Virus Genes 30:2, 267–278, 2005

Amino Acid Variation within the Fusion Protein of Respiratory Syncytial Virus Subtype A and B Strains during Annual Epidemics in South Africa†

ELIZABETH AGENBACH, CAROLINE T. TIEMESSEN & MARIETJIE VENTER*
National Institute for Communicable Diseases, Private bag X4, Modderfonteinroad, Sandringham 2131, South Africa

Received August 11, 2004; Accepted October 5, 2004

Abstract. Recent evidence of positive selection within the cytotoxic T-cell (CTL) epitopes of the highly conserved nucleoprotein of influenza virus raised the question of whether the CTL epitopes of Respiratory syncytial virus (RSV) are also affected by immune driven change over annual epidemics. The fusion protein (F-protein) of RSV is highly conserved within the two subtypes (A and B) and the most important target for the protective response. The position of various neutralizing epitopes has been mapped and characterized between RSV subtypes. CTL epitopes have also recently been mapped for the F-protein of subtype A, however variation within these epitopes between and within the subtypes has not been determined. To address this question, the F-proteins of 18 strains representative of all subgroup A and B genotypes identified in South Africa over a period of 5 years were sequenced. F-protein sequences were highly conserved within and between South African genotypes, with most variability occurring at the nucleotide level. Most of the amino acid differences identified within neutralizing and CTL epitopes were conserved within the subtypes, and therefore does not indicate immune selection. However, out of three CTL epitopes previously identified in subtype A, two (restricted to HLA B*57 and HLA A *01) were conserved only within subtype A, while the third (restricted to Cw*12) contained both subtype- and genotype-specific changes. These results suggest that most of the identified CTL epitopes are subtype A-specific and may not be recognized in subtype B viruses, while the HLA Cw*12 restricted epitope may also not be recognized efficiently in GA5 strains.

Key words: CTL epitopes, F-protein, immune selection, neutralizing epitopes

Introduction

Respiratory syncytial virus (RSV) is the most important viral cause of acute lower respiratory tract infections (ALRTIs) in infants and young children in both industrialized and developing countries [1,2]. RSV belongs to the *Pneumovirus* genus in the family *Paramyxoviridae*. Two antigenic subtypes, subtype A and B exist [3–5], with distinct lineages or genotypes within them (Reviewed by [6]). The most important antigenic proteins include the attachment protein (G-protein) and the fusion protein (F-protein), which mediates attachment to host cells and viral penetration, respectively [7]. The RSV G-protein and F-protein are both targets for the neutralizing antibody response [8,9] while the F-protein is also a target for the cytotoxic T-Iymphocyte (CTL) response [10,11].

The G-protein of RSV is highly variable between and within subtypes A and B, and does not provide cross protection between subtypes (Reviewed by [6]). G-protein genotyping of annual epidemics have demonstrated that both subtypes can circulate concurrently, and that switching of

*Author for all correspondence:
E-mail: mariav@nicd.ac.za
†Genbank Accession numbers: AY526550-6567.

the predominant genotypes occurs [12–14]. Evidence suggests that positive selection may account for the extent of amino acid variability observed within the G-protein [15]. Seven subtype A genotypes (GA 1–7) and four subtype B genotypes (GB1–4) have been identified among isolates from North and South America and Europe ([13] (reviewed [6,16])) while an additional subtype A genotype (SAA1) and three new subtype B genotypes (SAB1, SAB2, SAB3) were identified in South Africa [14] using G-protein sequencing.

In contrast to the G-protein, the F-protein, is highly conserved between subtypes A and B, with an amino acid sequence identity of 91% [8,17]. Evidence suggests that the F-protein is the major target for the protective immune response following RSV infection [9], and provides cross-protection against subtypes A and B in experimental animals [18,19]. The F-protein is synthesized as an N-glycosylated precursor, the F0-protein, and cleaved into two disulfide-linked fragments (F1 and F2) of approximately 50 and 20 kDa, respectively [7]. The F0-protein is cleaved by a furin-type intracellular protease at cleavage-activation sites, exposing a hydrophobic domain on the F1-subunit, which may be important in membrane fusion, cell tropism and pathogenicity [20]. The majority of RSV-specific neutralizing monoclonal antibodies are directed to the F-protein [21,22] and it is also an important target for RSV-specific CTL [10,11].

Recent findings in influenza virus have suggested that immune selection occurs in the CTL epitopes of the highly conserved influenza nucleoprotein [23–25]. Evidence suggests that CTL are needed for the clearance and disease-free protection against RSV [26], which suggests that these cells would play an important role in an effective vaccine. Although positive selection does not occur in the CTL epitopes of the RSV nucleoprotein (N-protein) which is the most conserved of all RSV proteins [27], the emphasis that is placed on the F-protein in RSV vaccine development makes it important to determine if immune selection occurs in the neutralizing antibody and CTL epitopes in the F-protein of different RSV genotypes that appear over consecutive epidemics.

Multiple antigenic sites are present on the F-protein [3]. Limited sequence and antigenic variation exist between the F-protein genes from subtypes A and B although subtype-specific neutralizing epitopes have been identified [28–30]. Escape mutants with sequence changes in the F-protein have been identified in tissue culture that resulted in epitope loss in conserved epitopes or with point mutations in subtype A-specific epitopes [31].

The position of two RSV F-protein-specific T-helper cell epitopes [32] as well as a MHC class I H-2K(d)-restricted CTL epitope have been mapped in mice [33]. Human HLA class I-restricted CTL epitopes have also been mapped to the F-protein of RSV. Brandenburg et al. [34] identified HLA class I (B*57 and Cw*12)-restricted CTL epitopes in position 118–126 and 551–559, respectively, while Rock and Crowe [35] identified a HLA-A *01-restricted CTL epitope in position 109–118, as well as a number of putative epitopes for which the restriction has not been determined.

