by reverse genetics. In *Advances in virus research*, pp. 423-447.
- 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., Huang, Y. T. & Wertz, G. W. (1984). Nucleotide sequence of the gene encoding the fusion (F) glycoprotein of human respiratory syncytial virus. *Proc Natl Acad Sci U S A* 81, 7683-7.
- Connor, A. L., Bevitt, D. J. & Toms, G. L. (2001). Comparison of human respiratory syncytial virus A2 and 8/60 fusion glycoprotein gene sequences and mapping of sub-group specific antibody epitopes. *J Med Virol* 63, 168-77.
- Connors, M., Collins, P. & Firestone, C. (1992). Cotton rats previously immunized with chimeric RSV FG glycoprotein develop enhanced pulmonary pathology when infected with RSV, a phenomenon not encountered following immunization with vaccinia RSV recombinants or RSV. *Vaccine* 10, 475-484.
- Connors, M., Giese, N. A., Kulkarni, A. B., Firestone, C. Y., Morse, H. C., 3rd & Murphy, B. R. (1994). Enhanced pulmonary histopathology induced by respiratory syncytial virus (RSV) challenge of formalin-inactivated RSV-immunized BALB/c mice is abrogated by depletion of interleukin-4 (IL-4) and IL-10. *J Virol* 68, 5321-5.
- Corvaisier, C., Bourgeois, C. & Pothier, P. (1997). Cross-reactive and group specific immune responses to a neutralizing epitope of human respiratory syncytial virus fusion protein. *Arch Virol* 142, 1073-1086.
- Corvaisier, C., Guillemin, G., Bourgeois, C., Bour, J. B., Kohli, E. & Pothier, P. (1993). Identification of T-cell epitopes adjacent to neutralizing antigenic domains on the fusion protein of respiratory syncytial virus. *Res Virol* 144, 141-50.
- 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.
- Cristina, J., Moya, A., Arbiza, J., Russi, J., Hortal, M., Albo, C., Garcia-Barreno, B., Garcia, O., Melero, J. A. & Portela, A. (1991). Evolution of the G and P genes of human respiratory syncytial virus (subgroup A) studied by the RNase A mismatch cleavage method. *Virology* 184, 210-8.
- Crowe, J. (2002). Respiratory syncytial virus vaccine development. *Vaccine* 20 (suppl), S32-S37.

- Crowe, J., Collins, P., London, W., Chanock, R. & Murphy, B. (1993). A comparison in chimpanzees of the immunogenicity and efficacy of live attenuated respiratory syncytial virus (RSV) temperature-sensitive mutant vaccines and vaccinia virus recombinants that express the surface glycoprotein of RSV. *Vaccine* 11, 1395-1404.
- Crowe, J. E., Jr. (1998). Immune responses of infants to infection with respiratory viruses and live attenuated respiratory virus candidate vaccines. *Vaccine* 16, 1423-32.
- Deleage, G., Clerc, F., Roux, B. & Gautheron, D. (1988). ANTHEPROT : A package for protein sequence analysis using a microcomputer. *Cabios* 4, 351-356.
- Doggett, J. E. & Taylor-Robinson, D. (1965). Serological studies with respiratory syncytial virus. *Arch Gesamte Virusforsch* 15, 601-8.
- Dudas, R. A. & Karron, R. A. (1998). Respiratory syncytial virus vaccines. *Clin Microbiol Rev* 11, 430-9.
- Eleraky, N. Z., Kania, S. A. & Potgieter, L. N. (2001). The ovine respiratory syncytial virus F gene sequence and its diagnostic application. *J Vet Diagn Invest* 13, 455-61.
- 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. & Walsh, E. (1997). Safety and immunogenicity of a respiratory syncytial virus subunit vaccine ((PFP-2) in the institutionalized elderly. *Vaccine* 15, 1130-1132.
- Felsenstein, J. (1993). PHYLIP (Phylogeny Inference Package) version 3.5c. In *Distributed by the author. Department of Genetics, University of Washington, Seattle*. Department of Genetics, University of Washington, Seattle.
- Felsenstein, J. (1996). Inferring phylogenies from protein sequences by parsimony, distance and likelihood methods. *Methods Enzymol.* 266, 418-427.
- Folkerts, G., Busse, W. W., Nijkamp, F. P., Sorkness, R. & Gern, J. E. (1998). Virus-induced airway hyperresponsiveness and asthma. *Am J Respir Crit Care Med* 157, 1708-20.
- Garcia, J., Garcia-Barreno, B., Vivo, A. & Melero, J. A. (1993). Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein, and the 22K protein. *Virology* 195, 243-7.
- 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-Barreno, B., A, P., Delgado, T., Lopez, J. & Melero, J. A. (1990). Frame shifts mutations as a novel mechanism for the generation of neutralisation resistant mutants of human respiratory syncytial virus. *EMBO J* 9, 4181-4187.

- Garcia-Barreno, B., Delgado, T. & Melero, J. A. (1996). Identification of protein regions involved in the interaction of human respiratory syncytial virus phosphoprotein and nucleoprotein: significance for nucleocapsid assembly and formation of cytoplasmic inclusions. *J Virol* 70, 801-8.
- Garcia-Barreno, B., Delgado, T., Akerlind-Stopner, B., Norrby, E. & Melero, J. (1992). Location of the epitope recognized by monoclonal antibody 63G on the primary structure of human respiratory syncytial virus G glycoprotein and the ability of synthetic peptides containing this epitope to induce neutralizing antibodies. *Journal of General Virology* 73, 2625-2630.
- Garcia-Barreno, B., Palomo, C., Penas, C., Delgado, T., Perez-Brena, P. & Melero, J. A. (1989). Marked differences in the antigenic structure of human respiratory syncytial virus F and G glycoproteins. *J Virol* 63, 925-32.
- Gattiker, A., Gasteiger, E. & Bairoch, A. (2002). ScanProsite: a reference implementation of a PROSITE scanning tool. *Applied Bioinformatics* 1, 107-108.
- 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.
- Gimenez, H. B., Hardman, N., Keir, H. M. & Cash, P. (1986). Antigenic variation between human respiratory syncytial virus isolates. *J Gen Virol* 67 ( Pt 5), 863-70.
- 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.
- Gonzalez, I., Karron, R. & Eichelberger, M. (2000). Evaluation of the live attenuated cpts-248/404 RSV vaccine in combination with a subunit vaccine (PFP-2) in healthy young and older adults. *Vaccine* 18, 1763-1772.
- González-Reyes, L., Ruiz-Arguello, M., Garcia-Barreno, B., Calder, L., Lopez, J., Albar, J., Skehel, J., Wiley, D. & Melero, J. (2001). Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. *PNAS* 98, 9859-9864.
- Goulder, P. J., Lechner, F., Klennerman, P., McIntosh, K. & Walker, B. D. (2000). Characterization of a novel respiratory syncytial virus-specific human cytotoxic T-lymphocyte epitope. *J Virol* 74, 7694-7.
- Graham, B. (1995). Pathogenesis of respiratory syncytial virus vaccine-augmented pathology. *Am. J. Respir. Crit. Care Med* 152, 563-566.
- Graham, B. (1996). Immunological determinants of disease caused by respiratory syncytial virus. *Trends Microbiol* 4, 290-3.
- Graham, B., Bunton, L., Wright, P. & Karzon, D. (1991). Role of T-lymphocyte subsets in the pathogenesis of primary infection and re-challenge with respiratory syncytial virus. *J Clin Invest* 88, 1026.