The RSV F-protein is not thought to undergo positive selection over the complete protein [36]. Investigators were unable to identify any changes in known antibody epitopes among five strains isolated in the same year in Birmingham that belonged to different lineages of subtype A, or in four subtype A viruses isolated earlier in different locations [37]. However progress in the worldwide characterization of RSV genotypes and the recent identification of CTL epitopes justify further investigation in order to determine if the variation that does exist between strains is directed to these epitopic domains, and if immune-driven selection may occur between genotypes that appear over consecutive years. Most investigations of F-protein variation have so far focused on a limited number of isolates from the Americas and Europe [31,37–39], and on subtype A antibody epitopes [40–43]. No data are available about variation in the F-protein of the newly-identified subtype A and B genotypes from the Americas, Europe and Africa, or in the CTL epitopes which have all been identified using subtype A-specific peptides.

In this study, we compared the extent of nucleotide and amino acid variation in the F-protein neutralizing antibody and CTL epitopes between representative specimens of all genotypes identified in South Africa over five consecutive seasons. We investigated both the position of the mapped CTL epitopes [34,35] as well as predicted CTL epitopes restricted to common South African HLA types [27,44]. Although limited variability can be

expected between the F-protein sequences, it is not known if the variation that does exist results in evasion of the CTL response through positive selection in the F-protein-specific CTL epitopes. We also analyzed all known F-protein-specific neutralizing epitopes to determine the extent of variation between South African clinical specimens and the prototype strains currently used for vaccine development.

Materials and Methods

Clinical Specimens

A total of 18 clinical specimens consisting of nasopharyngeal aspirates from infants collected from 1997 to 2000 from across South Africa that have been genotyped by means of their G-protein in previous studies [14,45,46] and from Soweto in 2001 [47], were selected for analysis of the F-protein. Specimens were selected to be representative of all genotypes identified in South Africa to date and included different subgenotypes of each G-protein genotype where appropriate, for subtype A (genotype GA2 (GA2-SA010157K, GA2SA98V173); GA5 (GA5SA0101-28K, GA5SA97D1131); GA7 (GA7SA97D540, GA7SA99VR360) and SAA1 (SAA1SA98D707; SAA1SA98VR341)) and subtype B genotypes (genotype GB3 . (GB3SA98D796, GB3SA98D941); GB4 (GB4SA97D1107, GB4SA98VR468); SAB1 (SAB1SA01 0258K, SAB1SA0100258K); SAB2 (SAB2SA99V800), SAB3 (SAB3SA98VR192, SAB3SA010072K) and SAB4 (SA98D1656)).

RNA Extraction

RNA was extracted directly from NPA specimens using the High Pure^{TM} Viral RNA kit (Roche molecular biochemicals, Mannheim, Germany), according to the manufacturer's instructions. In specimens with high viscosity, the Iysates were homogenized using QIAshredder^{TM} homogenizer columns (QIAGEN, Hilden, Germany), according to the manufacturer's recommendations.

F-protein RT-PCR

Subtype-specific primers were designed to be complementary to the termini of the 5′ and 3′ coding regions of the precursor F-protein (F0) of the prototype strains A2 and B1 for subtypes A and B, respectively, using Dnasis for windows version 2.5 (Hitachi Software engineering Co, Ltd). The primer sequences were as follows: subtype A sense primer RSVFA(5654) 5′-GGGGCAAATAACAATGGAGTT-3′; subtype A antisense primer RSVFAR(7422) 5′-CA TTGTAAGAACATGATTAGGTGCT-3′; subtype B sense primer RSVFB(5632) 5′-AAGCAAGAAC-GAAATTAAACCTG-3′; subtype B antisense primer RSVFBR(7383) 5′-CTGAA TGCAATAT-TATTGATTCCA-3′. The PCR reactions were performed separately for subtypes A and B using the Titan^{TM} one tube RT-PCR system (Roche molecular biochemicals, Mannheim, Germany): 10 μl of RNA was added to 10 μl 5X reaction buffer, 10 mM of each dNTP, 20 pmol of each primer, 5 mM DTT-solution, 10 U RNase inhibitor, and 1 μl of the Titan^{TM} enzyme mix in a 50 μl reaction, using the following amplification cycle: 50°C for 30 min, 94°C for 2 min, (94°C for 30 s, 52°C for 30 s, 68°C for 1 min) × 39 cycles; 68°C for 7 min.

Nucleotide Sequencing

Subtype-specific sequencing primers were designed with Webprimer (http://genome-www2.stanford.edu/cqi-bin/SGD/web-primer) using the prototype A2 and B1 sequences [20,48] for subtype A and B, respectively. The primers for subtype A were as follows: Sense – RSVFA(5654) GGGGCAAATAACAATGGAGTT; RSVFA-(578) CAGCAAAGTGTTAGACCTCAA and RSVFA(1280) CAGCATCCAATAAAAATCG-TG; antisense – RSVFAR(419) CTTCTTTTC-CTTTTCTTGCTTA; RSVFAR(860) GT AACTTTGCTGTCTAACTATT; RSVFAR(1429) GTTCACCTTTTACATAGAGACT and RSV-FAR(7422) CATTGTAAGAACATGATTAGG-TGCT. The primers for subtype B were as follows: Sense – RSVFB(5632) AAGCAAGAACGAA-ATTAAACCTG; RSVFB(541) CTATCAAA-TGGGGTCAGTGTT; RSVFB(1004) CAAGG-ACTGATAGAGGATGGT and RSVB(1446) GTTTCCTTCTGATGAGTTTGA; antisense – RSVFBR(433)CTACACCT AACAAGAAGCCCA; RSVFBR(854)CACATAGTCACAACCATTAGA; RSVBR(1326) CACATAGTCACAACCATTAGA and RSVFBR(7383) CTGAATGCAATATTA-

TTGATTCCA. DNA sequencing products were purified with Centri-Sep spin columns (PE Biosystems, Foster City, USA). Nucleotide sequencing was carried out using the ABI Prism^{R}BigDye^{TM} Terminator Cycle Ready Reaction kit v2.0 (PE Biosystems, Foster City, USA). Primary sequence analysis was performed with Sequencher^{TM} version 4.0.5.

F-protein Sequence Analysis

The South African F-protein sequences were aligned with the prototype strains for subtype A (A2) and B (B1), respectively [20,48], using Clustal X 1.64b [49]. The nucleotide alignments were analyzed with MEGA version 2.1 [50] to identify specific mutations and positive selection and the proportion of differences (*P*-distance). The ratio of nonsynonymous substitutions per nonsynonymous site (K_a) to synonymous substitutions per synonymous site (K_s), were calculated according to the modified Nei-Gojobori method using MEGA version 2.1 [50]. K_a/K_s ratios of greater than 1 were considered as evidence for positive selection [51,52].

Amino Acid Analysis of Published CTL and Neutralizing Antibody Epitopes

The position of published CTL and T-helper and neutralizing epitopes were shaded manually with Genedoc version 2.6.002 (http://www.psc.edu/biomed/qenedoc/). The BIMAS CTL epitope prediction algorithm [53] was used to predict putative CTL epitopes for the F-protein of subtype A prototype strain (A2) [20]. Only HLA types shown to be dominant in at least 10% of the different ethnic groups of South Africa were used in the predictions ([27], personal communication Dr. Adrian Puren, NICD). Secondary structure predictions (α-helices, β-strands and turns) for the F-protein amino acid sequences were performed with NNPREDICT [54] (http://www.cmpharm.ucsf. edu/~nominnpredict.html).

Results

Nucleotide Analysis

The nucleotide *P*-distance for the F-protein of the South African strains ranged from 0.02 to 0.03 between the subtype A genotypes and from 0.00 to 0.02 between the subtype B genotypes with 82% identity between subtypes (*P*-distance 0.18). The overall *P*-distance within subtype A and B was 0.02 and 0.01, respectively, suggesting that the F-protein coding region was 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 synonymous substitutions per synonymous site (K_s) exceeded the nonsynonymous substitutions per nonsynonymous site. The K_a/K_s ratios both between the South African genotypes and relative to the prototype strains for both subtype A and B were less than 1 (results not shown) ($K_a/K_s = 0.07$ for all subtype A genotypes and 0.09 for all subtype B genotypes; and 0.12 between subtype A and B). This suggested an absence of positive selection between the F-protein genes of RSV strains identified in South Africa over five consecutive seasons and relative to the prototype strains.

Amino Acid Analysis

When comparing the South African F-protein sequences from both subtypes to the prototype strains, A2 and B1 for subtype A and B, respectively [20,48], the amino acid *P*-distances ranged from 0.02 to 0.03 for South African subtype A strains and equaled 0.01 for South African subtype B strains (Table 1). Subtype A and B shared 90% F-protein amino acid identity (*P*-distance = 0.1) suggesting that sequence variation was higher at nucleotide level ($P = 0.18$) than at an amino acid level ($P = 0.1$), and that most mutations were silent. Subtype-specific mutations and a limited number of intergenotypic mutations were observed across the F-protein (Fig. 1). The genotype-specific 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, may potentially affect the secondary structure since these mutations resulted in amino acids with altered biochemical properties. Similar findings were also reported in the subtype A F-protein sequences of isolates from south Birmingham [37]. Rixon et al. [55] reported that multiple glycosylated forms of the RSV F-protein are expressed in virus infected cells and that although the biological significance was unclear, it may affect the F-protein function.

Table 1. F-protein amino acid *P*-distance between the South African genotypes and the prototype strains (A2 and B1 for subtype A and B, respectively) [20,48]

	A2	GA2	GA5	GA7	SAA1	B1	GB3	GB4	SAB1	SAB2	SAB3	SAB4
A2		**15**	**19**	**16**	**21**							
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		**5**	**7**	**6**	**6**	**5**	**6**
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	

The number of amino acid differences relative to the prototype strains, are indicated in bold.

In the South African genotypes, all the cysteine residues, as well as the six N-glycosylation sites (one in the F1-subunit and five in F2) and one O-glycosylation site were conserved within the subtypes. The F1 terminus was also conserved between both subtypes (results not shown).

A few amino acid changes were identified that may potentially result in conformational changes (as predicted by α-helixes, β-strands and turns [54]) within the F-protein of South African genotypes (results not shown). 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 relative to the prototype strain (Fig. 1) in SAA1SA98D707 (D → E) (amino acid 344) and GB4SA97D1107 (A → T) (amino acid 518) resulted in changes from turns to α-helix predictions. The mutations in SAB3-SA98VR192 (A → V) and GA5 (L → F) (amino acids 39 and 547, respectively) resulted in changes from α helix to β-strand. In SAB1, a change from a turn to a β-strand was found (P → H) (amino acid 312), relative to the prototype strain. These changes in conformation of the F-protein correlated with changes observed in the hydropathic profiles (results not shown).

Variation in known Neutralizing Epitopes between the South African Isolates and Prototype Strains

The position of known F-protein neutralizing antibody epitopes (7C2 (amino acids 221–236) [56], 47F (amino acids 262–268) [43], L4 (amino acids 289–298) [41] and RS-348 (amino acids 205–225) [32]) were compared to the F-proteins of South African strains in an alignment relative to the prototype strains (Fig. 1). Epitopes RS-348 and 7C2 showed subtype-specific differences between subtype A and B but were conserved within the RSV subtypes while 47F were 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 belong to the same amino acid classification group, and may not necessarily influence epitope recognition. The positions of amino acid point mutations previously identified in escape mutants [29,31] (indicated by arrows above alignment in Fig. 1) were also conserved in all of the South African genotypes.