- Graham, B. S., Henderson, G. S., Tang, Y. W., Lu, X., Neuzil, K. M. & Colley, D. G. (1993). Priming immunization determines T helper cytokine mRNA expression patterns in lungs of mice challenged with respiratory syncytial virus. *J Immunol* 151, 2032-40.
- Groothuis, J., King, S., Hogerman, D., Paradiso, P. & Simoes, E. (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, 467-469.
- Groothuis, J. & Nishida, H. (2002). Prevention of respiratory syncytial virus infections in high-risk infants by monoclonal antibody (palivizumab). *Pediatr Inf* 44, 235-241.
- Groothuis, J. R. (2003). Safety of Palivizumab in Preterm Infants 29 to 32 Weeks' Gestational Age Without Chronic Lung Disease to Prevent Serious Respiratory Syncytial Virus Infection. *Eur J Clin Microbiol Infect Dis*.
- 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.
- Group, T. I. R. S. (1998). Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high risk infants. *Pediatrics* 102, 531-537.
- Gruber, C. & Levine, S. (1983). Respiratory syncytial virus polypeptides. III. The envelope-associated proteins. *J Gen Virol* 64, 825-832.
- Gruber, C. & Levine, S. (1985). Respiratory syncytial virus polypeptides. V. The kinetics of glycoprotein synthesis. *J Gen Virol* 66 ( Pt 6), 1241-7.
- Gupta, R., Birch, H., Rapacki, K., Brunak, S. & Hansen, J. E. (1999). O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res* 27, 370-2.
- Hall, C. B. (2000). Nosocomial respiratory syncytial virus infections: the "Cold War" has not ended. *Clin Infect Dis* 31, 590-6.
- 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.
- Hammond, M. G., Toit, E. D. d., Sanchez-Mazas, A., Andrien, M., Coluzzi, M., Pablo, M. d., Stefano, G. d., Kaplan, C., Kenedy, L. J., Louie, L. & Migot, F. (1997). HLA in sub-Saharan Africa. In *The 12th International Histocompatibility Workshop and conference.*, pp. 345-353. Edited by D.Charron. Paris: EDK.
- Hancock, G., Speelman, D., Frenchick, P., Mineo-Kuhn, M., Baggs, R. & Hahn, D. (1995). Formulation of the purified fusion protein of respiratory syncytial virus with the saponin QS-21 induces protective immune responses in Balb/c mice that are similar to those generated by experimental infection. *Vaccine* 13, 391-400.
- Hancock, G. E., Tebbey, P. W., Scheuer, C. A., Pryharski, K. S., Heers, K. M. & LaPierre, N. A. (2003). Immune responses to the nonglycosylated ectodomain of respiratory syncytial virus attachment

- glycoprotein mediate pulmonary eosinophilia in inbred strains of mice with different MHC haplotypes. *J Med Virol* 70, 301-8.
- Hansen, J., Lund, O., Tolstrup, N., Gooley, A., Williams, K. & Brunak, S. (1998). NetOglyc: Prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate J* 15, 115-130.
- Hardy, R. W. & Wertz, G. W. (1998). The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J Virol* 72, 520-6.
- Heminway, B. R., Yu, Y., Tanaka, Y., Perrine, K. G., Gustafson, E., Bernstein, J. M. & Galinski, M. S. (1994). Analysis of respiratory syncytial virus F, G, and SH proteins in cell fusion. *Virology* 200, 801-5.
- 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.* Feb;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.
- Hengst, U. & Kiefer, P. (2000). Domains of human respiratory syncytial virus P protein essential for homodimerization and for binding to N and NS1 protein. *Virus Genes* 20, 221-5.
- Hernandez, J., Martinez, M., Rocha, E., Domingo, E. & Mateu, M. (1992). Generation of a subtype specific neutralization epitope in foot and mouth disease virus of a different subtype. *J Gen Virol* 73, 213-216.
- HersHKovitz, M. & Leipe, D. (1998). Phylogenetic analysis. In *Bioinformatics: A practical guide to the analysis of genes and proteins*. Edited by A. Baxevanis & B. Ouellette. New York: Wiley-Liss, Inc.
- Howard, C. R. (1995). The structure of hepatitis B envelope and molecular variants of hepatitis B virus. *J Viral Hepat* 2, 165-70.
- Huang, Y. & Anderson, R. (2003). A single amino acid substitution in a recombinant G protein vaccine drastically curtails protective immunity against respiratory syncytial virus (RSV). *Vaccine* 21, 2500-5.
- Huang, Y. T., Collins, P & Wertz, G.W. (1984). *Abstr. Annu. Meet. Am. Soc. Microbiol* absr. T27, 233.
- Huang, Y. T. & Wertz, G. W. (1983). Respiratory syncytial virus mRNA coding assignments. *J Virol* 46, 667-72.
- Hurst, L. (2002). The Ka/Ks ratios: diagnosing the form of sequence evolution. *Trends Genet* 18, 486.
- 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.

- 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.
- Jin, H., Zhou, H., Cheng, X., Tang, R., Munoz, M. & Nguyen, N. (2000). Recombinant respiratory syncytial viruses with deletions in the NS1, NS2, SH, and M2-2 genes are attenuated in vitro and in vivo. *Virology* 273, 210-8.
- Johansen, J., Christensen, L. S., Hornsleth, A., Klug, B., Hansen, K. S. & Nir, M. (1997). Restriction pattern variability of respiratory syncytial virus during three consecutive epidemics in Denmark. *Apmis* 105, 303-8.
- Johnson, P. & Collins, A. (1988). The fusion glycoprotein of human respiratory syncytial virus of subgroups A and B: sequence conservation provides a structural basis for antigenic relatedness. *J Gen Virol* 69, 2623-2628.
- Johnson, P. J., Olmsted, R., Prince, G., Murphy, B., Alling, D., Walsh, E. & Collins, P. (1987). 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*, Oct; 61, 3163-6.
- 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 ( Pt 6), 1539-47.
- 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.
- Kamasaki, H., Tsutsumi, H., Seki, K. & Chiba, S. (2001). Genetic variability of respiratory syncytial virus subgroup B strain isolated during the last 20 years from the same region in Japan: existence of time-dependent linear genetic drifts. *Arch Virol* 146, 457-66.
- 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. (1997). Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. *Proc Natl Acad Sci U S A* 94, 13961-6.
- Kasel, J. A., Walsh, E. E., Frank, A. L., Baxter, B. D., Taber, L. H. & Glezen, W. P. (1987). Relation of serum antibody to glycoproteins of respiratory syncytial virus with immunity to infection in children. *Viral Immunol* 1, 199-205.

- Kaul, T., Faden, H. & Ogra, P. (1981). Effect of respiratory syncytial virus and virus-antibody complexes on the oxidative metabolism of human neutrophils. *Infect Immun* 32, 649-654.
- Kim, H., Leikin, S., Arroby, J., Brandt, C., Chanock, R. & Parrot, R. (1976). Cell-mediated immunity to respiratory syncytial virus induced by inactivated vaccine or by infection. *Pediatr Res* 10, 75-78.
- Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16, 111-120.
- King, J., Burke, A., Clemens, J., Nair, P. & Farley, J. (1993). Respiratory syncytial virus illnesses in human immunodeficiency virus and non-infected children. *Pediatr Infect Dis J* 12, 733-739.
- Kneller, D., Cohen, F. & Langridge, R. (1990). Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *J. Mol. Biol.* 214, 171-182.
- Kong, X., Shou, H., Liu, C. & Jiang, Z. (2001). Molecular epidemiology of respiratory syncytial virus. *Chin Med J (Engl)* 114, 364-8.
- Kumar, S., Tamura, K., Jakobsen, I. & Nei, M. (2000). MEGA: Molecular evolutionary genetics analyses. Version 2.0. University Park: Pennsylvania State University.
- Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157, 105-32.
- Lambert, D. M. (1988). Role of oligosaccharides in the structure and function of respiratory syncytial virus glycoproteins. *Virology* 164, 458-66.
- Lambert, D. M., Barney, S., Lambert, A. L., Guthrie, K., Medinas, R., Davis, D. E., Bucy, T., Erickson, J., Merutka, G. & Petteway, S. R., Jr. (1996). Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion. *Proc Natl Acad Sci U S A* 93, 2186-91.
- Langedijk, J., Brandenburg, A., Middel, W., Osterhaus, A., Melen, R. & Oirshot, J. v. (1997). A subtype specific peptide-based enzyme immunoassay for detection of antibodies tot the G protein of human respiratory syncytial virus is more sensitive than routine serological tests. *J Clin Microbiol* 35, 1656-1660.
- Levine, S., Klaiber-Franco, R. & Paradiso, P. (1987). Demonstration that glycoprotein G is the attachment protein of respiratory syncytial virus. *J Gen Virol* 68, 2521-2524.
- Li, W. (1997). Molecular Evolution: Sunderland, MA: Sinauer Associates.
- Li, X., Sambhara, S. & Li, C. (1998). Protection against respiratory syncytial virus infection by DNA immunisation. *J Exp Med* 188, 681-688.
- Li, X., Sambhara, S., Li, C. X., Ettore, L., Switzer, I., Cates, G., James, O., Parrington, M., Oomen, R., Du, R. P. & Klein, M. (2000). Plasmid DNA encoding the respiratory syncytial virus G protein is a promising vaccine candidate. *Virology* 269, 54-65.
- Lopez, J., Bustos, R., Orvell, C., Berois, M., Arbiza, J., Garcia-Barreno, B. & Molero, J. (1998). Antigenic Structure of Human Respiratory Syncytial Virus Fusion Glycoprotein. *Journal of Virology*, 6922-6928.

- Lopez, J. A., Penas, C., Garcia-Barreno, B., Melero, J. A. & Portela, A. (1990). Location of a highly conserved neutralizing epitope in the F glycoprotein of human respiratory syncytial virus. *J Virol* 64, 927-30.
- Lopez, J. A., Villanueva, N., Melero, J. A. & Portela, A. (1988). Nucleotide sequence of the fusion and phosphoprotein genes of human respiratory syncytial (RS) virus Long strain: evidence of subtype genetic heterogeneity. *Virus Res* 10, 249-61.
- 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. *Journal of Pediatrics* 137, 78-84.
- Madhi, S. A., Venter, M., Alexandra, R., Lewis, H., Kara, Y., Karshagen, W. F., Greef, M. & Lassen, C. (2003). Respiratory syncytial virus associated illness in high-risk children and national characterisation of the circulating virus genotype in South Africa. *J Clin Virol* 27, 180-9.
- 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.
- Martinez, I., Dopazo, J. & Melero, J. A. (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 ( Pt 10), 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.
- Martin-Gallardo, A., Fien, K., Hu, B., Farley, J., Seid, R., Collins, P., Hildreth, S. & Paradiso, P. (1991). Expression of the F Glycoprotein Gene from Human Respiratory syncytial virus in *Escherichia coli*: Mapping of a Fusion Inhibiting Epitope. *Virology* 184, 428-432.
- Matthews, J., Young, T., Tucker, S. & Mackay, J. (2000). The core of the respiratory syncytial virus fusion protein is a trimeric coiled coil. *J Virol* 74, 5911-5920.
- Meanwell, N. A. & Krystal, M. (2000). Respiratory syncytial virus: recent progress towards the discovery of effective prophylactic and therapeutic agents. *Drug Discov Today* 5, 241-252.
- Meissner, H., Welliver, R., Chartrand, S., Law, B., Weizman, L., Dorkin, H. & Rodriguez, W. (1999). Immunoprophylaxis with palivizumab, a humanized respiratory syncytial virus monoclonal antibody, for prevention of respiratory syncytial virus infection in high risks infants: a consensus opinion. *Pediatr Infect Dis J* 18, 223-231.
- Meissner, H. C. (2003). Selected populations at increased risk from respiratory syncytial virus infection. *Pediatr Infect Dis J* 22, S40-4; discussion S44-5.
- 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.

- Merz, D. C., Scheid, A. & Choppin, P. W. (1980). Importance of antibodies to the fusion glycoprotein of paramyxoviruses in the prevention of spread of infection. *J Exp Med* 151, 275-88.
- Morris, J., Blount, R. & Savage, R. (1956). Recovery of cytopathogenic agent from chimpanzees with coryza. *Proc Soc Exp Biol Med* 92, 544-594.
- Muelenaer, P. M., Henderson, F. W., Hemming, V. G., Walsh, E. E., Anderson, L. J., Prince, G. A. & Murphy, B. R. (1991). Group-specific serum antibody responses in children with primary and recurrent respiratory syncytial virus infections. *J Infect Dis* 164, 15-21.
- Mufson, M., Orvell, C., Rafnar, B. & Norrby, E. (1985). Two distinct subgroups of human respiratory syncytial virus. *J Gen Virol.* Oct;66, 2111-24.
- Munoz, F. M., Piedra, P. A. & Glezen, W. P. (2003). Safety and immunogenicity of respiratory syncytial virus purified fusion protein-2 vaccine in pregnant women. *Vaccine* 21, 3465-7.
- Murphy, B. R., Prince, G. A., Walsh, E. E., Kim, H. W., Parrott, R. H., Hemming, V. G., Rodriguez, W. J. & Chanock, R. M. (1986). Dissociation between serum neutralizing and glycoprotein antibody responses of infants and children who received inactivated respiratory syncytial virus vaccine. *J Clin Microbiol* 24, 197-202.
- 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 immunization. *Vaccine* 8, 497-502.
- Murphy, L. B., Loney, C., Murray, J., Bhella, D., Ashton, P. & Yeo, R. P. (2003). Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein. *Virology* 307, 143-53.
- Nadal, D. & Ogra, P. (1990). Development of local immunity: role in mechanisms of protection against or pathogenesis of respiratory syncytial virus infections. *Lung* 168 (suppl), 379-387.
- Nei, M. & Kumar, S. (2000). Molecular Evolution and Phylogenetics. *Oxford University press*.
- Noah, T. L., Henderson, F. W., Wortman, I. A., Devlin, R. B., Handy, J., Koren, H. S. & Becker, S. (1995). Nasal cytokine production in viral acute upper respiratory infection of childhood. *J Infect Dis* 171, 584-92.
- Norrby, E., Mufson, M. A., Alexander, H., Houghten, R. A. & Lerner, R. A. (1987). Site-directed serology with synthetic peptides representing the large glycoprotein G of respiratory syncytial virus. *Proc Natl Acad Sci U S A* 84, 6572-6.
- Oien, N., Brideau, R., Walsh, E. & Wathen, M. (1994). Induction of local and systemic immunity against human respiratory syncytial virus using a chimeric RG glycoprotein and cholera toxin B subunit. *Vaccine* 12, 731-735.
- Olmsted, R. & Collins, P. (1989a). The 1A protein of respiratory syncytial virus is an integral membrane protein present as multiple structurally distinct species. *J Virol* 63, 2019-2029.

- Olmsted, R., Murphy, B., Lawrence, L., Elango, N., Moss, B. & Collins, P. (1989b). Processing, surface expression and immunogenicity of carboxy-terminally truncated mutants of G protein of human respiratory syncytial virus. *J Virol* 63, 411-420.
- Openshaw, P. J. (1995). Immunopathological mechanisms in respiratory syncytial virus disease. *Springer Semin Immunopathol* 17, 187-201.
- Openshaw, P. J. (2002). Potential therapeutic implications of new insights into respiratory syncytial virus disease. *Respir Res* 3(suppl1), S15-S20.
- Palomo, C., Cane, P. A. & Melero, J. A. (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 ( Pt 3), 669-75.
- Parker, J., Guo, D. & Hodges, R. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25, 5425-5432.
- Parker, K. C., 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, 163-75.
- Pauling, L. & Corey, R. (1951). The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* 37, 205-211.
- Peisajovich, S. G., Samuel, O. & Shai, Y. (2000). Paramyxovirus F1 protein has two fusion peptides: implications for the mechanism of membrane fusion. *J Mol Biol* 296, 1353-65.
- Pemberton, R., Cannon, M., Openshaw, P., Ball, L., Wertz, G. & Askonas, B. (1987). Cytotoxic T cell specificity for respiratory syncytial virus proteins: Fusion protein is an important target antigen. *J Gen Virol* 68, 2177-2182.
- 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.
- Piedra, P., Grace, S. & Jewell, A. (1996). Purified fusion protein vaccine protects against lower respiratory tract illness during respiratory syncytial virus season in children with cystic fibrosis. *Pediatr Infect Dis J* 15, 23-31.
- Piedra, P. A. (2003). Clinical experience with respiratory syncytial virus vaccines. *Pediatr Infect Dis J* 22, S94-9.