CTL Epitope Variation

The F-protein CTL epitopes identified in humans to date include epitopes restricted to HLA*B57 and Cw*12 between amino acids 106–114 and 542–550, respectively [34] and a HLA-A*01-restricted epitope between amino acids 109–118 [35] and are shaded in dark grey in Fig. 2. Putative CTL epitopes identified experimentally by Rock and Crowe [35] but for which the restriction has not been determined, are located between amino acids 8–18, 93–103, 260–270, 273–283, 285–295, 374–384, 388–398, 519–520, 521–531 and are shaded in light grey in Fig. 2. The position of a MHC class I CD8+

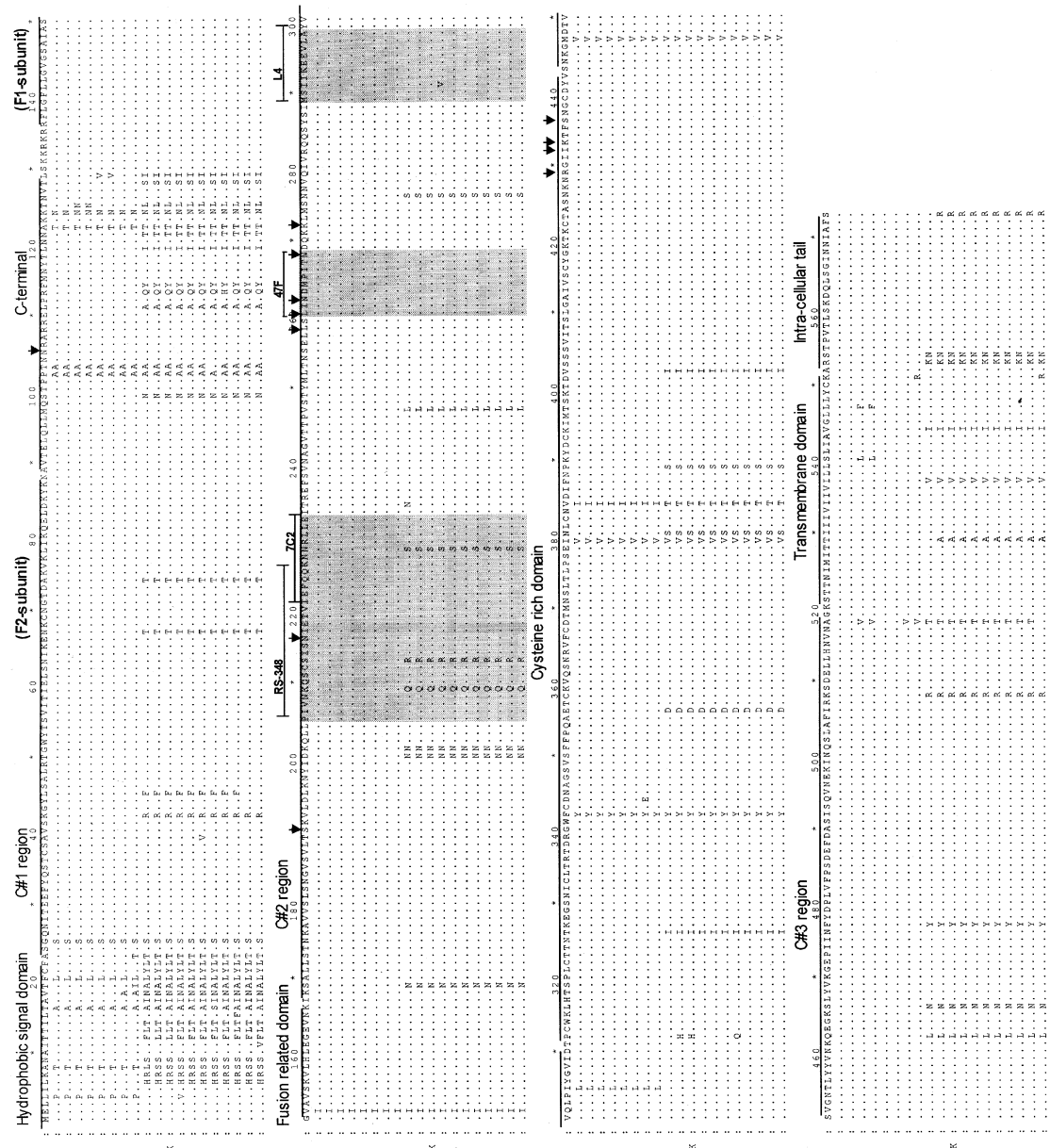

Fig. 1. Amino acid alignment between the F-protein of South African subtype A and B genotypes and the prototype strains indicating the position of the neutralizing epitopes and structural features. Published neutralizing epitopes (7C2, RS-348, 47F, L4) are shown in grey [32,41,43,56] and the epitope name indicated above the sequence. The positions of amino acid point mutations previously identified in escape mutants are indicated by arrows above the alignment [29,31]. The structural features common to Pneumovirus fusion proteins are indicated above the sequence [37]. The isolate name and genotype are shown at the left of the sequence.