- Plows, D. J. & Pringle, C. R. (1995). Variation in the fusion glycoprotein gene of human respiratory syncytial virus subgroup A. *Virus Genes* 11, 37-45.
- Power, U., Plotnicky-Gilquin, H. & Huss, T. (1997). Induction of protective immunity in rodents by vaccination with a prokaryotically expressed recombinant fusion protein containing a respiratory syncytial virus G protein fragment. *Virology* 230, 155-166.
- 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.
- Rajala, M. S., Sullender, W. M., Prasad, A. K., Dar, L. & Broor, S. (2003). Genetic variability among Group A and B respiratory syncytial virus isolates from a large referral hospital in New Delhi, India. *J Clin Microbiol* 41, 2311-6.
- Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanovic, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213-9.
- Rixon, H., Brown, C., Brown, G. & Sugrue, R. (2002). Multiple glycosylated forms of the respiratory syncytial virus fusion protein are expressed in virus-infected cells. *J Gen Virol* 83, 61-66.
- 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.
- Rock, M. & Crowe, J. (2003). Identification of a novel human leucocyte antigen-A\*01-restricted cytotoxic T-lymphocyte epitope in the respiratory syncytial virus fusion protein. *Immunology* 108, 474-480.
- Roeckl-Wiedmann, I., Liese, J. G., Grill, E., Fischer, B., Carr, D. & Belohradsky, B. G. (2003). Economic evaluation of possible prevention of RSV-related hospitalizations in premature infants in Germany. *Eur J Pediatr* 162, 237-44.
- Rueda, P., Garcia-Barreno, B. & Melero, J. A. (1994). Loss of conserved cysteine residues in the attachment (G) glycoprotein of two human respiratory syncytial virus escape mutants that contain multiple A-G substitutions (hypermutations). *Virology* 198, 653-662.
- Rueda, P., Palomo, C., Garcia-Barreno, B. & Malero, J. (1995). The three C-terminal residues of human respiratory syncytial virus G glycoprotein (Long strain) are essential for integrity of multiple epitopes distinguishable by antiidiotypic antibodies. *Viral Immunol* 8, 37-46.
- Rychlik, W. (1993). Selection of primers for Polymerase Chain Reaction. In *Methods in Molecular Biology*, pp. 31-40. Edited by B. A. White. Totowa, NJ: Humana Press Inc.

- Saitou, N. (1996). Reconstruction of gene trees from sequence data. *Methods Enzymol.* 266, 427-449.
- Salvato, M., Borrow, P., Shimomaye, E. & Oldstone, M. B. (1991). Molecular basis of viral persistence: a single amino acid change in the glycoprotein of lymphocytic choriomeningitis virus is associated with suppression of the antiviral cytotoxic T-lymphocyte response and establishment of persistence. *J Virol* 65, 1863-9.
- Schuler, G. (1998). Sequence alignment and database searching. In *Bioinformatics: A practical guide to the analysis of genes and proteins*. Edited by A. Baxevanis & B. Ouellette. New York: John Wiley & Sons, Inc.
- Seal, B., Sellers, H. & Meinersmann, R. (2000). Fusion protein predicted amino acid sequence of the first US avian pneumovirus isolate and lack of heterogeneity among other US isolates. *Virus Res.* 66, 139-147.
- 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.
- Shek, L. P. & Lee, B. W. (2003). Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatr Respir Rev* 4, 105-11.
- Simoes, E. A. & Carbonell-Estrany, X. (2003). Impact of severe disease caused by respiratory syncytial virus in children living in developed countries. *Pediatr Infect Dis J* 22, S13-8; discussion S18-20.
- Singh, M., Berger, B. & Kim, P. S. (1999). LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins. *J Mol Biol* 290, 1031-41.
- Singleton, R., Dooley, L., Bruden, D., Raelson, S. & Butler, J. C. (2003). Impact of palivizumab prophylaxis on respiratory syncytial virus hospitalizations in high risk Alaska Native infants. *Pediatr Infect Dis J* 22, 540-5.
- Stensballe, L. G., Devasundaram, J. K. & Simoes, E. A. (2003). Respiratory syncytial virus epidemics: the ups and downs of a seasonal virus. *Pediatr Infect Dis J* 22, S21-32.
- Steward, M. W. (2001). The development of a mimotope-based synthetic peptide vaccine against respiratory syncytial virus. *Biologicals* 29, 215-9.
- Storch, G., Anderson, L., Park, C., Tsou, C. & Dohner, D. (1991). Antigenic and genomic diversity within group A respiratory syncytial virus. *J Infect Dis.* Apr;163 858-61.
- Storch, G. A. & Park, C. S. (1987). Monoclonal antibodies demonstrate heterogeneity in the G glycoprotein of prototype strains and clinical isolates of respiratory syncytial virus. *J Med Virol* 22, 345-56.
- Stolt, E. J., Taylor, G., Ball, L. A., Anderson, K., Young, K. K., King, A. M. & Wertz, G. W. (1987). Immune and histopathological responses in animals vaccinated with recombinant vaccinia viruses that express individual genes of human respiratory syncytial virus. *J Virol* 61, 3855-61.
- Sullender, W. (2000). Respiratory syncytial virus Genetic and Antigenic diversity. *Clinical Microbiology Reviews* Jan, p1-15.

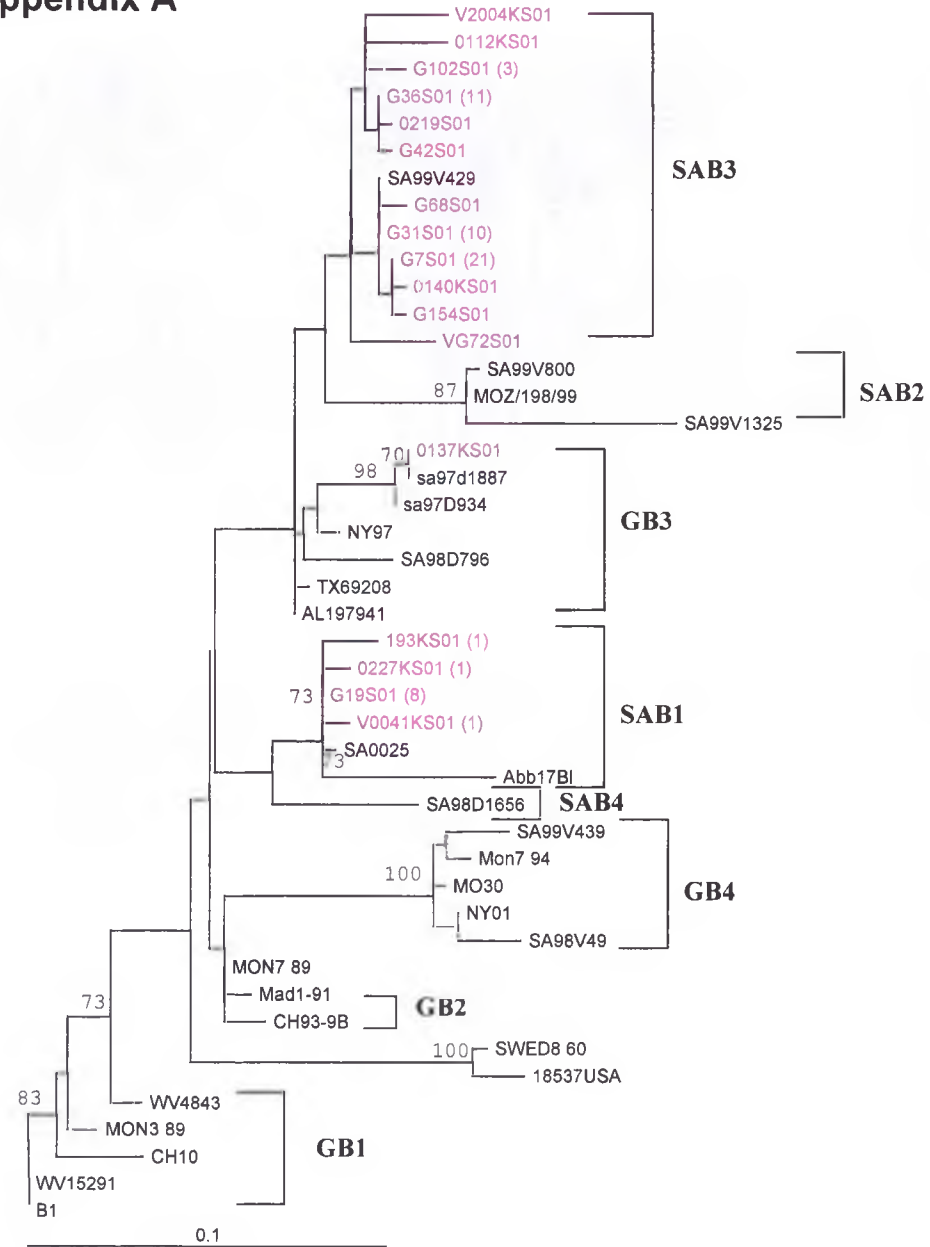
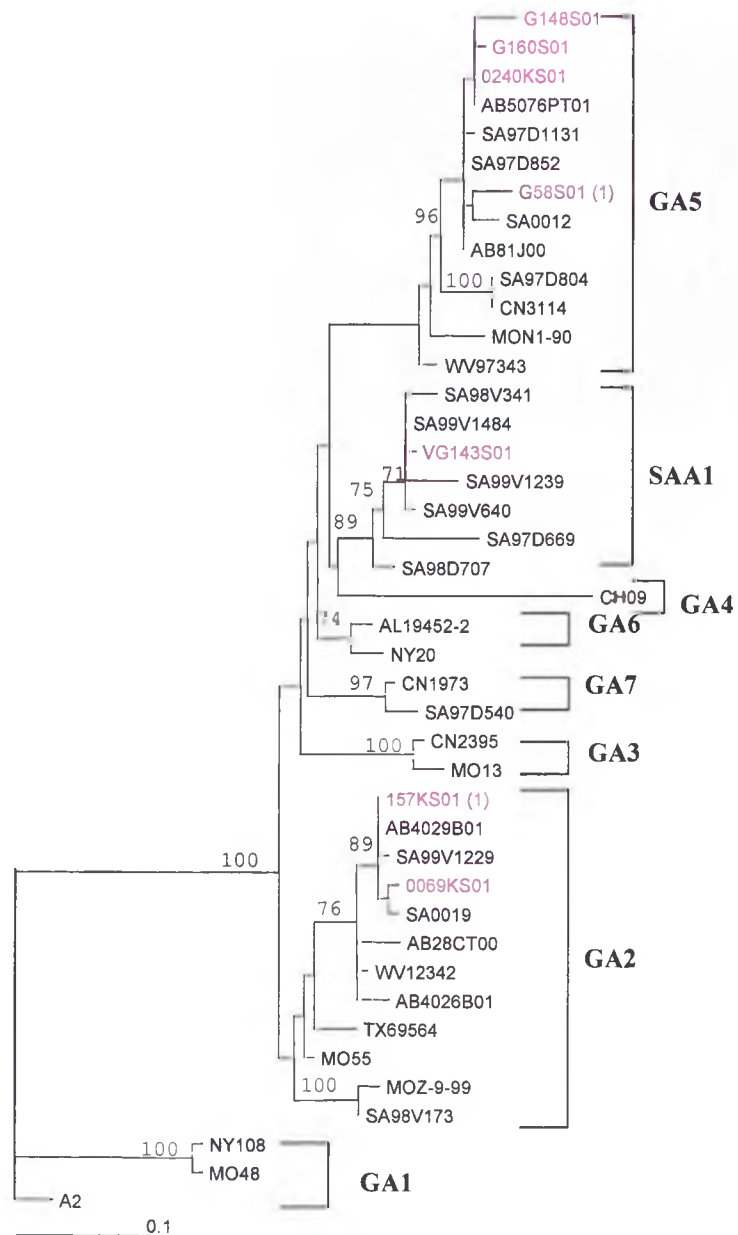
- Sullender, W. & Wertz G.W. (1991a). The unusual attachment glycoprotein of the respiratory syncytial viruses: structure, maturation and role in immunity. In *The Paramyxoviruses*, pp. 383-406. Edited by D. Kingsbury. New York: Plenum Press.
- Sullender, W. M., Anderson, K. & Wertz, G. W. (1990). The 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., Mufson, M., Prince, G., Anderson, L. J. & Wertz, G. W. (1998). Antigenic and genetic diversity among the attachment proteins of group A respiratory syncytial viruses that have caused repeat infections in children. *J. Infect. Dis.* 178 (4), 925-932., 925-932.
- Sullender, W. M., Mufson, M. A., Anderson, L. J. & Wertz, G. W. (1991b). 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.
- Sullender, W. M. & Wertz, G. W. (1991c). Synthetic oligonucleotide probes differentiate respiratory syncytial virus subgroups in a nucleic acid hybridization assay. *J Clin Microbiol* 29, 1255-7.
- Swofford, D., Olsen, G., Wadell, P. & Hillis, D. (1996). Phylogenetic inference. In *Molecular Systematic*, pp. 407-514.
- Takahashi, H., Merli, S., Putney, S., Houghten, R., Moss, B., Germain, R. & Berzofsky, J. (1989). A single amino acid change yields reciprocal CTL specificities for HIV1 gp160. *Science* 246, 118-121.
- Taylor, G., Stott, E. J., Bew, M., Fernie, B. F., Cote, P. J., Collins, A. P., Hughes, M. & Jebbett, J. (1984). Monoclonal antibodies protect against respiratory syncytial virus infection in mice. *Immunology* 52, 137-42.
- 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. (1999). Altered growth characteristics of recombinant respiratory syncytial viruses which do not produce NS2 protein. *J Virol* 73, 466-73.
- 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.
- Terajima, M., Jameson, J., Norman, J. E., Cruz, J. & Ennis, F. A. (1999). High-yield reassortant influenza vaccine production virus has a mutation at an HLA-A 2.1-restricted CD8+ CTL epitope on the NS1 protein. *Virology* 259, 135-40.
- 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.

- Tristram, D. A., Welliver, R. C., Hogerman, D. A., Hildreth, S. W. & Paradiso, P. (1994). Second-year surveillance of recipients of a respiratory syncytial virus (RSV) F protein subunit vaccine, PFP-1: evaluation of antibody persistence and possible disease enhancement. *Vaccine* 12, 551-6.
- Trudel, M., Nadon, F., Seguin, C., Payment, P. & Talbot, P. J. (1987). Respiratory syncytial virus fusion glycoprotein: further characterization of a major epitope involved in virus neutralization. *Can J Microbiol* 33, 933-8.
- Valarcher, J. F., Schelcher, F. & Bourhy, H. (2000). Evolution of bovine respiratory syncytial virus. *J Virol* 74, 10714-28.
- Valdes, O., Martinez, I., Valdivia, A., Cancio, R., Savon, C., Goyenechea, A. & Melero, J. A. (1998). Unusual antigenic and genetic characteristics of human respiratory syncytial viruses isolated in Cuba. *J Virol* 72, 7589-92.
- van Bleek, G. M., Poelen, M. C., van der Most, R., Brugghe, H. F., Timmermans, H. A., Boog, C. J., Hoogerhout, P., Otten, H. G. & van Els, C. A. (2003). Identification of immunodominant epitopes derived from the respiratory syncytial virus fusion protein that are recognized by human CD4 T cells. *J Virol* 77, 980-8.
- 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.
- Varga, S. M. & Braciale, T. J. (2002). RSV-induced immunopathology: dynamic interplay between the virus and host immune response. *Virology* 295, 203-7.
- Venter, M., Collinson, M. & Schoub, B. D. (2002). Molecular epidemiological analysis of community circulating respiratory syncytial virus in rural South Africa: Comparison of viruses and genotypes responsible for different disease manifestations. *J Med Virol* 68, 452-61.
- 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., Rock, M., Puren, A. J., Tiemessen, C. T. & Crowe, J. E., Jr. (2003). Respiratory Syncytial Virus Nucleoprotein-Specific Cytotoxic T-Cell Epitopes in a South African Population of Diverse HLA Types Are Conserved in Circulating Field Strains. *J Virol* 77, 7319-29.
- 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. & Hruska, J. (1983). Monoclonal antibodies to RSV proteins: Identification of the fusion protein. *J Virol* 47, 171-177.
- Walsh, E., Schlesinger, J. & Brandiss, M. (1986a). Purification and characterization of GP90, one of the envelope glycoproteins of respiratory syncytial virus. *J Gen Virol* 65, 761-767.