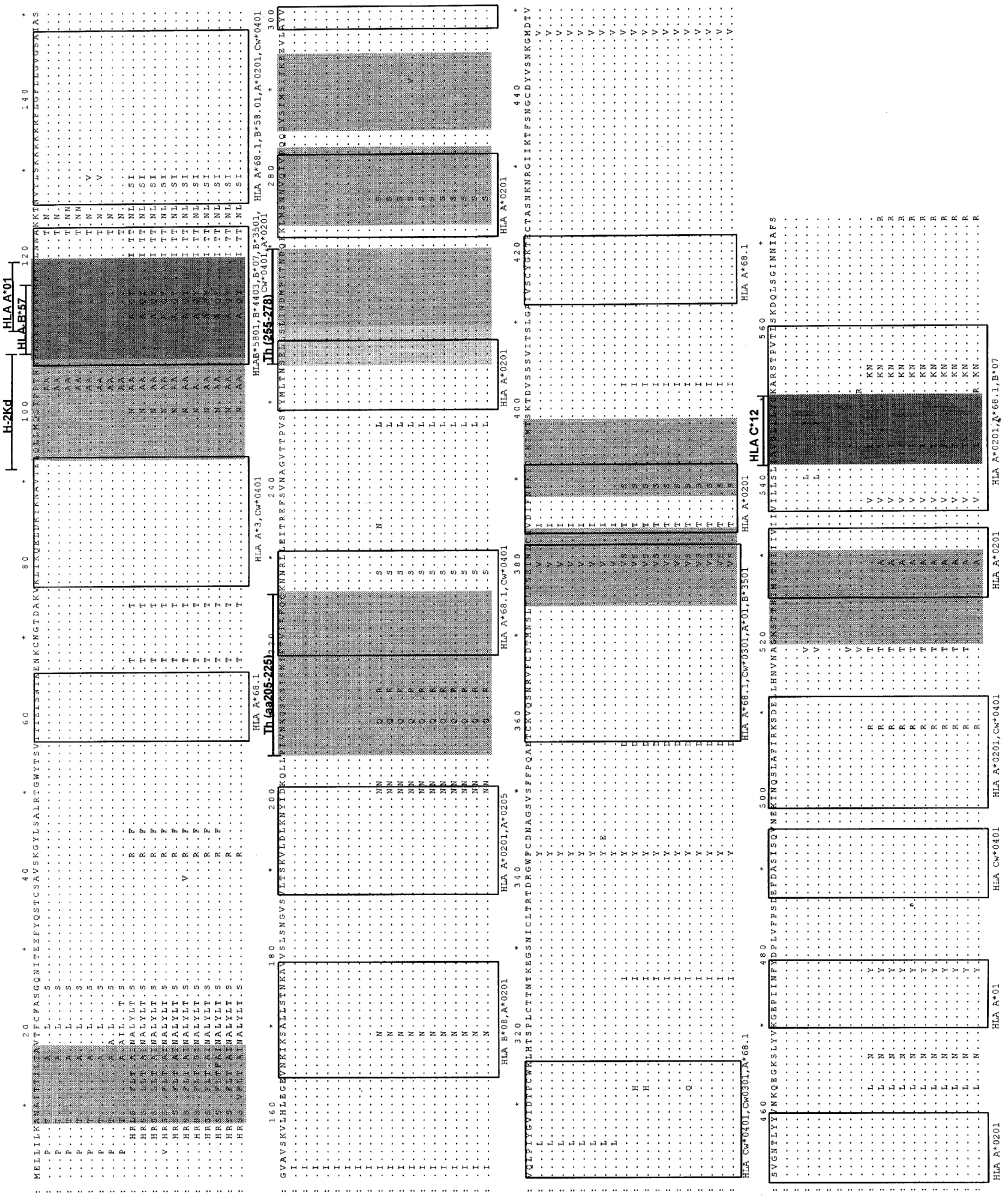

Fig. 2. The position of F-protein specific CTL epitopes identified for the RSV subtype A prototype strain shown relative to the South African genotypes. The known HLA B*57, Cw*12 and HLA-A*01-restricted CTL epitopes [34,35] are shaded in dark grey and indicated between hurdle bars. Putative CTL epitopes that have been mapped experimentally but for which the restriction has not been determined [35] are indicated in light grey, The position of a MHC class I CD8 + H-2K(d)-restricted-CTL epitope (amino acid 92–106) [33] and two H-2-restricted T-helper cell epitopes (amino acids 205–225, 255–278) identified in mice [30,32] are indicated above the alignment. The position of predicted CTL epitopes (listed in Table 2) that was identified for dominant South African HLA types using the BIMAS algorithm [53] are boxed and the HLA-restrictions are indicated below the sequence. The name and genotype of each strain are shown to the left of the sequences.

H-2K(d)-restricted CTL epitope (amino acid 92–106) [33] and two H-2-restricted T-helper cell epitopes (amino acids 205–225, 255–278) identified in mice [30,32] are indicated above Fig. 2 between hurdle bars. The positions of putative CTL epitopes that are restricted to HLA types that have been shown to be present in at least 10% of the different ethnic groups of South Africa ([27], personal communication, Dr. A Puren, NICD) were predicted with the BIMAS algorithm [53]. The predicted epitopes with the highest score results are listed in Table 2 and the corresponding positions boxed in Fig. 2.

The RSV F-protein specific HLA B*57-restricted [34], and HLA-A*01-restricted [35] CTL epitopes, were conserved for subtype A, but differed in at least three 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 [34] 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 [35] 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 8–18 and 93–103, that differed between the South African sequences and the prototype strains.

The H-2K-restricted CTL epitope identified in BALB/c mice [33] 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 205–225 in BALB/c mice [30]. Subtype-specific amino acid changes are visible at position 209 (K $\rightarrow$ Q) and 213 (S $\rightarrow$ R) between subtype A and B, which may affect recognition of these epitopes. The T-helper cell epitope at position 255–278 [32] was conserved between both subtypes. A number of putative CTL epitopes predicted with the BIMAS algorithm for the major South African HLA types listed in Table 2 were conserved between the RSV subtypes, although the majority of them had subtype-specific changes in subtype B. Genotype-specific changes are visible in three of the predicted CTL epitopic domains (Fig. 2).

Discussion

The F-protein is the major focus of prophylactic antibody and subunit or recombinant vaccine research (Reviewed by [58]) Despite the conserved nature of the F-protein, variation does occur between different strains and between the subtypes [37], although the significance of the changes for immune evasion is unclear. The discovery of new G-protein genotypes in Africa that are different from the prototype strains [6,14] raised the question of how the F-proteins are affected. Although previous reports have suggested that immune selection does not occur in known B-cell epitopes, these investigations made use of strains that were isolated during one season and of genotypes that are closely related to the prototype strains [37]. The recent identification of human CTL epitopes on the F-protein also makes it possible to determine if CTL immune-driven selection occurs over consecutive epidemics and relative to the prototype strains. To address these issues the extent of variation between the various genotypes identified in South Africa to date relative to the prototype strains that are used in vaccine development was determined and variation in the F-protein epitopic domains evaluated.

In general, the F-protein was found to be highly conserved between the South African strains, which confirmed previous findings [8,17]. Subtypes A and B shared 82% nucleotide identity within the F-protein of the South African genotypes and classical calculations for selection suggested an absence of positive selection across the complete F-protein, with most nucleotide changes resulting in silent mutations. Amino acid sequence homology was 90% between RSV subtypes A and B, which corresponded with previous reports [8].

However, genotype-specific mutations were identified across the F-protein of the South African isolates, and the secondary structure prediction identified minor amino acid changes

Table 2. Putative CTL epitopes predicted with the BIMAS algorithm for the major HLA types identified in South African ethnic groups

HLA type	Ethnic group	Score result	Start position	Sequence
A*68.1	African	60	57	ITIELSNIK
A*3	Caucasian	180	77	KLIKQELDK
Cw*0401	African/Caucasian	440	85	KYKNAVTEL
B*5801	African	79	106	RARRELPRF
B*4403	Caucasian	480	109	RELPRFMNY
B*3501	Caucasian	60	111	LPRFMNYTL
B*07	Caucasian	800	111	LPRFMNYTL
Cw*0401	African/Caucasian	88	111	LPRFMNYTL
A*0201	African/Caucasian	57	114	FMNYTLNNT
A*68.1	African	120	126	NVTLSKKRK
A*68.1	African	100	127	VTLSKKRKR
B*5801	African	88	129	LSKKRKRRF
Cw*0401	African/Caucasian	50	136	RFLGFLLGV
A*0201	African/Caucasian	178	140	FLLGVGSAL
B*08	Caucasian	80	164	VNKIKSALL
A*0201	African/Caucasian	257	170	ALLSTNKAV
A*0201	African/Caucasian	84	187	VLTSKVLDL
A*0201	African/Caucasian	267	191	KVLDLKNYI
A*0205	African/Caucasian	54	191	KVLDLKNYI
A*68.1	African	180	218	ETVIEFQQK
Cw*0401	African/Caucasian	200	222	EFQQKNNRL
CW*0401	African/Caucasian	240	249	TYMLTNSEL
A*0201	African/Caucasian	129	250	YMLTNSELL
A*0201	African/Caucasian	135	272	KLMSNNVQI
A*0201	African/Caucasian	81	273	LMSNNVQIV
Cw*0401	African/Caucasian	50	298	AYVVQLPLY
Cw*0301	Caucasian	125	301	VQLPLYGVI
A*68.1	African	240	307	GVIDTPCWK
A*68.1	African	150	356	ETCKVQSNR
Cw*0401	African/Caucasian	380	365	VFCDTMNSL
Cw*0301	Caucasian	50	373	LTLPSEINL
A*01	Caucasian	50	383	NVDIFNPKY
B*3501	Caucasian	180	388	NPKYDCKIM
A*0201	African/Caucasian	124	394	KIMTSKTDV
A*68.1	African	120	413	IVSCYGKTK
A*0201	African/Caucasian	77	451	SVGNTLYYV
A*01	Caucasian	113	470	KGEPIINFY
Cw*0401	African/Caucasian	75	487	EFDASISQV
A*0201	African/Caucasian	64	498	KINQSLAFI
Cw*0401	African/Caucasian	200	504	AFIRKSDEL
A*0201	African/Caucasian	81	525	IMITTIIIV
A*0201	African/Caucasian	138	536	VILLSLIAV
A*68.1	African	240	543	AVGLLLYCK
B*07	Caucasian	120	551	KARSTPVTL

Only the top scoring nine-mer peptides are listed (Top 10 scorers for each HLA type with a BIMAS score of >50) [53]. Only HLA types were included in the predictions that were present in at least 10% of the major ethnic groups of South Africa ([27], personal communication Dr. A Puren, NICD).

that may influence the secondary structure of the protein. It has been shown that single amino acid substitution can influence the neutralization response of various viruses including foot-and-mouth disease virus [58], hepatitis B virus [59] and human T-cell leukemia virus type 1 [60], or the CTL response in HIV [61], lymphocytic choriomeningitis virus [62] and influenza virus [63]. It is

therefore important to determine if amino acid changes that do occur within the F-protein alter antibody and CTL epitopes of RSV.

Three CTL epitopes previously identified in subtype A (restricted to HLA B*57, HLA Cw*12 [34] and HLA-A*01 [35]) 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 three 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 hydropathic 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 memory CTL directed against subtype A-specific epitopes may not necessarily recognize MHC class I presenting subtype B peptides. 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 [35] 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) [33] 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 recognized 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 [43] and the linear L4 epitope [41] were highly conserved between both subtypes. A genotype-specific amino acid substitution (I → V) is visible in the L4 epitope for SAB2 (Fig. 1), 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 were highly conserved between subtypes, these epitopes may be useful for the development of prophylactic monoclonal antibodies. The RS-348 epitope [32] and linear 7C2 epitope [56] were conserved within the subtype A genotypes with subtype-specific changes in subtype B. Corvaisier and colleagues found that two amino acid substitutions are present in the RS-348 epitope of all subtype B genotypes at position 209 (K → Q) and at position 213 (S → R) [29], which were the same for the South African F-protein sequences. No changes were visible in the position of these neutralizing epitopes in the South African strains relative to the prototype strains.

The B-cell and T-helper cell epitopes identified in mice between amino acids 205 and 225 [30], were conserved within the South African subtype A and B genotypes, respectively. Corvaisier et al. [32] 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 recognition of the T-cell epitope, since a strong cross-reactivity was obtained with *in vitro* lymphocyte proliferation assays. The T-helper cell epitope residing between amino acids 255 and 278 [32], was found to be conserved between all genotypes of both subtypes (Fig. 2). None of the amino acid changes identified in escape mutants [29,31] were found within the South African F-protein sequences. This may suggest that these changes are not viable in a clinical setting.