- Walsh, E. E., Brandriss, M. W. & Schlesinger, J. J. (1985). Purification and characterization of the RSV fusion protein. *J Gen Virol* 66, 409-415.
- Walsh, E. E., Cote, P. J., Fernie, B. F., Schlesinger, J. J. & Brandriss, M. W. (1986b). Analysis of the respiratory syncytial virus fusion protein using monoclonal and polyclonal antibodies. *J Gen Virol* 67 ( Pt 3), 505-13.
- Walsh, E. E., Hall, C. B., Briselli, M., Brandriss, M. W. & Schlesinger, J. J. (1987). Immunization with glycoprotein subunits of respiratory syncytial virus to protect cotton rats against viral infection. *J Infect Dis* 155, 1198-204.
- Walsh, E. E., Schlesinger, J. J. & Brandriss, M. W. (1984). Protection from respiratory syncytial virus infection in cotton rats by passive transfer of monoclonal antibodies. *Infect Immun* 43, 756-8.
- Waris, M. (1991). Pattern of respiratory syncytial virus epidemics in Finland: two-year cycles with alternating prevalence of groups A and B. *J Infect Dis* 163, 464-9.
- Wathen, M., Brideau, R., Thomsen, D. & Murphy, B. (1989). Characterization of a novel human respiratory syncytial virus chimeric FG glycoprotein expressed using a baculovirus vector. *J Gen Virol* 70, 2625-2635.
- Weber, M., Mulholland, E. & Greenwood, B. (1998). Respiratory syncytial virus infection in tropical and developing countries. Review. *Trop Med Int Health* 3, 268-80.
- Welliver, R., Kaul, T. & Ogra, P. (1980). The appearance of cell-bound IgE in respiratory-tract epithelium after respiratory syncytial virus infection. *N Engl J Med* 303, 1198-1202.
- Welliver, R. C. (1999). Immunologic mechanisms of virus-induced wheezing and asthma. *J Pediatr* 135, 14-20.
- 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., Bukreyev, A., Teng, M. N., Firestone, C. Y., St Claire, M., Elkins, W. R., Collins, P. L. & Murphy, B. R. (1999). Recombinant respiratory syncytial virus bearing a deletion of either the NS2 or SH gene is attenuated in chimpanzees. *J Virol* 73, 3438-42.
- 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.
- Wolf, E., Kim, P. & Berger, B. (1997). MultiCoil: A Program for Predicting Two- and Three-Stranded Coiled Coils. *Protein Science* 6, 1179-1189.
- 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.
- Yunus, A., Khattar, S., Collins, P. & Samal, S. (2001). Rescue of bovine respiratory syncytial virus from cloned cDNA: entire genome sequence of BRSV strain A51908. *Virus Genes* 23, 157-164.

- 
- Zhao, X., Singh, M., Malashkevich, V. & Kim, P. (2000). Structural characterization of the human respiratory syncytial virus fusion protein core. *PNAS* 97, 14172-14177.
- Zheng, H., Storch, G. A., Zang, C., Peret, T. C., Park, C. S. & Anderson, L. J. (1999). Genetic variability in envelope-associated protein genes of closely related group A strains of respiratory syncytial virus. *Virus Res* 59, 89-99.

## Appendix A



Maximum likelihood phylogenetic trees constructed for subtype A and B from the Soweto 2001 outbreak at the Chris Hani Baragwanath hospital, using Phylip version 3.5c (Felsenstein, 1993). The trees are rooted with the outgroups (A2 for subgroup A and B1 for subgroup B). The Soweto sequences are indicated in pink.

## Appendix B

**Partial sequences used in phylogenetic analysis of the Soweto 2001 G-protein (Chapter 3).** Unpublished refer to unpublished data of M. Venter.

| SUBTYPE A  |                  |          | SUBTYPE B  |                  |          |
|------------|------------------|----------|------------|------------------|----------|
| Sequence   | Accession number | Genotype | Sequence   | Accession number | Genotype |
| A2         | M11486           | GA1      | B1         | AF013254         | GB1      |
| MO48       | AF233914         |          | Wv15291    | M73542           |          |
| NY108      | AF233917         |          | Ch10       | 9747732          |          |
| SA98V173   | AF348806         | GA2      | Mon389     | 9934694          | GB2      |
| AB28CT00   | unpublished      |          | Wv4843     | M73540           |          |
| AB4026B01  | AY146435         |          | CH93-9B    | AF065251         |          |
| SA0004     | unpublished      |          | Mon 7 89   | 9934694          | GB3      |
| AB4029B01  | AY146436         |          | Mad 1-91   | 9934694          |          |
| SA99V1229  | unpublished      |          | SA98D796   | unpublished      |          |
| SA0020     | unpublished      |          | SA97D934   | AF348817         | GB3      |
| SA0019     | unpublished      |          | SA97D1887  | unpublished      |          |
| MO13       | AF233911         | GA3      | AI197941   | AF233925         |          |
| CN2395     | AF233905         | GA4      | TX69208    | AF233933         | SAB1     |
| CH09       | AF065254         |          | NY97       | AF233932         |          |
| AB5076PT01 | AY146437         |          | SA0025     | AF348825         |          |
| SA97D852   | unpublished      | GA5      | Abb17BI    | AY146438         | SAB2     |
| SA99V1395  | unpublished      |          | SA99V800   | AF348821         |          |
| Ab81J00    | AY146426         |          | SA99V1325  | AF348822         |          |
| SA0012     | unpublished      |          | Moz/198/99 | AF309676         | SAB3     |
| SA97D1131  | unpublished      |          | SA99V429   | AF348813         |          |
| AB3059CT01 | AY146433         |          | SA98D1656  | AF348826         |          |
| SA97D804   | AF348802         |          | 18537USA   | D00394           | SAB4     |
| WV97343    | M55633           |          | Swed8 60   | M73545           |          |
| CN3114     | AF233908         |          |            |                  |          |
| NY20       | AF233918         | GA6      |            |                  |          |
| AL19452-2  | AF233901         | GA7      |            |                  |          |
| SA97D540   | unpublished      |          |            |                  |          |
| CN1973     | AF233904         |          |            |                  |          |
| SA98D707   | AF348810         | SAA1     |            |                  |          |
| SA97D669   | AF348809         |          |            |                  |          |
| SA99V640   | unpublished      |          |            |                  |          |
| SA99V1484  | unpublished      |          |            |                  |          |
| AG38-00    | AF548797         |          |            |                  |          |
| SA98V341   | unpublished      |          |            |                  |          |
| SA99V1239  | AF348808         |          |            |                  |          |

## Appendix C

**Nucleotide alignment of the unique subtype A (A) and B (B) sequences identified in the Soweto 2001 outbreak. The alignments are relative to the prototype strains (blue) (A2 for subtype A and B1 for subtype B), indicated above the reference sequences. The isolate name is shown on the left of the sequences and the genotype and sequence group to the right. Genotypes are marked in different colours. Identical nucleotides are indicated by (\*) symbols.**