Previous reports suggest that positive selection occurs in the G-protein of RSV over consecutive seasons in South Africa [14,45]. In contrast to this,

sequence analysis of the N-protein suggested that recently identified CTL epitopes remain conserved and that positive selection does not occur in the N-protein's CTL epitopes [27]. 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 [65]. It has also been reported that immune selection occurs in the CTL epitopes of the highly conserved influenza virus nucleoprotein [23–25]. With the exception of one genotype-specific mutation in known subtype A CTL epitopes and in three predicted CTL epitopes, most of the amino acid differences identified in this study, in both the B-cell and T-cell epitopes of the F-protein were subtype specific. The 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. Many of the subtype-specific differences are located within CTL epitopic domains which could suggest evolutionary advantage for these differences in the F-protein of the two subtypes. The differences in CTL epitopes between the two subtypes may contribute to the phenomena of subtype switching that has been recorded during annual epidemics as a way to avoid herd immunity and cause repeat RSV infections in the same individual. The genotype-specific changes identified in some of the subtype A-specific CTL epitopes may justify further investigation although this is not conclusive of selective pressure.

In conclusion, 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. Most of the amino acid differences identified within neutralizing and CTL epitopes were conserved within the subtypes, and therefore does not indicate immune selection. However most of the identified CTL epitopes are subtype A-specific and may not be recognized for subtype B viruses, while the HLA Cw*12 restricted epitope at position 551–559 may not be recognized as efficiently in GA5 strains. Differences in CTL epitopes of subtype A and B viruses may be an additional way to overcome herd immunity during annual epidemics. This is the first study that addresses CTL epitope variation in the F-protein and amino acid variation in the F-proteins of the newly-identified RSV genotypes and may contribute to the development of effective vaccines for this and other populations affected by the same genotypes.

Acknowledgments

We thank the Poliomyelitis Research Foundation of South Africa for financial support of this project. We thank Amelia Buys, Michelle Morgan, Areeva Oliver and Lorraine Cranston at the National Institute for Communicable Diseases, South African for diagnostic work; and the staff of the Agincourt Health and Population Unit, University of the Witwatersrand for specimen collection. We also thank Dr. Shabir Madhi and the staff at Chris Hani Baragwanath hospital for specimen collection and diagnosis on specimens from the Chris Hani Baragwanath hospital, Soweto [14]. We would like to thank Dr. Adrian Puren for providing information about HLA types of the different South African population groups.

References

1. Simoes E.A. and Carbonell-Estrany X. Pediatr Infect Dis J *22*, S13–S18; discussion S18–S20, 2003.
2. Selwyn B., Rev Infect Dis *12*, 870–888, 1990.
3. Anderson L., Hierholzer J., Tsou C., Hendry R., Fernie B., Stone Y., and Mclntosh K., J Infect Dis *151*, 626–633, 1985.
4. Gimenez H.B., Hardman N., Keir H.M., and Cash P., J Gen Virol 67(Pt 5), 863–870, 1986.
5. Mufson M., Orvell C., Rafnar B., and Norrby E. J Gen Virol *66*, 2111–24, 1985.
6. Cane P.A., Rev Med Virol *11*, 103–16, 2001.
7. Walsh E. and Graham B. in Dolin R. Wright P.F. (eds), *Viral Infections of the Respiratory Tract*., Marcel Dekker, Inc., New York, N.Y., 1999, pp.161–203.
8. Johnson P.J., Olmsted R., Prince G., Murphy B., Alling O., Walsh E., and Collins P., J Virol *61*, 3163–3166, 1987.
9. Lopez J.A., Villanueva N., Melero J.A., and Portela A., Virus Res *10*, 249–261, 1988.
10. Pemberton R., Cannon M., Openshaw P., Ball L., Wertz G., and Askonas B., J Gen Virol *68*, 2177–2182, 1987.
11. Cherrie A.H., Anderson K., Wertz G.W., and Openshaw P.J., J Virol *66*, 2102–2110, 1992.
12. Peret T.C., Hall C.B., Schnabel K.C., Golub J.A., and Anderson L.J., J Gen Virol *79*, 2221–2229, 1998.
13. Peret T.C., Hall C.B., Hammond G.W., Piedra P.A., Storch G.A., Sullender W.M., Tsou C., and Anderson L.J., J Infect Dis *181*, 1891–1896, 2000.

14. Venter M., Madhi S.A., Tiemessen C.T., and Schoub B.D, J Gen Virol *82*, 2117–2124, 2001.
15. Woelk C.H. and Holmes E.C., J Mol Evol *52*, 182–192, 2001.
16. Sullender W., Clin Microbial Rev *Jan*, 1–15, 2000.
17. Johnson P. and collins A., J Gen Virol *69*, 2623–2628, 1988.
18. Taylor G., Stott E.J., Bew M., Fernie B.F., Cote P.J., Collins A.P., Hughes M., and Jebbett J., Immunology *52*, 137–142, 1984.
19. Walsh E.E., Schlesinger J.J., and Brandriss M.W., Infect Immun *43*, 756–758, 1984.
20. Collins P.L., Huang Y.T., and Wertz G.W., Proc Natl Acad Sci USA *81*, 7683–7687, 1984.
21. Walsh E. and Hruska J., J Virol *47*, 171–177, 1983.
22. Walsh E.E., Brandriss M.W., and Schlesinger J.J., J Gen Virol *66*, 409–415, 1985.
23. Boon A.C., de Mutsert G., Graus Y.M., Fouchier R.A., Sintnicolaas K., Osterhaus A.D., and Rimmelzwaan G.F., J Virol *76*, 2567–2572, 2002.
24. Price G.E., Ou R., Jiang H., Huang L., and Moskophidis D., J Exp Med *191*, 1853–1867, 2000.
25. Voeten J.T., Bestebroer T.M., Nieuwkoop N.J., Fouchier R.A., Osterhaus A.D., and Rimmelzwaan G.F., J Virol *74*, 6800–6807, 2000.
26. Munoz J.L., McCarthy C.A., Clark M.E., and Hall C.B., J Virol *65*, 4494–4497, 1991.
27. Venter M., Rock M., Puren A.J., Tiemessen C.T., and Crowe J.E., Jr., J Virol *77*, 7319–7329, 2003.
28. Connor A.L., Bevitt D.J., and Toms G.L, J Med Virol *63*, 168–177, 2001.
29. Arbiza J., Taylor G., Lopez J.A., Furze J., Wyld S., Whyte P., Stott E.J., Wertz G., Sullender W., Trudel M, et al., J Gen Virol *73*(pt 9), 2225–2234, 1992.
30. Corvaisier C., Bourgeois C., and Pothier P., Arch Virol *142*, 1073–1086, 1997.
31. Lopez J., Bustos R., Orvell C., Berois M., Arbiza J., Garcia-Barreno B., and Molero J., J Virol *72*, 6922–6928, 1998.
32. Corvaisier C., Guillemin G., Bourgeois C., Bour J.B., Kohli E., and Pothier P., Res Virol *144*, 141–150, 1993.
33. Jiang S., Borthwick N.J., Morrison P., Gao G.F., and Steward M.W., J Gen Virol *83*, 429–438, 2002.
34. Brandenburg A.H., de Waal L., Timmerman H.H., Hoogerhout P., de Swart R.L., and Osterhaus A.D., J Virol *74*, 10240–10244, 2000.
35. Rock M., and Crowe J., Immunology *108*, 474–480, 2003.
36. Zheng H., Storch G.A., Zang C., Peret T.C., Park C.S., and Anderson L.J., Virus Res *59*, 89–99, 1999.
37. Plows D.J. and Pringle C.R., Virus Genes *11*, 37–45, 1995.
38. Sullender W.M., Sun L., and Anderson L.J., J Clin Microbiol *31*, 1224–1231, 1993.
39. Garcia-Barreno B., Palomo C., Penas C., Delgado T., Perez-Brena P., and Melero J.A., J Virol *63*, 925–932, 1989.
40. Beeler J. and Coelingh Kv.W., J Virol *63*, 2941–2950, 1989.
41. Martin-Gallardo A., Fien K., Hu B., Farley J., Seid R., Collins P., Hildreth S., and Paradiso P., Virology *184*, 428–432, 1991.
42. Walsh E., Cote P., Fernie B., Schlesinger J., and Brandriss M., J Gen Virol *67*, 505–513, 1986;
43. Lopez J.A., Penas C., Garcia-Barreno B., Melero J.A., and Portela A., J Virol *64*, 927–930, 1990.
44. Hammond M., Toit E., Sanchez-Mazas A., Andrien M., Coluzzi M., Pablo M., Stefano G., Kaplan C., Kenedy L., Louie L., and Migot F., HLA in sub-Saharan Africa. *The 12th International Histocompatibility Workshop and Conference*. Paris, 1997.
45. Venter M., Collinson M., and Schoub B.D., J Med Virol *68*, 452–461, 2002.
46. Madhi S.A., Venter M., Alexandra R., Lewis H., Kara Y., Karshagen W.F., Greef M., and Lassen C., J Clin Virol *27*, 180–189, 2003.
47. Agenbach E., National Institute for Communicable Diseases. University of Witwatersrand, Johannesburg, 2004, p. 150.
48. 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., and Sidhu M.S., Proc Natl Acad Sci USA *94*, 13961–13966, 1997.
49. Thompson J.D., Gibson T J., Plewniak F., Jeanmougin F., and Higgins D.G., Nucleic Acids Res *25*, 4876–4882, 1997.
50. Kumar S., Tamura K., Jakobsen I., and Nei M., MEGA: Molecular Evolutionary Genetics Analyses, Version 2.0. Pennsylvania State University, University Park, 2000.
51. Hurst L., Trends Genet *18*, 486, 2002.
52. Nei M. and Kumar S., *Molecular Evolution and Phylogenetics*. Oxford University press, 2000.
53. Parker K.C., Bednarek M.A., and Coligan J.E., J Immunol *152*, 163–175, 1994.
54. Kneller D., Cohen F., and Langridge R., J Mol Biol *214*, 171–182, 1990.
55. Rixon H., Brown C., Brown G., and Sugrue R., J Gen Virol *83*, 61–66, 2002.
56. Trudel M., Nadon F., Seguin C., Payment P., and Talbot P.J., Can J Microbiol *33*, 933–938, 1987.
57. Meanwell N.A. and Krystal M., Drug Discov Today *5*,241–252, 2000.
58. Hernandez J., Martinez M., Rocha E., Domingo E., and Mateu M., J Gen Virol *73*, 213–216,1992.
59. Howard C.R., J Viral Hepat *2*, 165–70,1995.
60. Blanchard S., Astier-Gin T., Tallet B., Moynet D., Londos-Gagliardi D., and Guillemain B., J Virol *73*, 9369–9376, 1999.
61. Takahashi H., Merli S., Putney S., Houghten R., Moss B., Germain R., and Berzofsky J., Science *246*, 118–121, 1989.
62. Salvato M., Borrow P., Shimomaye E., and Oldstone M.B., J Virol *65*, 1863–1869, 1991.
63. Terajima M., Jameson J., Norman J.E., Cruz J., and Ennis F.A., Virology *259*, 135–140 1999.
64. Price D.A., Goulder P.J., Klenerman P., Sewell A.K., Easterbrook P.J., Troop M., Bangham C.R., and Phillips R.E., Proc Natl Acad Sci USA *94*, 1890–1895,1997.