### Subtype A:

|          |                                                                   |     |          |
|----------|-------------------------------------------------------------------|-----|----------|
| A2       | AAAAAAGATCCCAAACCTCAAACCACTAAATCAAAGGAAGTACCCACCACCAAGCCCACA      | 60  | GA1      |
| 0069KS01 | AAAAAAGATCTCAAACCTCAAACCACTAAATCAAAGGAAGTACCCACCACCAAGCCCACA      |     |          |
| 157KS01  | AAAAAAGATCCCAAACCTCAAACCACTAAACCAAAACCAAGGAAGTACCCACCACCAAGCCCACA |     |          |
| 0240KS01 | AAAAAAGATCTCAAACCTCAAACCTACAAAACCAAGGAAGCACCTACCACCAAGCCCACA      | 60  | GA5_III  |
| G148S01  | AAAAAAGATCTCAAACCTCAAACCTACAAAACCAAGGAAGCACCTACCACCAAGCCCACA      | 60  | GA5_IV   |
| G160S01  | AAAAAAGATCTCAAACCTCAAACCTACAAAACCAAGGAAGCACCTACCACCAAGCCCACA      | 60  | GA5_V    |
| G58S01   | AAAAAAGATCTCAAACCTCAAACCACTAAACCAAGGAAGTACCTACCACCAAGCCCACA       | 60  | GA5_VI   |
| VG143S01 | AAAAAAGATCTCAAACCTCAAACCACTAAACCAAGGAAGTACCTACCACCAAGCCCACA       | 60  | SAA1_VII |
|          | *****                                                             |     |          |
| A2       | GAAAGGCCAACCATCAACACCACCAAAACAAACATCATAACTACACTACTCACCTCCAAC      | 120 | GA1      |
| 0069KS01 | GAAAGGCCAACCATCAACACCACCAAAACAAACATCATAACTACACTACTCACCTCCAAC      |     |          |
| 157KS01  | GAAAGGCCAACCATCAACACCACCAAAACAAACATCATAACTACACTACTCACCTCCAAC      |     |          |
| 0240KS01 | GAAAGGCCAACCATCAACATCACCAAAACCAACATCAGAACTACACTGCTCACCAACAGT      | 120 | GA5_III  |
| G148S01  | GAAAGGCCAACCATCAACATCACCAAAACCAACATCAGAACTACACTGCTCACCAACAGT      | 120 | GA5_IV   |
| G160S01  | GAAAGGCCAACCATCAACATCACCAAAACCAACATCAGAACTACACTGCTCACCAACAGT      | 120 | GA5_V    |
| G58S01   | GACAAGCCAACCATCAACATCACCAAAACCAACATCAGAACTACACTGCTCACCAACAGT      | 120 | GA5_VI   |
| VG143S01 | GAAAAATCAACCATCAACACCACCAAGACAACATCAGAACTACTCTGCTCACCAACAAC       | 120 | SAA1_VII |
|          | ** * *****                                                        |     |          |
| A2       | ACCACAGGAAATCCAGAAGTCCACAAGTCAAATGGAACCTTCCACTCAACTTCCTCCGAA      | 180 | GA1      |
| 0069KS01 | ACCACAGGAAATCCAGAAGTCCACAAGTCAAATGGAACCTTCCACTCAACTTCCTCCGAA      |     |          |
| 157KS01  | ACCACAGGAAATCCAGAAGTCCACAAGTCAAATGGAACCTTCCACTCAACTTCCTCCGAA      |     |          |
| 0240KS01 | ACCACAGGAAATCTAGAACACACAAGTCAAGAGGAAACCTCCATTCAACCTCCTCCGAA       | 180 | GA5_III  |
| G148S01  | ACCACAGGAAATCTAGAACACACAAGTCAAGAGGAAACCTCCATTCAACCTCCTCCGAA       | 180 | GA5_IV   |
| G160S01  | ACCACAGGAAATCTAGAACACACAAGTCAAGAGGAAACCTCCATTCAACCTCCTCCGAA       | 180 | GA5_V    |
| G58S01   | ACCACAGGAAATCTAGAACACACAAGTCAAGAGGAAACCTCCATTCAACCTCCTCCGAA       | 180 | GA5_VI   |
| VG143S01 | ACCACAGGAAATCCAGAAGTCCACAAGTCAAATGGAACCTTCCACTCAACTTCCTCCGAA      | 180 | SAA1_VII |
|          | *****                                                             |     |          |
| A2       | GGCAATCCAAGCCCTTCTCAAGTCTCTACAACATCCGAGTACCCATCACAACTTCATCT       | 240 | GA1      |
| 0069KS01 | GGCAATCCAAGCCCTTCTCAAGTCTCTACAACATCCGAGTACCCATCACAACTTCATCT       |     |          |
| 157KS01  | GGCAATCTAAGCCCTTCTCAAGTCTCTACAACATCCGAGTACCCATCACAACTTCATCT       |     |          |
| 0240KS01 | GGCAATACAAGCCCTTCACAAATCTATACAACATCCGAGTACCTATCACAACTTCATCT       | 240 | GA5_III  |
| G148S01  | GGCAATACAAGCCCTTCACAAATCTATACAACATCCGAGTACCTATCACAACTTCATCT       | 240 | GA5_IV   |
| G160S01  | GGCAATACAAGCCCTTCACAAATCTATACAACATCCGAGTACCTATCACAACTTCATCT       | 240 | GA5_V    |
| G58S01   | GGCAATACAAGCCCTTCACAAATCTATACAACATCCGAGTACCTATCACAACTTCATCT       | 240 | GA5_VI   |
| VG143S01 | GGCAATCCAAGCCCTTCACAAAGTCTATACAACATCCGAGTACCTATCACAACTTCATCT      | 240 | SAA1_VII |
|          | *****                                                             |     |          |
| A2       | CCACCCAACACACCACGCCAGTAGTACTT                                     | 270 | GA1      |
| 0069KS01 | CCATCCAACACACAACAAAGTGTATTA                                       |     |          |
| 157KS01  | CCATCCAACACACAACAAAGTGTATTA                                       |     |          |
| 0240KS01 | CCATCCAACATAACAGACCAGTAGTCATTA                                    | 270 | GA5_III  |
| G148S01  | CCATCCAACATAACAGACCAGTAGTCATTA                                    | 270 | GA5_IV   |
| G160S01  | CCATCCAACATAACAGACCAGTAGTCATTA                                    | 270 | GA5_V    |
| G58S01   | CCATCCAACATAACAGACCAGTAGTCATTA                                    | 270 | GA5_VI   |
| VG143S01 | CCATCCAACACAACAACTAGTAGTCATTA                                     | 270 | SAA1_VII |
|          | ***                                                               |     |          |

Appendix C continues  
Subtype B:

|                      |                                                               |     |            |
|----------------------|---------------------------------------------------------------|-----|------------|
| B1                   | AAAAGAGACCCAAAAACACCAGCCAAAACGACGAAAAAGAACTACCACCAACCCAAACA   | 60  | GB1        |
| 0137KS01             | AAAAGAGACCCAAAAATACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | GB3_VIII   |
| G19S01               | AAAAGAGACCCAAAAACACCAGCCAAAACGCTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB1_IX    |
| V0041KS01            | AAAAGAGACCCAAAAACACCAGCCAAAACGCTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB1_X     |
| 193KS01              | AAAAGAGACCCAAAAACACCAGCCAAAACGCTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB1_XI    |
| 0227KNS              | AAAAGAGACCCAAAAACACCAGCCAAAACGCTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB1_XII   |
| G36S01               | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XIII  |
| G7S01                | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XIV   |
| G31S01               | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XV    |
| V2004KS01            | AAAAGAGACCCAAAAACACCAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XVI   |
| G42S01               | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XVII  |
| G68S01               | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XVIII |
| VG72S01              | AAAAGAGACCCAAAAACACCAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XIX   |
| 0112KS01             | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XX    |
| G102S01              | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XXI   |
| 0140KS01             | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XXII  |
| 0219S01              | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XXIII |
| G154S01              | AAAAGAGACCCAAAAACACTAGCCAAAACACTGAAAAAGAAAACACCACCAACCCAAACA  | 60  | SAB3_XXIV  |
| ***** * *****        |                                                               |     |            |
| B1                   | AAAAAACCAACCCCTCAGGACACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC  | 120 | GB1        |
| 0137KS01             | AAAAAACCAACCCCTAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC  | 120 | GB3_VIII   |
| G19S01               | AAAAAACCAACCCCTCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC | 120 | SAB1_IX    |
| V0041KS01            | AAAAAACCAACCCCTCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC | 120 | SAB1_X     |
| 193KS01              | AAAAAACCAACCCCTCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC | 120 | SAB1_XI    |
| 0227KNS              | AAAAAACCAACCCCTCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC | 120 | SAB1_XII   |
| G36S01               | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XIII  |
| G7S01                | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XIV   |
| G31S01               | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XV    |
| V2004KS01            | AAAAAACCAACCTCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC  | 120 | SAB3_XVI   |
| G42S01               | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XVII  |
| G68S01               | AAAAAACCAATCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XVIII |
| VG72S01              | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XIX   |
| 0112KS01             | AAAAAACCAACCCCAAGATCAGAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XX    |
| G102S01              | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XXI   |
| 0140KS01             | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XXII  |
| 0219S01              | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XXIII |
| G154S01              | AAAAAACCAACCCCAAGACCACAGAAAGAGACACCAAGCACTCACAATCCACTGTGCTC   | 120 | SAB3_XXIV  |
| ***** ** * * * ***** |                                                               |     |            |
| B1                   | GACACAACCACATTAGAACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | GB1        |
| 0137KS01             | GACACAACCACATCAAAACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | GB3_VIII   |
| G19S01               | GACACAACCACATCAAAACACACAACCCCTACAGCAATCCCTCCACTCAACTACCCCGAA  | 180 | SAB1_IX    |
| V0041KS01            | GACACAACCACATCAATACACACAGCCCTACAGCAATCCCTCCACTCAACTACCCCGAA   | 180 | SAB1_X     |
| 193KS01              | GACACAACCACACCAAAACACACAACCCCAACAGCAATCCCTCCACTCAACTACCCCGAA  | 180 | SAB1_XI    |
| 0227KNS              | GACACAACCACATCAAAACACACAACCCCTACAGCAATCCCTCCACTCAACTACCCCGAA  | 180 | SAB1_XII   |
| G36S01               | GACACGGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XIII  |
| G7S01                | GACACAGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XIV   |
| G31S01               | GACACAGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XV    |
| V2004KS01            | GACACGACCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XVI   |
| G42S01               | GACACGACCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XVII  |
| G68S01               | GACACAGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XVIII |
| VG72S01              | GACACAGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XIX   |
| 0112KS01             | GACACGGCTACATCGGGACACACAATCCAACAGCAATCCCTTACTCAACCAACCCCGAA   | 180 | SAB3_XX    |
| G102S01              | GACACGGCCACACCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XXI   |
| 0140KS01             | GACACAGCCACATCGGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XXII  |
| 0219S01              | GACACGGCCACATCGGGATACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XXIII |
| G154S01              | GACACAGCCACATCAGGACACACAATCCAACAGCAATCCCTCCACTCAACCAACCCCGAA  | 180 | SAB3_XXIV  |
| ***** * *** * *****  |                                                               |     |            |

**Appendix C continues**

|           |                                                              |     |            |
|-----------|--------------------------------------------------------------|-----|------------|
| B1        | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | GB3_V:II   |
| 0137KS01  | AACACACCCAACTCCAAACAAACACCCACATCATCCGAGCCCTCCACATCAAATTCACC  | 240 | GB3_VIII   |
| G19S01    | AACACACCCAACTTTCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC | 240 | SAB1_IX    |
| V0041KS01 | AACACACCCAACTTTCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC | 240 | SAB1_X     |
| 193KS01   | AACACACCCAACTTCACACAAACACCCACAGCATCCGAGCCCTCCACATTAAATTCACC  | 240 | SAB1_XI    |
| 0227KNS   | AAAACACCCAACTTTCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC | 240 | SAB1_XII   |
| G36S01    | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XIII  |
| G7S01     | AACACACCCAACTCCACACAAATACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XIV   |
| G31S01    | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XV    |
| V2004KS01 | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XVI   |
| G42S01    | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XVII  |
| G68S01    | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XVIII |
| VG72S01   | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XIX   |
| 0112KS01  | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XX    |
| G102S01   | AACACACCCAACTCCAAACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XXI   |
| 0140KS01  | AACACACCCAACTCCACACAAATACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XXII  |
| 0219S01   | AACACACCCAACTCCACACAAACACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XXIII |
| G154S01   | AACACACCCAACTCCACACAAATACCCACAGCATCCGAGCCCTCCACATCAAATTCACC  | 240 | SAB3_XXIV  |

|           |                                 |     |            |
|-----------|---------------------------------|-----|------------|
| B1        | CAAAATACCCAATCACATGCTTAGTTATTC  | 270 | GB1        |
| 0137KS01  | CAAAAACCTCCAGTCATATGCTTAGTTATTC | 270 | GB3 VIII   |
| G19S01    | CAAGAAGCCTAGTCATATGCTTAGTTATTC  | 270 | SAB1 IX    |
| V0041KS01 | CAAGAAGCCTAGTCATATGCTTAGTTATTC  | 270 | SAB1 X     |
| 193KS01   | CAAGAAGCCTAGTCATATGCTTAGTTATTC  | 270 | SAB1 XI    |
| 0227KNS   | CTAGAAGCCTAGTCATATGCTTAGTTATTC  | 270 | SAB1 XII   |
| G36S01    | TAAAAACTCTAGTCATATGCTTAGTTATTC  | 270 | SAB3 XIII  |
| G7S01     | CAAAAACCTCCAGTCATATGCTTAGTTATTC | 270 | SAB3 XIV   |
| G31S01    | CAAAAACCTCCAGTCATATGCTTAGTTATTC | 270 | SAB3 XV    |
| V2004KS01 | CAAAAACCTCTAGTCATATGCTTAGTTATTC | 270 | SAB3 XVI   |
| G42S01    | TAAAAACTCTAGTCATATGCTTAGTTATTC  | 270 | SAB3 XVII  |
| G68S01    | CAAAAACCTCCAGTCATATGCTTAGTTATTC | 270 | SAB3 XVIII |
| VG72S01   | CAAAAACCTCTAATCATATGCTTAGTTATTC | 270 | SAB3 XIX   |
| 0112KS01  | CAAAAACCTCCAGTCATATGCTTAGTTATTA | 270 | SAB3 XX    |
| G102S01   | CAAAAACCTCTAGTCATATGCTTAGTTATTC | 270 | SAB3 XXI   |
| 0140KS01  | CAAAAACCTCCAGTTATATGCTTAGTTATTC | 270 | SAB3 XXII  |
| 0219S01   | TAAAAACTCTAGTCATATGCTTAGTTATTC  | 270 | SAB3 XXIII |
| G154S01   | CAAAAACCTCCAGTCATATGCTTAGTTATTC | 270 | SAB3 XXIV  |

## Appendix D

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)

Ref: R14/49 Venter

CLEARANCE CERTIFICATE

PROTOCOL NUMBER M01-09-35

PROJECT

Molecular Epidemiology of Respiratory  
Syncytial Virus In South Africa

INVESTIGATORS

Mrs M Venter

DEPARTMENT

Department of Virology, National Institute for Virology

DATE CONSIDERED

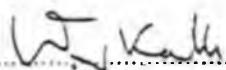
01-09-28

DECISION OF THE COMMITTEE \*

Approved unconditionally

DATE 01-10-01

CHAIRMAN



(Professor P E Cleaton-Jones)

\* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Prof B Schoub

Dept of Department of Virology, National Institute of Virology

Works2\bin\0154\HumEth97\wdb\01-09-35

=====

### DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE .....SIGNATURE .....

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES